

DISS. ETH Nr. 22026

DECISIONS AND THEIR UNINTENDED CONSEQUENCES

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZÜRICH
(Dr. sc. ETH Zürich)

presented by
PAVLIN MAVRODIEV

MASTER OF SCIENCE ETH IN MTEC ETH ZÜRICH CH

born on SEPTEMBER 23, 1984

citizen of BULGARIA

accepted on the recommendation of
PROF. DR. DR. FRANK SCHWEITZER
PROF. DR. RYAN MURPHY

2014

ETH zürich

Acknowledgments

As this is the last page that I write in this thesis, I find myself oddly at a loss for words. During these three years, and up to the very moment of completing this dissertation, too many people have helped me in this journey that I do not know where to start. Admittedly, my biggest gratitude goes to Frank Schweitzer, my thesis advisor. Everything I know about science is due to him. If it were not for his continual support in all matters, academic and personal, this thesis would not have materialised. Next, I cannot express enough my gratitude to all my friends in the Chair of Systems Design who were always eager to provide valuable advice. Claudio, Nico, David, Rene, Ingo, Antonios, Mario, Emre, Rahel you all made me feel privileged to know you both personally and intellectually.

On a personal level, I am particularly indebted to my dancing group, *Ludi-Mladi*, where I met so many amazing people and shared so many wonderful moments that I consider them as my second family. The positive emotions I experience every Monday in our meetings is what kept me going at times when the only thing before me was a dead end. I also want to thank a special person who, unknowingly, has brought joy in my life.

Contents

Contents	i
Abstract	v
1 Introduction	1
1.1 Unintended consequences of individual decisions	2
1.2 Focus of the Thesis	6
1.3 Contribution	8
I Decisions in Human Societies	13
2 Consequences of less information in outcomes of social dilemmas	15
2.1 Introduction	16
2.2 A model of social herding	20
2.2.1 Basic Model	20
2.2.2 Results of Computer Simulations	30
2.2.3 Mean-field investigations	33
2.3 Conclusions	38
3 Wisdom of Crowds	41
3.1 Introduction	42
3.2 Experimental set-up	45
3.3 Results	46

3.4	Agent-based model of social influence	49
3.4.1	Analytical approach	52
3.4.2	Simulation results	61
3.5	Conclusion	68
4	Designing wise crowds: the importance of ranking and competition	72
4.1	Introduction	73
4.2	Experiment	74
4.2.1	Set-up	74
4.2.2	Data analysis	77
4.3	Results	79
4.4	Conclusion	87
5	A model for the collapse of Online Social Networks	89
5.1	Introduction	90
5.2	Social Resilience in OSN	91
5.2.1	Quantifying Social Resilience	91
5.2.2	Generalised k-core decomposition	93
5.2.3	A rational model for OSN users	93
5.3	Data on Online Social Networks	95
5.4	Not power-law degree distributions	96
5.5	Empirics of OSN Resilience	99
5.5.1	K-core decomposition	99
5.5.2	Resilience comparison	100
5.6	The decline of Friendster	102
5.7	Conclusion	103
6	Cost of participation increases resilience of Online Social Networks	105
6.1	Introduction	106
6.2	A reputation model	108

6.2.1	Costs and benefits	109
6.2.2	Quasistationary Equilibrium	112
6.2.3	Network dynamics	117
6.3	Results of computer simulations	118
6.3.1	Resilience	118
6.3.2	Core size and largest eigenvalue	120
6.4	Conclusion	122
 II Decisions in Animal Societies		125
 7 Mitigating negative consequences		127
7.1	Introduction	128
7.2	Coordination problem in Bechstein bats	128
7.3	Mechanisms for information transfer	130
7.3.1	Data and methods	130
7.3.2	Leading-following networks	137
7.4	Conclusion	156
 8 Conclusions		159
8.1	Summary	159
8.2	Scientific contributions and future research	162
 Bibliography		167
 Appendix		189
 9		190
9.1	Derivation of $\langle \ln x(t) \rangle$	190
9.2	Derivation of $\langle \sigma(0)^m \delta(t)^n \rangle$	195

10		198
10.1	Results	198
11		200
11.1	Media Coverage	200
11.1.1	Researchers conduct “autopsy” of social network Friendster	201
11.1.2	The Friendster Autopsy: How a Social Network Dies	202
12		204
12.1	Absolute and relative reputation	204
12.2	Impact of cycles on λ_1	206
12.3	Cascades of users leaving the OSN	208
12.4	Optimal cost level	211
12.5	Network structure	213
13		214
13.1	Indications of Swarming	214
13.2	Eigenvector centrality	216
Curriculum Vitae		218

Abstract

All individuals who live in groups, whether they be humans or animals, rely on collective decision-making to establish and sustain viable social organisations. While the benefits of effective collective decisions are widely recognised (e.g. functioning democracies), it is the unexpected collective effects of many individual decisions that deserve attention, as they bear far-reaching consequences for our social lives. Drawing from diverse contexts, this thesis presents examples of such unintended effects and, in the spirit of complex systems, offers a way by which we can understand these effects and, sometimes, use them to our advantage. In the first part, we focus on contemporary decision-making scenarios in human societies. How does social influence affect collective decisions and can we control its effects? How can we use social herding as a mechanism to promote cooperation without explicit enforcement? Under what conditions can user actions, innocuous at first sight, cause the collapse of an online community? Using formal tools and agent-based models, we study the interaction mechanisms underlying the complexity inherent in these questions. In the second part, we shift our focus to the mitigation of unintended negative consequences. We study two colonies of Bechstein bats, whose survival is predicated on solving a coordination problem under limited information. We follow up on existing field work and apply concepts from network theory to reveal the individual contribution in maintaining the needed group cohesion. Finally, we combine agent-based modelling and network analysis to infer simple interaction rules that reproduce the observed collective coordination. We emphasise that these mechanistic rules can serve as a guide for the design of future experimental studies on collective-decision making in Bechstein bats.

Kurzfassung

Alle Individuen, die in Gruppen leben, seien es Menschen oder Tiere, sind darauf angewiesen, Entscheidungen im Kollektiv zu treffen, um tragfähige soziale Organisationen sowohl zu etablieren als auch zu erhalten. Während die Vorzüge effektiver Kollektiventscheidungen weitestgehend bekannt sind (wie zum Beispiel funktionierende Demokratien), sind es doch die unerwarteten Auswirkungen individueller Entscheidungen, die unsere Aufmerksamkeit verdienen, da sie weitreichende Konsequenzen für unser Sozialleben mit sich bringen. Die vorliegende Dissertation präsentiert Beispiele aus diversen Kontexten für solch unintendierten Auswirkungen und zeigt ganz im Geiste komplexer Systeme einen Weg für deren Verständnis auf. In manchen Fällen lassen sich diese sogar zu unserem Vorteil nutzen. Im ersten Teil fokussieren wir uns auf zeitgenössische Entscheidungsfindungsszenarien in menschlichen Gesellschaften. Wie beeinflussen soziale Dimensionen Kollektiventscheidungen? Können wir deren Auswirkungen kontrollieren? Wie können wir soziale Herdenbildungseffekte nutzen, um Kooperation zu fördern ohne diese direkt zu oktroyieren? Unter welchen Bedingungen können die Handlungen der Nutzer einer Online-Gemeinschaft zu deren Kollaps führen? Mithilfe formaler Argumente und Agenten basierter Modelle untersuchen wir die Interaktionsmechanismen, die der Komplexität der behandelten Fragestellungen zugrunde liegen. Im zweiten Teil richten wir unseren Fokus auf die Schadensminderung unintendierter negativer Konsequenzen. Wir betrachten zwei Bechstein-Fledermaus-Kolonien, deren Überleben von der Lösung eines Koordinationsproblems bei limitierter Information abhängt. Wir bauen auf existierender Feldarbeit auf und wenden Konzepte aus der Netzwerktheorie an, um die individuellen Beiträge zur Erhaltung des Gruppenzusammenhalts zu beleuchten. Schließlich kombinieren wir Agenten basierte Modellierung und Netzwerkanalyse für den Rückschluss auf einfache Verhaltensregeln, die die beobachtete kollektive Koordination reproduzieren. Wir möchten hervorheben, dass diese mechanistischen Regeln als Richtlinien für die Gestaltung künftiger experimenteller Aufbauten dienen können, die kollektive Entscheidungsfindungsprozesse der Bechstein Fledermäuse untersuchen wollen.

Chapter 1

Introduction

“History has proved again and again that whenever mankind interferes with a less developed civilisation, no matter how well intentioned that interference may be, the results are invariably disastrous.”

CAPT. JEAN-LUC PICARD

Star Trek: The Next Generation (Season 1, Episode 22, 1988)

The *Prime Directive*, eloquently described by captain Picard, would prevent him from granting knowledge and technology to other inhabited worlds or from providing a life-saving medicine even when the alternative is the extinction of civilizations. The Prime Directive is an impressive intellectual achievement of the fictional 24th-century human civilisation. It is impressive for two main reasons. First, it contains the realisation that individual actions can potentially bear unintended effects on affected systems, regardless of the original intention. Second in the face of this realisation, it imposes a wisdom of restraint, which has been passionately challenged, defended, and occasionally violated in the Star Trek universe¹.

While the Prime Directive is a restriction on fictional interactions with less advanced civilisations, its philosophy is by no means detached from our reality². For one, it is a direct allegory to real human interactions, such as those between modern cultures and indigenous peoples. Above all, however, it is a moral choice of non-interference rooted in awareness

¹Indeed, debating the Prime Directive or its variants, such as the Temporal Prime Directive, is a recurring theme and a main factor for the mainstream popularity of the Star Trek series.

²Or novel for that matter. In fact, one theoretical explanation for the apparent contradiction between the high probability of existence of alien civilisations and humanity’s lack of contact with them (known as the Fermi paradox by the physicist Enrico Fermi) is the zoo hypothesis [17] – an active avoidance by the supposed advanced beings to allow for our natural evolution and sociocultural development.

and recognition of the limited predictability of the long-term effects of individual decisions, policies and regulations on the functioning of a society. Limited predictability is what should make us circumspect that individual actions may engender adverse macroscopic outcomes when intervening in a social system, despite our good intentions. Indeed, many non-interventionist policies, be it political or environmental, are based on avoidance or minimisation of such unintended effects.

Recognising the practical relevance of this cautious stance, in this dissertation I will explore some yet uncharted, unintended consequences of individual decisions in humans and non-human social animals. My goal would be to develop means of dealing with the unintended in situations in which simply restraining from action is not feasible. We will be answering questions such as “What are the expected and unexpected effects of social influence in a population of decision-makers?”, “What kind of incentives promote cooperative behaviour at the expense of free-riding?”, “How can individual behaviour cause the collapse of an online social network and how can we increase its resilience?”, and “How do social systems solve inherent coordination problems, crucial to their survival?”.

All these questions have in common the premise that the constituent members of the social systems involved are not aware of the full extent of their individual decisions on the group as a whole. For instance, members of an online community do not necessarily (and in many cases are not able to) expect that their actions, limited at first glance to their circle of friends, may have a profound effect on the entire community, sometimes as dramatic as causing its collapse. Or a group of decision makers may trust and expect that more information leads to better group decisions, while completely omitting the possibility that they collectively drift in the wrong direction. Even more curious are scenarios in which individuals are naturally faced with hard-to-predict group outcomes, yet the survival of the whole population depends on finding a viable resolution.

It is this type of surprising macroscopic effects, engendered by actions at an individual level, that we will try to model, understand, and design in this thesis.

1.1 Unintended consequences of individual decisions

The idea of unintended consequences has historical roots in philosophy [8] and ethics (consequentialism, [230] – p.92). The early works dealt primarily with accountability, as the debate was on whether to hold agents responsible for the unexpected effects of their

actions³. However, it was the prominent sociologist Rober Merton, who in the beginning of the 20th century, popularized the phenomenon in the social sciences and initiated the study of its causes [174]. Merton posits that unintended consequences are group-level outcomes that are not planned for in advance by any group member, yet they result from their individual decisions.

Causes. In April 2007, a consortium of tech giants, including Intel, Microsoft, Panasonic and IBM, together with six major Hollywood studios, started issuing cease and desist letters to various websites publishing a 16-byte number, beginning with “09 F9”⁴. This number represented a cryptographic key that could be used to decode copy-protected HD-DVDs and Blu-ray discs. When Digg.com – a popular news-rating website – decided to comply with the request and remove all mentions of the key (including banning users who re-posted it), the Digg community revolted and literally crashed the site with “09 F9” posts. The indignation against such blatant censorship transcended the confines of Digg.com and eventually resulted in eternal Internet fame for the key in question. What started as an attempt to suppress information that was virtually unknown to all but a small group of DVD security specialists and the hacker community, ended up in one of the most emblematic illustrations of the Streisand effect⁵ – “09 F9” is now in song lyrics⁶, printed on t-shirts⁷ and coffee mugs⁸.

Evidently, none of the tech and movie giants anticipated the spectacular way in which their actions backfired. However, the notion of unintended consequences has even more fundamental roots – it is contained in one of the building blocks of modern economics, Adam Smith’s “invisible hand”. This famous metaphor illustrates the idea that profit-maximizing individuals can unintentionally create a self-organised and self-regulating marketplace to the benefit of society. As Smith maintains, “by pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it”.

What gives rise to such unintended consequences? After all, can we not, given enough planning and diligence, predict all foreseeable outcomes of our decisions? The answer, I

³For example, consequentialism posits that the virtue of an act can only be judged by its results. Hence, agents must be held responsible even for unintended effects. This view was opposed by moral philosophers, such as Adam Smith and G.E.M Anscombe (see [8, 230])

⁴The complete number is “09 F9 11 02 9D 74 E3 5B D8 41 56 C5 63 56 88 C0”, but is commonly referred to as “09 F9”

⁵Referring to singer Barbara Streisand whose attempt to censor photos of her mansion attracted extra publicity that eventually made the photos widely shared in the Internet

⁶<https://www.youtube.com/watch?v=L9HaNbsIfp0>

⁷<http://www.cafepress.com/+09-f9+t-shirts>

⁸<http://www.cafepress.com/+09-f9+mugs>

believe, is negative.

One reason can be attributed to the imperfect nature of human beings themselves. Merton summarizes four limitations [174] – ignorance, error⁹, conflict of interests¹⁰, basic values¹¹ and self-defeating prophecies¹². Arguably, the most pertinent and wide-spread of these human traits is the first one. Ignorance implies that we may have insufficient knowledge about what motivates individual behaviour in a social system. Certainly, as far as the social sciences are concerned, theories and models are often based on assumptions of human behaviour known to be simplistic. Reality, sadly, shows that we know less about establishing global peace and cooperation, than we do about the fundamental properties of matter. In non-human, social animals, scientists are even more constraint to simply observing behaviour and inferring motivations in retrospect. Due to our inherent intellectual limitations, it becomes impossible to anticipate all possible outcomes.

While Merton was concerned with factors that limit or hamper correct anticipation, he acknowledged that even if cognitive limitations were eliminated, the collective output of many individual actors could still have effects counter to the original intention of each individual. This perspective spurred scientific work as well, and became known as “compositional effects” which emphasize the importance of studying the interdependency between individual actors [30].

The notion of compositional effects is related to the theory of *complex systems*. Stemming from this theory is the second fundamental cause – the world’s inherent *complexity* [108]. Social systems, such as human and animal societies, are examples of complex systems, for they (i) consist of a large number of interacting individuals and (ii) show emergent properties and behaviours that cannot be observed in their constituent parts [125]. Our brain gives us consciousness, yet consciousness does not reside in individual neurons, nor do we understand its emergence from studying the intricate web of nerve cells [54, 152]. In his famous 1841 book, “Extraordinary Popular Delusions and the Madness of Crowds”, the Scottish journalist, Charles Mackay, argues that, “Men, [...], think in herds; it will be seen that they go mad in herds, while they only recover their senses slowly, and one by one.” [164], and reasons that interacting groups of individuals can exhibit inexplicable

⁹E.g. in assessing the situation, selecting or/and implementing a course of action

¹⁰ Due to misaligned incentives, decision-makers do not wish to consider further consequences of their immediate actions

¹¹When a system of basic societal values (e.g. Protestant ethic, capitalism, etc.) requires certain specific behaviour, decision-makers are not concerned with objectively analysing the full consequences of their actions, as they obtain subjective satisfaction of conforming.

¹²Predicting a problem may lead people to find solutions in advance, thereby preventing its occurrence, which is unintended in the original prediction

“madness”, such as witch-hunts, economic bubbles¹³, crusades, lynching, superstition, etc., even though individuals alone are much less prone to such extreme behaviour. The French social psychologist, Gustav Le Bon, further claims that, “an individual immersed [...] in a crowd soon finds himself – either in consequence of magnetic influence given out by the crowd or from some other cause of which we are ignorant – in a special state, which much resembles the state of fascination in which the hypnotized individual finds himself in the hands of the hypnotizer” [25].

Complexity, begot by a large number of interactions among a system’s constituent parts, and between the system itself and its environment, is this “other cause of which we are ignorant”. Complexity should thus be seen as a fundamental system property that causes a complex system to often respond to intervention in a highly *non-linear* manner¹⁴. Non-linearity implies that affecting a few elements, or their interactions, can cause global ripple effects that feedback on the system itself, and thus set a natural limit to the a-priori predictability of the system’s behaviour at any aggregation level. In other words, some system qualities are fundamentally irreducible to the system’s constituent parts, i.e. cannot be studied by the principle of superposition [152]. If we were to accept this notion of limited predictability, then many unintended consequences in our social lives are a natural outcome of the dynamics of complex systems. Consequently, we also have to recognise that unintended consequences cannot be avoided completely. This implies that we should expect that despite our best efforts to achieve a desired outcome, unintended consequences are likely to occur.

In fact, in many aspects of our lives, they often do occur. Examples range from increase in marijuana use among adolescents due to higher minimum drinking age regulations [63], reduced trust among board members in independent corporate boards¹⁵ [151] and unintended influence of facilitators in group discussions that can drastically affect the group decision [93], through parent-child conflict caused by advertising [196], increase in mortality risk to illegal immigrants due to more stringent border control [52], to the counter-intuitive surge of cyber warfare with increased cyber security [14], and even the “revenge” of the unintended consequences of the mechanical, chemical, biological, and medical advances that have been the hallmark of our modern age [250]. We should also

¹³During the tulip mania in the 17th century in the Netherlands, the price of a single tulip bulb reached more than 10 times the annual income of a skilled craftsman.

¹⁴Example for a non-linear response is when scaling an input to the system by a given factor causes more than proportionate change in the output. Or when adding a new incentive in a group discussion drastically shifts the group decision.

¹⁵In corporate governance theory, it has been proposed that corporate boards in which independence, skepticism, and rigorous loyalty to shareholder interests are the dominating values perform better than boards composed primarily of company insiders.

note that unintended consequences can be both positive and negative. And we often find the two types at the same time. In studying the 1999 UN peacekeeping operations in Kosovo and their unintended effects on the local economy, Ammitzboell finds positive impacts in terms of more job opportunities and increased demand for services and goods [5]. These, however, are coupled with rise in basic commodity prices, an increase in salary disparities, and invasion of the international private sector in taking up the increased demand.

As these examples demonstrate, the effects of unintended consequences can transcend the domain from which they originated, and affect other systems of social life. The opposite is also true. Individual decisions can affect the primary decision-making domain, unbeknownst to anyone in the group. These systemic effects, in turn, can feedback on the behaviour of individuals themselves. For example, it has been proposed that segregation in large cities is the unintended consequence of minor individual preference to be similar to one's neighbours [214]. Also known as homophily¹⁶, this psychological tendency to associate with similar others leads to individuals grouping together, which further attracts others of the same type. Over time, the whole population segregates into homogeneous clusters, even though no individual had such intentions¹⁷. Another example is panic stampedes – the collective movement of crowds, without a clear sense of direction or purpose, as a response to a disaster (e.g. fire in a crowded theater) [104]. This phenomenon arises when individuals react maladaptively (from systemic perspective) to crisis situations (e.g. due to fear) and start moving faster and pushing others, e.g. making a dash to the nearest exit. The increased interaction drives others to assume the same behaviour. In turn, the mass movement feedbacks on each individual and further intensifies his behaviour, ultimately building up jamming and clogging the exits.

1.2 Focus of the Thesis

This dissertation is about exploring, some yet uncharted, unintended consequences of individual decisions in humans and non-human social animals. In humans, we will show how evolved traits, such as herding, can lead to surprising positive and negative group effects in the contexts of promoting cooperation and the wisdom of crowds. Moreover, we will investigate the extent to which individuals affect each others' decisions, also known as *social influence*, and its impact on these effects. Building on this, we will then proceed

¹⁶See [94] and references therein

¹⁷The implications of this are profound, e.g. ghetto formations in large cities would not be attributable to malicious discriminatory behaviour, but rather to the unrecognized long-term effects of homophily.

with identifying rules for individual behaviour which, much like the invisible hand of Adam Smith, lead a population to favourable collective outcomes. In animals, we will investigate social systems with *inherently* conflicting individual preferences. To survive, such groups need to reconcile the tension between personal and group benefits. The second part of the thesis is dedicated to investigating the behavioral rules that make this possible.

Why are we interested in identifying and designing individual-level behavioural mechanisms? First, most common and natural observations are at the level of the individual, not the social system (although the problem of interest usually resides on the system level). Second, some of the most interesting, unexpected outcomes involve a large number of individuals, who act on their own, yet are interrelated within their own social circles in which they interact and thus mutually influence each other. In the absence of centralised control, it is not immediately obvious why many decisions have group-level outcomes that have not been specifically planned for, such as fads and fashion, bank runs, panic stampedes, pedestrian crowd dynamics, cultures and norms, business cycles, etc., and which external or internal factors contribute to them.

In the social sciences this individual-based approach to understanding group-level phenomena was popularized by the American sociologist James Coleman [46, 47]. Coleman proposed a framework, coined “the micro-macro link”, which is illustrated in Figure 1.

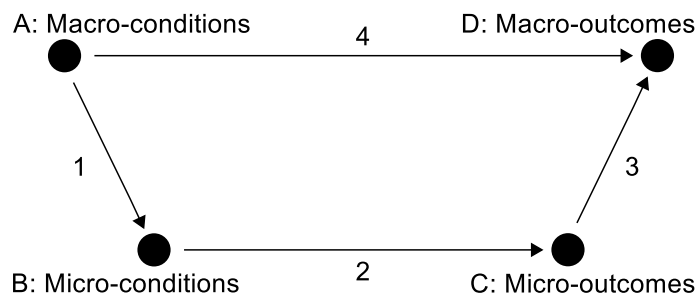


Figure 1.1: Coleman’s framework of linking micro-level phenomena to macro-level outcomes. Reproduced from [203].

Nodes A and D represent premises describing macro-conditions and macro-outcomes respectively. A particular political system, governmental structure or religious doctrine can be characteristic of system-wide macro-conditions, whereas revolutions or order can be examples of possible macro-outcomes. Arrow 4 is a proposition of an empirical regularity that associates macro-conditions with macro-outcomes – e.g. “the larger a group, the more difficult it is to sustain cooperative behaviour” or “despotic leaders prompt social unrest and riots”. Node D and arrow 4 are thus explananda. Node B represents assumptions about rules of individual behaviour, such as values, available strategies and incentives. Arrow 1 symbolises assumptions on how social conditions affect these regularities. For

example, economic conditions, governmental institutions or political organisation can be conceived to influence individuals' values, feasible strategies and available information. Once a viable description of conditions operating on an individual level has been proposed, arrow 2 describes how these are turned into actual individual behaviour (node C). Hence, arrow 2 represents a micro-theory of individual behaviour. For instance, given two available individual strategies — cooperation and defection — a micro-theory, such as the rational choice theory would predict that individuals always defect in the standard Prisoner's dilemma [187, 202]. Finally arrow 3 offers assumptions on how observed individual behaviour is combined to a particular global outcome (e.g. voting systems).

What is lacking from Figure 1.1 is a feedback between macro-outcomes and micro-outcomes. It is reasonable that not only individuals create group phenomena, but also that their subsequent decisions are, in turn, affected by the newly-formed state of the system. For instance, initial contributors to a public good may be discouraged by the prevailing number of free-riders, and eventually decide to stop their contributions. Social networks are another classical example – the decision to join one, is among other things, also a function of their current user base. Hence, to properly account for this linkage and to formalize the framework, we will use a computational class of models known as *agent-based models* (ABMs).

The prime idea behind agent-based models is to abstract the constituents of a system (and their properties) into self-sufficient modules, or agents, and to impose rules of interaction among them [218](Chapter 1). The idea reflects the conceptual thinking towards the study of complex systems – it is through the interaction of many individual elements that the emergent properties of such systems can be understood; in direct analog, we see unintended consequences as an emerging feature of social systems, be it human or animal. The proper abstraction of the real system must, as in any, modelling approach, conform to the goals of the model. If the main goal is prediction, such as in weather forecasting or in a flight simulator, the encoding of the real system is necessarily thorough, including as much detail as possible. In contrast, the goal of agent-based models is to find the minimum set of agent properties and interaction rules that produce a certain emergent behaviour. In doing so, we include as much detail as necessary. This is the main challenge of our approach, as there is no ready-made solution for isolating the right ingredients.

1.3 Contribution

In the spirit of ABMs and complex systems theory, we model the emergence of unintended consequences as a result of individual-level interactions. We believe this approach

adds a valuable systematic treatment to the phenomenon. Historically, unintended consequences have been treated extensively in variety of specific contexts – from theology (e.g. theodicy¹⁸ and predestination¹⁹) to technology [174]. In fact, it is the diversity of context that hampers a general understanding of the phenomenon. In addition, there is a strong tendency in sociology to model specific effects as a “linear” consequence of a-priori premises [197]²⁰. Examples are the transformation of hard work into achievement, the relation between years of formal education and wages, the link between parents’ aspirations and children’s vocation, the belief that inflation and unemployment are always negatively correlated, and the conviction that modernization should necessarily lead to secularisation [31]. Reality, however, is far more complex and such neat logical deductions cannot always be drawn. The emergence of unintended consequences tends to be addressed by sociologists mostly retrospectively after all outcomes (including the unexpected) have been realized [197]. And despite the awareness of compositional effects, the emphasis falls back into the discovery of new “natural” laws that could have predicted the unintended in hindsight [31, 276].

As a result, the phenomenon of unintended consequences is still a productive topic in social sciences, and is not well understood [66, 175]. By focusing exclusively on modelling interactions among simple agents, our ABM approach can contribute to identifying genetic conditions for the emergence of unintended consequences that could hopefully lend to generalizations across systems.

The thesis consists of two parts. Part I focuses on contemporary decision-making scenarios in human societies. First, we start with the most prominent case in social sciences – *social dilemmas*. Social dilemmas are situations in which private interest is at odds with public interest. The dilemma arises because each individual receives a higher benefit from maximising his own interest, yet if everyone did that, all would be worse off at the end ^{21,22}. In Chapter 2, we formalize the social dilemma in a game-theoretic context and identify the conditions under which herding, a basic human trait, can promote global cooperation when selfish behaviour is in the individual best interest. This result is contrasted to Chapter 3, in which we demonstrate some unexpected negative effects of herding and

¹⁸In philosophy of religion, this is an attempt to answer the question, “Why God permits evil?”. If unintended consequences are some form of “evil” then theodicy can provide an alternative cause for them.

¹⁹In theology this is a doctrine according to which all events have been willed by God.

²⁰Linear is to be understood in the more general sense of consequences following rationally from antecedents.

²¹Example: it is in the best individual interest not to pay taxes, yet provision of many public goods crucially depends on tax income

²²Formally, a social dilemma is defined in the context of game theory. It is a situation when the Nash equilibrium is not Pareto optimal.

their dependence on the strength of social influence. In the same chapter, we generalize the insights about herding and social influence with the help of an agent-based model. Interestingly, the model predicts that the polarity of the effects depends on the composition of the population, rather than on the nature of herding. This assumption is tested in Chapter 4 with the help of a laboratory experiment.

The last two chapters deal with decisions in present-day scenarios, namely in online social networks (OSNs). Fundamentally, users in OSNs face two major choices - joining and staying in or leaving the network. New and existing users contribute to an OSN's user base, which is a common and accepted criteria for measuring its overall success. However, under certain circumstances, and dependent on the networks topology, users who decide to leave can create an unexpected cascade effect of others leaving that could lead to the demise of the whole community. In Chapter 5, we show that this mechanism might have been the culprit in the collapse of Friendster – an online social network, founded in 2002, that at its peak attracted close to 112 million users and even rejected a \$30 million bid from Google. Despite its success, Friendster inexplicably lost around 80% of its user base in about a year. Therefore, from the perspective of OSN administrators, resilience to such events is just as important as the user base. We proceed with Chapter 6 and introduce a mechanism that interestingly increases resilience in reputation networks ²³.

Part II builds on Part I by considering a system facing an intrinsic coordination problem – individual decisions take place under conditions of conflicting personal preferences and limited information about others' choices. In this system, individuals are not aware of the full consequences of their actions, nevertheless they must prevent the highly negative outcomes associated with failure to coordinate collectively, namely that of group splitting. As this is an animal system under observation, our aim is to infer the decision-making mechanisms that allow such groups to cope and maintain group cohesion.

Chapter 7 introduces our system – a colony of Bechstein bats. Bats are an example of a fission-fusion society; the colony splits during the night to forage and explore available roosts, and assembles back into groups of varying size and composition during the day to roost. Living in groups is known to provide evolutionary benefits [145, 244], but roosting in groups is vital for the colony survival as it serves crucial thermoregulatory purposes [199]. However, since individuals forage alone or in pairs, they accumulate different information and preferences about discovered roosts²⁴. Consequently, bats are not aware of the roosting

²³Reputation networks are those in which users receive reputation from their followers. Examples are social networks such as Twitter, product review communities such as Amazon and Doyoo, and many others.

²⁴Preferences depend on the experience at the time of visiting a roost, e.g. presence of other animals is perceived as negative, while warmth is positive.

intentions of all others and do not come back to the roosting area all at the same time. On one hand, they have certain knowledge of and inclinations toward potential roosts that they would like to occupy for the day. On the other hand, due to their limited information, they are not aware of the roosting intentions of others. If individuals followed only their personal preferences, they run the risk of roosting either alone or in a too small of a group, should they be among the very few who have decided for that particular roost. Hence, the collective interests of the colony requires individuals to forfeit some of their personal preference in favour of forming feasible roosting groups. This is the essence of the coordination problem, whose resolution will be the topic of Chapter 7.

Part I

Decisions in Human Societies

“Atreides’ power must never be marginalized by the chaos of democracy.”

REVEREND MOTHER ALIA ATREIDES
Dune Messiah (1969)

Chapter 2

Consequences of less information in outcomes of social dilemmas

Summary

How can we promote cooperation in social dilemmas where defection is the evolutionary stable strategy? Social dilemmas are sometimes studied in the game-theoretical framework of the Prisoner's dilemma game. Most theoretical approaches in this direction have proposed various modifications to the game so that cooperative behaviour does not get universally exploited. These approaches predominantly rely on mechanisms that in general add more information in terms of e.g. changing the payoff structure, extension of the time horizon, assuming spatial interactions, introducing alternative strategies, etc. In this chapter we propose a novel mechanism that instead of assuming extra information requires less. We assume that agents adjust their strategies based on both payoff considerations and the predominant strategy in their local neighbourhood. The latter represents social herding. By means of detailed analytical and numerical investigations, we show that restricting payoff-related information to some agents, and designing the "proper" non-linear herding incentives can grow an initially small cluster of cooperators into invading a sea of defectors.

Based on Schweitzer, F, Mavrodiev P., and Tessone C.J. *How can social herding enhance cooperation?*, in *Advances of Complex Systems*, vol. 16, number 04n05, 2013. F.S. inspired the approach in the paper and wrote most of the published paper. C.J.T. created Figures 2.5, 2.6 and 2.7, and provided technical advise. P.M. was responsible for the technical analysis in the paper and the remaining figures. However, P.M re-wrote most of the original text to make it suitable for this dissertation.

2.1 Introduction

Social Dilemmas In the winter of 1978, a village in the Netherlands was struck by unusually heavy snow, which cut off electricity, heating, television, etc., and effectively disconnected the village from the rest of the country. Fortunately, one villager possessed a generator that, if used frugally, could provide sufficient electricity for the village until normal operations were restored. Not before long the generator collapsed, as most people were abusing the opportunity to live comfortably – watching television, having multiple lights on, heating water, etc. The inhabitants of the village were able to repair the generator, and appointed inspectors to prevent abuses. Eventually, the generator collapsed again, due to overuse of energy [257].

Situations like this are referred to as social dilemmas, because individual interests are contrary to the interests of the society as a whole. Examples abound. In the long run everyone would benefit from a cleaner environment, though not many are willing to voluntarily reduce their carbon footprint. Selling a share of stock would benefit the individual investor, yet if the majority did that, the stock price would plummet to the detriment of all investors. Parties involved in an arms race find it unilaterally beneficial to steadily increase military power, which eventually decreases national security and leads to an outcome termed MAD (mutually assured destruction [231]). Families in developing countries achieve the highest social payoff by having as many children as possible, yet the resulting overpopulation makes a social security or old-age benefit system impossible, which eventually worsens overall living conditions. Athletes who resort to doping would always have an advantage, regardless of whether their competitors' choose the same.

In social dilemmas it is rational for each individual to pursue his own gain, regardless of the actions of others, even though the collective societal gain (and in turn the long-term individual gain) suffers as a result. Often, the rational pursuit of one's own gain involves entering into conflict, which uses up resources and is thus socially inefficient. "Rational", here, does not refer to the colloquial usage of conveying sanity. Rather we consider the term in its economic meaning of utility maximising agents [48] – individuals are able to quantify the benefits and costs associated with alternative choices and act in a way that maximises the difference. In this way, the benefits of defection, i.e. pursuing one's selfish interests, are transferred fully to the individual, whereas the costs to the society are shared by all, which inevitably makes such behaviour economically attractive. Eventually, as everyone reaches this conclusion, the society gets locked in a state of ubiquitous ruin. This inexorable fate, which almost seems like part of the natural order of things, has

motivated people to refer to social dilemmas as *the tragedy of the commons* [97]¹.

Arguably, we humans may not be rational, in such an absolute sense, for every decision in our lives. However, it is still individually tempting to defect from the socially optimal choice, as doing so would not pose a perceivable risk to the collective provided the majority cooperates. Indeed, the provision of many public goods, e.g. city parks, roads, etc., is not crucially affected by a few selfish individuals who do not pay local taxes. The issue arises, of course, when everyone believes that they are in the defecting minority.

While many societies throughout history have faced social dilemmas, it is these dilemmas that are particularly global and pressing that have attracted the most attention, such as in areas of resource depletion [33, 121], pollution [165] and overpopulation [216]. Often, the scale of the problem and the inability to form reliable commitments to cooperative behaviour between agents, prevents recognising and mitigating the collective risk. For example, in the context of climate change, it is particularly telling that some of the major contributors to greenhouse emissions – USA, China, and India – have not ratified the Kyoto protocol. The distant time horizon of the issue and bleak results of similar agreements up to now, contrasts the immediacy of industrial demand.

The main challenge to cooperation is that it has an immediate negative economic effect on individual actors. It requires an “upfront” investment in terms of contribution to a public good. For example, latest estimates show that to considerably reduce the risk of climate change, global greenhouse emissions must be reduced by $\approx 50\%$ by 2050 [178]. A reduction of this magnitude is a significant economic cost for any country. Therefore, given their pervasive nature and serious consequences, it would serve policymakers well to understand how to encourage people to cooperate in social dilemmas. To this end, we must recast social dilemmas in a rigorous methodological framework, which hopefully allows us to formalise and design mechanisms for promoting cooperation.

Game theoretical formulation A particularly useful framework for analysing individual decisions in social dilemmas, and their collective outcomes, is provided by game theory. Game theory assumes that individuals are rational actors motivated to maximise their economic utilities. The canonical example of modelling a social dilemma in a game-theoretic context is the so called Prisoner’s dilemma game[202].

The original formulation of the game, by the mathematician Albert Tucker to a group of psychology students at Stanford University, is as follows [198]. Two men, charged

¹As Hardin clarifies in [97], “tragedy” should not be seen as a state of unhappiness, but as “the solemnity of the remorseless working of things”. He further explains that “this inevitableness of destiny can only be illustrated in terms of human life by incidents which in fact involve unhappiness.”

with joint violation of a law, are held separately by the police. The police does not have conclusive incriminating evidence, hence it seeks confessions from the men. Each is told that (i) if he confesses to the crime, he will be given a reward of one unit, while the other will be fined two units, and (ii) if both confess each will be fined one unit. At the same time each man has good reasons to believe that if neither confesses both will go clear. The normal form of this game is shown in Table 2.1a

	Prisoner 2			Prisoner 2	
	Confess	Not confess		Cooperate	Defect
Confess	-1, -1	1, -2	Cooperate	3, 3	0, 5
Not confess	-2, 1	0, 0	Defect	5, 0	1, 1
	(a) Original formulation			(b) Modern version	

Table 2.1: The normal form of the Prisoner’s dilemma game, as originally presented by Albert Tucker (a), and its modern formulation (b).

If each man is rational and follows his immediate self-interest, the pure strategy “confess” dominates the pure strategy “not confess” – it maximises the payoff regardless of what the other does. This creates a unique equilibrium given by both men choosing to confess. From the perspective of the prisoners, this is a non-cooperative solution². If both had formed a coalition binding each other to not confess, their individual payoffs would have been maximised. Modern versions of the game (Table 2.1b) generalise the setting by e.g. explicitly defining cooperative and defective behaviour or allowing for repeated interactions. These refinements are used to study real-life problems in various fields, such as environmental science [67, 235], psychology [4, 11, 146], economics [7, 36, 51], animal behaviour [45, 239], etc.

Promoting cooperation Clearly, cooperation is a crucial ingredient in the evolution and sustainability of any biological and social system. At the most fundamental level, multi-cell organisms would not have developed were it not for cooperating single cells. Humans survived because they engaged in group-living, which implies foregoing individual interests in the name of the collective (e.g. sharing of food, cooperative hunting, etc.). Yet, natural selection with its maxim – “survival of the fittest” – favours competition [256]. Therefore, the evolution of life, in its essence, is a complex interplay between these two extreme behaviours³. Consequently, the standard game-theoretical formulation

²Of course, the opposite is true from the perspective of the police.

³There are attempts to explain seemingly irrational cooperative behaviour as “masked” competition. The most prominent is the selfish gene theory proposed by Richard Dawkins [61]. According to it, natural

of selfish agents pursuing their self interest is only an approximation for studying social dilemmas. Humans do, in fact, cooperate to some degree, in contrast to the expected ubiquitous defection [69]. At the same time, in many situations, the degree of cooperation is insufficient to prevent the tragedy of the commons, especially when the scope and scale of the problem increases. As a result, a large body of literature has expanded the canonical Prisoner’s dilemma framework to account for observed cooperative behaviour, and to propose mechanisms by which to promote it further. These works are summarised below.

- *changes of the payoff structure*: lowering the costs of cooperation to make it more attractive in the first place as a form of “buying cooperation” [221]
- *extension of the time horizon*: considering either repeated interactions, memory for the strategy of the counterparts, calculating payoffs over a longer time interval or anticipating opponents’ responses to one’s own actions [13]
- *considering spatial interaction*: the threshold for the outbreak of cooperation is lowered if individual interactions are constrained to nearest or second-nearest neighbours (as opposed to randomly chosen partners), or if individuals can migrate between different spatial domains, and thus avoid exploitation [220, 247]
- *co-evolution of strategies and interaction*: individuals can adopt the strategy of their best-performing neighbour and are allowed to break links with defecting partners [275]
- *introduction of additional strategies*: instead of unconditional cooperation/defection, the set of available strategies can be extended to include retaliation against defectors, e.g. TIT-FOR-TAT, or voluntary participation. [12, 100, 109]

Particularly for biological systems, other additional mechanisms have been considered [184], such as altruism, the role of kinship relations, selection mechanisms on the group level, etc.

Below, we present a new mechanism for promoting cooperation – *social herding*. Social herding is subtle, as it is an evolved behaviour in the presence of limited information [200]. At a fundamental level, herding provides safety as it increases the effective vigilance of the individual (e.g., herds can confuse or intimidate a predator or can more effectively find

selection acts on the level of genes, i.e. it transcends the individual organism. From this perspective, cooperation is rational if it promotes gene replication globally, even at the expense of individual well-being.

cover where none exists), helps in finding more fertile feeding patches and brings together prospective mates [172, 264]. From a more modern perspective, herding behaviour has been shown to be a factor in economic decisions and in the dynamics of financial markets [22, 27, 50]. It has further been proposed that in the absence of pertinent information, the rational decision is to imitate the behaviour of others in one’s social neighbourhood [21].

2.2 A model of social herding

The basic premise of our model is that evaluating the benefit of a certain strategy is not based on selfish considerations alone, but also considers awareness of the behaviour of others, without knowledge of how beneficial this behaviour is. In a game-theoretical setting this means that the complete payoff structure is unknown, as the payoffs obtained by one’s interaction neighbours are private. Agents, however, possess knowledge on the frequency of strategies either globally, or in their interaction neighbourhood, and choose their own strategy based on a combination of payoff-related and frequency-related information. In our model, we assume that the influence of these two components is weighed by a parameter ζ , which represents the level of social herding. In particular, $\zeta \rightarrow 0$ results in purely payoff-driven decisions, $\zeta \rightarrow 1$ in pure social herding.

2.2.1 Basic Model

Social herding has already been discussed as part of decision-making in binary opinion dynamics models⁴ [39, 114, 176, 222, 225, 248, 252, 258], where opinions are not necessarily related to payoffs but rather to social norms. From this perspective, agents may adopt the opinion of a majority in order to minimise social conflicts, but a utility-based preference for either of these opinions is not explicitly included. The so called *linear voter model*, where the probability to choose a particular opinion is directly proportional to its frequency, is a prominent example of this approach. For homogeneous networks and finite system sizes it is known to result in consensus, i.e. the existence of only one opinion, asymptotically, but the dominating opinion cannot be determined a-priori under general conditions.

To illustrate this point, consider a population of agents. Each agent, i , must choose a binary opinion, $\theta \in \{0, 1\}$, which could represent a choice between cooperation and defection. For the linear voter model, starting e.g. with the frequency f_1 of opinion $\theta = 1$ and $f_0 = 1 - f_1$ of opinion $\theta = 0$, the probability that the final consensus state is $\theta_i = 1$,

⁴In these models, agents choose between one of two possible opinions, typically quantified as 0 or 1.

for all i , is f_1 [112].

Hence, a simple majority rule of social herding, as expressed in the linear voter model, may not promote cooperation at all, if the initial conditions favour defection. Moreover, if the network of interactions is heterogeneous, e.g. in a complex network where agents have different degrees, the probability to reach each of the ordered configurations is given by the initial fractions weighted by the degree [243, 269]. Furthermore, the network topology can slow down the ordering dynamics and may even lead to a disordered system where no consensus is reached in the thermodynamic limit [234].

We, thus, turn to the class of *nonlinear voter models* [219]. As we show analytically, arbitrary choice for the non-linearity and the level of social herding does not lead to a transition towards cooperation. Instead, what is needed is a minimum level of social herding combined with the “right” non-linearity.

What do we gain from combining the standard game-theoretic strategic behaviour and a social herding mechanism? First, it is the realisation that less pertinent information for an informed decision can have positive unexpected effects for the population. The intuition is that by withholding information about the actual payoffs of others, a system planner “forces” individuals to resort to an evolved behaviour, which under the right conditions spreads cooperation. Moreover, herding is by nature indiscriminate, i.e. it does not favour a-priori a particular behaviour. Therefore, we are not forcing a desired outcome into the model design.

This realisation is surprising, as common wisdom would suggest that it is always better to have more information, e.g. to choose among more alternatives, to determine their consequences in advance, to know all payoffs and thus to *reduce the risk* associated with making the wrong individual decision. For example, we will show in Chapter 3, that in experiments on the *wisdom of crowd* effect, more information about the guesses of others, combined with social influence, leads to worse predictions [157]. Also, in a network formation model of agents sharing knowledge, it was shown that *best response*, i.e. the choice of partners based on knowing all alternatives, resulted in a worse global performance as compared to a situation where only the immediate best partner was considered [140]. As we point out in the analysis below, escaping the trap of defection crucially depends on using less of the available information, or on having a considerable fraction of less informed agents.

Second, from our insights we can derive mechanisms to improve the outcome in systems of strategically interacting agents. Mechanism design can be seen as the engineering part of economics. It allows to propose rules, or algorithms, for interactions that prevent the system from getting trapped in suboptimal states. Some of these algorithms, such as the

nowadays famous “Gale-Shapley” algorithm [76], are basically related to combinatorial optimisation problems. They propose a solution *for* the agents *without* involving the agents in finding it on their own. On the other hand, agent-based models (see also Section 1.2) aim at proposing new ways of *interaction* at the individual level, in order to arrive at more favourable solutions at the collective level. Our model gives a lucid example of this kind of systems design, by proposing a different way of combining information an individual agent already has. This still leaves room for the forces of self-organisation to act, but restricts the possible negative outcomes.

Combining social herding and strategic interaction. Consider a system with N agents. Each agent $i \in 1 \dots N$ is characterised by two individual variables which may change over time: $\theta_i(t)$ describes an agent’s strategic behaviour when interacting with other agents, whereas $\zeta_i(t)$ quantifies how much an agent is prone to social influence. We adopt the definition of *social influence* as the psychological tendency of individuals to adhere to and behave according to the expectations of their local neighbourhood [126]. In this sense, our approach belongs to a wider class of models which do not restrict herding behaviour to perfectly rational agents [191].

In a game-theoretic context, θ_i refers to the strategy of a utility maximising agent, chosen from a (discrete) set σ of possible strategies. We use the standard game theoretical setting of the Prisoner’s Dilemma, i.e. $\sigma \in \{0, 1\}$, where the strategic behaviour $\sigma = 0$ refers to defection (D) and $\sigma = 1$ to cooperation (C).

We assume that each agent plays a 2-person single-shot game with n other agents from its neighbourhood. The completion of these n games is called a round. From each of these interactions the agent receives a payoff which depends both on the strategic behaviour of the agent itself and on the behaviour of the opponents. The game structure describing a single interaction between two agents can be summarised by the standard payoff matrix of a 2-person game:

$$\begin{array}{c|cc} & \theta_j = 1 & \theta_j = 0 \\ \hline \theta_i = 1 & R/R & S/T \\ \theta_i = 0 & T/S & P/P \end{array}$$

Suppose, agent i has chosen to cooperate. Its payoff is R if the counter-party j has also chosen to cooperate (without knowing about the decision of agent i), but S if j defects. On the other hand, if agent i has chosen to defect, then it will receive the payoff T if agent j cooperates, while it will receive P if agent j defects.

While our discussion is restricted to the Prisoner’s Dilemma, we note that the approach can be extended to other games that result from different values of R , S , T and P [221].

For the particular case of the Prisoner's Dilemma, the payoffs have to fulfil the following inequalities

$$T > R > P > S ; \quad 2R > S + T \quad (2.1)$$

and we assume the standard values: $T = 5$, $R = 3$, $P = 1$, $S = 0$ (see also Table 2.1b).

Let us define the degree of cooperation on the collective level by the total number of cooperating agents, $N_1(t)$ relative to the total population N . Since the number of agents is constant, the global frequencies f_σ of cooperating and defecting agents are given by

$$\begin{aligned} N &= \sum_{\sigma} N_{\sigma} = N_0 + N_1 = \text{const.} ; \quad \sigma \in \{0, 1\}, \\ f_{\sigma} &= \frac{N_{\sigma}}{N} ; \quad f \equiv f_1 = 1 - f_0. \end{aligned} \quad (2.2)$$

In the following, we the variable f stands for the global frequency of cooperators.

The interaction of each agent with n other agents in a 2-person game results in $\binom{N}{n}$ different possibilities to choose a partner. As a result of these interactions that may occur *independently*, but *simultaneously* [101, 221], agent i receives a total payoff $A_i(\theta_i)$ which depends both on its own strategy θ_i and the strategies of the n different partners. Let us assume that n_0 of these partners have chosen to defect, whereas $n_1 = n - n_0$ partners have chosen to cooperate. The total payoff from these n interactions reads

$$A_i(\theta_i) = \delta_{1,\theta_i} [n_1 R + n_0 S] + \delta_{0,\theta_i} [n_1 T + n_0 P], \quad (2.3)$$

where $\delta_{x,y}$ is the Kronecker delta, which is 1 only for $x = y$ and 0 otherwise. Dividing by n gives the scaled total payoff

$$a_i(\theta_i, f_i) = \frac{A_i(\theta_i)}{n} = \delta_{1,\theta_i} [f_i R + (1 - f_i) S] + \delta_{0,\theta_i} [f_i T + (1 - f_i) P], \quad (2.4)$$

where $f_i = n_1/n = 1 - n_0/n$ gives the fraction of cooperators agent i interacts with. Assuming e.g. that agent i interacts with its neighbours, f_i gives the *local* frequency of cooperators. If, on the other hand, agent i interacts with n randomly chosen agents, the probability to choose a cooperator is directly proportional to the global fraction f . In the so-called *mean-field approach* we set $f_i \equiv f$.

Strategic considerations imply that agent i 's payoffs are not determined solely from its own strategy θ_i , but also incorporate the scaled payoff $a_i(\theta_i, f_i)$ expected from the interaction with f_i cooperators. A nonlinear function $\mathcal{G}(a_i)$ shall consider the way agent i combines the information about the possible payoffs $a_i(\theta_i, f_i)$ and $a_i(1 - \theta_i, f_i)$ resulting from its

strategic choice. In a very general way, we assume the following form for $\mathcal{G}(a_i)$

$$\mathcal{G}(a_i) = \frac{\exp[\beta_i a_i(\theta_i, f_i)]}{\exp[\beta_i a_i(\theta_i, f_i)] + \exp[\beta_i a_i(1 - \theta_i, f_i)]}. \quad (2.5)$$

Eq. 2.5 has the form of a logit-function well established in decision theory [24, 173, 253]. The parameter β_i allows agents to individually weigh differences between the payoffs. $\beta_i \rightarrow 0$ represents the limit of random choice, $\mathcal{G}(a_i) \rightarrow 1/2$, whereas $\beta_i \rightarrow \infty$ means that even small differences in payoffs lead to an immediate switch in strategies. For small values of β_i , $\mathcal{G}(a_i)$ approaches 1, if the expected payoff times the $a_i(\theta_i, f_i)$ from strategy θ_i is much larger than the expected payoff $a_i(1 - \theta_i, f_i)$ from the opposite strategy $1 - \theta_i$, and it tends to zero in the opposite case. If both payoffs become comparable, $\mathcal{G}(a_i)$ is about 1/2. Intermediate values of β_i allow for a smooth transition between the two strategic cases.

We note that for sufficiently small values of β_i , Eq. 2.5 can be approximated by the linear function

$$\mathcal{G}(a_i) \approx \frac{1}{2} \left[1 + \frac{\beta_i}{2} \{a_i(\theta_i, f_i) - a_i(1 - \theta_i, f_i)\} \right], \quad (2.6)$$

i.e. agents pay attention to the *difference* between the two possible payoffs.

The situation becomes different if agents are unable to calculate the expected payoffs. In this case, our model assumes that they would rather pay attention to the actions of the majority and would tend to imitate without knowing the full payoff consequences. Thus, agents only respond to frequency information in a similar manner as in Eq. 2.5

$$\mathcal{F}(f_{\theta_i}) = \frac{\exp[2\beta_i \kappa_i(f_{\theta_i}) f_{\theta_i}]}{\exp[2\beta_i \kappa_i(f_{\theta_i}) f_{\theta_i}] + \exp[-2\beta_i \kappa_i(f_{\theta_i}) f_{\theta_i}]} - \frac{1}{2}. \quad (2.7)$$

f_{θ_i} describes the local frequency of agents playing strategy θ_i in the neighbourhood of agent i , and $f_{1-\theta_i} = 1 - f_{\theta_i}$ is the local frequency of agents playing the opposite strategy. With both frequencies being equal, $\mathcal{F}(f_{\theta_i}) = \mathcal{F}(f_{1-\theta_i}) = 1/2$. Again, for sufficiently small β_i a linear approximation of Eq. 2.7 yields,

$$\mathcal{F}(f_{\theta_i}) \approx \beta_i \kappa_i(f_{\theta_i}) f_{\theta_i}. \quad (2.8)$$

$\kappa_i(f_{\theta_i})$ is a nonlinear response function that considers the weighted influence of the frequency. This function may be the result of a particular incentive scheme or may depend on the time an agent has kept its current strategy or opinion [236, 237]. We emphasise that for the so-called linear voter model, $\kappa_i(f_{\theta_i})$ is simply a constant κ that does not depend on the frequency. So $\beta_i \kappa$ can be scaled to one, which means that for the linear voter model

we simply arrive at $\mathcal{F}(f_{\theta_i}) = f_{\theta_i}$. Thus, the response of agent i is directly proportional to the local frequency of agents playing strategy θ_i .

After having defined the agent's response to both payoff considerations and majority influence, we use the individual parameter ζ_i to weigh these two different influences. Specifically, we define the transition rate for agent i to switch from strategy $(1-\theta_i)$ to the opposite strategy θ_i as follows

$$w(\theta_i|(1-\theta_i), f_i, \zeta_i) = (1 - \zeta_i) \mathcal{G}(a_i) + \zeta_i \mathcal{F}(f_{\theta_i}). \quad (2.9)$$

For $\zeta_i \rightarrow 0$, we recover the limit case of strategic interactions in the standard setting of the Prisoner's dilemma. For $\zeta_i \rightarrow 1$, we arrive at the limit case of pure social herding, i.e. imitation behaviour without awareness of the resulting consequences.

Specifying the transition rates. Before describing the system's dynamics by means of a master equation, it is useful to specify the transition rates of Eq. 2.9 more precisely. The transition rates apply for a *frequency dependent* process, i.e. they do not depend on the specific sequence of interactions. In our model, we fix the number of independent, but simultaneous 2-person games to $n = 4$. This is convenient for comparing random and local interactions on a regular lattice. Hence, the relevant frequencies have only discrete values $f_i \equiv k_i/n$ where $k_i \equiv n_1 = 0, 1, 2, 3, 4$ is the actual number of cooperators agent i is interacting with. On the other hand, random interactions can be approximated by the so-called mean field approximation, where $f_i = f$ is the global fraction of cooperators.

Dropping the individual index i for the moment, we have to distinguish between two different transition rates, $c_k(\zeta) = w(1|0, k, \zeta)$, the transition from defection to cooperation dependent on k cooperating agents, and $d_k(\zeta) = w(0|1, k, \zeta)$, the transition from cooperation to defection under the same conditions. Both of these rates are comprised of two parts, one resulting from strategic behaviour (\tilde{c}_k, \tilde{d}_k), the other one resulting from social herding (\hat{c}_k, \hat{d}_k)

$$c_k(\zeta) = (1 - \zeta)\tilde{c}_k + \zeta\hat{c}_k; \quad d_k(\zeta) = (1 - \zeta)\tilde{d}_k + \zeta\hat{d}_k. \quad (2.10)$$

Since, strategically it always pays off to be a defector, we can simplify the above equation by noting that $\tilde{c}_k = 0$ and $\tilde{d}_k = 1$ for all k . Hence

$$c_k(\zeta) = \zeta\hat{c}_k; \quad d_k(\zeta) = 1 - \zeta + \zeta\hat{d}_k. \quad (2.11)$$

As an illustration, the social herding components, (\hat{c}_k, \hat{d}_k) , for the linear voter model, with

$n = 4$, are shown below and illustrated in Figure 2.1.

$$\hat{c}_k = \frac{k}{4}\beta\kappa_k; \quad \hat{d}_k = 1 - \hat{c}_k = 1 - \frac{k}{4}\beta\kappa_k. \quad (2.12)$$

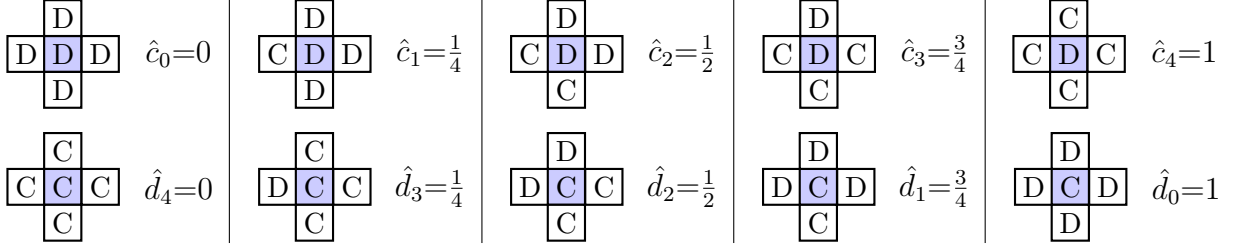


Figure 2.1: The social herding components, \hat{c}_k and \hat{d}_k dependent on the number of cooperators in one’s neighbourhood, k . Coloured squares represent the focal individual. Note that full cooperation and full defection are absorbing states, as the probability to switch to alternative strategies is 0.

In order to use nonlinearities in the frequency response, it is convenient to specify the components \hat{c}_k and \hat{d}_k by discrete values $\alpha_0, \alpha_1, \alpha_2$ as shown in Table 2.13

$f = k/n$	\tilde{c}_k	\tilde{d}_k	\hat{c}_k	\hat{d}_k
0	0	1	α_0	$1 - \alpha_0$
1/4	0	1	α_1	$1 - \alpha_1$
2/4	0	1	α_2	$1 - \alpha_2$
3/4	0	1	$1 - \alpha_1$	α_1
1	0	1	$1 - \alpha_0$	α_0

(2.13)

The parameter α_0 describes the transition of a cooperator (defector) towards defection (cooperation) if surrounded by cooperators (defectors) solely based on *social herding*. Because agents with opposing strategies are absent in the neighbourhood, α_0 should be consequently zero, even if there is a strong *strategic* incentive for a cooperator to switch to defection if surrounded by cooperators. Hence, considering only social herding, pure cooperation and pure defection are “absorbing” states for the dynamics of the system. This can be avoided by choosing $\alpha_0 = \varepsilon$, a very small value that allows for occasional random changes in strategies, due to e.g. noise [219], but here we fix $\alpha_0 = 0$.

Possible combinations of (α_1, α_2) define a parameter space to distinguish between different forms of social herding, as shown in Fig. 2.2 (left). Positive frequency dependence (pf) means that the probability to change to the opposite strategy monotonously increases with the frequency of that strategy in the local neighbourhood, also known as “majority voting”. Negative frequency dependence (nf) means the opposite, i.e. the switching probability

monotonously decreases with the local frequency, also known as “minority voting”. On the other hand, “positive allee” (pa) and “negative allee” (na) define parameter regions with non-monotonous dependence. For example, (pa) means an increase of the probability as long as the opposite strategy is not the majority, also known as voting against the trend, while (na) describes constellations with a strong amplification of minority strategies. We note that the so-called “voter point” representing the linear voter model – where $\alpha_1 = 1/4$ and $\alpha_2 = 2\alpha_1 = 1/2$ are strictly proportional to k – is on the border between the (pf) and (pa) parameter regions. For our investigations, we will consider a scenario where the non-linearity is only represented by α_2 , whereas α_1 is chosen according to the linear voter model. Four possible cases which refer to the (pf), (pa), (na) and the linear voter model are shown in Fig. 2.2 (right).

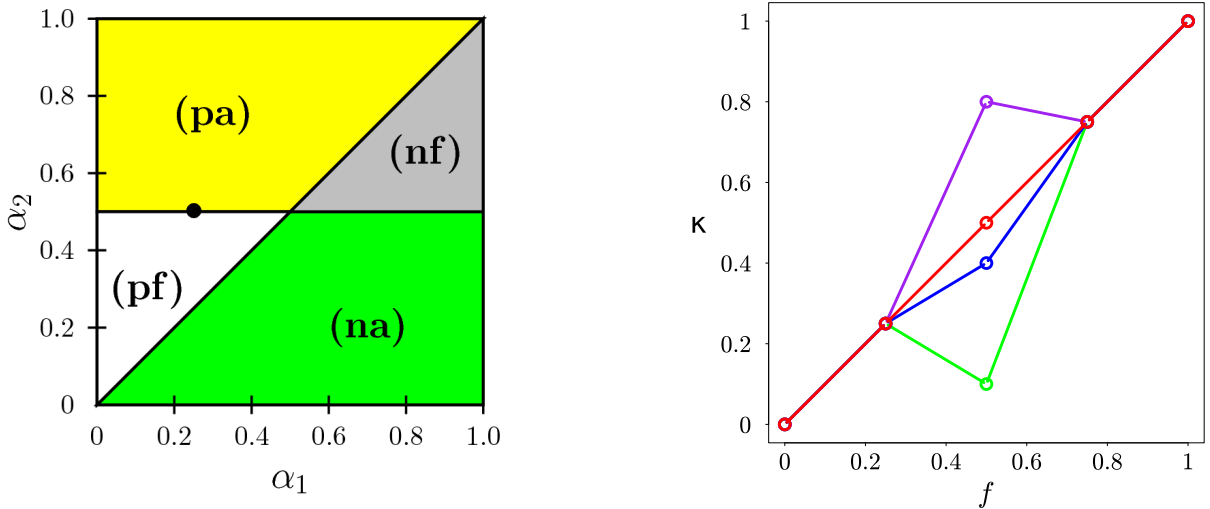


Figure 2.2: (left) Parameter space (α_1, α_2) to define the nonlinearity in social herding (see also Table 2.13). The different regions are explained in the text. We use the (pa) region, defined by Eq. (2.19). (right) Linear voter model (red line) and deviations controlled by α_2 at $f = 0.5$.

What remains to be specified are the payoff-related terms \tilde{c}_k and \tilde{d}_k that follow directly from Eq. 2.5. Here, we assume the deterministic limit $\beta_i \rightarrow 0$, for which we get $\mathcal{G}(a_i) = \Theta[a_i(\theta_i, f_i) - a_i(1 - \theta_i, f_i)]$ with $\Theta[y]$ being the Heavyside function, which is 1 if $y > 0$ and 0 otherwise. Hence, $\mathcal{G}(a_i)$ is either 1 or 0 dependent on whether the payoff from an alternative strategy is larger or smaller than the payoff from the current strategy. Taking into account the payoff relations, Eq. 2.1, we can easily verify that the expected payoffs in Eq. 2.4 for defectors, $a(0, f)$, are *always* larger than the corresponding ones for cooperators, $a(1, f)$, regardless of the local fraction of cooperators. This is the well-known result that in the standard non-repeated Prisoner’s dilemma defection is the evolutionary

stable strategy. Hence, in the deterministic limit of strategic interactions, we always have $\tilde{c}_k = 0$ and $\tilde{d}_k = 1$. From a systemic point of view, that can rightly be assumed as the worst-case scenario, because the system will always end up in pure defection. We wish to identify conditions where inducing social herding would allow the system to not only avoid this trap, but to also converge to cooperating majority.

The observant reader may have noticed that we interpreted β_i differently for social herding (where we merely assumed it was small) and for strategic interactions (where $\beta \rightarrow 0$). This is not a contradiction. As β quantifies randomness or noise, we can assume that the payoff-related attention is much higher and less prone to errors than the response to the behaviour of one's neighbours. In general, we may distinguish between $\tilde{\beta}_i$ and $\hat{\beta}_i$ for the different responses, but this complication is not applied here.

Dynamics of strategy change. So far, we have defined “rules” for agents to change their strategy, dependent on both strategic considerations and social herding. Most agent-based models, at this point, would continue with extensive computer simulations to probe the parameter space for some non-trivial results. We will certainly follow with computer simulations as well, however, we are also interested in some analytic insights that would allow us to predict the system's dynamics without testing every possible parameter combination. For this reason, we need to formally specify the dynamics of strategy change on two different levels – (a) on the micro level of the individual agent, and (b) on the macro level, describing the fraction of cooperators in the system.

For the micro level, we use a stochastic approach. Here, the variable of interest is the probability $p_i(\theta_i, t)$ that agent i chooses strategy θ_i at time t . As explained before, this probability depends on the strategies present in the local neighbourhood of agent i expressed by the vector $\underline{\theta}_i = \{\theta_{i_1}, \theta_{i_2}, \dots, \theta_{i_n}\}$. Hence, $p_i(\theta_i, t)$ is defined as the marginal distribution

$$p_i(\theta_i, t) = \sum_{\underline{\theta}'_i} p(\theta_i, \underline{\theta}'_i, t). \quad (2.14)$$

The summation is over all possible distributions $\underline{\theta}'_i$. Specific realisations of these distributions are denoted as $\underline{\sigma}$. For $n = 4$, there are 2^n possible realisations. For the time-dependent change of $p_i(\theta_i, t)$ we assume the following master equation

$$\frac{d}{dt} p_i(\theta_i, t) = \sum_{\underline{\theta}'_i} \left[w(\theta_i | (1 - \theta_i), \underline{\theta}'_i) p(1 - \theta_i, \underline{\theta}'_i, t) - w(1 - \theta_i | \theta_i, \underline{\theta}'_i) p(\theta_i, \underline{\theta}'_i, t) \right]. \quad (2.15)$$

This equation considers all possible processes that may lead to an increase or decrease in the probability that agent i uses strategy θ_i given the neighbourhood distribution $\underline{\theta}_i$,

with the transition rates $w(\theta_i|(1-\theta_i), \underline{\theta}'_i)$, $w(1-\theta_i|\theta_i, \underline{\theta}_i)$. Note that these are not the transition rates defined in Eq. 2.9, which depend only on the local frequency f_i , but not on the neighbourhood distribution $\underline{\theta}_i$. In order to map the two, we have to consider how many specific realisations of the distribution $\underline{\theta}_i$ may lead to the same f_i . Taking as example $\underline{\sigma} = \{0010\}$, there are exactly $\binom{4}{1}$ different possibilities to realise $f_i = 1/n$. Hence, transforming the master equation Eq. 2.15 that depends on the neighbourhood distribution $\underline{\theta}_i$ into one that only contains the respective local frequency f_i results in a combinatorial prefactor of $\binom{n}{k}$. Using again the specific notations c_k and d_k for the transition rates from Eq. 2.10, we can rewrite the master equation 2.15 as

$$\frac{d}{dt}p_i(1, \zeta, t) = \sum_{k=0}^n \binom{n}{k} \left[c_k(\zeta) p(0, k/n, \zeta, t) - d_k(\zeta) p(1, k/n, \zeta, t) \right]. \quad (2.16)$$

The corresponding master equation for $p_i(0, \zeta, t) = 1 - p_i(1, \zeta, t)$ follows likewise. Note that in Eq. 2.16, we have chosen the individual parameter ζ_i to be a constant ζ . While the local frequency $f_i = k/n$ changes over time due to concurrent strategy choices of neighbouring agents, ζ is assumed to be a global control parameter the impact of which will be discussed together with the computer simulations.

With this, we have a bottom-up description of the collective dynamics given by N stochastic equations, Eq. 2.16, which are coupled, because agents' neighbourhoods overlap.

For the macro level, we have to deal with the probability $P(f, \zeta, t)$ of finding a given fraction of cooperators, f , at time t , assuming the social herding factor ζ . The dynamics can again be specified by a stochastic equation

$$\frac{d}{dt}P(f, \zeta, t) = \sum_{f'} \left[W(f|f', \zeta) P(f', \zeta, t) - W(f'|f, \zeta) P(f, \zeta, t) \right]. \quad (2.17)$$

f' denotes all possible deviations from a given value f that can be reached during one time step by means of the transition rates $W(f'|f, \zeta)$. These are not identical with the individual transition rates in Eq. 2.9, but are aggregated rates that take into account all possible ways to change f . The smallest change of $f \equiv N_1/N$ in Eq. 2.2 is the addition or subtraction of a single cooperator, i.e. $f' \in \{(N_1 + 1)/N; (N_1 - 1)/N\}$. The individual equivalent for such process is given by Eq. 2.10, where the terms $c_k(\zeta)$ describe the transition of a single defector into a cooperator, and the $d_k(\zeta)$ the opposite transition. Hence, we find for the

aggregated transition rates

$$\begin{aligned}
W(f + 1/N|f, \zeta) &\equiv W_+(f, \zeta) = \sum_{k=0}^n \binom{n}{k} f^k (1-f)^{n-k} c_k(\zeta) \\
W(f - 1/N|f, \zeta) &\equiv W_-(f, \zeta) = \sum_{k=0}^n \binom{n}{k} f^k (1-f)^{n-k} d_k(\zeta).
\end{aligned} \tag{2.18}$$

The combinatorial prefactors preceding $c_k(\zeta)$ and $d_k(\zeta)$ result from the various ways to choose agents with $n = 4$ neighbours, k of which are cooperators, given the global fraction of cooperators f . Here, we use the so-called mean-field approach that replaces the frequencies f_i of the individual neighbourhoods by the global value f . With the specific values for $c_k(\zeta)$ and $d_k(\zeta)$ given by Eqs. 2.11 and 2.13, the dynamics on the systemic level is also completely specified. In the following, we will use the dynamics on the micro level for carrying out computer simulations, while the dynamics on the macro level will be used for analytic investigations.

2.2.2 Results of Computer Simulations

We use the dynamics specified in Eq. 2.16 to run agent-based computer simulations for different sets of parameters. According to Eqs. 2.10) and 2.13, we only need to vary the weight $0 \leq \zeta \leq 1$ and the parameters $0 \leq (\alpha_1, \alpha_2) \leq 1$ characterising the social herding behaviour. Regarding strategic decisions, everything is already defined, and with $\tilde{c}_k = 0$, $\tilde{d}_k = 1$ defection remains the only choice. This “worst case scenario” can only be changed by a considerable amount of social herding, in which agents copy the strategies of their neighbours, regardless of the payoffs assigned to them. This is shown in Figure 2.3. Below a critical level for social herding, $\zeta \approx 0.7$, only defection remains. For $\zeta > 0.7$ we observe different levels of long-term cooperation that depend on the combination of ζ and α_2 . If $\zeta > 0.8$, cooperation becomes a majority, i.e. $f > 0.5$, but only for large values of ζ and α_2 full cooperation, $f \rightarrow 1$, is achieved. This issue is further investigated below.

The role of the non-linearity in social herding, expressed in terms of α_1 , α_2 , is further investigated in Figure 2.4, given a supercritical level of social herding. We see that there is an *optimal non-linearity* that enhances cooperation – α_1 , α_2 have to be chosen such that they belong to the area of (pa) effects. This area is defined by the following inequalities (see also Eq. 2.13),

$$0 \leq \alpha_1 \leq \alpha_2; (1 - \alpha_1) \leq \alpha_2 \leq 1. \tag{2.19}$$

It describes a response where the transition toward a given strategy *increases* with the

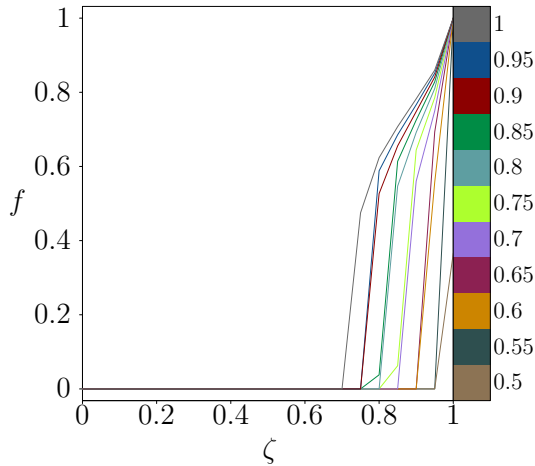


Figure 2.3: Global fraction of cooperation f dependent on the level of social herding ζ . $\alpha_1 = 0.25$ is fixed, α_2 varies between 0.5 and 1.0 according to the color scale. System size $N=400$.

frequency of that strategy as long as that strategy is *not* the majority, i.e. minority strategies are eventually favoured. A special case where α_1 is taken from the linear voter model, whereas α_2 is larger than 0.5 is shown in Figure 2.2 (right). We note in particular that social herding according to the *linear* voter model will *not* allow the transition toward cooperation (more on this in the analytic results in the next section). Further, all forms of the transition rates that *monotonously* increase with the frequency, indicated by the (pf) area, will *not* lead to cooperation. Social herding in this case only amplifies defection.

In addition, to estimate the impact of the *system size*, N , on the fraction of cooperators, we present two cuts through the parameter space of α_1 , α_2 in Figure 2.5, i.e. α_1 varies while α_2 is fixed to either a high or a low value. In Figure 2.5 the fraction of cooperators is shown for four different values of N . Differences in the symbols are hardly noticeable, which means that f is largely independent of N for system sizes of 100 and larger. Finite size effects on the fraction of cooperation play a role only for rather small systems ($N < 50$).

Assuming the right choice of parameters for the transition to cooperation, in Figure 2.6 we investigate how the dynamics evolve in space. We choose a two-dimensional regular lattice with Von-Neumann neighbourhood, where each agent interacts with $n = 4$ immediate neighbours. First, let us assume a small cluster of cooperating agents, as shown on the left-most snapshot of Figure 2.6. *Without* social herding, this cluster would immediately disappear in the next time step, because all agents will choose defection, as this is the rational choice that maximises payoffs. Instead, we observe an invasion of the cooperating strategy into the domain of defectors. The cooperating agents, however, do not form

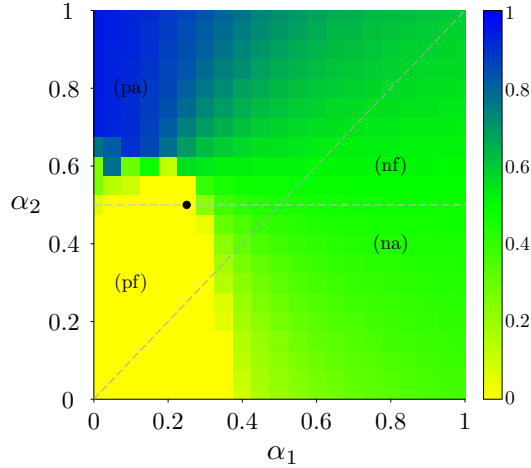


Figure 2.4: Fraction of cooperation (color scale) dependent on the nonlinearities in social herding, defined by α_1 , α_2 . Fixed level of social herding $\zeta=0.95$. The four different areas are defined in Fig. 2.2(left). \bullet indicates the linear voter model. System size $N=400$.

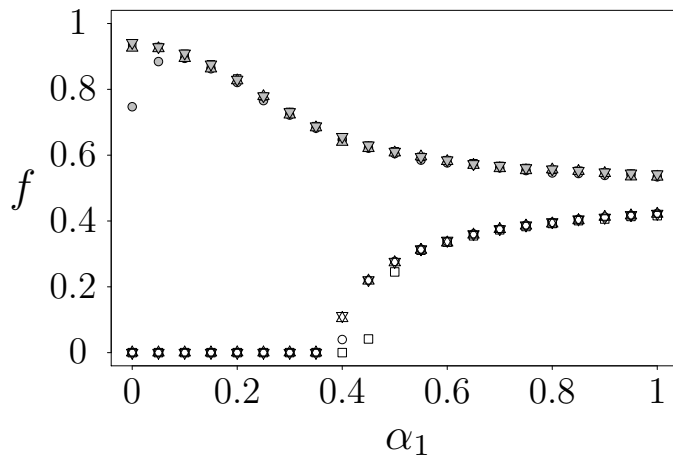


Figure 2.5: Fraction of cooperation, f dependent on the nonlinearity α_1 , with fixed $\alpha_2=0.75$ (upper curve, filled symbols) and $\alpha_2=0.25$ (lower curve, empty symbols). Different system sizes: $N=100$ (\square), $N=400$ (\circ), $N=1000$ (\triangle), $N=10000$ (∇). The fact that they are almost indistinguishable indicates that N does not have an impact. Other parameters: $\zeta=0.95$

compact clusters. A minority fraction of defectors will always survive and their spatial distribution in small clusters across the domain of cooperators continues to change in time. Therefore, the system never reaches a stationary state in space. The global fraction of both strategies, on the other hand, does reach an equilibrium on average.

We further note that there is a critical size for the initial cluster of cooperators to grow.

This has been already discussed in detail for the pure Prisoner’s dilemma on a regular lattice [101, 221], and in opinion dynamics models [252]. Now, the addition of supercritical social herding reduces these requirements. It is worth mentioning that, starting from random initial conditions in a spatially extended system, we find that a vanishingly small initial density of cooperators is enough to trigger convergence to the final state. The reason is that, provided the system is large enough one cluster of cooperators, larger than the critical size, will appear by chance. This cluster will be sufficient to trigger the outbreak of cooperation. Here, however, we do not pursue further this discussion. Instead, the initial conditions and parameter constellations for the *outbreak of cooperation* are further discussed for the mean-field case, in the next section.

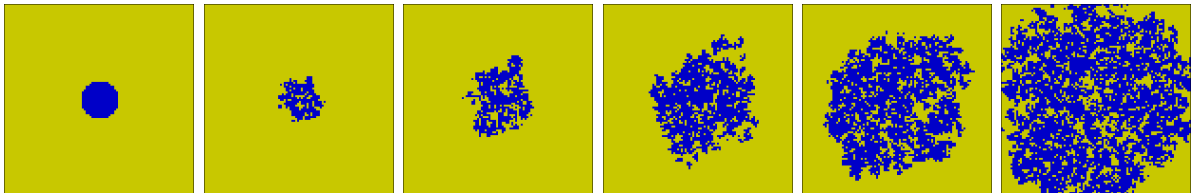


Figure 2.6: Snapshots of the transition toward cooperation at times $t=0, 10, 20, 50, 150, 500$. $N=10^4$ agents are placed on a regular lattice and interact each with their $n=4$ spatial neighbours. Dark colour (blue) indicates cooperators, light colour (yellow) defectors. Parameters $\alpha_1 = 0.25$, $\alpha_2 = 0.7$, $\zeta = 0.95$.

2.2.3 Mean-field investigations

Calculating the effort We showed by means of computer simulations that there is indeed a way of using social herding to boost cooperation. Now, let us illustrate this finding by some analytic considerations. As a first step, we wish to calculate the “effort” required to transfer the system into a majority of cooperators. Informally, effort should be understood as the amount of resources that a system planner needs to allocate, e.g. in terms of incentive schemes, punishment mechanisms, costs for information restriction, etc., to establish cooperative behaviour in a considerable fraction of the population. Considering only the strategic dimension, this effort should be very high because there is a strong incentive to defect. On the other hand, inducing social herding may help since it mitigates the perceived payoff differences between defection and cooperation.

A formal approach to calculate the effort starts from the master equation (Eq. 2.17) on the systemic level, in the mean-field limit. The detailed balance condition, which is a specific form of the equilibrium condition $dP(f, t)/dt = 0$, requires that the net probability fluxes

are balanced, i.e.

$$W(f|f - 1/N, \zeta) P^0(f - 1/N, \zeta) = W(f - 1/N|f, \zeta) P^0(f, \zeta), \quad (2.20)$$

where $P^0(f, \zeta)$ denotes the equilibrium probability distribution which is independent of t . This equation is recursive and, using $f = N_1/N$, Eq. 2.2, can be re-formulated as:

$$P^0(f, \zeta) = P^0(0, \zeta) \prod_{i=1}^{N_1} \frac{W\left(\frac{i}{N} \mid \frac{i-1}{N}, \zeta\right)}{W\left(\frac{i-1}{N} \mid \frac{i}{N}, \zeta\right)}. \quad (2.21)$$

The normalisation constant $P^0(0, \zeta)$ can be found by enforcing $\sum_{i=0}^N P^0(i/N, \zeta) = 1$ and the transition rates are given by Eq. 2.18. We visualise the equilibrium probability distribution by means of a potential $\Omega(f, \zeta)$ that has its minimum where $P^0(f, \zeta)$ has its maximum, i.e. it represents the effort of reaching a given equilibrium state

$$P^0(f, \zeta) = \exp\{-\Omega(f, \zeta)\}, \quad (2.22)$$

where Ω is given by

$$\Omega(f, \zeta) = -\ln P^0(0, \zeta) - \sum_{i=1}^{N_1} \ln \left[\frac{W\left(\frac{i}{N} \mid \frac{i-1}{N}, \zeta\right)}{W\left(\frac{i-1}{N} \mid \frac{i}{N}, \zeta\right)} \right]. \quad (2.23)$$

Figure 2.7 shows the effort $\Omega(f, \zeta)$ as a function of the global fraction of cooperators f and the level of social herding ζ , which is our control parameter. We observe that for very low values of ζ the effort is a monotonously increasing function of the frequency f . Given a fraction of cooperators, $f = 0.2$, and small ζ , it becomes more and more difficult, or unlikely, to find a larger fraction of cooperators (red line). Considering instead a high level of social herding, e.g ζ about 0.85, there is a *monotonous decrease* of the effort with an increasing fraction of cooperators. I.e. starting from a supercritical level of social herding, the outbreak and the increase of cooperation becomes very likely (green line).

The observant reader will notice that for large ζ Figure 2.7 there is *nonmonotonous dependence* of the effort on the fraction of cooperators. There is a critical region around of $f \approx 0.2$ below which large-scale defection becomes the most probable state, regardless of the level of social herding. This relates to the critical cluster size of cooperators in Figure 2.6. However, there is a noticeable difference underlying both results. Figure 2.7 is based on mean-field approximation, i.e. there is no spatial correlation between interacting agents, whereas Figure 2.6 assumes a spatial neighbourhood defined by the regular lattice. In fact, it is known that spatial interaction enhances cooperation [209, 221, 223].

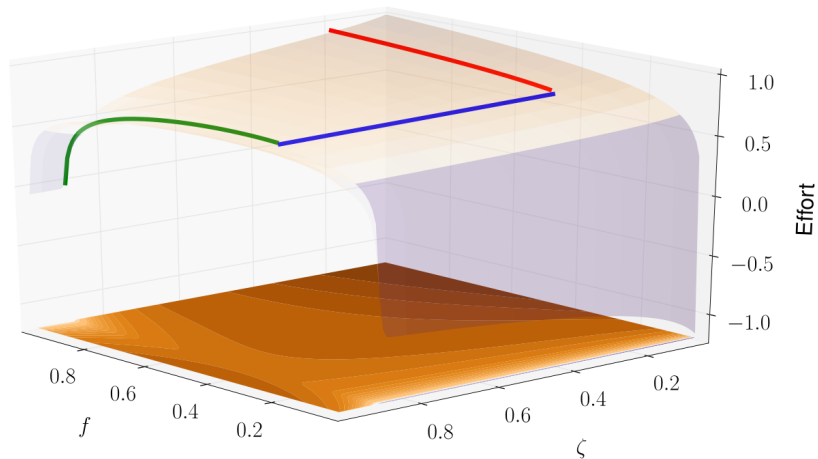


Figure 2.7: Effort $\Omega(f, \zeta)$, Eq. (2.23) dependent on the global fraction of cooperators f and the level of social herding ζ . The non-linearity is specified by $\alpha_1=0.25$, $\alpha_2=0.85$.

Already small, randomly formed clusters of cooperators are sufficient for the outbreak of cooperation, whereas random interaction results in a much larger threshold.

Competition dynamics Eventually, we can also derive a deterministic dynamics for the global fraction of cooperators, $f(t)$, in the mean-field limit. One approach is to start from the stochastic dynamics on the microscopic level, $p_i(\theta_i, t)$ in Eq. 2.15, as already shown in [219]. Alternatively, one can consider the stochastic dynamics on the macroscopic level, $P(f, \zeta, t)$ from Eq. 2.17, in which case the expected value for the global fraction of cooperators is given by

$$\langle f(\zeta, t) \rangle = \sum_{f'} P(f', \zeta, t), \quad (2.24)$$

where f' denote all possible realisations of f . Using the master equation Eq. 2.17, we arrive at the deterministic dynamics

$$\frac{d\langle f(\zeta, t) \rangle}{dt} = W_+(f, \zeta) (1 - \langle f \rangle) - W_-(f, \zeta) \langle f \rangle, \quad (2.25)$$

where the aggregated transition rates $W_+(f, \zeta)$, $W_-(f, \zeta)$ are given by Eq. 2.18. Assuming a narrow probability distribution in equilibrium, $P^0(f, \zeta)$, the expected value $\langle f^0(\zeta) \rangle$ can be approximated by the maxima of $P^0(f, \zeta)$. In particular, the deterministic dynamics

will converge to those areas where $P^0(f, \zeta)$ is largest, or where $\Omega(f, \zeta)$ has its minima (see Figure 2.7). While we do not argue about the specific global dynamics at intermediate times (which can be governed by stochastic influences, particularly in the early stages), we can see the late stage of the dynamics as a “quasi-stationary” motion along the valley in the potential landscape in Figure 2.7, provided ζ is large enough.

We can rewrite Eq. 2.25, which describes the “replication” of cooperators at the global scale, to a form that resembles the well known replicator dynamics from evolutionary game theory

$$\frac{d\langle f(\zeta, t) \rangle}{dt} = \langle f \rangle (1 - \langle f \rangle) \left[E_1(f, \zeta) - E_0(f, \zeta) \right]. \quad (2.26)$$

The two terms E_1 and E_0 are the “fitness” values associated with the two different strategies. The fraction of cooperation will grow if the fitness of cooperation $E_1(f, \zeta)$ is larger than the fitness of defection $E_0(f, \zeta)$. The fitness of both strategies depends on the global level of cooperation and the level of social herding

$$E_1(f, \zeta) = \frac{W_+(f, \zeta)}{f}; \quad E_0(f, \zeta) = \frac{W_-(f, \zeta)}{1 - f}. \quad (2.27)$$

To evaluate the fitness associated with a given strategy, one should consider the strictly nonlinear dependence of the transition rates on f (Eq. 2.18). Figure 2.8 shows the difference $E_1 - E_0$ in the whole range of f and ζ . We emphasise that this graph holds for fixed values of the non-linearity parameters α_1, α_2 . Hence it adds another dimension to Figure 2.4, which was obtained for a fixed herding level ζ . Additionally, Figure 2.8 clearly shows the influence of the initial fraction of cooperators, $f(0)$, in the mean-field case. Assuming a fixed value of $\zeta=0.85$, we see that the fraction of cooperators $f(t)$ can be increased in time only if $f(0)$ is between 0.15 and 0.6. While the lower bound has an intuitive meaning as the minimum threshold to start cooperation, the upper bound is less obvious. It results from the influence of the *nonlinear* social herding. We recall that social herding is indiscriminate, i.e it does not assume any “value” related to the strategies. Once cooperation becomes too prevalent, the fitness of defection increases with the amount of potential cooperators to exploit. Hence, for the example considered, the maximum fraction of cooperators is given by $f = 0.6$. A higher level of social herding, or different values for the nonlinearities, may increase this fraction up to about one, i.e. full cooperation.

Another way of expressing the dynamics of Eq. 2.26 is through

$$\frac{d\langle f(\zeta, t) \rangle}{dt} = \langle f(\zeta, y) \rangle (E_1 - \langle E \rangle); \quad \langle E \rangle = \sum_{\sigma} E_{\sigma} \langle f_{\sigma} \rangle = E_1 \langle f \rangle + E_0(1 - \langle f \rangle). \quad (2.28)$$

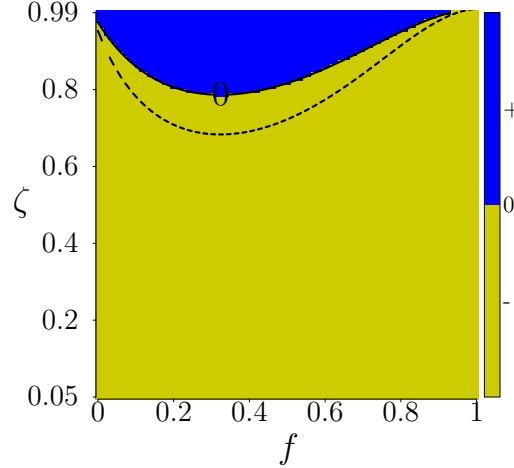


Figure 2.8: Difference of the fitness values $E_1(f, \zeta) - E_0(f, \zeta)$ dependent on the fraction of cooperators, $f \in [0.02, 0.99]$, and the level of social herding, $\zeta \in [0.05, 0.99]$. Dashed-line denotes the minimum level of social herding required for a given initial fraction, $f(0)$, such that the growth of cooperation starts. Contour line at 0 denotes the minimum level of social herding that ensures positive fraction of cooperators in the long term. Non-linearity parameters: $\alpha_1 = 0.25$, $\alpha_2 = 0.8$.

As long as E_1 is larger than the average fitness, $\langle E \rangle$, the fraction of cooperators in the system is able to grow, but one has to recognise that, because of the time dependence of $\langle f(t) \rangle$ and its implicit feedback on E_σ , $\langle E(t) \rangle$ evolves over time as well. Hence, Eq. 2.28 describes a nonlinear selection process for each of the strategies, dependent on the parameters describing strategic interaction and social herding.

For some special cases, we are able to derive closed form solutions of the competition dynamics expressed by Eqs. (2.25)-(2.28). In the absence of any social herding, $\zeta = 0$, we only need to consider the transition rates from strategic interaction, $\tilde{c}_k = 0$ and $\tilde{d}_k = 1$. This results in $E_1(f, \zeta = 0) = 0$ and $E_0(f, \zeta = 0) = 1$, i.e. $\langle f(t) \rangle = f(0) \exp(-t)$, which means that cooperation dies out, exponentially. In the opposite case, $\zeta = 1$, Eq. 2.26 can be solved for the case of the linear voter model (Eq. 2.12) for which $\hat{c}_k = k/4$ and $\hat{d}_k = 1 - (k/4)$. We then find $E_1(f, \zeta = 1) = E_0(f, \zeta = 1)$, that is the fitness of both “strategies” is the same since, for $\zeta = 1$, cooperation and defection represent mere labels void of any strategic meaning. The average fraction of cooperators is given by $\langle f(t) \rangle = f(0)$, i.e. the initial fraction of cooperators is conserved on average. This is known as one of the puzzles associated with the linear voter model – individual realisations of the dynamics, e.g. using stochastic simulations, always lead to convergence with $f \rightarrow 0$ or $f \rightarrow 1$, but averaging over many runs reveals that the frequency with which cooperators or defectors dominate

is equal to their initial fraction $f(0)$.

In the general case of $0 < \zeta < 1$, we can derive a closed-form solution for $\langle f \rangle(\zeta, t)$

$$\langle f(\zeta, t) \rangle = f(0) \exp[(\zeta - 1)t]. \quad (2.29)$$

which is similar to the purely strategic interaction case, except that the time scale for the extinction of cooperators is stretched by the factor $(\zeta - 1)$. This is an important result because it demonstrates that *linear* social herding will not prevent the extinction of cooperation, not even for large ζ . In order to turn defection into cooperation, we essentially need a high level of non-linearity in the herding mechanism, i.e. the right ζ and α_2 .

Considering a non-linearity where $\alpha_1 = 1/4$ but $\alpha_2 \neq 2/4$, we find from Eq. 2.25

$$\frac{d\langle f(\zeta, t) \rangle}{dt} = \langle f \rangle \left[\zeta \left[1 + 3 \langle f \rangle (1 - \langle f \rangle)^2 (2\alpha_2 - 1) \right] - 1 \right]. \quad (2.30)$$

For $\alpha_2 = 2/4$, the solution reduces to Eq. 2.29, whereas for $\zeta = 1$ we arrive at the mean-field equation for the nonlinear voter model only [219]. In order to make full cooperation, $\langle f \rangle = 1$, a stable fixed point, the following condition for α_2 has to be met

$$\frac{1}{2} + \frac{1 - \zeta}{6\zeta \langle f \rangle (1 - \langle f \rangle)^2} < \alpha_2 \leq 1, \quad (2.31)$$

which implies $1/[1 + 3 \langle f \rangle (1 - \langle f \rangle)^2] < \zeta < 1$. This inequality can be only satisfied for a considerable high level of social herding. The feasible range of (f, ζ) values that is consistent with a given value of α_2 , e.g. $\alpha_2=0.8$, is shown in Figure 2.8. The maximum range resulting from $\alpha_2 = 1$ is also shown in the same figure by the dashed line. We note again that, even if Eq. 2.31 is fulfilled, the dynamics does not necessarily converge to $f \rightarrow 1$. Dependent on the parameters $\{\zeta, \alpha_1, \alpha_2\}$ also lower equilibrium fractions of cooperators may be reached, in which case there will be *coexistence* of cooperators and defectors.

2.3 Conclusions

In this chapter, we have explored a new route towards solving the standard social dilemma of promoting cooperation. This route differs from many other attempts, most of which are rooted in traditional or evolutionary game theory, where the transition toward cooperation is induced by specific neighbourhood relations, repeated interactions, discounted payoffs over long time horizons, indirect reciprocity, favourable strategy mutations, enforcement

of social norms, etc. [105, 220, 247, 249, 251]. All of these propositions either improve the payoff of the cooperating strategy or provide, in one or another way, *additional information* agents may consider when making a strategic decision.

Our approach is much simpler. It does not modify payoffs at all, but only considers the information agents already have when they simultaneously play the Prisoner’s dilemma game with their n neighbours (spatial or randomly chosen). This information is the local fraction of cooperators, $f_i = n_1/n$, and defectors, $(1 - f_i)$, of an agent i . As f_i is inevitably observed, it does not constitute new information. In our model, we assume that agents respond to f_i in two different ways summarised in Eq. 2.9. From a strategic point of view, they choose the strategy θ_i that brings the highest payoff $a_i(\theta_i, f_i)$, whereas from a herding perspective, they simply respond to the local frequency of each strategy in a nonlinear manner, $\mathcal{F}(f_{\theta_i})$. As we argued, social herding can be induced by providing *less* payoff-related information, in which case both strategies are seen as equally valuable.

The parameter ζ_i gives a weight to these two perspectives. In Eq. 2.9, we have assumed ζ_i to be an individual parameter, so that agents, dependent on their internal preferences or access to knowledge (such as a known payoffs), can apply different weights. Here, however, we did not explore this source of heterogeneity. Instead, we kept ζ as a global parameter equal for all agents. This limit case is equivalent to assuming a population of agents, a fraction ζ of which follows *only* social herding, whereas a fraction $(1 - \zeta)$ are driven entirely by strategic considerations. In this way, we can interpret our main result about the critical ζ required to turn a population of defectors into cooperators in a more general manner: ζ can be seen as the minimal fraction of herding agents needed to enable transition towards cooperation. With respect to the access to information, we can interpret this finding as follows: if payoff-related information is known to all agents, they will –in the standard Prisoner’s dilemma setting– collectively choose defection. However, if only a small fraction of agents (about 20%) (see Figure 2.3) is granted access to such information, the system can be driven towards a state where cooperation is the predominant strategy. Therefore *less* information (or a larger fraction of uninformed agents) will lead to *more* cooperation. This result, though theoretical, has already been empirically demonstrated for a different context in [56], where it was shown that uninformed individuals, given favourable interactions, can promote democratic consensus in animal groups.

Our conclusion is derived for the case of an evolutionary Prisoner’s dilemma game with fixed payoffs and a fixed four-player neighbourhood, where agents follow a stochastic better response dynamics. The transition toward cooperation relies on choosing the right nonlinear social herding in response to the local (or global) fraction of cooperators. We have demonstrated that the *linear* response, where the probability of choosing a strategy is

directly proportional to the local (or global) fraction of that strategy, irrevocably leads to large-scale defection. However, when the fraction of cooperators and defectors are equal, a small non-linearity in α_2 turns the situation around. This finding tells us that social herding matters most in tie situations, similar to another class of group decision models [75]. To design a mechanism that induces social herding only in case of ties provides a markedly “cost-efficient” solution, since we would not need to enforce a decision against the majority. Agents can still follow the strategy of the majority – just in the undecided case, when their local neighbourhood is split, we need to ensure that the symmetry is broken in favour of the cooperative direction.

Chapter 3

Wisdom of Crowds

Summary

In this chapter we present empirical evidence that demonstrates negative effects of social influence on the wisdom of crowds phenomenon, in situations where one would normally not expect them. The main contribution of the chapter, however, consists of the theoretical investigation to explain these negative effects. We develop an agent-based model that not only is able to qualitatively reproduce the empirical observations, but also to suggest testable predictions on the general effects of social influence. In particular, the model asserts that it is the starting configuration of the crowd, in terms of initial collective error and diversity, that ultimately determines whether social influence would be beneficial or detrimental for the wisdom of crowds.

Based on (1) Lorenz, J., Raihut H., Schweitzer F., Helbing D. *How social influence can undermine the wisdom of crowds effect*, in Proceedings of the National Academy of Sciences, vol. 108, number 22, 2011, (2) Mavrodiev P., Tessone C.J., and Schweitzer, F. *Quantifying the effects of social influence* in Scientific Reports, vol.3, 2013, (3) Mavrodiev P., Tessone C.J., and Schweitzer, F. *Effects of Social Influence on the Wisdom of Crowds*, Collective Intelligence conference 2012, 2012, and (4) Mavrodiev P., Tessone C.J., and Schweitzer, F. *Social Influence and the Wisdom of Crowds*, manuscript pending submission. P.M did not participate in the empirical study (1). For (2), (3) and (4) P.M. wrote the full text and produced all the quantitative results.

3.1 Introduction

In the previous chapter we showed how restricting the amount of payoff-related information and assuming herding behaviour can promote cooperation in environments, in which defection is an evolutionary stable strategy. As with most examples of unintended consequences, positive and negative effects go hand in hand. Therefore, it is trivial to concoct a scenario in which herding and imitation would be harmful – simply ensure that the “wrong” non-linearity is chosen in our previous model and defection wins. However, we will demonstrate detrimental effects of herding in situations where one would normally not expect them. More specifically, this chapter will reveal the undermining effects of social influence for the *wisdom of crowds*. At the end, we will design an empirically-inspired agent-based model to derive generic conditions that determine the polarity of the social influence effect.

Crowds and wisdom of crowds. According to the 15th century writer John Heywood two heads are better than one. This well-known epigram illustrates the idea that two individuals are able to solve a problem that a single individual cannot. In present days, we have extended the aphorism to the so-called *wisdom of crowds* effect [246] – the aggregate individual opinion is more accurate than the average individual. Since the early empirical evidence that demonstrated this phenomenon [77], many studies from different disciplines have reproduced it. Empirical evidence from social psychology [65, 91, 137, 138, 159, 241], finance and prediction markets [115, 205], problem-solving experiments [160, 167, 272], online communities [135, 136], internet search engines, prediction markets, forecasting literature [9, 44, 183, 268], and even life-saving situations [139], show that aggregated judgements often outperform individual ones. The wisdom of crowds can even be observed in the absence of a crowd per se – individuals can combine their own guesses on a given issue to the same effect [84, 107, 204, 260].

Formally, the wisdom of crowds refers to the phenomenon that the aggregate prediction or forecast of a group of individuals can be surprisingly more accurate than most individuals in the groups, and sometimes – than any of them. For the sake of illustration, imagine that two professors wish to guess the test score of a certain student. Based on idiosyncratic knowledge they both have about the student, the two professors guess 15 and 20 points respectively. Imagine further that the student actually scored 17.5 points. In this case, both professors are off by 2.5 points, whereas the average¹ of their guesses is exactly 17.5, showing that the aggregate guess is more accurate than any of the individuals. As,

¹In this example, we have implicitly quantified the aggregate opinion as the arithmetic average.

this example shows, when individuals equally overestimate or underestimate the truth, the imperfections in their opinions cancel out and only useful information remains ([246], p. 10).

The notion that collecting and aggregating diverse opinions may produce more accurate results than individual (even expert) ones is as intriguing as the circumstances which promote it. Surowiecki listed four criteria required to form a wise crowd – diversity of opinions, independence of one’s judgement from the others, minimum level of individual expertise, and a quantitative mechanism to aggregate private judgements to collective action.

Diversity has been identified as instrumental in providing creative perspectives to problem-solving, thus preventing the population from getting trapped in locally suboptimal solutions [116]. In fact, the “diversity prediction theorem”² [189] shows that diversity weighs as much as individual ability in determining collective accuracy. From a mathematical point of view, diversity is required in order to balance out uncorrelated imperfections in opinions through aggregation. Intuitively, as no single individual is aware of all traits of a given problem at hand, diversity helps people combine their idiosyncratic perceptions so that together they gain a wider perspective. Diversity, however, is not the pinnacle of optimal decision-making. No amount of diversity can help if the population is completely ignorant on a given issue, or if opinions are diverse, but heavily skewed. Thus, the composition of diversity is as important as diversity itself [26].

Independence of opinions is another distinguishing feature of the wisdom of crowds phenomenon. Independence refers to the absence of *social influence* – the pervasive tendency of individuals to conform to the behaviour and expectations of others [126]. When individual opinions are public knowledge³, social influence processes, such as peer pressure and conformity, may induce individuals to seek the approval of others in decision-making scenarios, *at the expense* of improved collective accuracy. Even though, in certain circumstances conformity can produce positive social outcomes, e.g. norm compliance [95] and law abidance [126], it is considered detrimental to the wisdom of crowds, as it may compel one to act against his best judgement. A prime example is Asch’s famous conformity experiment [10], in which he showed that a group of confederates can create enough conformity pressure to persuade individuals into publicly reporting incorrectly the lengths of two identical lines, even though the individuals correctly identified the lines being of the same length. Importantly, this is not overt coercion. After the experiment, test subjects

²The theorem states that collective accuracy equals average individual error minus the variance in opinions, i.e. group diversity.

³Or can be inferred from the available information, even approximately.

shared that they either lost confidence in their own judgement when it was at odds with the majority, or that they preferred to avoid potential conflict instead of giving an honest answer. Another example is the well-studied psychological phenomenon of group-think [233, 255] in which the desire to minimise conflict, preserve group harmony and reach consensus compels group members to make decisions without critical evaluation of alternative ideas or viewpoints, often resulting in inferior solutions [180].

Conformity pressure manifests also in situations in which herding can be perceived as rational [62]. For example, in bank runs, investors can see other investors running the bank (through long queues) and deduce, based on imperfect information, that regardless of the availability of funds in the bank, they can be left empty-handed if enough others engage in this behaviour. Hence, it is individually rational to join the run, as it maximises the expected payoff (alternatively it minimises the risk of a considerable loss).

Cost of conformity. In all these examples a major factor that makes social influence detrimental for the group decision is that conformity is costless, whereas upholding one's opinion is not, as it bears the risk of creating conflict or reduce expected payoffs. Furthermore, when knowing the opinions of others, individuals can easily lose confidence in their private judgements and erroneously follow others who seem more knowledgeable or are more influential [90].

However, let us consider a decision-making scenario in which individuals (i) have to repeatedly guess a well-defined objective answer, (ii) have monetary incentives to give their best estimate, and (iii) are aware of the estimates of others. In such a scenario, one is entirely motivated by personal gain to give the guess closest to the truth. Consequently, there is no pressure or expectation to reach group consensus nor is there any type of ostracism for minority judgements. Consensus, is not only unnecessary, but may be costly, because following majority opinions, at the expense of one's private knowledge, bears the risks of following the wrong crowd and lowering one's payoff.

Would social influence be beneficial for the wisdom of crowds in this case? More precisely, how would individuals adjust their guesses if they know the judgements of others? A policy maker who deliberates the design of a prediction market for example, could convincingly argue that since we have excluded any explicit or implicit incentives to conform and have exclusively linked individual accuracy to monetary payoffs, then the more information individuals have at their disposal the more precise they will be. Therefore, the information of others will only be used for improving one's own opinion and not for the sake of building consensus. Individuals, then, will imitate those who are genuinely perceived to possess superior knowledge or expertise. Hence, from a policy maker perspective, given the right

incentives design, more information can only improve the crowd wisdom.

We will present empirical evidence for the opposite. Even in situations in which social influence is intuitively expected to improve the quality of individual decisions, the resulting herding behaviour leads the population to collectively drift away from the truth over time. Even more, individuals become at the same time more confident in their own accuracy; instead of wisdom of crowds, this effect could more appropriately be called the “madness of crowds”. The next two sections outline the experimental set-up and its results. Following that, we will employ an agent-based model to investigate the general conditions that lead to this discrepancy.

3.2 Experimental set-up

The experiment involved a total of 144 subjects chosen among students at ETH Zürich. There were 12 experimental sessions, conducted at different times, with 12 participants in each session. During a session the subjects were asked a total of 6 quantitative questions regarding geographical facts and crime statistics⁴. Each question had to be answered over five time periods. The questions were designed in such a way that individuals were not likely to know the exact answer, but could still formulate educated guesses based on related knowledge and experience. Figure 3.1 (left) shows the structure of the experiment.

In the first (initial) time period, all subjects responded to the particular question on their own. After all 12 subjects gave their estimates, each had to answer the same question over 4 additional periods for a total of 5 rounds. Social influence was introduced in terms of so called *information conditions* (or information regimes) – experimental conditions categorised by the type of information subjects had available about the guesses of others. In the “no” information condition, no individual was aware of others’ guesses throughout the 5 periods. This represented the control group. In the “aggregate” condition, each subject was provided with the arithmetic average of everyone else’s answers in the previous round. Finally, in the “full” information condition, subjects learnt all opinions from all previous rounds so far. In each session, two questions were posed in the no, two in the aggregated, and two in the full information condition. The 12 subjects were randomly assigned to the three information conditions in the beginning of the session⁵.

As we already mentioned, an important component of the experiment was the reward structure (Figure 3.1, right). While answering a given question, participants received in-

⁴Example: What is the population density of Switzerland?

⁵The detailed experimental protocol presented to the subjects is available at <http://www.pnas.org/content/suppl/2011/05/10/1008636108.DCSupplemental>

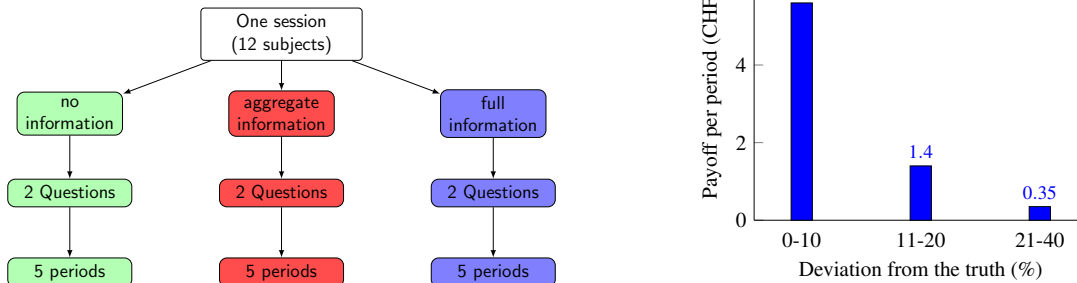


Figure 3.1: Experiment structure: **(left)** The experiment consisted of 12 sessions each composed of 12 subjects. In each session, the group of 12 subjects had to repeatedly answer two questions in the no information, two in the aggregate and two in the control condition, for a total of six questions. The order of the questions was randomised across sessions. After each of the five periods subjects were asked the same question again and could revise their answers depending on the information available to them. **(right)** Shown are individual payoffs in Swiss francs (CHF) dependent on the percentage deviation of the provided guess from the truth.

dividual payoffs depending on their deviations from the correct answer – reward-bearing deviations were defined as 10%, 20% and 40% intervals around the truth. The rewards were calculated after each round, so that at all times individuals were motivated to use the available information for optimal decisions⁶. However, the correct answer and the rewards earned were disclosed at the end of the experiment to avoid giving away a-priori knowledge about the truth. This reward structure also eliminated the benefits of strategic behaviour, such as misleading or cooperating with others, because it did not affect individuals’ payments – subjects had incentives to use only their own knowledge and interpretation of others’ opinions to find the truth.

3.3 Results

One of the requirements for quantifying the wisdom of crowds is a suitable aggregation measure. Ideally, an aggregated measure would indicate the most central opinion in the population. For opinion distributions that are Gaussian-like, the simple unweighted arithmetic average may be a good choice, although there are other averaging methods, which could perform better in special cases [60, 86, 103]. The opinion distributions in the experi-

⁶To put the per-round payoffs into perspective, a subject who managed to be consistently within 10% of the truth for all six questions in his session would have earned 5.6 CHF \times 5 rounds \times 6 questions = 168 CHF. The same calculation for someone who was always within the 40% range yields 42 CHF.

ment were heavily right-skewed with a majority of low estimates and a minority spread on a fat right tail, much like log-normal distributions. The skew made the arithmetic average inappropriate as a measure of centrality, since it was closer to the truth than individuals' first estimates in only 21.3% of the cases. However, taking the natural logarithm of the estimates, resulted in a more bell-shaped distribution. Consequently, the arithmetic average of the log-transformed opinions (which equals the logarithm of the geometric mean of the original data) performed much better – it was closer to the true value than individuals' first estimates in 77.1% of the cases. This provided a justification for log-transforming all opinions over all sessions before further analysis⁷.

Three main quantities were used to evaluate the aggregate performance of the crowd – “collective error”, “group diversity” and “wisdom of crowds indicator”. Let N be the number of individuals and $x_i(t)$ be the answer of individual i in round t . The collective error, $\mathcal{E}(t)$, is defined as the squared deviation of the average opinion in round t from the true value, \mathcal{T} . Since all x_i 's were log-transformed, the arithmetic average is $\langle \ln x(t) \rangle$ ⁸, which equals the logarithm of the geometric mean of the original data. The collective error in period t then equals:

$$\mathcal{E}(t) = (\ln \mathcal{T} - \langle \ln x(t) \rangle)^2. \quad (3.1)$$

The group diversity in period t , $\mathcal{D}(t)$, is the variance of the opinion distribution:

$$\mathcal{D}(t) = \frac{1}{N} \sum_{i=1}^N (\ln x_i(t) - \langle \ln x(t) \rangle)^2. \quad (3.2)$$

Finally, we define a *wisdom of crowds indicator*, $\mathcal{W}(t)$, to quantify the wisdom of crowds effect:

$$\mathcal{W}(t) = \max\{i | \hat{x}_i(t) \leq \mathcal{T} \leq \hat{x}_{N-i+1}(t)\}. \quad (3.3)$$

where \hat{x}_i 's are the original (i.e. not log-transformed) *sorted* opinions. The rationale behind this definition is that the phenomenon of wisdom of crowds requires both collective accuracy and a group of individuals diverse enough to constitute a crowd. If all predictions are narrowly distributed around the wrong value, then a system planner, such as a regulator or a government, would gain confidence from such prevalent unanimity, in advise that is

⁷More specifically, the transformed guesses were obtained by dividing the the logarithm of the original estimates by the true values, so that answers across different questions could be compared

⁸The notation $\langle X \rangle$ stands for $(1/N) \sum_{i=1}^N X_i$

actually misleading.

Eq. 3.3 accounts for both collective accuracy and group diversity simultaneously by evaluating how many estimates around the median are needed to encompass the truth. The indicator has a maximum of $N/2$ when the truth lies between the most central estimates (or is the most central estimate) and a minimum of zero when the truth is outside the range of all estimates. In this way, a group with low collective error, but small diversity may still be less wise than another, less accurate but more diverse crowd.

Social influence effect One way in which herding and imitation affects the wisdom of crowds is a statistical effect termed *social influence effect*. It results in diminishing group diversity without significant decrease in collective error. This means that, on average, groups could not benefit from the information exchange, but instead engaged in a convergence process that did not improve their collective accuracy. Figure 3.2 illustrates the social influence effect.

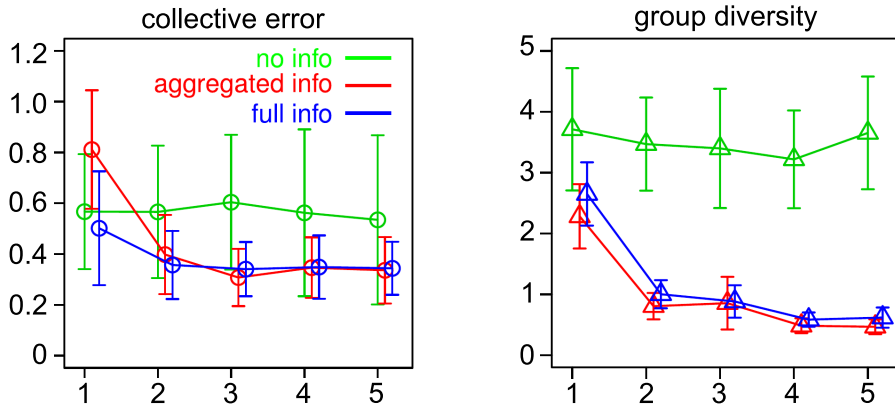


Figure 3.2: Social influence effect: average group diversity and average collective error for all information conditions over the five experimental rounds. Each data point represents 24 groups, aggregated over all questions (2 questions/information condition/session \times 12 sessions, see Figure 3.1 left). Data points are computed on the logarithm of estimates and normalised by the respective true value; error-bars represent 10% confidence intervals. Figure from [158] used with permission.

It is evident that diversity decreases in the aggregate and full information conditions, whereas the collective error does so only slightly. This social influence effect is further supported by Kolmogorov-Smirnov tests, which confirm that the impact of social influence is highly significant for the group diversity, while only weakly so for the collective error⁹.

⁹For more details on the statistical verification of the social influence effect, refer to SI Appendix of [158]

Range reduction effect *Range reduction* refers to expansion of the core range of estimates needed to enclose the true value. This effect occurs when the truth is gradually displaced to outer regions of the estimate distribution (Figure 3.3). As a result, the number of estimates, or more generally opinions, that need to be accounted for to ensure that the correct solution is within the scope of the group, increases. Range reduction is more likely to take place in the presence of the social influence effect. The corresponding decrease of diversity, even when collective accuracy improves, can still lead to shifting of the truth to the fringe of the estimate distribution. In essence, the group would be less “wise” as individuals would cluster around a wrong value and provide a false clue for the location of the truth.

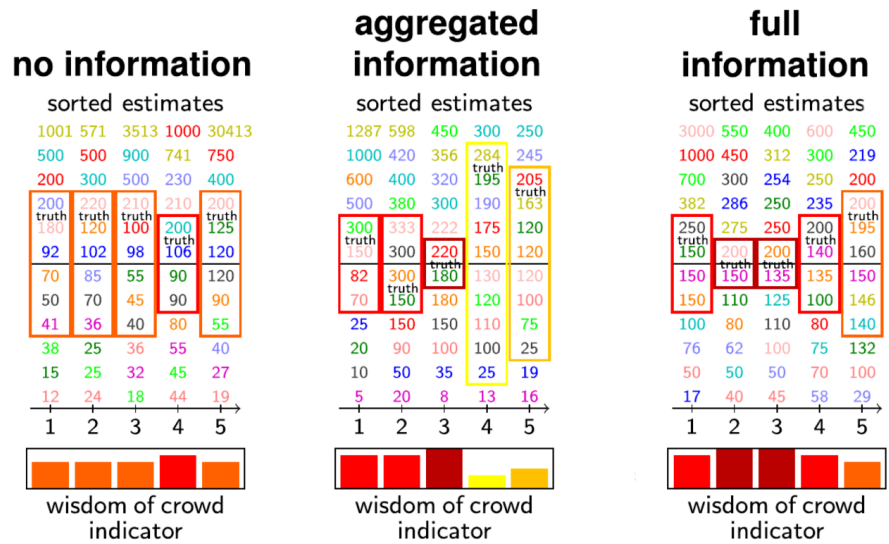
Range reduction is quantified with the wisdom of crowds indicator in Eq. 3.3. The indicator considers the group to be wisest if the truth lies between the two most central estimates (in context of the experiment, between the sixth and seventh largest of 12 estimates). If the four most central values are needed to enclose the truth, the level of wisdom is lower, and if the six most central values are needed, even lower. If the truth is outside the range of the group, there is no wisdom of crowds effect and the indicator is 0.

Figure 3.3 (top) shows bar plots of the wisdom of crowds indicator for typical responses in the three experimental treatments. The figure demonstrates that the wisdom of crowds indicator tends to decline over time under social influence. The effect is substantial and statistically significant for all questions, as shown by the regression model Figure 3.3 (bottom). The indicator is one unit lower when information exchange is present compared with the control condition. A decrease of one unit means that one has to consider one additional person in the upper range and one – in the lower range of the sorted estimates so that the truth is included in the selected range. Note that the reduction is stronger under the aggregated condition than under full information.

The agent-based model introduced below reproduces the social influence and range reduction effects in an artificial population under all experimental conditions. Based on this framework we will show that the decline of crowd wisdom cannot be imputed to social influence alone.

3.4 Agent-based model of social influence

Consider a population of N individuals engaged in a decision-making scenario. Each individual possesses a real-valued opinion, $x_i(t)$, that evolves according to the following



	groups' wisdom-of-crowd indicator
intercept	3.92*** (15.1)
aggregated information	-1.39** (-3.29)
full information	-0.98* (-2.21)
<i>N</i>	288

t statistics in parentheses (robust std. errors), * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 3.3: Range reduction effect. (Top) Sorted estimates over successive rounds for five typical responses to the question, “How many murders were registered in Switzerland in 2006?”. Coloured rectangles indicate the most inner estimates that still include the true value. The range of the box depicts the wisdom of crowds indicator. The maximum value of 6 is attained when the truth is between the most central guesses. A value of 0 indicates that the truth is outside the range of all estimates. Coloured numbers represent individuals. (Bottom) Linear regression on the groups’ wisdom of crowds indicator. The predictor is the experimental treatment, represented by a binary variable. The control condition is the reference category, represented by the intercept of the linear model. The indicator is calculated for the second, third, fourth, and fifth rounds, as the first round had no information feedback and thus does not constitute experimental treatment. Figure from [158] used with permission.

process

$$\frac{d}{dt}x_i(t) = \sum_{j=1, j \neq i}^N \mathcal{F}(x_j, x_i) + \beta_i(x_i(0) - x_i(t)) + D\xi_i(t). \quad (3.4)$$

The first term represents coupling to an individual’s environment and describes the impact of all current opinions in the population on an agent’s future estimates. We refer to it as the *social influence term*. We formalise it by assuming a function \mathcal{F} , which specifies the size of the influence of agent j on the focal agent i . Summing over all j ’s, then, produces the total impact of the population on agent i . By assuming different functional forms for \mathcal{F} we can model various types of information spreading and interactions structure.

The second term models an individual’s tendency to uphold his original opinion. We refer to it as *individual conviction*, and its strength is given by the parameter β_i .

The third term reflects the idea that individuals may change their opinions because they incorporate other information known to them *ex ante*. This term does not come from interactions, but originates purely from internal mechanisms. We model it as Gaussian white noise, $\xi_i(t)$, with unit variance. D is the corresponding noise intensity. It might be argued that individual conviction is likely to be a stable personality trait, at least during the duration of a decision-making task. This would be equivalent to considering $\beta_i x_i(0) + D\xi_i(t)$ a constant, different for each individual and generated by a given distribution. This modelling approach would imply considering *heterogeneity* at the individual level. Alternatively, one can fix $\beta_i = \beta$ and add a random fluctuation term. Both approaches are equivalent as long as the random terms are drawn from the same distribution. However, for tractability we prefer the latter.

From a physicist’s point of view, the dynamics in Eq. 3.4 resembles Brownian particles interacting in a potential field [218]. By changing the coupling term, one can reproduce different responses to the field. In social psychology, we can also see Eq. 3.4 as a formalisation of Lewin’s heuristics that individual action is a function of idiosyncratic perception of available information and the influence of a “field”, that is the influence of the environment [211].

The general framework postulated in Eq. 3.4 can be applied easily to the experimental set-up from the previous section. Let us define a true value, \mathcal{T} , which individuals have to repeatedly guess. The collective error at time t is thus given by

$$\mathcal{E} = (\ln \mathcal{T} - \langle \ln x(t) \rangle)^2, \quad (3.5)$$

where, we have log-transformed the estimates, $x_i(t)$, to reflect the observed skew in the

experimental data¹⁰. Note that $\langle \ln x(t) \rangle = \ln(\langle x(t) \rangle_{\text{GM}})$, i.e. the arithmetic average of the transformed estimates, $\langle \ln x(t) \rangle$, is also the logarithm of the geometric mean of the original x_i 's.

The group diversity, $\mathcal{D}(t)$, is the variance of the transformed estimates

$$\mathcal{D} = \frac{1}{N} \sum_{i=1}^N (\ln x_i(t) - \langle \ln x(t) \rangle)^2 = \langle \ln^2 x(t) \rangle - \langle \ln x(t) \rangle^2. \quad (3.6)$$

We now define a group's *configuration*, or *state*, to be the pair of collective error and diversity, i.e. the pair $\{\mathcal{E}(t), \mathcal{D}(t)\}$. The starting configuration is then expressed as $\{\mathcal{E}(0), \mathcal{D}(0)\}$. Note the equivalence between holding the strength of social influence or/and individual conviction constant while varying the starting configuration of the population, and fixing the starting configuration while varying the strength of the social influence or/and individual conviction terms. The first reflects the statement: "the same strength of social influence/individual conviction can lead to a different long-term state of the population, depending on its starting configuration". The second implies that: "a population with a fixed starting configuration can be driven to different end states depending on the strength of the social influence/individual conviction". In other words, a group can reach the same long-term state either by fixing the initial state and varying the force which drives it, or by fixing the force and adjusting the initial state. Here, we are going to keep the starting configuration fixed. This allows us to assume various weights on the two terms in Eq. 3.4, and thus to investigate the interplay between social influence and individual conviction.

Finally, by specifying a precise form for the social influence component, $\mathcal{F}(\cdot)$, we can model the information exchange in the three experimental conditions. This is presented in the next section, where we additionally show an analytical analysis of the no- and aggregate-information regimes. The full information regime will be defined, but due to tractability issues, we study it only with the help of computer simulations in Section 3.4.2.

3.4.1 Analytical approach

No information regime. In this information regime agents do not interact nor do they exchange information. Hence, their estimates are only influenced by the strength of their individual conviction. Zeroing the social coupling among the agents in Eq. 3.4 yields

$$\frac{d}{dt} x_i(t) = \beta_i (x_i(0) - x_i(t)) + D\xi_i(t). \quad (3.7)$$

¹⁰The notation $\langle \cdot \rangle$ denotes ensemble average.

This is a standard Ornstein-Uhlenbeck process, whose solution is given by

$$x_i(t) = x_i(0)e^{-\beta t} + x_i(0) (1 - e^{-\beta t}) + D \int_0^t e^{\beta(s-t)} \xi_i(s) ds. \quad (3.8)$$

Therefore, for large t , the time average of the individual estimates, $\overline{x_i(t)}$, would eventually revert to its initial value, $x_i(0)$, with decreasing variance.

As there is no social influence in this regime regime, agents do not have incentives to deviate from their original opinions, up to small random fluctuations. In this sense, it is trivial to describe the dynamics of the collective error and group diversity analytically. From 3.8 it follows that each $x_i(t)$ oscillates around $x_i(0)$, hence

$$\lim_{t \rightarrow +\infty} \langle \ln x(t) \rangle = \langle \ln x(0) \rangle.$$

The log of the geometric mean is quasi-stationary at its starting value, fluctuating randomly around it. The collective error and group diversity exhibit the same behaviour (Eqs. 3.5 and 3.6)

$$\begin{aligned} \lim_{t \rightarrow +\infty} \mathcal{E} &= (\ln \mathcal{T} - \langle \ln x(0) \rangle)^2. \\ \lim_{t \rightarrow +\infty} \mathcal{D} &= \text{Var}(\ln x(0)). \end{aligned} \quad (3.9)$$

Aggregate information regime In the aggregate information regime agents observe the average of all estimates over time, which is equivalent to a mean-field scenario. We assume the following form for the social influence component, \mathcal{F} , in Eq. 3.4

$$\mathcal{F}(x_j, x_i) = (x_j - x_i)w_{ij}. \quad (3.10)$$

The influence that agent i perceives from agent j is a product of the opinion difference between i and j , and the weight, w_{ij} , which i gives to the opinion of j . In a mean-field context all estimates are weighted equally, i.e

$$w_{ij} = \frac{\alpha_i}{N}, \quad (3.11)$$

with α_i being the strength of the coupling. Therefore, Eq. 3.4 reduces to the form

$$\frac{dx_i(t)}{dt} = \alpha_i (\langle x(t-1) \rangle - x_i(t)) + \beta_i (x_i(0) - x_i(t)) + D\xi_i(t). \quad (3.12)$$

Note that our mean-field approach predicts a linear relation between the change in one’s estimate and the distance of the previous estimate from the mean. This can actually be tested rigorously from the experimental data. Referring to the description in Section 3.2, we aggregated individual responses across all subjects and sessions, and estimated the following linear model relating the distance of one’s estimate from the mean, and the adjustment in the next period

$$\Delta x_i(t) = \gamma_0 + \gamma_1(\langle x(t-1) \rangle - x_i(t-1)) + \varepsilon, \quad (3.13)$$

with the associated null hypothesis $\mathcal{H}_0 : \gamma_1 = 0$ and two-sided alternative $\mathcal{H}_1 : \gamma_1 \neq 0$. Table 3.1 shows the results from estimating this model for each of the six experimental questions.

		Estimate	Std. Errors	Robust std. errors	<i>t</i> -value	<i>p</i> -value	samples	df
Q1	γ_0	-176.46	14.98	15.55	-11.35	$< 2.2 \times 10^{-16}$	192	190
	γ_1	0.97	0.02	0.1	9.57	$< 2.2 \times 10^{-16}$		
Q2	γ_0	35.33	12.6	12.9	2.74	0.007	192	190
	γ_1	0.27	0.05	0.09	2.89	0.004		
Q3	γ_0	-1321.5	828.2	853	-1.55	0.12	192	190
	β_1	0.83	0.05	0.1	6.25	2.7×10^{-9}		
Q4	γ_0	-146.3	23.2	23.7	-6.2	3.8×10^{-9}	192	190
	γ_1	0.6	0.01	0.03	18.8	$< 2.2 \times 10^{-16}$		
Q5	γ_0	6.8	14.8	15.1	0.5	0.66	188	186
	γ_1	0.4	0.04	0.1	3.72	0.0003		
Q6	γ_0	-821	10^3	1387	-0.6	0.55	188	186
	γ_1	0.46	0.02	0.03	15.3	$< 2 \times 10^{-16}$		

Table 3.1: Robust linear regression of Eq. 3.13 for each of the six questions (Q). Uncorrected standard errors are reported for comparison only. Last column shows degrees of freedom.

We focus primarily on the estimation of γ_1 , as the constant term, γ_0 , is heavily influenced by a few outliers, and thus exhibits large standard errors even when significant. From the reported *p*-values, we see that the impact of the distance to the mean opinion, $\langle x(t-1) \rangle - x_i(t-1)$, is highly significant across all questions (with low rob. std. errors) in explaining one’s own opinion change; the size of the effect is also considerable. This holds for questions with correct answers that differ by about 10 orders of magnitude, hence the effect is not simply a first-order approximation of a non-linear regime around a narrow range of $\langle x \rangle - x_i$. Finally, an important consequence of the linear model is that the same mathematical

regularity underlies individual reactions to social influence, despite individual differences in subjects, e.g. emotions, conviction in one's own opinion or beliefs about the competency of others. This suggests that once initial guesses are formed, diversity among subjects does not play a role in the adjustment of subsequent guesses. For this reason, we take α_i in Eq. 3.12 as a constant. In [169] we have provided more technical details on the methodology and justification of the linear model.

We can now average Eq. 3.12 over the whole population to obtain

$$\frac{d\langle x(t) \rangle}{dt} = \beta (\langle x(0) \rangle - \langle x(t-1) \rangle) + \frac{D}{\sqrt{N}} \langle \xi(t) \rangle. \quad (3.14)$$

which is again an Ornstein-Uhlenbeck process, with solution

$$\langle x(t) \rangle = \langle x(0) \rangle e^{-\beta t} + \langle x(0) \rangle (1 - e^{-\beta t}) + \frac{D}{\sqrt{N}} \int_0^t e^{\beta(s-t)} \langle \xi(s) \rangle ds. \quad (3.15)$$

As in Eq. 3.8, the time average of the ensemble average of the estimates, $\overline{\langle x(t) \rangle}$, equals $\langle x(0) \rangle$ for large t .

To derive the precise dynamics of the collective error from Eq. 3.5 in this information regime, we first need to express $\langle \ln x(t) \rangle$. Rewrite Eq. 3.12 as

$$\begin{aligned} \frac{d \ln x_i(t)}{dt} &= \frac{\alpha (\langle x(t-1) \rangle - x_i(t)) + \beta (x_i(0) - x_i(t)) + D\xi_i(t)}{x_i(t)} = \\ &= \frac{\alpha \langle x(t-1) \rangle}{x_i(t)} - \alpha + \frac{\beta x_i(0)}{x_i(t)} - \beta + \frac{D\xi_i(t)}{x_i(t)}. \end{aligned} \quad (3.16)$$

The ensemble average becomes

$$\frac{d \langle \ln x(t) \rangle}{dt} = \alpha \left\langle \frac{\langle x(t-1) \rangle}{x_i(t)} \right\rangle - \alpha + \beta \left\langle \frac{x_i(0)}{x_i(t)} \right\rangle - \beta + \frac{D}{\sqrt{N}} \left\langle \frac{\xi_i(t)}{x_i(t)} \right\rangle. \quad (3.17)$$

In Appendix 9.1, we show that with the above transformation, the closed-form solution of Eq. 3.17 is given by

$$\begin{aligned} \langle \ln x(t) \rangle &= \langle \ln x(0) \rangle + \\ &+ \sum_{n=1}^{\infty} \frac{(-1)^n \alpha^n \langle \sigma(0)^n \rangle}{\langle x(0) \rangle^n n(\alpha + \beta)^n} \left[(1 - \alpha^n)(1 - e^{-(\alpha+\beta)nt}) + (\alpha + \beta)^n - (\beta + \alpha e^{-(\alpha+\beta)t})^n \right]. \end{aligned} \quad (3.18)$$

Or in the long term for $t \rightarrow \infty$

$$\langle \ln x \rangle_{\text{LT}} = \langle \ln x(0) \rangle + \sum_{n=1}^{\infty} \frac{(-1)^n \alpha^n \langle \sigma(0)^n \rangle}{\langle x(0) \rangle^n n(\alpha + \beta)^n} \left[1 + (\alpha + \beta)^n - \alpha^n - \beta^n \right]. \quad (3.19)$$

The dynamics of the collective error is, thus, obtained by plugging in the above solution into Eq. 3.5.

Observe that $\langle \ln x \rangle_{\text{LT}}$ crucially depends on the distribution of initial estimates, and in particular on its central moments, $\langle \sigma(0)^n \rangle$. Let Ω_n be the n^{th} term of the summation in Eq 3.19. The sum is convergent if and only if $|\Omega_{n+1}/\Omega_n| < 1$, i.e.

$$\left| \frac{\langle \sigma(0)^{n+1} \rangle}{\langle \sigma(0)^n \rangle} \right| < \frac{(n+1) \langle x(0) \rangle (\alpha + \beta)}{n\alpha} \cdot \frac{1 + (\alpha + \beta)^n - \alpha^n - \beta^n}{1 + (\alpha + \beta)^{n+1} - \alpha^{n+1} - \beta^{n+1}}.$$

Furthermore if $\alpha + \beta > 1$, it is sufficient that

$$\left| \frac{\langle \sigma(0)^{n+1} \rangle}{\langle \sigma(0)^n \rangle} \right| < \frac{(n+1) \langle x(0) \rangle (\alpha + \beta)}{n\alpha}, \quad (3.20)$$

to guarantee that the long-term geometric mean converges to a fixed point. This fixed point will always be larger than $\langle \ln x(0) \rangle$, as the second term in Eq. 3.19 is positive.

For obtaining the dynamics of the group diversity (Eq. 3.6) we use the delta method as an approximation. The method is in essence a first order Taylor expansion of the form

$$\text{Var}[f(X)] \approx (f'(\langle X \rangle))^2 \text{Var}[X]. \quad (3.21)$$

It will be a poor approximation in cases where $f(X)$ is highly non-linear, which is not the case for $f(X) = \ln X$

$$\mathcal{D}(t) = \frac{1}{\langle x(t-1) \rangle^2} \langle \delta(t)^2 \rangle. \quad (3.22)$$

Plugging in $\langle \delta(t)^2 \rangle$ from Eq. 9.4 yields

$$\mathcal{D}(t) = \frac{\langle \sigma(0)^2 \rangle}{[\langle x(t-1) \rangle (\alpha + \beta)]^2} [\beta + \alpha e^{-(\alpha + \beta)t}]^2, \quad (3.23)$$

which is necessarily always positive. The long-term group diversity to which the population converges in our model is given by

$$\mathcal{D}_{\text{LT}} = \frac{\langle \sigma(0)^2 \rangle \beta^2}{[\langle x(0) \rangle (\alpha + \beta)]^2}. \quad (3.24)$$

It can further be shown that the group diversity increases with the degree of individual conviction and the variance of the starting estimate distribution, and decreases with the strength of social influence, as expected

$$\frac{d}{d\alpha} \mathcal{D}_{\text{LT}} = -\frac{2 \langle \sigma(0)^2 \rangle \beta^2}{\langle x(0) \rangle^2 (\alpha + \beta)^3} < 0. \quad (3.25)$$

$$\frac{d}{d\beta} \mathcal{D}_{\text{LT}} = \frac{2\alpha\beta \langle \sigma(0)^2 \rangle}{\langle x(0) \rangle^2 (\alpha + \beta)^3} > 0. \quad (3.26)$$

$$\frac{d}{d\langle \sigma(0)^2 \rangle} \mathcal{D}_{\text{LT}} = \frac{\beta^2}{\langle x(0) \rangle^2 (\alpha + \beta)^2} > 0. \quad (3.27)$$

We now have closed-form solutions for the collective error and group diversity; obtaining one for the wisdom of crowds indicator is not feasible, hence we will investigate it numerically in the next section. At this point we use Eqs. 3.19 and 3.24 to explore the aggregate regime analytically, by running a parameter sweep on $\{\alpha, \beta\}$ for various starting configurations. As mentioned in Section 3.4 we keep the initial estimate distribution constant, and vary the location of the truth, \mathcal{T} . This creates different collective errors in the beginning and allows us to isolate the effects of social influence and individual conviction from the initial composition of the crowd.

In particular, to reflect the positive skewness of the initial estimate distributions observed experimentally (see Section 3.2), we require that the third central moment $\langle \sigma(0)^3 \rangle$ is positive. It follows that all subsequent odd central moments will be non-negative, in addition to the even moments, which are always non-negative by definition. A distribution, which fulfils these requirements is the log-normal distribution, $\mathcal{L}(\mu, \hat{\sigma}^2)$, whose central moments are strictly non-negative. Therefore, in evaluating the analytical description of the aggregate regime, and later when simulating the model in Section 3.4.2, we assume that the initial estimates were generated by a log-normal distribution with $\mu = -3$ and $\hat{\sigma}^2 = 0.7$, i.e. $\langle \ln x(0) \rangle = -3$, $\mathcal{D}(0) = 0.7$. Since μ and $\hat{\sigma}^2$, are related to the arithmetic mean and variance of a log-normally distributed variable, we have $\langle x(0) \rangle = 0.07$ and $\langle \sigma(0)^2 \rangle = 0.005$.

Finally, from the properties of $\mathcal{L}(\mu, \hat{\sigma}^2)$, it follows that the first four central moments are decreasing¹¹. Therefore, without loss of generality, we fix $\langle \sigma(0)^3 \rangle = 0.001$, $\langle \sigma(0)^4 \rangle = 0.00004$, and assume that higher moments vanish. Consequently, condition 3.20 is satisfied.

At this point, the observant reader may object, stating that theoretically higher central moments of the log-normal distribution diverge. We argue that in practice, data generated from such distribution will always have finite moments, as accurately sampling higher moments requires increasing sample sizes; reproducing all divergent moments is only possible in the limit of infinitely large samples. For system sizes, up to the order of 1000, central moments of $\mathcal{L}(-3, 0.7)$ higher than four, become negligibly small. As we never even approach such large sizes in this chapter, our assumption is justified.

Figure 3.4(A-C) shows the long-term collective errors and group diversity (D) for three initial values of the collective error, $\mathcal{E}(0)$. The group diversity behaves the same for all three starting configurations, as it does not depend on \mathcal{T} . As expected, in the presence of social influence, agents' estimates tend to converge, and the long-term group diversity declines (Figure 3.4D). Moreover, the degree of decline increases with the strength of social influence. Note that individual conviction acts in the opposite way – for a fixed strength of social influence, it increases diversity in the group.

We can see this opposing interplay between social influence and individual conviction, for the long-term collective error, as well. This is represented by the transitions from blue to red regions in Figure 3.4(A-C). However, the impact on the long-term configuration to which the group converges is equivocal. In Figure 3.4(C), stronger individual conviction is detrimental for \mathcal{E}_{LT} , which is not true for all other cases. This is an indication that the net effect of social influence on the collective error depends on the initial configuration of the population.

Due to the rightward-only motion of $\langle \ln x(t) \rangle$ described previously, the collective error always increases when $\langle \ln x(0) \rangle > \ln(\mathcal{T})$. Such a configuration is shown in Figure 3.4(A). It essentially represents a scenario where social influence not only does not improve the accuracy of the crowd, but can significantly decrease it, reducing the group diversity at the same time. The same consequences await the group even when $\langle \ln x(0) \rangle < \ln(\mathcal{T})$, albeit for a reduced parameter range. In Figure 3.4(B), the aggregate opinion of the group starts from a relatively accurate state, and ends up at a larger long-term collective error for values of $\{\alpha, \beta\}$ to the right of the 0.01 contour line. In effect, these two configurations reproduce the empirical result from Section 3.2 of the negative effect of social influence. In particular, we see diminishing group diversity without improvement of the collective

¹¹The central moments of a log-normal distribution can be computed in a number of ways, e.g. recursively as outlined in [229]

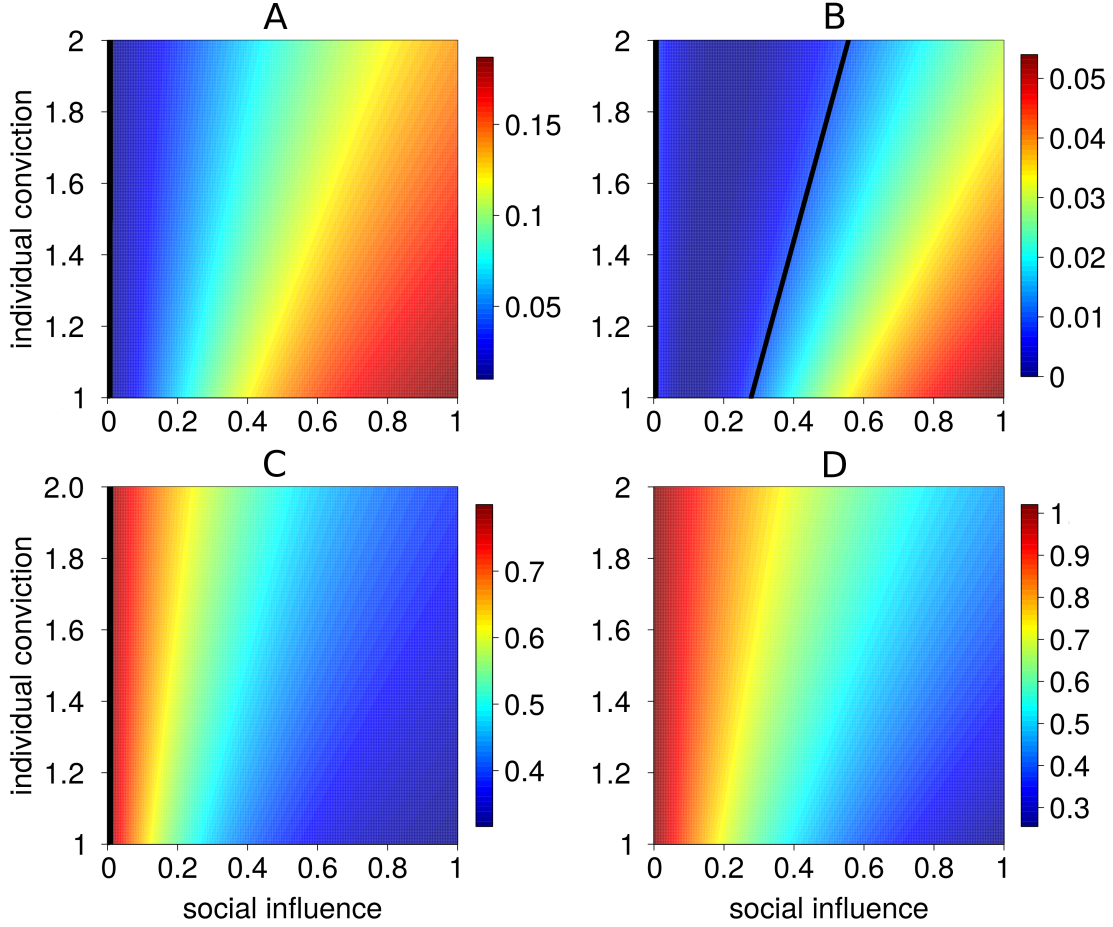


Figure 3.4: Ambiguous effect of social influence on the long-term collective error, \mathcal{E}_{LT} . Plotted in each figure is the strength of social influence, α , vs. the strength of individual conviction β . Colour indicate the long-term collective error (A,B and C), and long-term group diversity (D). Note that colour legends are not consistent across plots. Depending on the initial distance to the truth, social influence can lead a population closer or farther from it in the long term. Initial configurations are given by (A): $\mathcal{E}(0) = 0.02$, $\ln \mathcal{T} = -3.12$, $\langle \ln x(0) \rangle = -3$, $\mathcal{D}(0) = 0.7$, (B): $\mathcal{E}(0) = 0.01$, $\ln \mathcal{T} = -2.1.9$, $\langle \ln x(0) \rangle = -3$, $\mathcal{D}(0) = 0.7$, and (C): $\mathcal{E}(0) = 0.8$, $\ln \mathcal{T} = -2.1$, $\langle \ln x(0) \rangle = -3$, $\mathcal{D}(0) = 0.7$. Black contour lines indicate regions where the long-term collective error equals its starting value. The mismatch between long-term group diversity at $\alpha = 0$, and the expected value of $\mathcal{D}(0) = 0.7$ is due to the Taylor approximation used in deriving Eq. (3.24).

accuracy. This is precisely the social influence effect that Lorenz et. el. reported in their experiment.

Interestingly, however, our model suggests other starting configurations in which social influence brings a clear advantage. Consider the relatively inaccurate initial population in Figure 3.4(C). Virtually for the entire parameter range, the long-term collective error is

lower than that at the beginning. Even more, with the weakest individual conviction and strongest social influence, the group actually converges to the most accurate long-term state. Such a positive outcome is also achieved in Figure 3.4(B). Here, we find a non-linear social influence effect – increasing the strength, α up to small values not larger than 0.2, leads to the lowest \mathcal{E}_{LT} (darker blue regions), but beyond the right-most contour line the effect is reversed and the population ends up collectively more inaccurate than at the start.

Full information regime. In the full information regime, agents are aware of the estimates of others from all previous periods. However, accumulating information over several periods and incorporating it into future decisions tends to become increasingly difficult due to cognitive limitations such as information overload. Therefore, we assume that agents consider only the estimates from the last immediately preceding round. The social influence component from Eq. 3.4, thus has the following general form

$$\mathcal{F}(x_j, x_i) = w_{ij}(x_j(t-1) - x_i(t-1)). \quad (3.28)$$

The distance between the estimates of two agents, i and j , is scaled by the weight, w_{ij} , that agent i attributes to the estimate of j ¹². We use the following weight function

$$w_{ij} \propto \frac{\mathcal{N}_i}{1 + \exp\left(\frac{|x_j - x_i|}{\alpha}\right)}, \quad (3.29)$$

with a normalization constant \mathcal{N}_i , such that $\sum_{j=1}^N w_{ij} = 1$

$$\mathcal{N}_i = \left[\sum_{k=1}^N \frac{1}{1 + \exp\left(\frac{|x_k - x_i|}{\alpha}\right)} \right]^{-1}.$$

We see that the weight is inversely proportional to the distance between the two estimates. This is similar to bounded confidence models [96, 102] in opinion dynamics, which postulate that inter-individual influence decreases with the disparity between opinions. Typically, such models assume a maximum opinion distance, beyond which agents are deemed too

¹²To simplify notation, from now on we omit the time index in x_i

“different” to interact, and thus to influence each other. Here, we do not have such a discrete threshold, but rather a discounting force which increases with the disparity between estimates. The end result, however, is the same – due to the exponential discount rate, agents distant in estimate space will exert virtually no influence on each other.

The weight function is also in line with the social psychology concept of cognitive dissonance [72]. According to this concept, people experience a discomfort in holding opposing ideas. Consequently, they seek evidence that supports their views and avoid conflicting information. In the same spirit, agents seek proof of being correct and attribute more weight to those with similar estimates.

Finally, the social influence excited by agent j on agent i becomes

$$\mathcal{F}(x_j, x_i) = \frac{\left[\sum_{k=1}^N \frac{1}{1 + \exp\left(\frac{|x_k - x_i|}{\alpha}\right)} \right]^{-1}}{1 + \exp\left(\frac{|x_j - x_i|}{\alpha}\right)} (x_j - x_i).$$

The parameter α_i can again be interpreted as the strength of social influence. Small values for α_i represent individuals who need more evidence supporting alternative opinions and quickly discount distant ones; in other words they are less susceptible to others’ opinions. Conversely large α_i characterises agents who have more tolerance towards opposing views and as a result are influenced by a wider range of opinions.

Summing up over the whole population yields the evolution of agent i ’s estimate

$$\frac{dx_i(t)}{dt} = \sum_{j=1}^N \mathcal{F}(x_j, x_i) + \beta(x_i(0) - x_i(t)) + D\xi_i(t). \quad (3.30)$$

The above dynamics does not easily lend itself to analytical treatment, hence we will simulate it in Section 3.4.2.

3.4.2 Simulation results

Here we present computer simulations of Eq. 3.4 and calculate the resulting long-term collective error, group diversity, and wisdom-of-crowds indicator for all three information regimes. In the aggregate regime, we additionally compare the simulations to the analytical results in the previous section.

The Euler method with a constant time step of $\Delta t = 10^{-2}$ was used to numerically integrate the dynamics of individual estimates in the no- and aggregate-information regimes (Eqs. 3.7 and 3.12), i.e.

$$x_i(t) = x_i(t - \Delta t) + \Delta t \beta [x_i(0) - x_i(t - \Delta t)] + \sqrt{D^2 \Delta t} \cdot \text{GRND}, \quad (3.31)$$

for the no-information case, and

$$x_i(t) = x_i(t - \Delta t) + \Delta t \alpha [\langle x(t - \Delta t) \rangle - x_i(t - \Delta t)] + \Delta t \beta [x_i(0) - x_i(t - \Delta t)] + \sqrt{D^2 \Delta t} \cdot \text{GRND}, \quad (3.32)$$

for the aggregate-information scheme. GRND is a Gaussian random number with mean of 0 and standard deviation of 1. The noise intensity $D = 10^{-3}$.

Due to the multitude of non-linear interactions in the full-information regime, we used the 4th order Runge-Kutta method with time step $\Delta t = 10^{-2}$ to integrate Eq. 3.30. The initial estimates were generated by sampling $N = 100$ random numbers from a log-normal distribution, $\mathcal{L}(\mu, \hat{\sigma}^2)$, with $\mu = -3$ and $\hat{\sigma}^2 = 0.7$. The sample is shown in Figure 3.5.

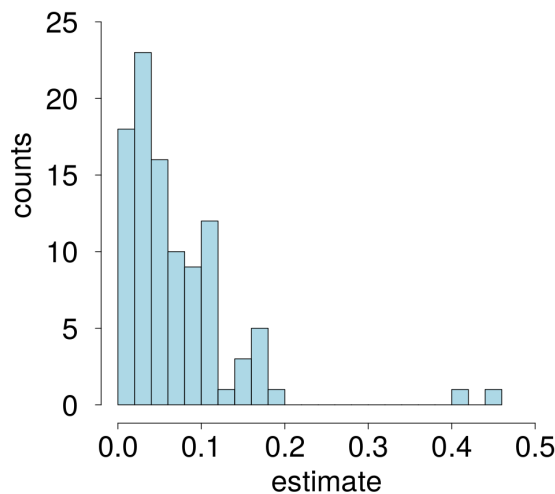


Figure 3.5: Initial distribution of estimates for the three information regimes. It has positive skewness, as required, and its first four central moments are $\langle \sigma(0)^2 \rangle = 0.005$, $\langle \sigma(0)^3 \rangle = 0.001$ and $\langle \sigma(0)^4 \rangle = 0.00004$, i.e. the same as the values we assumed in Section 3.4.1 for the aggregate regime. Higher central moments are decreasing (fifth moment is in the order of 10^{-6}) and eventually vanish.

No information regime. We showed analytically that in this regime the group can be considered quasi static – agents have no incentives to change their estimates (up to random fluctuations), nor do they have any information on which to base such changes. Hence, the group is expected, on average, to maintain its original accuracy, diversity, and wisdom, regardless of the strength of individual conviction.

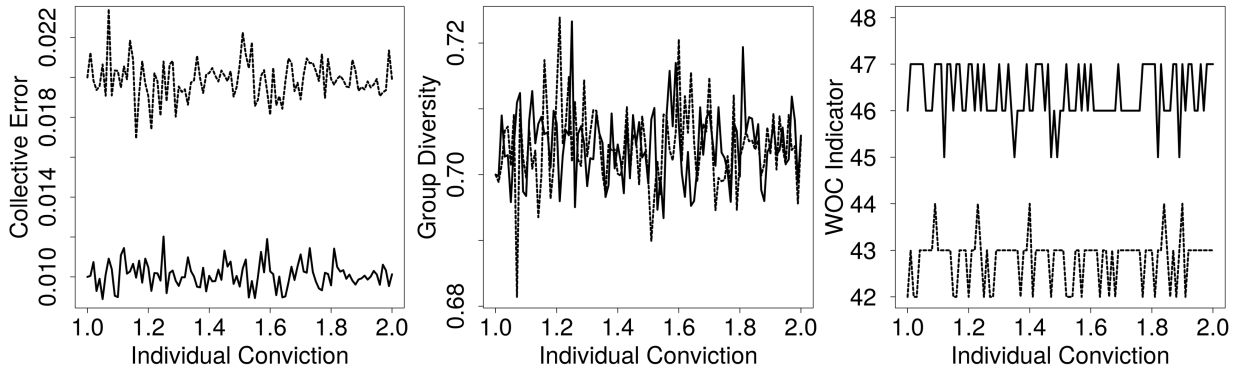


Figure 3.6: Simulation of Eq. (3.7). In the no information regime, the long-term collective error \mathcal{E}_{LT} (left), group diversity \mathcal{D}_{LT} (middle), and wisdom of crowds indicator \mathcal{W}_{LT} fluctuate around their starting values, regardless of the strength of individual conviction, β . In each plot and for each β , we have simulated the no-information regime for $T = 10^4$ time steps. Plotted is the value of the respective aggregate measure after the last time step. Parameters: $\ln \mathcal{T} = -2.1.9$, $\mathcal{E}(0) = 0.01$ (solid lines), $\ln \mathcal{T} = -3.12$, $\mathcal{E}(0) = 0.02$ (dashed-lines)

Figure 3.6 shows a simulation of the no-information regime for two different starting configurations – one at $\{\mathcal{E}(0) = 0.01, \mathcal{D}(0) = 0.7\}$ (solid lines), and the other at $\{\mathcal{E}(0) = 0.02, \mathcal{D}(0) = 0.7\}$ (dashed-lines). The long-term behaviour of the three aggregate quantities of interest are characterised by random fluctuations around their starting values, which is in good agreement with the empirical results in Figures 3.2 and 3.3.

Aggregate information regime. Figure 3.7 presents the simulation of the aggregate regime. We have kept all model parameters identical to those in the analytical investigation of the aggregate regime in Section 3.4.1. There is an excellent quantitative match with the analytical results. The small deviations are likely due to our zeroing of central moments higher than four. Additionally, the simulated group diversity shows the largest mismatch, which is likely a consequence of the first-order Taylor approximation we used to derive it.

Qualitatively, we can observe identical long-term regimes. As $\langle \ln x(0) \rangle < \ln \mathcal{T}$ in (C), \mathcal{E}_{LT} decreases constantly due to the rightward-only motion of $\langle \ln x(t) \rangle$. Therefore (C) illustrates the positive effect of any non-zero amount of social influence on collective accuracy.

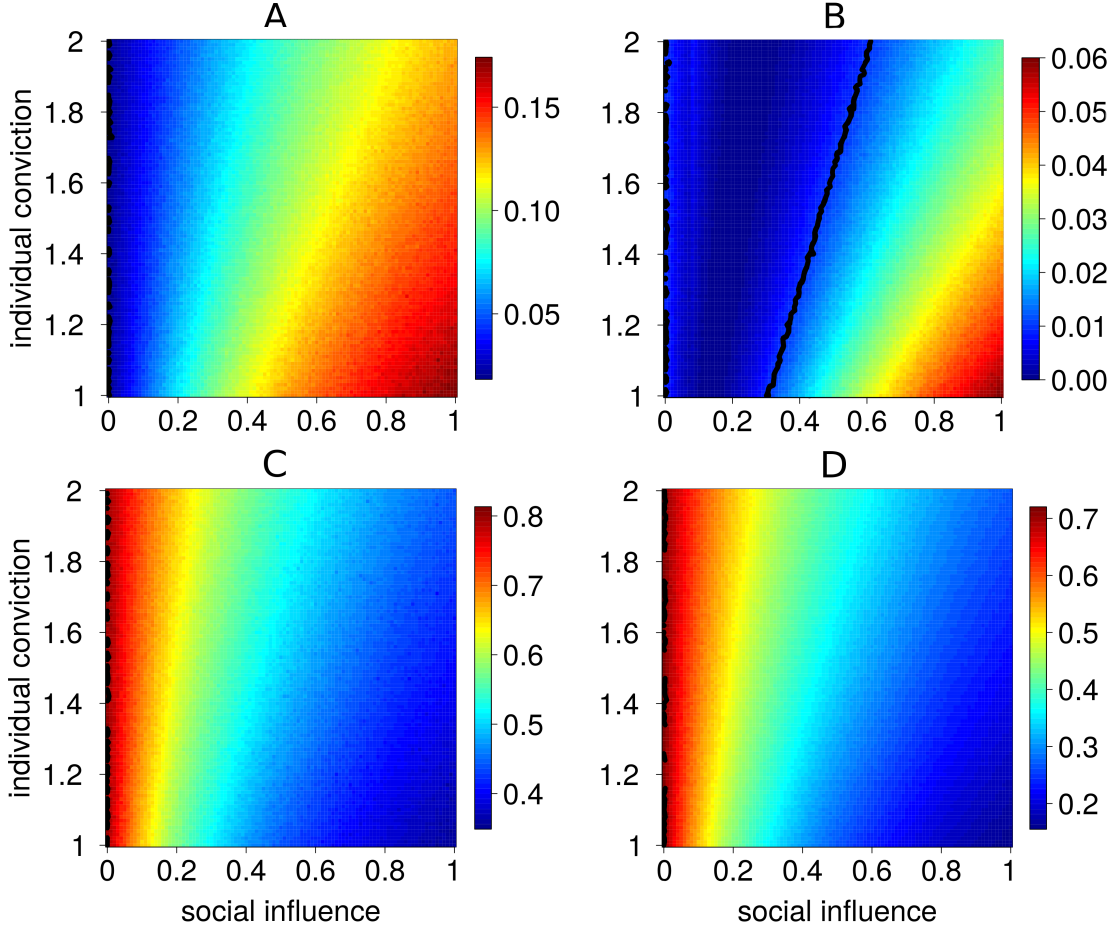


Figure 3.7: Simulation of the long-term collective error, \mathcal{E}_{LT} , and group diversity, \mathcal{D}_{LT} , in the aggregate regime. The interpretation of the plots is the same as in Figure 3.4. Colours indicate \mathcal{E}_{LT} (A,B and C), and \mathcal{D}_{LT} (D). Note that colour legends are not consistent across plots. Depending on the initial distance to the truth, social influence can lead a population closer or farther from it in the long term. Initial configuration are given by **(A)**: $\mathcal{E}(0) = 0.02$, $\ln \mathcal{T} = -3.12$, $\langle \ln x(0) \rangle = -3$, $\mathcal{D}(0) = 0.7$, **(B)**: $\mathcal{E}(0) = 0.01$, $\ln \mathcal{T} = -2.1.9$, $\langle \ln x(0) \rangle = -3$, $\mathcal{D}(0) = 0.7$, and **(C)**: $\mathcal{E}(0) = 0.8$, $\ln \mathcal{T} = -2.1$, $\langle \ln x(0) \rangle = -3$, $\mathcal{D}(0) = 0.7$. Black contour lines indicate regions where the long-term collective error equals its starting value.

In (A) and beyond the second 0.01 contour line in (B), we have the empirical finding that social influence does not decrease the collective error, while at the same time reducing group diversity. Finally, the range between the two contour lines in (B), represents the non-linear effect that up to a certain point, moderate amount of social influence can indeed enhance group accuracy.

Turning our attention now to the wisdom of crowds indicator, \mathcal{W} , we plot its long-term behaviour in Figure 3.8. Let us focus on the left-most plot. This is the starting configuration

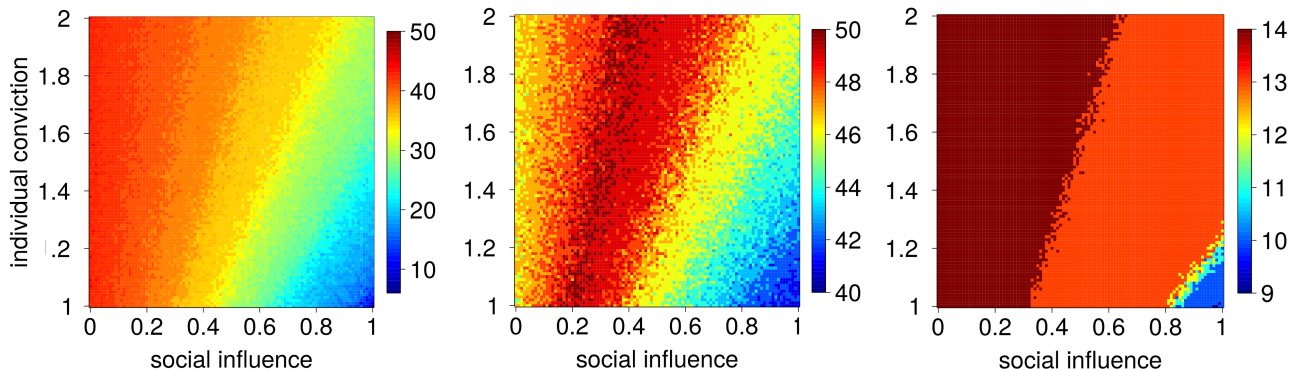


Figure 3.8: Simulation of the wisdom of crowds indicator, \mathcal{W}_{LT} , in the aggregate information regime. Colour indicates \mathcal{W}_{LT} . Note that colour legends are not consistent across plots. The three plots represent the three different starting configuration from Figures 3.4 and 3.7. **Left:** $\mathcal{E}(0) = 0.02$, $\ln \mathcal{T} = -3.12$, $\langle \ln x(0) \rangle = -3$, $\mathcal{D}(0) = 0.7$, $\mathcal{W}(0) = 43$. **Middle:** $\mathcal{E}(0) = 0.01$, $\ln \mathcal{T} = -2.1.9$, $\langle \ln x(0) \rangle = -3$, $\mathcal{D}(0) = 0.7$, $\mathcal{W}(0) = 46$. **Right:** $\mathcal{E}(0) = 0.8$, $\ln \mathcal{T} = -2.1$, $\langle \ln x(0) \rangle = -3$, $\mathcal{D}(0) = 0.7$, $\mathcal{W}(0) = 14$.

from Figure 3.7(A), characterised by relatively small initial collective error, $\mathcal{E}(0) = 0.02$. However, since $\langle \ln x(0) \rangle = -3 > \ln \mathcal{T} = -3.12$, $\mathcal{E}(t)$ will monotonously increase, hence the distribution of estimates will slowly drift away from the truth and converge to the wrong value. The consequence on the wisdom of crowds indicator would be that the range of estimates needed to bracket the truth becomes wider with time, therefore the long-term wisdom of the group will necessarily be smaller than at the beginning. As the transition from red to blue regions shows, the extent to which \mathcal{W}_{LT} decreases depends on the strength of social influence and individual conviction.

The same convergence toward a wrong value can be observed in the middle plot, which is the starting configuration of Figure 3.7(B). The parameter range where $\mathcal{E}_{LT} > \mathcal{E}(0)$ (blue to yellow regions) indicates that the wisdom of crowds is lower than at the beginning. The reason is that the strength of social influence is high enough (or alternatively the strength of individual conviction is low) to not only move the group toward the truth (since $\langle \ln x(0) \rangle = -3 < \ln \mathcal{T} = -2.9$), but to “overshoot” it and drive the population beyond it. Note that with strong social influence and weak individual conviction, e.g. $\alpha = 1$, $\beta = 1$, the group ends up as wise as the configuration from Figure 3.7(A). We can, thus, predict that by increasing α beyond 1, \mathcal{W}_{LT} will diminish further. These considerations indicate that we have essentially reproduced the range reduction effect described by Lorenz et.al. – the truth is displaced to peripheral regions of the opinion distribution and the group becomes narrowly centred around a wrong value.

However, as with the collective error, we find that social influence can be beneficial for the long-term wisdom of crowds indicator. Within almost the same $\{\alpha, \beta\}$ range as in

Figure 3.7(B), \mathcal{W}_{LT} actually grows with α . It even achieves its theoretical maximum of $N/2 = 50$, meaning that the truth is bracketed by the two most central estimates (darkest red region). Therefore, despite the diminished group diversity, it is the favourable initial conditions that allow for such moderate values of α to distribute the final estimates in a narrow range around the truth.

So far we only discussed initial configurations in which the initial collective error is close to the truth from either side. Let us, instead, take the group from Figure 3.8(right) with $\langle \ln x(0) \rangle = -3$, $\mathcal{T} = -2.1$ and $\mathcal{E}(0) = 0.8$. For this group the long-term collective error always decreases with the strength of social influence (Figure 3.7C). At the same time, however, the group becomes more homogeneous. The net effect for \mathcal{W}_{LT} would be negative, as the truth would be further displaced to outer regions of the estimates distribution with increasing α . In other words, the convergence toward the true value, as measured by the diminishing \mathcal{E}_{LT} , is not strong enough to compensate the loss in diversity. Quantifying such groups as less wise reflects our rationale for what constitutes a wise crowd. It is not enough for the group to be correct on average if, at the same time, most individuals are narrowly centred away from the truth. Such groups consist of likely-minded individuals and do not possess the necessary diversity to adapt to new problems.

Finally, recall the opposing effect of individual conviction and social influence in Figure 3.4. It is also present for the wisdom of crowds indicator – regions with larger values are reached with increasing individual conviction and vice versa.

Full information regime. It is quite interesting to see that changing the nature of the social coupling – from a mean-field to a fully connected system, does not affect, qualitatively, our insights about the long-term behaviour of the population. This is particularly useful, as it allows us to construe general conclusions for all three information regimes.

In Figures 3.9 and 3.10, we have simulated the three familiar starting configurations. The same interplay between positive and negative effects of social influence on one hand, and the initial collective error and diversity on the hand, can be identified for \mathcal{E}_{LT} , \mathcal{D}_{LT} and \mathcal{W}_{LT} .

An important difference to the aggregate regime, however, is that the same strength of social influence, as quantified by the parameter α , does not have the same long-term effect on the group level. In fact, for a given combination of $\{\alpha, \beta\}$, the dynamics of \mathcal{E}_{LT} , \mathcal{D}_{LT} and \mathcal{W}_{LT} is accelerated compared to the aggregate-regime. The reason is that with full information the number of influence sources acting on an individual are all other $N - 1$ individuals, and not only the average estimate. Essentially, the magnitude of the estimate change is higher the more information an individual has about the guesses of others. The

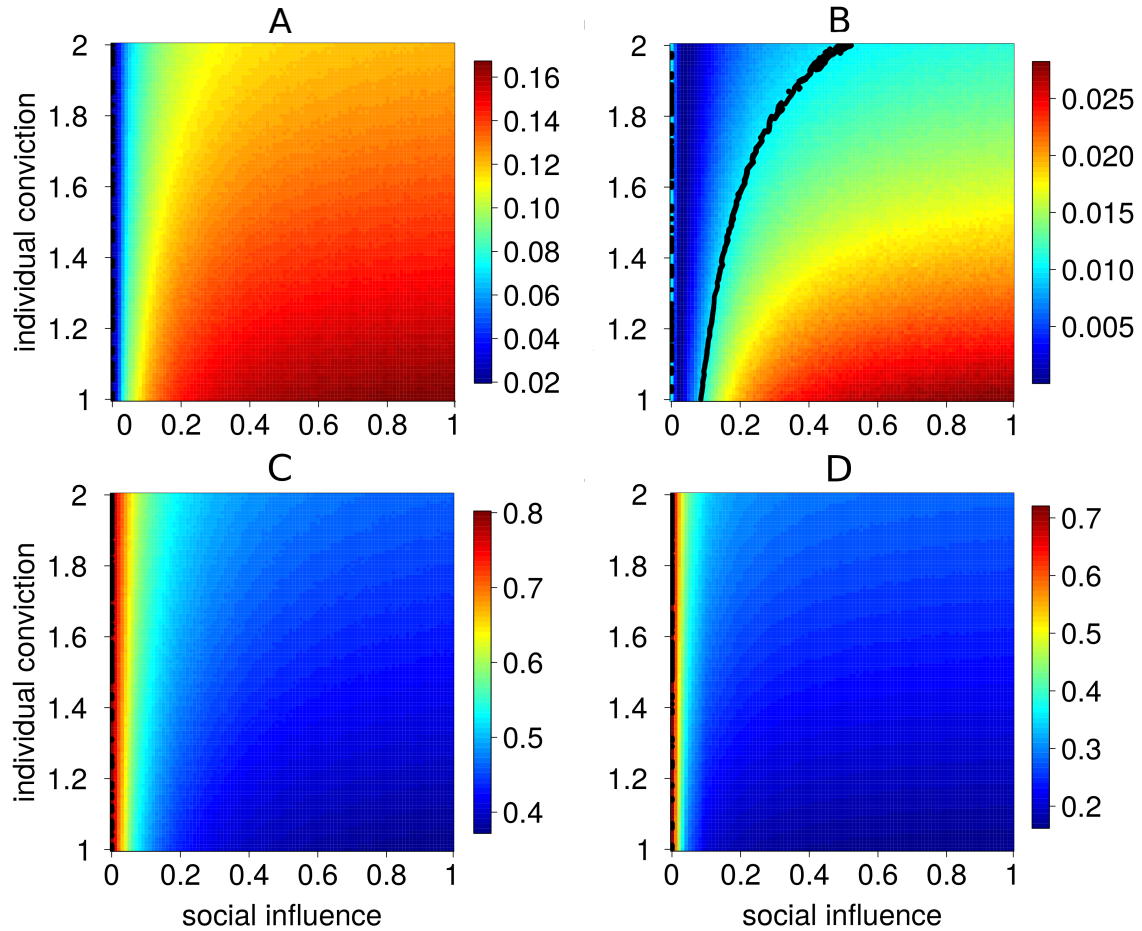


Figure 3.9: Simulation of the long-term collective error, \mathcal{E}_{LT} , and group diversity, \mathcal{D}_{LT} , in the full-information regime. The interpretation of the plots is the same as in Figure 3.4. Colours indicate \mathcal{E}_{LT} (A,B and C), and \mathcal{D}_{LT} (D). Note that colour legends are not consistent across plots. Depending on the initial distance to the truth, social influence can lead a population closer or farther from it in the long term. Initial configuration are given by **(A)**: $\mathcal{E}(0) = 0.02$, $\ln \mathcal{T} = -3.12$, $\langle \ln x(0) \rangle = -3$, $\mathcal{D}(0) = 0.7$, **(B)**: $\mathcal{E}(0) = 0.01$, $\ln \mathcal{T} = -2.1.9$, $\langle \ln x(0) \rangle = -3$, $\mathcal{D}(0) = 0.7$, and **(C)**: $\mathcal{E}(0) = 0.8$, $\ln \mathcal{T} = -2.1$, $\langle \ln x(0) \rangle = -3$, $\mathcal{D}(0) = 0.7$. Black contour lines indicate regions where the long-term collective error equals its starting value.

result of this acceleration is a general contraction of the regions in which social influence is beneficial. Considering, for example, Figures 3.7(B) and 3.9(B), we see that the blue area between the two contour lines is smaller with full information. Similar conclusion can be made when comparing the wisdom of crowds indicator.

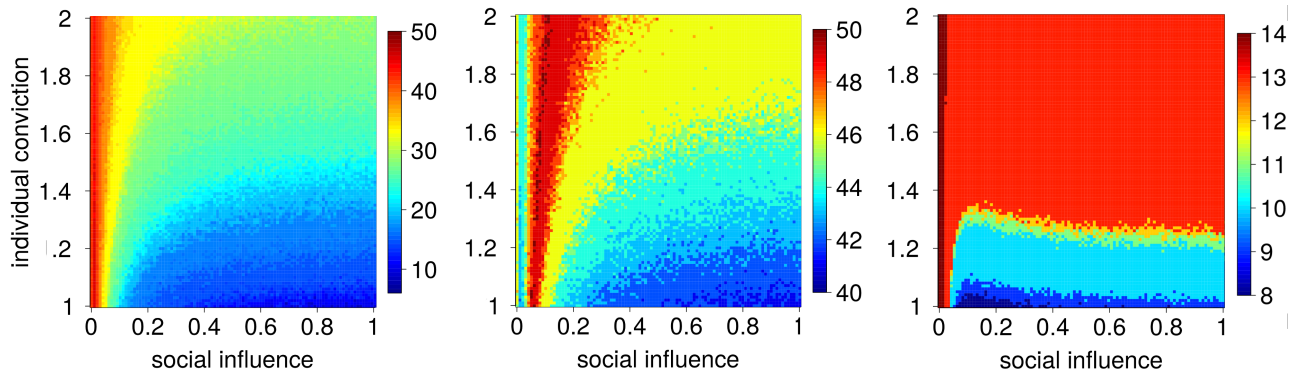


Figure 3.10: Simulation of the wisdom of crowds indicator, \mathcal{W}_{LT} , in the full-information regime. Colour indicates \mathcal{W}_{LT} . Note that colour legends are not consistent across plots. The three plots represent the three different starting configuration from Figure 3.4. **Left:** $\mathcal{E}(0) = 0.02$, $\ln \mathcal{T} = -3.12$, $\langle \ln x(0) \rangle = -3$, $\mathcal{D}(0) = 0.7$, $\mathcal{W}(0) = 43$. **Middle:** $\mathcal{E}(0) = 0.01$, $\ln \mathcal{T} = -2.1.9$, $\langle \ln x(0) \rangle = -3$, $\mathcal{D}(0) = 0.7$, $\mathcal{W}(0) = 46$. **Right:** $\mathcal{E}(0) = 0.8$, $\ln \mathcal{T} = -2.1$, $\langle \ln x(0) \rangle = -3$, $\mathcal{D}(0) = 0.7$, $\mathcal{W}(0) = 14$

3.5 Conclusion

The contributions of this chapter can be summarised as two main points. First, we demonstrated some unexpected negative consequences of herding behaviour that serve as a contrast to the message from Chapter 2. Second, we developed an agent-based model that helped us obtain a deeper understanding of these consequences. The model insights also allow us to create a testable assumption about the general effects of information about others on collective accuracy. Indeed, this is the topic of the next chapter.

In the beginning of the chapter we introduced the wisdom of crowds. This is the phenomenon that the aggregate of a diverse set of opinions tends to be closer to the true value than any single estimate in the group. The wisdom of crowds is supported and observed by numerous anecdotal, empirical and theoretical investigations, across a variety of settings [99, 153, 205, 246]. Importantly, the nature of the phenomenon is not psychological, but statistical, as it is based on a mathematical aggregation of individual opinions. Nevertheless, social influence can affect individual decision-making and, thus, also the statistical aggregate, as well. Social influence causes herding behaviour when individuals perceive an implicit pressure to conform, especially when their opinions are made public. Often, such herding comes at the expense of an individual’s own judgement and is detrimental to the collective wisdom. However, when conformity pressure is absent and

payoffs are unilaterally linked to individual accuracy, one could argue that social influence would lead to herding behaviour only when the individual believes it is in his best interest. From this perspective, the more information individuals have at their disposal, the more informed decisions they will make, resulting in wise crowds

In Section 3.2, we presented evidence from a guessing-task that investigated the extent to which information about others' guesses influences individuals decision-making and affects the wisdom of crowds. The study was designed with the explicit intention to eliminate conformity pressure and to reward individual selfishness. It was found that instead of promoting, social information undermined the crowd wisdom. Individuals engaged in a herding process that slowly reduced diversity in the group without improving its collective accuracy (social influence effect). As a consequence, the truth was displaced to peripheral regions of the estimate distribution and the group became narrowly centred around the wrong answer (range reduction effect). Simultaneously, as the group converged away from the truth, individuals became more confident in their answers. One can, thus, conclude, that more information makes us believe we are taking better decisions, whereas in fact collectively we become increasingly misguided.

In the rest of this chapter, we aimed at understanding how social information, when introduced with incentives that reward individual accuracy, still leads to herding behaviour that undermines the wisdom of crowds. This is a pressing question as social information is practically ubiquitous – groups are embedded in social contexts, which invariably couple the individuals within them. Democracies assume and rely on extensive public discussions to form opinions and create policies. Our behaviour as consumers, investors, voters, etc., is influenced by discussions with, among others, friends, colleagues and experts. From a policy-maker's perspective, this translates to whether a government can reliably harness the wisdom of crowds in practice, i.e. in settings where social influence is unavoidable [49]. Understanding the mechanisms by which social influence positively or negatively affects the wisdom of crowds, becomes then important for evaluating the trustworthiness of crowd predictions.

To this end, Section 3.4 introduced a general agent-based model of decision-making that can be applied to the three information conditions in the empirical study. The model consists of a population of agents endowed with a minimum set of cognitive abilities. The agents continuously revise their estimations, based on their beliefs in their own estimations (individual conviction term) and the influence they perceive from the rest of the population (social influence term). We quantified the long-term dynamics of three indicators measuring the performance of the population: (i) the collective error, (ii) group diversity and (iii) wisdom of crowds indicator. Analytically and numerically, we demonstrated that

groups whose initial average opinion is relatively far from the truth benefit from stronger social influence. Large initial collective error can also be decreased by larger initial group diversity. Social influence, however, is detrimental to groups, which start at a relatively accurate configuration and thus suffer from an excessive drift of the aggregate estimate. In these cases, the initially small collective error quickly reduces to zero, but then continues to increase even beyond its starting value, due to the persistent motion of the aggregate estimate. Small to moderate amounts of social influence are thus more beneficial for such groups. Finally, other initial conditions exist, where even trace amounts of social influence lead to deterioration in the long-term collective error. This constitutes the social influence effect from Lorenz *et al.* [158]. The above discussion applies analogously to the wisdom of crowds indicator – the quantification of how far the median estimate of the population is from the truth. We found starting configurations which lead groups to be less wise in the long-term, for any amount of social influence, i.e. the range reduction effect from Lorenz *et al.* [158], and configurations where groups end up wiser in the presence of moderate strength of social influence.

Based on these observations, the main result of our model is that social influence, and the herding behaviour it causes, does not directly influence the wisdom of crowds. Rather it is the starting configuration of the population, in terms of its initial collective error and group diversity, which determines the long-term benefits or harms of social influence. This insight is testable, and is a topic of Chapter 4. It also suggests how crowds can be driven to better collective outcomes by modulating the strength of social influence. Given some intuition about how inaccurate the crowd initially is, we suggest that a policy-maker either promotes social influence processes or increases the strength of individual conviction. For example, the number of discussion rounds in the Delphi method can be increased, when the group consists primarily of non-informed individuals, to promote influence processes [156]. Conversely, one or two rounds would suffice for a group of experts to decide as a wise crowd.

It is important to stress that this result is applicable only when individuals do not possess direct knowledge about the objective truth (except for idiosyncratic knowledge that helps in forming initial opinions), nor do they learn or receive information that can directly lead them towards it. In other words, there is no feedback between an agent’s opinion, at any given time, and his or her distance from the truth. Consequently, social information about the average or the whole group affects relevant system-wide properties (the geometric mean in our case), but not the collective error or wisdom of crowds indicator. Therefore, even when modelling social influence differently, our result would qualitatively hold, as long as no feedback between the objective truth and agents’ performance is present.

Lastly, the current model presumes the existence of an objective ground truth for the problem at hand, and ignores learning. Arguably, this is not the case in many real-world situations (e.g. financial markets, political polls, etc.). However, individuals in our model do not possess perfect knowledge of this truth. We have shown that the long-term configuration of the group is driven solely by herding begot by social information. The ground truth is needed ex-post only to quantify the state of the population, in terms of the collective error and the wisdom of crowds indicator. Therefore, our proposition that it is the crowd's starting configuration that ultimately determines the effect of social influence can be generalised to these scenarios without an objective truth as well.

Chapter 4

Designing wise crowds: the importance of ranking and competition

Summary

In this chapter we continue our investigations on the wisdom of crowds phenomenon in two main directions. First we focus on crowds composed of individuals without individual expertise in forming even approximate guesses for the correct solution. As we argue, individual expertise is one of the prerequisites for the phenomenon, hence without explicit intervention the collective judgement of such crowds is unreliable. By means of experimental study we show two mechanisms that can restore the wisdom of crowds in these circumstances – ranking and competition. In other words, making individuals aware of their relative performance and incentivising them to perform better can produce collective wisdom *without* consistent individual success. This means that the collective effect is truly the unintended result of individual actions. Second we demonstrate the importance of the crowd's initial configuration for the convergence time to the correct solution. We find that initial diversity can more than compensate initial inaccuracy, so that less wise, but diverse, crowds in the beginning end up outperforming initially more accurate but homogeneous groups.

Based on Mavrodiev, P., Tessone C.J., and Schweitzer F.S. *Designing wise crowds: the importance of ranking and competition*, manuscript pending submission. P.M and C.T. designed the experimental study. P.M programmed the software for the study. P.M. and C.T. conducted the study. P.M. gathered the data, analysed the results and wrote the paper.

4.1 Introduction

We finished the previous chapter with an assumption relating the initial configuration of a group of decision-makers to the group's wisdom when social influence is present. In particular, we claimed that depending on the initial accuracy and diversity of the group, social information affects the wisdom of crowds phenomenon differently. Testing this conjecture is one contribution of the empirical study we introduce in the next section.

Recall that the wisdom of crowds is a statistical phenomenon. As such, it presupposes the existence of an objective truth and quantifiable opinions. Once, these are in place, the concept is well-defined and we can make computations. However, for a crowd to actually be quantified as wise, this is not enough. As we know from Chapter 3, further requirements are (i) independence of individual opinions, (ii) diversity and (iii) individual expertise. Extant research has so far focused on how the phenomenon is adversely affected when the first two requirements are not satisfied. Regarding the first, already in this thesis we presented novel evidence and a theoretical model, which illustrated that even trace amounts of inter-dependence through social information can, under the right circumstances, create convergence toward the wrong value, even when incentives are intended to prevent that. As for diversity, its prominence has spurred creative theoretical [116] and empirical investigations [189] that convincingly argue that groups of individuals with a wider range of perceptions and abilities, possess greater problem-solving skills, are more adaptable to new problems, and should, therefore, be considered wiser than groups consisting of like-minded individuals.

This brings us to the other contribution of the current chapter. We will investigate decision-making scenarios in which the third of the above requirements does not hold, i.e. in which crowds do not possess individual expertise. Lack of individual expertise implies only a vague awareness of the solution space and inability to form educated guesses. As a result, individual guesses could be so disparate that the average opinion may be orders of magnitude away from the correct answer. It has been shown that groups consisting of such uninformed individuals destabilise the capacity for collective intelligence [119, 208], and are unreliable, as the resulting aggregate decision can converge to any unfavourable outcome [182]. Instead of giving up on such groups, however, we would like to find mechanisms that recover the wisdom of crowds in these situations. In the spirit of complex systems, we wish to design group interactions in such a way that a group of uninformed individuals manages to reliably find the right solution, while preserving its diversity.

To illustrate a real-life scenario in which a group lacking individual expertise may be

engaged in decision-making, consider the *Safecast*¹ project. This is a community-driven project, that was kick-started to monitor leftover radiation levels in eastern Japan after the disaster in the Fukushima nuclear power plant. Using handheld devices, citizens living in affected regions, provided accurate, real-time feeds of radiation readings that were used to create a radiation map². In fact, the project was so successful that it is being extended to provide a radiation map for the whole planet. Imagine that, in addition to simply reporting radiation readings, we wish to measure the highest radiation level present in a certain area. This may be needed to assess certain safety risks, e.g. cancer risk [32]. Arguably, individual expertise is hard to acquire for such situations. Even though the half-life of a radioactive element can hint at its long-term concentration, the exact distribution in an open area is largely unknown due to haphazard absorption, such as the formation of radiation hotspots in the soil system [150] or absorption by local marine animals [35]. As a result, even expert judgements about the highest radiation level can be unreliable. Consequently, an aggregated group opinion may not even bracket the correct solution at all. Can we still utilise crowds in this case? We answer affirmatively, and propose an individual-level interaction mechanism that may allow a group to determine such unknown quantity accurately.

In the next section, we introduce a laboratory experiment in which individuals were asked to repeatedly search for an unknown objective solution in a continuous space. Section 4.3 presents the results of the experiment and Section 4.4 concludes the chapter.

4.2 Experiment

4.2.1 Set-up

Our experiment was conducted in a controlled laboratory setting³ with real monetary payoffs. The subjects played a custom-developed computer game, in which they had to guess, over several game rounds, the location of a hidden point, positioned on a circle (Figure 4.1).

A total of 185 students from ETH Zurich and University of Zurich, Zurich, Switzerland were selected as participants. The experiment consisted of nine sessions, split into two categories – four sessions in the experimental category and five sessions in the control category. All sessions consisted of 20 participants, and subjects were aware of which

¹The *Safecast* project - <http://blog.safecast.org/>, last accessed 19.03.2014

²*Japan Geigermap* - <http://japan.failedrobot.com/#>, last accessed 19.03.2014

³<http://www.descil.ethz.ch/>

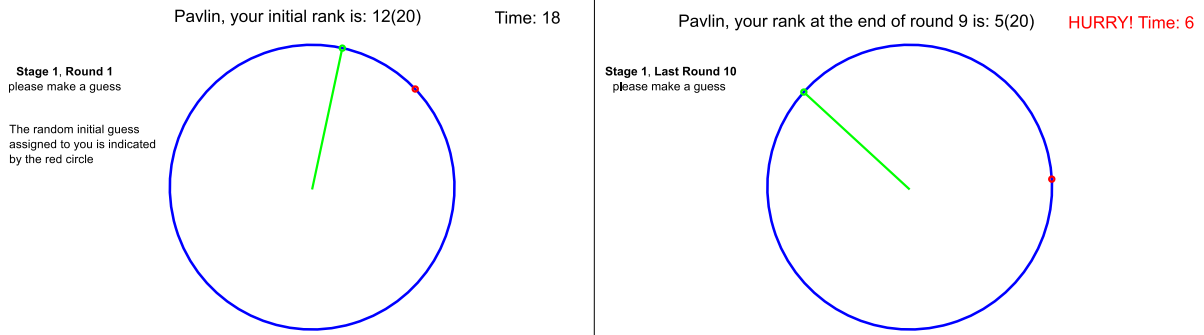


Figure 4.1: The circle as presented to the subjects’ screens. Before making a guess, subjects could see their current rank, together with the total number of players, on top. A guess was made by rotating the green line to a desired position and clicking. The red point shows the individual’s estimate in the previous round or the randomly assigned guess before the first round (see main text). The current stage and round were displayed to the left of the circle. The left- and right-hand plots show two different screen views: one in the beginning of a stage when initial estimates and ranks were assigned, and the other – at the beginning of the last round. A countdown of the remaining time subjects had to make a guess was also displayed on the top right; when less than 10 seconds remained, we changed the message and the colour to raise attention.

category they belonged to.

One session lasted approximately 1 hour. It was divided into 8 stages comprised of 10 rounds each. At the end of each round, participants’ ranks relative to others were displayed above the circle. The individual currently closest to the hidden point had rank 1, and the one farthest away – rank 20. Participants could use this rank information, as they saw fit, before making a guess for the next round. To ensure that subjects were alert, we enforced a time limit of 30 seconds for making a guess, after which a random guess was generated. The end of the final 10th round marked the end of the stage. The game was then reset and played for seven more stages in the given session.

Before round 1 of each stage, initial guesses were generated for every individual by sampling from a wrapped normal distribution, $\mathcal{WN}(\mu, \sigma)$. The parameters of the distribution were different across stages. In this way, we could control the initial configuration of the crowd in terms of initial collective error and diversity (see Table 4.1). Moreover, after their generation, the 20 random initial estimates for each stage were offset by 20 uniformly distributed random numbers in $[-180^\circ, +180^\circ]$. The offset estimates were then shown to the subjects as their starting guesses for this stage. The same offsets were applied to all subsequent guesses for the stage. By effectively decoupling individuals’ perception of the circle and the internal representation of their guesses, we eliminate any spatial biases that subjects may have had, e.g. a preferred direction. Due to this set-up, subjects had no

prior information about, or experience for, neither the correct answer nor successful search strategies. By imposing a circular landscape, we artificially limit the magnitude of one's response. This resembles real-life situations, where the response magnitude is necessarily bounded.

The study followed a strict no-deception policy. Subjects received monetary benefits in Swiss Francs (CHF) after the end of each stage, dependent on their final rank for that stage. Rank 1 received 10.6 CHF, rank 2 – 8.1 CHF, rank 3 – 5 CHF, ranks 4 to 6 – 3.8 CHF, ranks 7 to 9 – 3.1 CHF, ranks 10 to 13 – 1.9 CHF, ranks 14 to 17 – 1.2 CHF and ranks 18 to 20 – 0.6 CHF. Therefore, an individual who consistently finished first in all 8 stages would receive 85 CHF.

Since rank directly determined payoffs, we can see the game as competition to provide the guess closest to the true position of the hidden point. We note, however, that knowing one's rank, though constituting social information about others' actions, is not a direct clue for the location of the hidden point, as it cannot be used to gain knowledge about the solution space. Indeed, one could be ranked first and still be relatively far from the correct position. For this reason, to encourage subjects to do their best in searching for the hidden point, we chose payoffs heavily skewed in favour of the most successful individuals – $\approx 40\%$ of the total rewards for a stage were given to the lowest three ranks.

In the four experimental sessions, all subjects were human participants, guessing according to the rules described above. In the five control sessions, all but one participant were computer agents guessing randomly, i.e. without regard for their ranks. The human participant was aware of this set-up and was offered the same monetary incentives to minimise her own rank, as subjects in the experimental sessions. The control sessions served to isolate the influence of competition on the performance of the group. Since the computer agents did not take rank information into account, the human subject was effectively unchallenged.

Upon arrival at the laboratory, each subject was randomly assigned a cubicle number and seated at an isolated cubicle in front of a computer with no visual, verbal, or chat contact with others. The game was conducted in a web browser deprived of user controls, so that participants could not get distracted by navigating away. Subjects were told all details of the experimental procedures by printed instructions, including their belonging to either the experimental or control category. A freestyle play with the circle in the beginning ensured that everyone had understood the mechanics of the game and felt comfortable with it.

4.2.2 Data analysis

Aggregation of individual opinions. An appropriate aggregation mechanism should result in a collective opinion that reflects the prevailing tendency in the group. A common choice is the unweighted arithmetic average, but a number of alternative measures exists, which are less sensitive to outliers, e.g. the median or the geometric mean. In Figure 4.2, we compared the circular mean and the circular median⁴

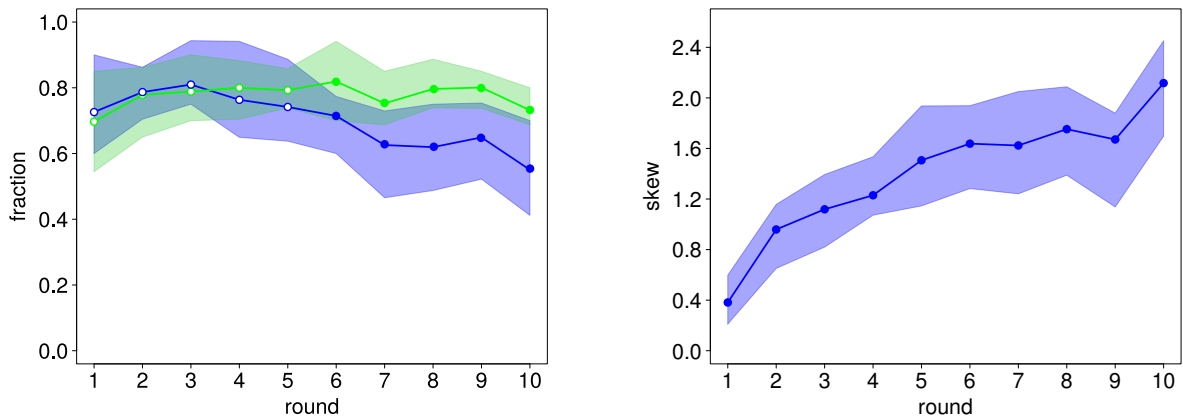


Figure 4.2: **Left:** Comparison of the circular mean and circular median as aggregation measures. For each experimental round, we computed the fraction of users who are farther from the truth than either the mean (blue) or the median (green). The fractions were then aggregated across all 8 stages and 4 sessions, resulting in 32 data points per round. The lines show the mean of all fractions in the corresponding round. Confidence bands are calculated between the first and third quartile. The initial round 0 is not included, as it does not represent a choice. To test the null hypothesis that the mean and median distributions are different we use a bootstrapped Wilcoxon rank-sum test with 10^5 bootstraps. Rejecting the null implies that the median distribution tends to have larger values than the mean distribution. Filled circles indicate statistical significance at 1%. **Right:** The skewness of the absolute distances of individual estimates from the truth. Data aggregation and confidence bands are the same as on the left. The solid blue line shows the mean skewness for a given round.

For each round, aggregated over all stages and sessions, we computed the fraction of individuals that are farther from the true value than each measure. The median is consistently closer to the truth than $\approx 82\%$ of the individuals' estimates, as opposed to the mean which deteriorates after the third round. This is because the absolute distance be-

⁴Circular statistics is necessary to account for the periodicity of the circle. For example, two guesses at 120° and -120° have 0° arithmetic mean and median. This is misleading, because the guesses are closer to 180° than to 0° .

tween individual estimates and the truth became increasingly right-skewed in the course of each stage (Figure 4.2, right). In other words, a majority of low to medium ranked individuals approached the truth, while a minority remained relatively far from it. As a result, the mean is biased toward this high ranked minority, and so it does not reflect the predominant opinion. The median, on the other hand, is resistant against such bias, therefore, we adopt it as an aggregation mechanism.

Data structure. For a set of n individuals, we denote the *original* guess of individual i , for any given round, as x_i . The internal representation of this guess, ϕ_i , is computed by offsetting x_i by a random number $\xi_i \in [-180^\circ, +180^\circ]$, i.e. $\phi_i = x_i - \xi_i$. The aggregate opinion for each round r is the circular median of all ϕ_i 's, denoted by $\langle \phi \rangle_{\text{med}}(r)$. A useful measure for the distance between two angles α and β is ([168], p. 18)

$$1 - \cos(\alpha - \beta). \quad (4.1)$$

Therefore, we define the collective error, $\mathcal{E}(r)$, as the distance between $\langle \phi \rangle_{\text{med}}(r)$ and the truth at 0°

$$\mathcal{E}(r) = 1 - \cos[\langle \phi \rangle_{\text{med}}(r)]. \quad (4.2)$$

The group diversity, \mathcal{D} , is the average deviation from $\langle \phi \rangle_{\text{med}}$

$$\mathcal{D} = \frac{1}{N} \sum_{i=1}^N [1 - \cos(\phi_i - \langle \phi \rangle_{\text{med}})]. \quad (4.3)$$

The initial ϕ_i 's for each stage were sampled from a wrapped normal distribution, \mathcal{WN} , with mean μ , and standard deviation σ . Since the circular median coincides with the circular mean for \mathcal{WN} , the initial collective error is given by $\mathcal{E}_0 = 1 - \cos(\mu)$. Moreover, the initial group diversity is well approximated by $\mathcal{D}_0 = \sigma^2$. We refer to the pair $(\mathcal{E}_0, \mathcal{D}_0)$ as a group's *initial configuration* for a given game stage. Table 4.1 summarises the initial configurations. For simplicity, the sampling from the wrapped normal distribution is not exact. Instead $\mu = \langle \phi \rangle_{\text{med}}(0)$ and σ^2 are taken within the ranges indicated in the table. Despite being an approximation, this approach does not prevent us from distinguishing between different initial configurations and from quantifying their effects on the wisdom of crowds.

We use five types of starting configurations. The first type ($\langle \phi_{\text{med}} \rangle \in (0^\circ, 30^\circ]$, $\mathcal{D}_0 \ll 0.1$) represents populations that are initially accurate and homogeneous. In comparison, the second type ($\langle \phi_{\text{med}} \rangle \in (30^\circ, 50^\circ]$, $\mathcal{D}_0 \ll 0.1$) represents less accurate but still homogeneous groups. The third type, ($\langle \phi_{\text{med}} \rangle \in (0^\circ, 30^\circ]$, $\mathcal{D}_0 > 0.1$), retains the initial accuracy but

	Session 1		Session 2		Session 3		Session 4	
	$\langle\phi\rangle_{\text{med}}(0)$	\mathcal{D}_0	$\langle\phi\rangle_{\text{med}}(0)$	\mathcal{D}_0	$\langle\phi\rangle_{\text{med}}(0)$	\mathcal{D}_0	$\langle\phi\rangle_{\text{med}}(0)$	\mathcal{D}_0
Stage 1	$(0^\circ, 30^\circ]$	$\ll 0.1$	$(0^\circ, 30^\circ]$	$\ll 0.1$	$(0^\circ, 30^\circ]$	$\ll 0.1$	$(0^\circ, 30^\circ]$	$\ll 0.1$
Stage 2	$(0^\circ, 30^\circ]$	$\ll 0.1$	$(0^\circ, 30^\circ]$	$\ll 0.1$	$(0^\circ, 30^\circ]$	$\ll 0.1$	$(0^\circ, 30^\circ]$	$\ll 0.1$
Stage 3	$(30^\circ, 50^\circ]$	$\ll 0.1$	$(30^\circ, 50^\circ]$	$\ll 0.1$	$(30^\circ, 50^\circ]$	$\ll 0.1$	$(30^\circ, 50^\circ]$	$\ll 0.1$
Stage 4	$(30^\circ, 50^\circ]$	$\ll 0.1$	$(30^\circ, 50^\circ]$	$\ll 0.1$	$(30^\circ, 50^\circ]$	$\ll 0.1$	$(30^\circ, 50^\circ]$	$\ll 0.1$
Stage 5	$(0^\circ, 30^\circ]$	> 0.1	$(30^\circ, 50^\circ]$	> 0.1	$(0^\circ, 30^\circ]$	> 0.1	$(50^\circ, 90^\circ]$	> 0.1
Stage 6	$(0^\circ, 30^\circ]$	> 0.1	$(30^\circ, 50^\circ]$	> 0.1	$(50^\circ, 90^\circ]$	> 0.1	$(30^\circ, 50^\circ]$	> 0.1
Stage 7	$(0^\circ, 30^\circ]$	> 0.1	$(30^\circ, 50^\circ]$	> 0.1	$(30^\circ, 50^\circ]$	> 0.1	$(30^\circ, 50^\circ]$	> 0.1
Stage 8	$(50^\circ, 90^\circ]$	> 0.1	$(0^\circ, 30^\circ]$	> 0.1	$(0^\circ, 30^\circ]$	> 0.1	$(0^\circ, 30^\circ]$	> 0.1

Table 4.1: Initial configurations for each stage in all sessions. For the sake of clarity, we show the initial circular median, $\langle\phi\rangle_{\text{med}}$. \mathcal{E}_0 can be calculated as $1 - \cos[\langle\phi\rangle_{\text{med}}(0)]$.

does away with homogeneity. Finally, the last two types represent increasingly inaccurate and heterogeneous groups.

Finally, the offsets ξ_i 's are sampled from a uniform distribution $U(-180^\circ, +180^\circ)$. The offsets are fixed for the duration of the stage. The initial guesses actually displayed to the subjects, and used to calculate the starting ranks for each stage, are thus computed as $x_i = \phi_i + \xi_i$. With this shift, we eliminate spatial bias by effectively decoupling subjects' choices from the internal representation of these choices.

4.3 Results

Wisdom of crowds. How successful was the group in finding the hidden point? Moreover did all individuals converge toward consensus or was there considerable diversity in opinions? Figure 4.3 shows the evolution of the collective error and group diversity in the four experimental sessions. The collective error steadily decreases in all experimental sessions. Moreover, after round 5, on average, the collective error is approximately 10^{-3} , which means that the aggregate estimate is less than 3° away from the true position of the hidden point. This represents a remarkable collective accuracy. Additionally, the group diversity remains high throughout a stage, and is considerable even at the last round when the group is less than 1° from the truth.

To what can we attribute such rapid and consistent decrease of the collective error? Figure 4.4 (left) shows the dynamics of the mean individual error during each round as a function of individual rank. The figure illustrates that lower ranks show stronger decrease in individual error from the initial round to the end of a stage. This improvement gradually declines with higher ranks, until it becomes negative for the last two ranks – on average

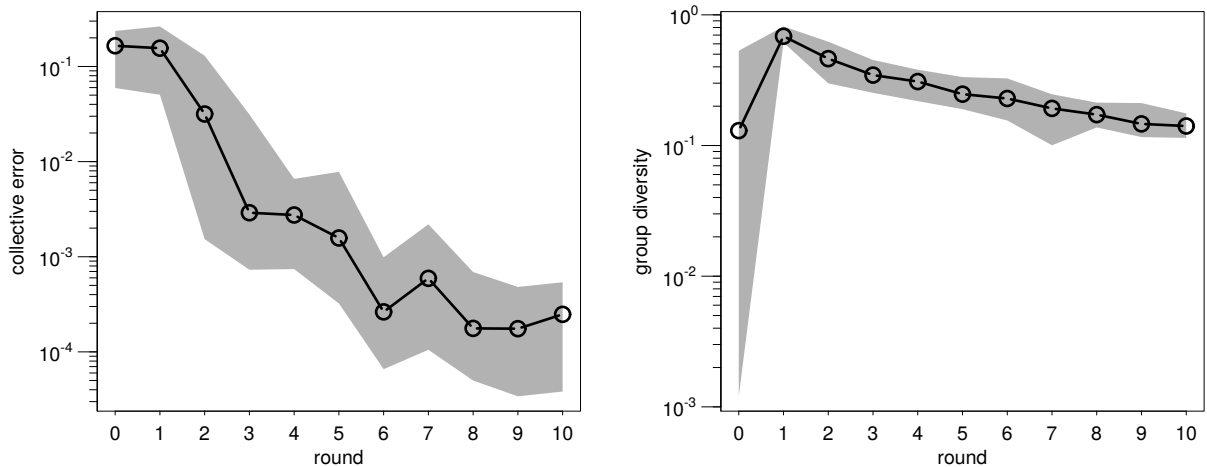


Figure 4.3: The Wisdom of Crowds phenomenon: low collective error and high group diversity. **Left:** Decreasing collective error. For each round, the collective error of all individual estimates is calculated and aggregated over all stages and experimental sessions. With 8 stages and 4 experimental sessions, this yields $8 \times 4 = 32$ data samples for the collective error per round. Solid line shows the median of all samples in each round. **Right:** Consistently high group diversity. Solid line and aggregation procedure as on the left. Confidence bounds in grey are within the first and third quartile for both plots. Round 0 refers to the initial round in which subjects were assigned random estimates.

individuals occupying these ranks actually drift farther from the truth compared to the random positions assigned to them in the beginning. In general, however, the majority of the group settles closely to the truth – the top 15 ranks converge to just within 20° (grey line) of the hidden point after round 10. The results suggest that the wisdom of crowds effect is due to a majority of the group steadily approaching the truth, while a minority remains as far away as possible.

However, the consistent collective convergence to the truth, as indicated by the rank dynamics, does not imply consistent individual success. Figure 4.4 (right), shows that the identity of the best-performing individual changes in more than half of the cases up to round seven, and in about one third of all cases in the remaining three rounds. Considering that as early as round 3 the best individual is on average within 1° of the hidden point, such frequent changes in identity indicate that it was hard for any subject to find a robust winning strategy. This is supported further by calculating the probability of transitioning to a particular rank, given one’s previous rank.

The left plot of Figure 4.5 presents the frequency of changing to any rank at the next round $t + 1$, given the individual’s rank, R_t , at round t . We refer to this frequency as the transition probability to R_{t+1} conditional on R_t . The colour pattern around the main diagonal indicates that most of the times individuals’ ranks were similar to the ranks from

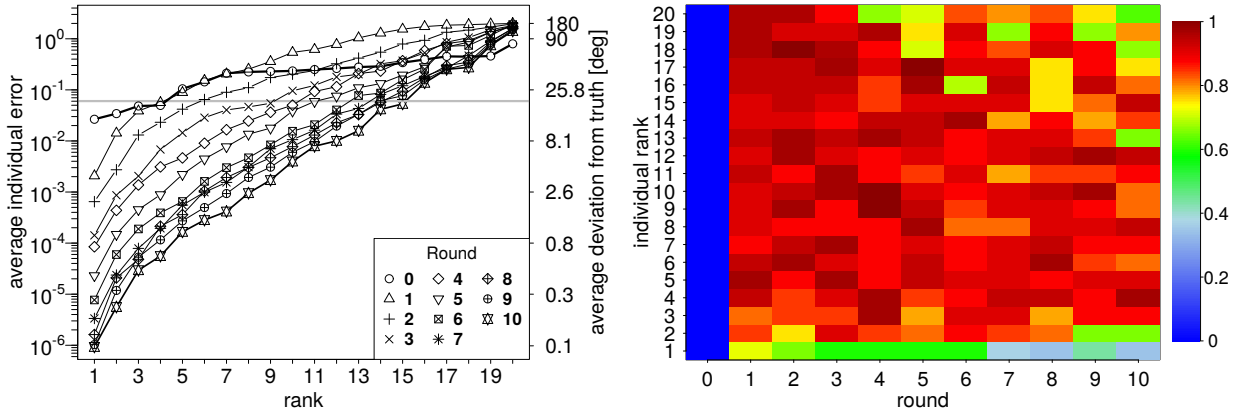


Figure 4.4: **Left:** A majority of the group attains remarkable individual accuracy, while a minority remains far from the truth. Shown are the dynamics of the average individual error during the progression of a stage, as a function of rank. Data aggregation is over all stages of the 4 experimental sessions, i.e. $8 \times 4 = 32$ data samples per round and individual rank. Lines with different point symbols represent the 10 rounds of a stage. The two bold lines serve as visual cues for the initial and final rounds. For illustration purposes, the average individual error is also shown in terms of degrees of deviation from the truth. The grey horizontal line denotes deviation of 20° , corresponding to error of $1 - \cos(20^\circ) = 0.06$. **Right:** While, on average, most ranks increase their accuracy during a stage, the identity of the individual at a given rank changes frequently, even for the 1st rank. Data aggregated as on the left. The colours indicate the fraction of cases (out of 32) when the identity of the individual at a specific rank changed from the previous round. Even the top rank changed his/her identity at the last round in about 40% of the cases.

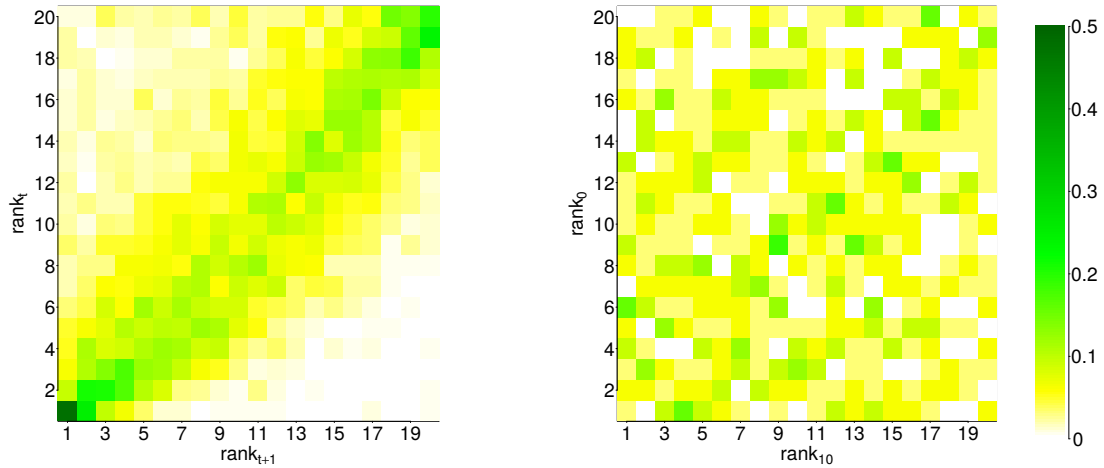


Figure 4.5: Transition probabilities for **(left)** an individual to occupy a given rank in round $t + 1$ conditional on his current rank in round $t + 1$ and **(right)** an individual to finish a stage at a given rank given his randomly assigned initial rank in round 0. Data aggregation is over all experimental sessions: **(left)** 4 sessions \times 8 stages \times 10 rounds = 320 data points, **(right)** 4 sessions \times 8 final stages = 32 data points.

the immediately preceding round, up to small fluctuations. We can also observe slightly higher transition probabilities in the upper diagonal, which confirms our observation that the majority of individuals tend to improve on average during a stage. The right-hand side of the figure is particularly useful in demonstrating that initial ranks, R_0 , are not related to final ranks, R_{10} , at least not in any discernible way. This reassures us that the game was fair in that individuals could not gain advantage from their starting positions.

To quantify these observations, we compute the mutual information (MI) in both plots of Figure 4.5. More precisely, we calculate the mutual information between R_t , R_{t+1} , and the conditional probability of transitioning between them. This quantity, also termed *total correlation* in information theory, measures the extent to which knowing the value of one variable allows the inference of the other two. Referring to the left-hand side of the figure, this amounts to asking the question, “Given my rank now, to what extent can I predict the probability of being at rank X in the next round?”. Similarly, for the right-hand side we can ask, “Given my initially assigned rank, to what extent can I predict the probability of finishing at rank X at the end?”. The advantage of computing MI is that it is sensitive not only to linear correlations, but also to non-linear dependencies not captured in the co-variance [142]. Formally, the mutual information between several variables equals $\sum_{i=1}^n H(x_i) - H(x_1, x_2, \dots, x_n)$, where the first term is the amount of information (i.e. entropy measured by $H(\cdot)$) that the variables would possess if they were totally independent. The second term is the *actual* amount of information that the variable set contains. The difference, then, represents the amount of information shared between all variables (redundancy) and quantifies a certain structure or dependency embodied in the set of variables. The MI equals 0 if the variables are completely independent, and a maximum of $\sum_{i=1}^n H(x_i) - \max H(x_i)$ if one variable is completely determined by another and is, therefore, redundant. We use this upper bound to normalise our measure of MI, so that it is between 0 and 1. The normalised mutual information for the left-hand side of Figure 4.5 is 0.61 and for the right-hand side – 0.34. This already corroborates our observation that there is more structural dependency on the left side. To measure its statistical significance, we shuffled the transition probabilities in both plots 10^4 times, and computed the resulting MI for each shuffled sample. Calculating the percentage of the shuffled MI values that are larger than the original MI allows us to derive a p-value for the null hypothesis that the structure observed in each plot can be obtained by chance. The p-values are 0.03 and 0.4 for the left- and right-hand side, respectively. Therefore, the diagonal pattern is significant in showing a general tendency of individuals to remain with similar ranks in successive rounds. On the contrary, the pattern of start-to-end transition probabilities evidences that an individual’s final position is statistically not related to his starting rank.

Competition. So far, our focus has fallen on the macro-level dynamics of this guessing game. We observed that the groups from all experimental sessions converge to configurations characterised by very low collective error and considerable diversity. Moreover, no single individual is able to consistently outperform the aggregate opinion. As these are the paramount features of the wisdom of crowds phenomenon, we may be tempted to conclude that the mechanism of ranking alone is sufficient in overcoming the inability to form educated guesses. However, to fully understand how groups can collectively outperform any single individual, we need to investigate the micro-level dynamics of estimate formation.

To this end, we group the 20 ranks in 6 groups, or performance categories, as follows – {rank 1, ranks 2-4, ranks 5-7, ranks 8-12, ranks 13-15, ranks 16-20}. The motivation for such grouping is to reduce the complexity associated with analysing the behaviour of 20 separate individuals, while still retaining the salient characteristics of the dependency between rank and response. Next, we compute the ratio between the ranks of all individuals within a given performance category in the beginning of a given round, and the ranks of these individuals in the end of that round (equivalently in the beginning of the next round), i.e. after they have reacted by adjusting their guesses. We call this ratio the average *relative improvement* of this category in the given round. For example if the first-ranked individuals in the beginning of round 5 always declined to rank 2 in the beginning of round 6, the relative improvement of performance category 1 in round 1 would be $1/2 = 0.5$. Similarly, a relative improvement of 2 for performance category 16–20 in round 10, indicates that individuals with ranks between 16 and 20 in the beginning of the last 10th round finished on average with ranks 8 to 10.

The left-hand side of Figure 4.6 shows the average relative improvement of all performance categories during the progression of a stage.

There is a clear indication that individuals in the worst performance category (black colour) show persistent relative improvement larger than one for all ten rounds. Ranks 13–15 perform neutrally hovering around the black borderline, which implies that individuals in this group are on average likely to remain in it. As we move to higher performance categories, we can see a general decline of relative improvement below one, with the average highest-ranked individual consistently losing his position in the first 6 rounds. This is to an extent expected as lower-ranked individuals, being already quite accurate, have fewer opportunities to improve and are more likely to fall behind.

Additionally, the relative improvement of these performance categories hints to the micro-dynamics of estimate change. It suggests that individuals become increasingly frustrated the higher their ranks are and deviate more from their previous guesses. Therefore, those with higher ranks would cover larger portion of the solution space and would tend to

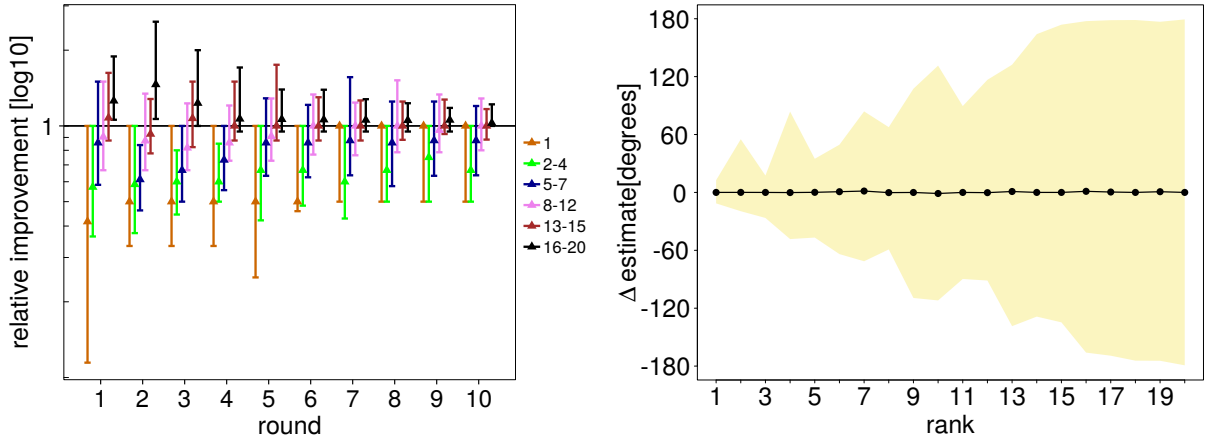


Figure 4.6: **Left:** Average relative improvement for the six performance categories. Error bars span the inter-quartile range; symbols indicate the median. Black horizontal line denotes relative improvement of 1, which means that individuals remain within their respective performance category. Data for each category were aggregated over all 4 experimental sessions: 4 sessions \times 8 stages \times {#ranks in a given category}. **Right:** Average estimate deviation for all ranks. The deviation is computed as the difference between the guess of an individual at a particular rank and the guess of the same individual in the next round. Each rank contains 4 sessions \times 8 stages \times 10 rounds = 320 data points. Coloured bands span the inter-quartile range; point symbols indicate the median.

increase their relative improvement on average. Conversely, individuals in the top performance categories are satisfied with their expected rewards, and find it riskier to deviate considerably from their previous guesses. As a result, they end up being surpassed by others, who through greater exploration produce more accurate guesses. This micro-dynamics is supported by the right-hand side of Figure 4.6 which shows the average estimate deviation of each rank for all experimental sessions. It is evident that on average lower-ranked individuals constrict their exploration to a narrow range around their current position, while those ranked among the last, cover virtually the complete solution space. Figure 10.1 in Appendix 10.1 confirms this conclusion even when considering a more fine-grained view of the per-round individual responses.

Based on the above discussion we can conjecture that ranking alone would not suffice in establishing the wisdom of crowds effect in our experiment. What seems to be needed in addition is a *competition* mechanism that keeps individuals engaged at all times by compelling them to widen their exploration range, as soon as they fall into a lower performance category. To verify this hypothesis we compare the collective performance of the four experimental and the five control sessions. Recall that in each control session, only one participant was human; the other 19 subjects were computer agents guessing randomly. The human subject was aware of this set-up and was offered the same monetary

rewards as participants in the experimental sessions.

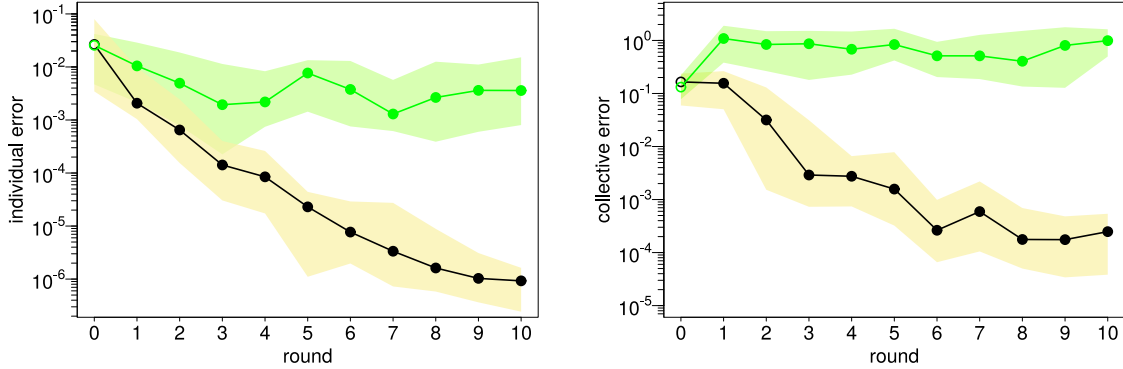


Figure 4.7: Comparison of individual and collective performance between the experimental (yellow) and control sessions (green) over the 10 rounds of a stage. Round 0 represents the initial round of random assignment. Confidence bands span the inter-quartile range. A bootstrapped Wilcoxon rank-sum test (10^4 bootstraps) was used to test the null hypothesis of difference between the two distributions in both plots. Points represent the medians. Full point symbols indicate significant difference at the 1% level. **Left:** Individual collective error for the best performing agents (i.e. those ranked first) in both categories. **Right:** Collective error in the experimental and control sessions.

Since the computer agents did not take their ranks into consideration, they provided no competition to the human participants. Therefore, it is not surprising that in each control session, the human subject earned the highest cumulative reward at the end of all 8 stages, and was thus the most successful participant. Moreover, averaged over all control sessions, the human was among the top three ranks at the end of 7, out of 8, stages. In contrast, the best performing individual in the experimental sessions was on average among the top three ranks at the end of only 3 stages.

However, the individual success (in terms of rank minimisation) in the control sessions did not translate to collective wisdom. Figure 4.7 compares the experimental and control sessions by illustrating the individual error of the best performing participant (left), and the collective error (right) in each round. Focusing on the right-hand side, we see that, without competition, the collective error in the control sessions never decreases, and actually increases on average after the initial round. At the end of round 10, it settles to a value of more than 3 orders of magnitude larger than the corresponding collective error in the experimental sessions. Referring to the left-hand side, the top-ranked human subjects in the control sessions, did not show considerable improvement in accuracy in all 10 rounds. It is particularly interesting that the individual accuracy of these subjects at the end of the 10th round was four orders of magnitude worse than the corresponding

individual accuracy of the top-ranked participant in the experimental sessions. Even more striking is that at the end of a stage, the collective accuracy of the group in presence of competition, exceeded by about 2 orders of magnitude the *individual* accuracy of the top-ranked human subjects in the control sessions.

Based on the above discussion, we conclude that indeed the stark decline in individual accuracy and collective wisdom arises from the absence of competition. The rationale is that without competition from others, the human participants are never effectively challenged. Once they achieve the top rank, there is no pressure for them to deviate from their guesses. Computer agents can outperform them only by chance, which happens rarely. Even when it does happen, the effects are short-lived, as the computer agents will not maintain their superior position. As a result, the exploration range of the humans participants remains narrow.

Effects of the initial configuration. Now that we have a good understanding of the two ingredients that create the wisdom of crowds effect, let us turn our attention to the influence of the starting configuration. In Figure 4.8, we compare several qualitatively different groups with the goal of isolating the effects of the initial collective error and group diversity on the wisdom of crowds.

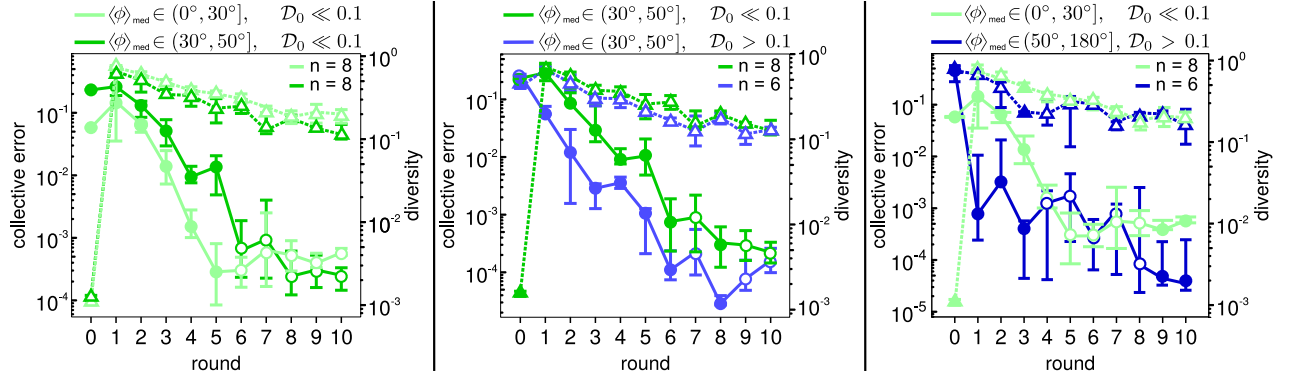


Figure 4.8: Effects of initial conditions on the wisdom of crowds. The three plots compare the dynamics of the collective error (solid lines) and group diversity (dotted lines) during a stage among groups with different starting configurations. The starting configurations are displayed above each plot. For clarity, we show the initial value of the circular median $\langle \phi \rangle_{\text{med}}$; \mathcal{E}_0 can be calculated as $1 - \cos \langle \phi \rangle_{\text{med}}$. The sample size for each round depends on the number of game stages (see Table 4.1) that correspond to the particular starting configuration and is shown as inset. Due to the small sample sizes, we tested equality of distributions with a bootstrapped Wilcoxon rank-sum (10^4 bootstraps). Point symbols represent the median collective error and diversity at a given round. Furthermore, full point symbols indicate significance at the 5% level.

The left-hand side juxtaposes two, initially homogeneous and relatively accurate crowds (referring to Table 4.1 these are the first two configuration types). Unsurprisingly, the crowd that is more accurate in the beginning, maintains a lower collective error for half the duration of a stage; after that the second group “catches up” and both settle at a very accurate aggregate opinion about 1.4° from the truth. We also note that due to the ranking mechanism and competition dynamics both groups display high and statistically similar group diversity for all 10 rounds. Therefore, when two groups both have low diversity, the initial accuracy is what ultimately determines for which one the wisdom of crowds effect is stronger.

Next, in the middle plot, we compare groups having the same initial collective error, but one exhibiting higher heterogeneity. Again, the dynamics of the group diversity is statistically the same for both. Interestingly, however, the more diverse group shows consistently lower collective error most of the time, especially in the first half of the stage when, in general, groups progress the most. That higher initial diversity is conducive to higher accuracy is further supported on the right-hand side. There, we clearly see how diversity can more than compensate for collective inaccuracy – the initially less “wise” group quickly overcame its disadvantage and even outperformed the “wiser” group in the first few rounds of the stage. Eventually, both groups settled to states characterised by low collective error and high diversity.

4.4 Conclusion

This chapter explored further the wisdom of crowds phenomenon in two main directions. Following up on our results regarding the unintended effects of social influence, we first delved into the *design* of individual-level interactions that promote the wisdom of crowds bottoms-up, much like Adam Smith’s invisible hand. Our focus was on crowds in which individuals lack any basis on which to form educated opinions. Even though, the impact of such uninformed individuals has been recognised in the field of opinion dynamics, the emphasis has fallen primarily on studying how they affect the stationary opinion distribution, e.g. establishment of extremist opinions [119, 185] or promoting democratic consensus [56]. To the best of our knowledge, with regards to the wisdom of crowds, the absence of essential individual expertise has not been seriously studied. We presented empirical evidence to demonstrate that ranking individuals based on relative performance is a subtle, yet powerful means of rapidly recovering the wisdom of crowds in these situations. If incentives to compete are in place, ranking constituted the micro-level interaction mechanism that promoted remarkable collective accuracy and considerable group diversity

in virtually all experimental groups. This result also casts further support on the main theme from the previous two chapters that social information can be beneficial, should it be introduced in the right way.

In a second direction, we revisited our conjecture regarding the influence of the group's initial composition on the wisdom of crowds. We found two advantageous effects of the initial group diversity. First, given two groups with equal initial collective error, the more diverse group approaches faster the right solution. The reason is that simply by the virtue of being more heterogeneous, such groups are more likely to start with top-ranked individuals whose opinions are closer to the correct solution, though unbeknownst to them. When such individuals are outperformed, due to the competition from higher ranks (see Figure 4.6), those who replace them are necessarily even closer to the solution; as a result, the collective error diminishes rapidly. Second, diversity helps groups which are less "wise" in the beginning to outperform "wiser" groups with more homogeneous opinions. Again, the reason can be attributed to the basic competition dynamics that drives group progress.

We further note that the above results cannot be attributed to the circular search landscape in the experiment. Though it is likely that the results will not hold for an unbounded landscape (e.g. a line), in reality physical confines always exist. More importantly, the performance of the groups in the control sessions clearly indicates that without competition the wisdom of crowds phenomenon does not occur, even though the solution space is the same.

Finally, our experiment can be seen as a practical improvement of the theoretical framework proposed by [116]. Hong et al. have demonstrated that given a large enough population of diverse problem solving agents, a team of randomly drawn individuals will find the correct solution with certainty, even in the absence of individual expertise. The reason is that in the limit of large group sizes, and assuming the presence of all possible search strategies, there will always be an agent who can find a better solution. Eventually, as no agent is able to find a further improvement, the group converges to the right solution. In practice, however, groups are of limited size and individuals do not necessarily cover a considerable part of the space of possible search strategies. Therefore, we need to find practical mechanisms to drive a group to the correct solution consistently and in a reasonable time.

Chapter 5

A model for the collapse of Online Social Networks

Summary

Due to the availability of large-scale online datasets, online social networks (OSNs) represent a unique opportunity to study unintended effects of individual decisions. As the structure of an OSN is a result of individual interactions, we can see the rise and fall of these online communities as outcomes of collective decisions. Here, we propose a rational model of user behaviour in an OSN according to which users quit the network when the costs of staying exceed the benefits. Importantly, the benefits depend on the number of one's friends, hence leaving users can cause further cascades of users leaving who lost connections to their friends. The ability of a community to limit the size of such cascades is generally referred to as resilience. To quantify resilience, we use the k -core decomposition, to identify subsets of the network in which all users have at least k friends. We use this methodology to analyse the resilience of five online communities and to compare it to their stories of success or failure. Interestingly we find that resilience does not always translate to commercial success, which indicates that network topology alone is not enough to guarantee the sustainability of a community.

Based on Garcia, D., Mavrodiev P., Schweitzer F. *Social Resilience in Online Communities: The Autopsy of Friendster*, in Proceedings of the first ACM Conference in Online Social Networks, pp. 39–50, ACM, 2013. P.M. and D.G. developed the concept of resilience in terms of k -core decomposition. P.M. developed and wrote the theoretical model of user behaviour, the degree distribution analysis and also contributed to the data analysis. D.G. wrote the remainder of the original paper.

5.1 Introduction

In this and the following chapter we continue our investigations on collective decisions in humans with a different focus. We move from social influence, cooperation and wisdom of crowds to unintended consequences in online social networks (OSNs). OSNs, such as Facebook, Youtube, Twitter, Myspace and LiveJournal¹ have gained massive popularity in the last decade. According to some reports, social media has surpassed email as the most popular online activity, and together with blogging, accounts for nearly 10% of the total time spent on the Internet [18]. The main appeal of OSNs is in mediating social interactions by reducing considerably the efforts in maintaining and growing one's social network. Eventually, the large user-base has created an ecosystem that goes beyond the original intent of making and interacting with friends. OSNs, nowadays, play a major role in information dissemination from traditional media, marketing and political campaigns, self-promotion, grassroots movements, etc.

From an academic perspective, OSNs are an invaluable and rich source for studying online collective human behaviour and comparing it to the offline world [212]. Influential works have studied psychological phenomena in online communities, such as emergence of social norms [78], emotional and gender biases [81, 83], psychological well-being [238], and addiction [64]. Furthermore, the behaviour of members of an OSN can provide valuable insights on processes with direct analogues in the offline world, e.g. information propagation and social influence [110, 273], building trust [89], adoption behaviour [186], group creation and maintenance [127, 149, 271], and disease trends [2].

Since OSNs are communication media that connect millions of people, they also represent communities – groups of people who stay together, formed by inter-individual interactions. From this perspective, the community aspect of OSNs is an emergent phenomenon and cannot be reduced back to the behaviour of individuals. While the dynamics of growth in such communities are an established research subject [15, 127], there are still many open questions regarding their decline, in particular related to large OSNs [270]. The most paradigmatic example is **Friendster**, one of the first and largest OSNs, which at its peak attracted close to 112 million users and even rejected a \$30 million bid from Google. Despite its success, Friendster rapidly collapsed after inexplicably losing around 80% of its user base in about a year. Why do online communities collapse? What are the reasons behind the decision of users to leave an OSN? What is the role of the underlying social network in this decision? Investigating these questions is the goal of this chapter.

¹In reality, the list of contemporary OSNs is indeed impressive: http://en.wikipedia.org/wiki/List_of_social_networking_websites

As the structure of an OSN is a result of individual interactions, we can see the rise and fall of these online communities as outcomes of collective decisions. In the spirit of the thesis, we propose that the collapse of an OSN can be studied as an unintended consequence of individual actions. Fundamentally, users in an OSN face two major choices – joining and staying in or leaving the network. These decisions are, to a large extent, determined by the number of one’s friends and their own engagement [15]. Therefore, users who decide to leave decrease the utility of being a community member of their friends. This may trigger the latter to also leave, which can result in further cascades of departing users that may ultimately endanger the whole community. It is these unexpected cascades, triggered by individual decisions to leave, which we propose as a mechanism for an OSN collapse. In this context, we refer to the ability of an online community to withstand such cascades as *resilience*.

In the rest of the chapter, we provide an approach to quantify social resilience to the cascades of user departures from OSNs. We start from a theoretical perspective that, under the assumption of rational behaviour, allows us to define a new metric for the relation between network topology and massive user leaves. We apply this metric to high quality datasets from `Friendster` and `LiveJournal`, comparing their social resilience with partial datasets from `Facebook`, `Orkut`, and `Myspace`. We conclude the chapter by focusing on the time evolution of `Friendster`’s collapse and find a good match with our proposed mechanism.

5.2 Social Resilience in OSN

5.2.1 Quantifying Social Resilience

One approach to quantify social resilience is by natural removal of nodes based on some local property, for example degree [155]. By studying the network connectivity after such removals, one can identify nodes with critical importance for keeping the community connected. Importantly, by focusing on local properties we can only quantify the direct effects that a node removal has on the connectivity of the network.

To account for the indirect effects of further cascades of departures, we propose an extension based on the k -core decomposition [226]. A k -core of a network is a sub-network in which all nodes have a degree $\geq k$. The k -core decomposition is a procedure of finding all k -cores, $\forall k > 0$, by repeatedly pruning nodes with degrees $< k$. Therefore, it captures not only the direct, but also the indirect impact of users leaving the network. As an illustration consider Figure 5.1, which shows targeted removal of nodes with degrees < 3 .

On one hand, starting from the network in A and removing all nodes with degrees < 3 ,

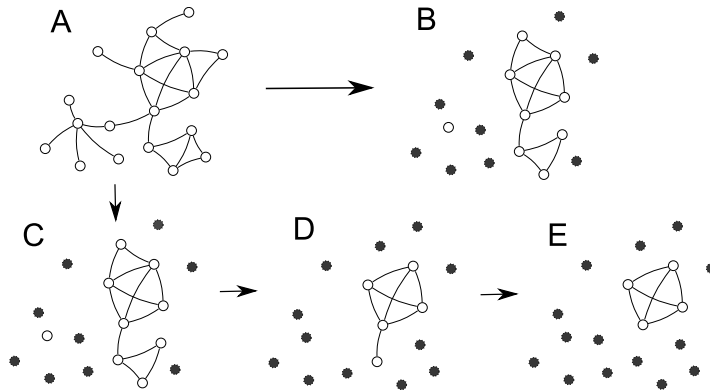


Figure 5.1: Effects of node removals on network connectivity as captured by degree only (A \rightarrow B) and k -core decomposition (A \rightarrow C \rightarrow D \rightarrow E)

produces the network in B. The dark-coloured nodes in B have been removed (and thus are disconnected), and the final network consists of the 9 white nodes. The transition A \rightarrow B shows only the direct effects of users with < 3 friends leaving.

On the other hand, starting again from A, and applying the k -core procedure, will repeatedly remove nodes until only those with degrees ≥ 3 remain. The first step, A \rightarrow C, removes the same light-grey nodes as before. Continuing, C \rightarrow D, removes those nodes that have been left with < 3 neighbours in C, and disconnects them as well. The final step, D \rightarrow E, finishes the process by disconnecting the last white node in D that was left with < 3 friends. As a result, the final network is the fully connected network of the 4 white nodes.

Hence, supposing that users leave a community when they are left with less than 3 friends, the k -core decomposition captures the full cascading effect that departing users have on the network as a whole.

We proceed by formalising social resilience based on a *generalised* k -core decomposition. To this end, we present a theoretical model in which rational users decide simultaneously either to stay in the network or to leave it. These decisions are based on maximising a utility function that weighs the benefits of membership against the associated costs. We show that the equilibrium network which maximises the total payoff in the community, corresponds to a generalised k -core decomposition of the network.

5.2.2 Generalised k -core decomposition

Following [98], we extend the traditional k -core decomposition by recognising that the pruning criterion need not be limited to degree only. Let us define a *property* function $\mathcal{B}_i(H)$ that given a sub-network $H \subseteq G$ associates a value, $n_i \in \mathbb{R}$, to node i . A generalised k -core of a network G is, then, defined as a sub-network $H \subseteq G$, such that $\mathcal{B}_i(H) \geq k$, $\forall i \in H$ and $k \in \mathbb{Z}$. The general form of B_i allows us to model different pruning mechanisms. For example, the traditional definition of the k -core can be recovered in the following way – for every node i take its immediate neighbourhood, \mathcal{N}_i , and fix $B_i(H) := |\mathcal{N}_i|$, $\forall H \subseteq G$. Other authors have also shown that considering weighted links in \mathcal{B}_i can more accurately reveal nodes with higher spreading potential in weighted networks [79].

Note that by definition higher order cores are nested within lower order cores. We use this to define that a node i has *coreness* k_s if it is contained in a core of order k , but not in a core of order $k' > k$.

5.2.3 A rational model for OSN users

Assume that users in a given network, G , incur a constant integer cost, $c > 0$, for the effort they must spend to remain engaged. Accordingly, they receive a benefit or payoff from their friends in the network. Let the benefit of user i be the property function $\mathcal{B}_i(H)$ with $i \in H$. Assume non-increasing marginal benefits with respect to the size of H , i.e. $\mathcal{B}_i''(H) \leq 0$, otherwise costs are irrelevant as any cost level could be trivially overcome by increasing the size of H . This assumption is also supported by other empirical investigations of large social networks which show that the probability of a user to leave is concave with the number of friends who already left [15, 270].

Users can execute one of two possible moves – **stay** or **leave**. The utility of a user i , is $U_i = 0$, if he chose **leave** or $U_i = \mathcal{B}_i(H) - c$, for **stay**. Finally, users are utility-maximising, therefore they will choose **stay** as long as $U_i > 0$.

It is easily seen that the equilibrium network, G^* , which maximises the total utility, $U(G) = \sum_i U_i$, is composed of users who choose **stay** when $c < k_s^i$, and **leave** otherwise. In other words, node i should remain engaged in the network as long as the cost, c , does not exceed its generalised coreness, k_s . In this sense, G^* corresponds to the generalised k -core of G .

To illustrate that G^* is indeed an equilibrium network, we need to show that no user has an incentive to unilaterally join it or leave it. Consider a node, $j \in G^*$ who chooses **stay**. This node would belong to a generalised k -core, k_s^j , and by definition, $B_j(H) - k_s^j \geq 0$.

Since, j stayed in the network, it must be that $c < k_s^j$, therefore $B_j(H) - c > 0$. So, j will be forfeiting positive utility, had he decides to leave. In the same manner, consider another node $l \notin G^*$ who chooses **leave**, thus his coreness $k_s^l \leq c$. All his friends with the same coreness would have left the network, therefore the only benefit that l could obtain from staying would come from his connections with nodes in higher cores. The benefit, B_l , from such connections must not exceed k_s^l , otherwise l would have belonged to a higher core in the first place. Since $k_s^l \leq c$ we have $B_l < c$. This implies that l necessarily obtains negative utility from staying, so he has no incentives to do so. Moreover, G^* is optimal, as we showed that any change from the equilibrium actions of any user inevitably lowers his utility and decreases the total utility in the network. We also argue that it is reasonable to expect this equilibrium network to be reached in an actual setting, since it maximises the utility of all users simultaneously, as well as the welfare of the network provider.

In the rest of the chapter, we approximate B_i as proportional to the number of i 's direct friends, N_i , i.e. $B_i = bN_i$, for some $b \in \mathbb{Z}$. Taking k_s^i to be the coreness of i , by definition it holds that $bN_i \geq k_s^i$. The maximum cost, c , that i would tolerate as a member of the community must be strictly smaller than its coreness, hence $bN_i > c$ and $N_i > c/b$. The last result implies that the minimum number of friends that a node i needs to remain engaged must be strictly larger than c/b . Therefore, the coreness of a participating user i must be at least $c/b + 1$, i.e. $k_s^i \geq K$, where $K = (c/b) + 1$.

Based on the above discussion, we see that a user will remain in a network with a high c/b ratio if its coreness k_s is high. This is because, by definition i is part of a connected network of nodes with large minimum degrees and hence large benefits.

In contrast, simply having a large degree does not imply that a user will obtain large utility from staying. Note that a high-degree node may nevertheless have low coreness. This means that i would be part of a sub-network in which all nodes have low minimum degrees. As a result a lower c/b ratio would suffice to start a cascade of users departing, that can quickly leave i with no friends and thus drive it to leave too.

With the above theoretical framework, we define social resilience of a community as the size of the K -core. In other words, this is the size of the network that remains after all users with $k_s \leq c/b$ have been forced out. This definition allows us to quantify social resilience and reliably compare it across communities even for unknown c/b ratios, as shown in Section 5.5.

5.3 Data on Online Social Networks

For our empirical study of social network resilience, we use datasets from five different OSNs. The choice of these datasets aims at spanning a variety of success stories across OSNs, including successful and failed communities, as well as communities currently in decline. The size, data gathering methods, and references are summarised in Table 5.1, and outlined in the following.

name	date	status	users	links	source
Livejournal	1999	successful	5.2M	28M	[179]
Friendster	2002	failed	117M	2580M	Internet Archive
Myspace	2003	in decline	100K	6.8M	[3]
Orkut	2004	in decline	3M	223M	[179]
Facebook	2004	successful	3M	23M	[88]

Table 5.1: Outline of OSN and datasets.

Friendster. The most recent dataset we take into account is the one retrieved by the Internet Archive, with the purpose of preserving **Friendster**’s information before its discontinuation. This dataset provides a high-quality snapshot of the large amount of user information that was publicly available on the site, including friend lists and interest-based groups [271]. Here, we provide the first analysis of the social network topology of **Friendster** as a whole.

Since some user profiles in **Friendster** were private, this dataset does not include their connections. However, these private users would be listed as contacts in the list of their friends who were not private. We symmetrised the **Friendster** dataset by adding these additional links.

LiveJournal. In **LiveJournal**, users keep personal blogs and define different types of friendship links. The information retrieval method for the creation of this dataset combined user id sampling with neighbourhood exploration [179], covering more than 95% of the whole community. We choose this **LiveJournal** dataset for its overall quality, as it provides a view of practically the whole OSN.

Note that the design of **LiveJournal** as an OSN deviates from the other four communities analysed here. First, **LiveJournal** is a blog community, in which the social network functionality plays a secondary role. Second, **LiveJournal** social links are directed, in the sense that one user can be friend of another without being friended back. In our analysis,

we only include reciprocal links, referring to previous research on its k-core decomposition [134]. By including this dataset, we aim at comparing how different interaction mechanisms and platform designs influence social resilience.

Orkut. Among declining social networking sites, we include a partial dataset on **Orkut** [179], which was estimated to cover 11.3% of the whole community. Far from the quality of the two previous datasets, we include **Orkut** in our analysis due to its platform design, as this dataset includes users that did not have a limit on their amount of friends. Furthermore, **Orkut** has a story of local success in Brazil, losing popularity against other sites at the time of writing of this article.

Myspace. One of the most famous OSN in decline is **Myspace**, which was the leading OSN before **Facebook**'s success [87]. We include a relatively small dataset of 100000 users of **Myspace** [3], which was aimed to sample its degree distribution. This dataset was crawled through a Breadth-First Search method, providing a partial and possibly biased dataset of **Myspace**. We include this dataset as an exercise to study the influence of sampling biases in the analysis of social resilience.

Facebook. We want to complete the spectrum of success of OSN, from the collapse of **Friendster** to the big success of **Facebook**. The last dataset we include is a special crawl which aims at an unbiased, yet partial dataset as close as possible to the whole community [88]. This dataset was retrieved with a special technique based on random walks, keeping unvaried some network statistics, including **Facebook**'s degree distribution.

5.4 Not power-law degree distributions

The first step in our analysis explores the degree distributions of each OSN. The reason to do so is the epidemic properties of complex networks. Under the assumptions of epidemic models, networks with power-law degree distributions do not have an epidemic threshold [193], i.e. a “sickness” would survive within the network for an unbound amount of time and eventually infect most of the nodes. Such sickness could be a meme or a social norm, but could also be the decision of leaving the community. Therefore, we need to assess the possibility of a power-law degree distribution, as it would pose an alternative explanation for the massive cascades of user departures.

Numerous previous works have reported power-law degree distributions in social networks

[3, 42, 155, 179]. Nevertheless, most of these works rely on goodness of fit statistics, and do not provide a clear test of the power-law hypothesis. The hypothesis states that the degree distribution follows the equation

$$p(d) = \frac{\alpha - 1}{\text{deg}_{\min}} \left(\frac{d}{\text{deg}_{\min}} \right)^{-\alpha} \quad \text{for } d \geq \text{deg}_{\min}.$$

This is usually described as $p(d) \propto d^{-\alpha}$, and often argued as valid if metrics such as R^2 , or F are high enough. While a high goodness of fit could be sufficient for some practical applications, the empirical test of the power-law hypothesis can only be tested, and eventually rejected, through the result of a statistical test, assuming a reasonable confidence level.

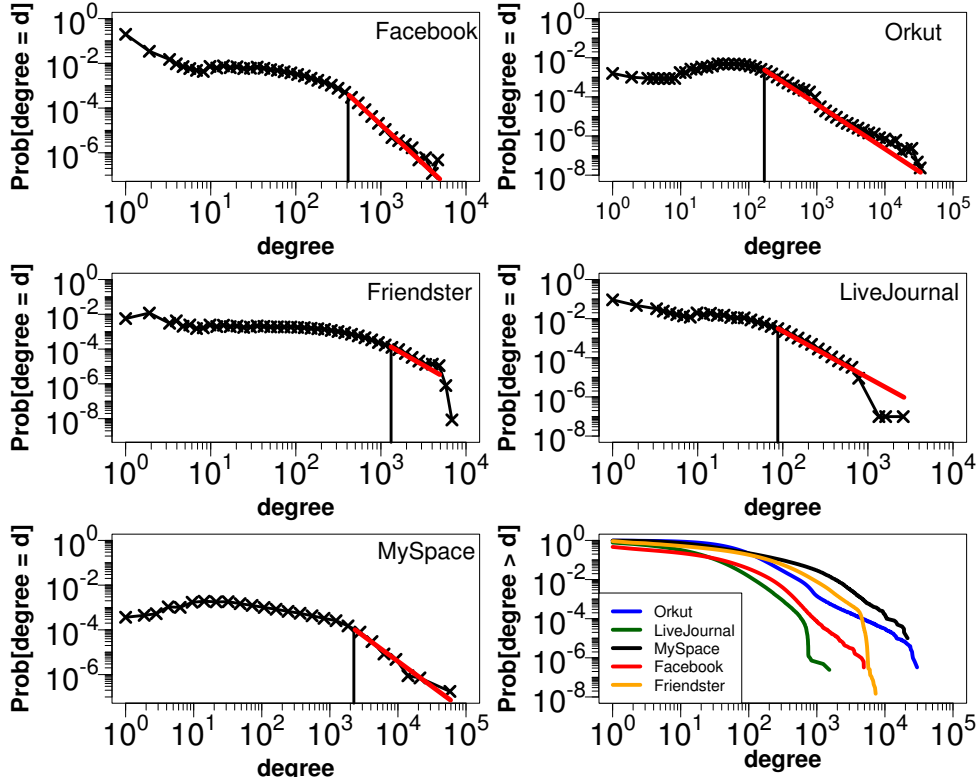


Figure 5.2: Complementary cumulative distribution functions and probability density functions of node degree in the five considered communities. Red lines show the ML power-law fits from $\widehat{\text{deg}}_{\min}$.

We followed the state-of-the-art methodology to test power laws [43], which roughly involves the following steps. First, we created Maximum Likelihood (ML) estimators $\hat{\alpha}$ and $\widehat{\text{deg}}_{\min}$ for $p(d)$. Second, we tested the empirical data above $\widehat{\text{deg}}_{\min}$ against the power law hypothesis and we recorded the corresponding KS-statistics (D). Third, we repeated the

KS test for 100 synthetic datasets that follow the fitted power law above $\widehat{\text{deg}}_{\min}$. The p-value is then the fraction of the synthetic D values that are larger than the empirical one. Thus, for each degree distribution, we have the ML estimates $\widehat{\text{deg}}_{\min}$ and $\hat{\alpha}$, which define the best case in terms of the KS test, with an associated D value, and the p-value. Ultimately, a power law hypothesis cannot be rejected if (i) the p-value of the KS-test is above a chosen significance level [43], and (ii) there is a sufficiently large amount of data points from deg_{\min} to deg_{\max} [242]. We found that the degree distributions of **Facebook**, **Friendster**, **Orkut** and **LiveJournal** have p-values well below any reasonable significance threshold, showing an extremely reliable empirical support to reject the power-law hypothesis (Table 5.2).

dataset	$\widehat{\text{deg}}_{\min}$	$\hat{\alpha}$	n_{tail}	D	p
Friendster	1311	3.6	2.9×10^5	4.59	$< 10^{-15}$
LiveJournal	88	3.3	81141	0.02	$< 10^{-15}$
Facebook	423	4.6	4918	0.14	$< 10^{-15}$
Orkut	171	3	2.8×10^5	0.02	$< 10^{-15}$
Myspace	2350	3.6	623	0.03	0.22

Table 5.2: Power law fits of the analysed datasets.

For the case of **Myspace**, a KS test gives a p-value of 0.22, which can be considered high enough to not reject the power-law hypothesis [43]. Therefore **Myspace** satisfies the first criterion, but when looking at the range of values from deg_{\min} to deg_{\max} (roughly one order of magnitude), and the low amount of data points included, this KS-test composes a merely anecdotal evidence of the extreme tail of Myspace. If accepted, the power-law distribution would explain just 0.623% of the **Myspace** dataset. In addition, the unsupervised breadth-first-search crawling method used for this dataset has been shown to have a bias that creates artificial power-law tails [1]. This leads to the conclusion that, while we cannot fully reject the power-law hypothesis, we can safely state that the dataset does not support the hypothesis otherwise. Figure 5.2 shows the degree distributions and their CCDF. For each OSN, we show how the typical log-log plot of the PDF is misleading, as a simple eye inspection would suggest power-law distributions, but a robust statistical analysis disproves this possibility.

5.5 Empirics of OSN Resilience

5.5.1 K-core decomposition

We computed the k -core decomposition for each of the OSN datasets we introduced in Section 5.3. Among those datasets, `Friendster` and `LiveJournal` cover the vast majority of their respective communities. Figure 5.3 shows a schematic representation of the k -cores of `Friendster` and `LiveJournal`. Each layer of the circles corresponds to the nodes with coreness k_s , with an area proportional to the amount of nodes with that coreness value. The colour of each layer ranges from light blue for $k_s = 1$, to red for $k_s = 304$. The distribution of colours reveals a qualitative difference between both communities: `Friendster` has many more nodes of high coreness than `LiveJournal`, which has a similar color range but a much larger fringe, i.e. the set of nodes with low k_s . This difference indicates that, to exist as a community, `LiveJournal` needs to have a much lower c/b than `Friendster`. This scenario is rather realistic, as `LiveJournal` is a blog community in which users create large amounts of original content. This leads to high benefits per social link as long as users have similar interests, which seems to be the key of `LiveJournal`'s relative success.

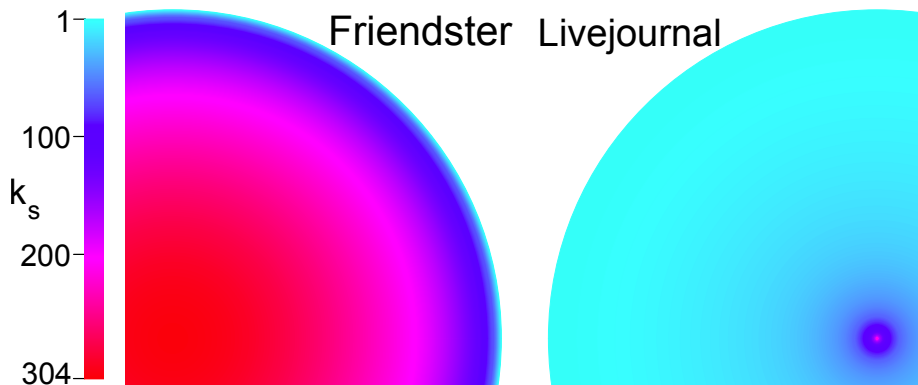


Figure 5.3: Overview of the k -core decomposition for `Friendster` and `LiveJournal`. Layers are coloured according to k_s , with areas proportional to the amount of nodes with such k_s .

Our theoretical argumentation, presented in Section 5.2.3, indicates that node coreness is a more reasonable estimator for resilience than node degree. A degree of at least k_s is a necessary condition for a coreness of k_s , but a high degree does not necessarily mean a high coreness. Taking `Friendster` an example, Figure 5.4 shows the boxplot for the distribution of k_s versus node degree, indicating the spread of k_s for nodes of similar degree. The empirical data shows that a high degree does not necessarily mean a high k_s ,

even finding nodes with very low k_s and very high degree. Nevertheless, it is clear that k_s is likely to increase with degree, but mapping degree to coreness would misrepresent the resilience of the community as a whole. By measuring coreness, we can detect that some nodes belong to the fringe despite their high degree, as the coreness integrates global information about the centrality of the node.

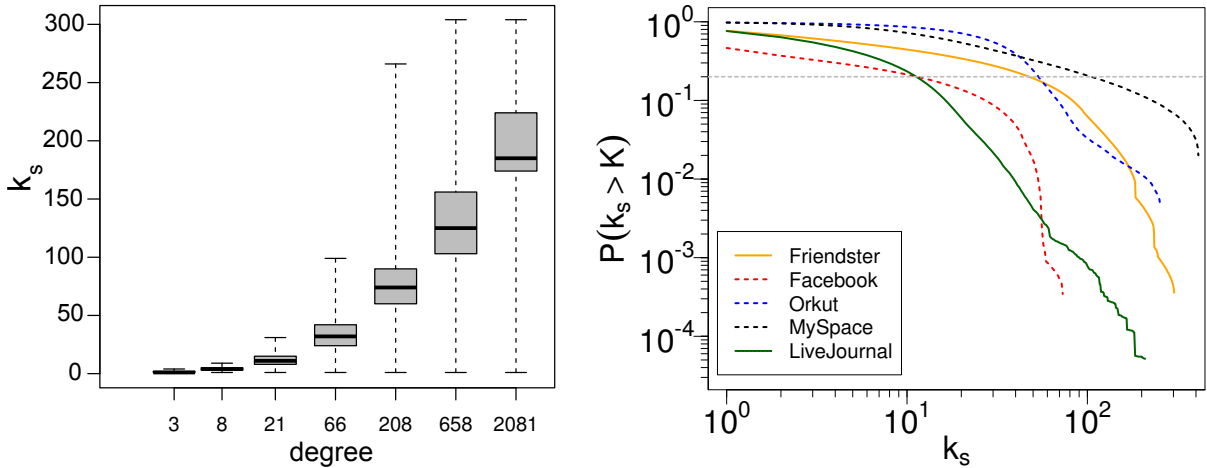


Figure 5.4: **Left:** boxplot of k-shell indices by degree for Friendster. Dark lines represent the mean, and dashed bars show extreme values. Boxes are arranged in the x-axis according to the middle value of their bin. **Right:** CCDF of k_s . The horizontal dashed line shows the cut at 0.2.

5.5.2 Resilience comparison

Extending the above observations, we computed the k-core decomposition of the three additional OSN, aiming at comparing their relation between their environment, measured through c/b , and the amount of users expected to be active under such conditions.

We focus our analysis on the Complementary Cumulative Density Function (CCDF) of each network, defined as $P(k_s > K)$. As shown in Section 5.2.3, the cost-benefit ratio c/b corresponds to a value K that determines the nodes that leave the network, which are those with coreness k_s below K . Under this conditions, the CCDF of k_s measures the amount of nodes that will remain in the network under a given c/b , allowing us to compare how each OSN would withstand the same values of cost and benefit.

The right panel of Figure 5.4 shows the log-log CCDF of the five OSN. The first two communities to compare are LiveJournal and Friendster, as the datasets on these two are the most reliable. First, the CCDF of Friendster is always above the CCDF of

LiveJournal. This is consistent with the structure shown in Figure 5.3, where it can be appreciated that **LiveJournal** has many more nodes in the fringe than **Friendster**. Second, both CCDF reach comparable maximum values, regardless of the fact that **Friendster** was 20 times larger than **LiveJournal**. Such skewness in the coreness of **LiveJournal** can be interpreted as a result of a higher competition for attention, as expected from a blog community in comparison with a pure social networking site, like **Friendster** was.

Focusing on the tails of the distributions, we can compare the patterns of resilience for environments with high K . The comparison between the resilience of these communities is heavily dependent of the value of K , as for example, **LiveJournal** is less resilient than **Facebook** for values of K between 10 and 50, but more resilient below and above such interval. A similar case can be seen between **Friendster** and **Orkut**, as their CCDFs cross at 60 and 200. Thus, **Friendster** would be more resilient than **Orkut** if K lies in that interval, while **Orkut** would have a larger fraction of active nodes if $K < 60$ or $K > 200$.

It is important to note that these comparisons are made between the reliable datasets of **Friendster** and **LiveJournal**, compared with partial datasets from the other communities. While our conclusions on the first two OSN can be seen as global findings on the community as a whole, the rest are limited to the size of the datasets available. A particularly clear example of the effect of the crawling bias is the distribution of coreness for **Myspace**, which shows an extreme resilience in comparison to all the other datasets, with the exception of **Orkut** for $K < 50$. As commented in Section 5.3, the method used for **Myspace** was very biased towards nodes of high degree, leaving an unrealistic picture of the resilience of the whole community. Additionally, the method used for **Facebook** seems to have delivered a degree distribution close to a random sample of **Facebook** users, but its restarting of random walkers leaves tendrils of nodes that accumulate on the 1-core. Hence the low starting value of the CCDF of **Facebook** could be an artefact of this crawling method.

Regardless of any crawling bias, we found that these networks have maximum coreness numbers much higher than previous results. The maximum k_s found for the network of instant messaging was limited to 68 [155], and close to 100 for the OSN **Cyworld** [42]. **LiveJournal** has a maximum k_s of 213, **Friendster** of 304, **Orkut** of 253, and **Myspace** as a very deep core of $k_s = 414$. The exception lies in the **Facebook** dataset, where we find a maximum k_s of 74. This evidence shows that OSN can have much tighter cores than the ones found in previous research, revealing that they contain small communities with very high resilience.

As a final comparison, we focus on the values of K for the catastrophic case of the networks losing 80% of their nodes, i.e. where the CCDF has a value of 0.2. The data shows that

both **Facebook** and **LiveJournal** would lose 80% of their users under a value of K close to 10. For the case of the unsuccessful communities of **Orkut** and **Friendster**, it requires a much worse environment, with values of K above 60. This way, the empirical data supports the idea that, under the same environmental conditions, both successful communities are less resilient than the three unsuccessful ones. This means that the topology of their social networks is not enough to explain their collapse, but indicates that bad decisions in design and interface changes can spread through the network and drive many users away.

In the next section, we describe a *post-hoc* case study of the way **Friendster** rose and collapsed, using the available timing information in the dataset.

5.6 The decline of Friendster

To conclude our analysis, we explored how the spread of departures captured in the k -core decomposition (see Section 5.2.3) can describe the collapse of **Friendster** as an OSN. As we do not have access to the precise amount of active users of **Friendster**, we proxy its value through the **Google** search volume of *www.friendster.com*. At some point in 2009, **Friendster** introduced changes in its user interface, coinciding with some technical problems, and the rise of popularity of **Facebook**². This led to the fast decrease of active users in the community, ending on its discontinuation in 2011.

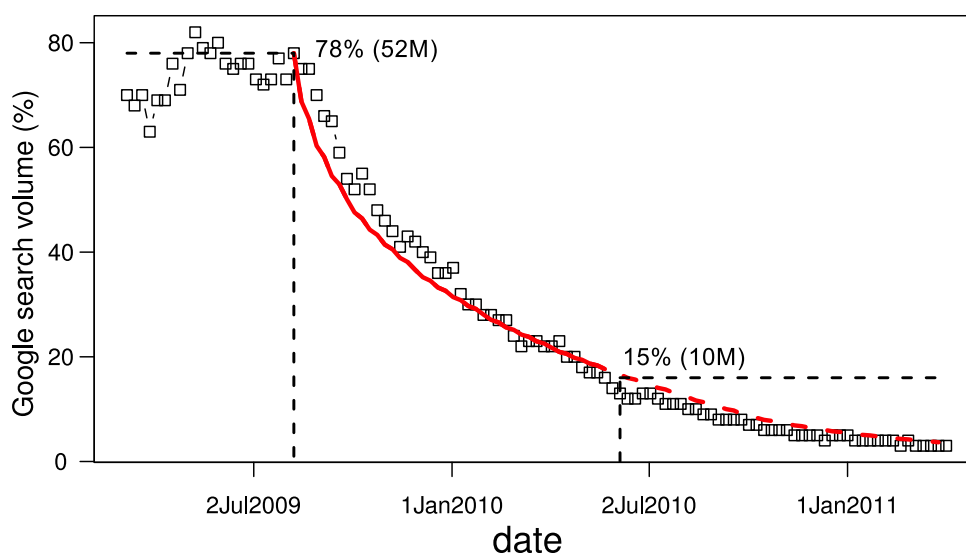


Figure 5.5: Weekly Google search trend volume for **Friendster**. The red line shows the estimation of the remaining users in a process of unravelling.

²www.time.com/time/business/article/0,8599,1707760,00.html

We scale the search volumes fixing 100% as the total amount of users with coreness above 0, 68 million. At the point when the collapse of **Friendster** started, the search volume indicates a popularity of 78% of its maximum. We take this point to start the simulation of a user departure cascade, with an initial amount of 52 million active users, i.e. users with coreness above 3. The second reference point we take is June 2010, when Friendster was reported to have 10 million active users ³, corresponding to 15% of the 68-million user reference explained above. The search volume on that date is 14%, showing the validity of the assumption that the maximum amount of active users corresponds to those with coreness above 0. Thus, these 10 million remaining users correspond to nodes with $k_s > 67$.

Given these two reference points, we can approximate the collapse of the network through its “unravelling” per k-core. Our assumption is that a critical coreness K_t starts at 3 and increases by 1 at a constant rate. Such K_t is the result of an increasing cost-to-benefit ratio, and thus all the nodes with $k_s < K_t$ would leave the community. Then, for each time step, the amount of remaining users would correspond to the CCDF shown in Figure 5.4. In our analysis, K increases at a rate of 6 per month, i.e. from 3 to 67 between our two reference point.

The red line of Figure 5.5 shows the remaining users under this process, with dashed values after the second reference point of June 2010. We can observe that the process approximates well the decay of **Friendster** from the start of its decline, to its total shutdown in 2011. The R^2 value for this fit is 0.972, leaving some slight underfit through 2009. This fit shows the match between two approximations: on one side the search volume as an estimation of the amount of active users, and on the other side the amount of remaining users when the c/b ratio increases constantly through time.

5.7 Conclusion

The rapid collapse of large OSNs, such as **Friendster**, often comes as a surprise. Our theoretical and practical analysis proposes compelling evidence that such events may occur as the unforeseen collective result of user behaviour. In the theoretical framework we introduced, the collapse of an online community is due to large-scale cascades of users leaving. Importantly, the cascades are not a coordinated activity. Rather, external shocks can get amplified by the network structure and propagate through a considerable part of the OSN. This has direct real-world repercussions for OSN administrators, and in fact the work presented in this chapter received considerable media attention (see Appendix 11.1).

³en.wikipedia.org/wiki/Friendster

For example, it suggests that technological modifications, e.g. changes in the graphical interface, should be done gradually and applied preferably to users with low coreness, to minimise the size and impact of potential departure cascades. OSNs administrators should also consider that the user-base is not a sufficient predictor for stability – the users need to be strongly connected too, and such types of connections should be encouraged.

We verified the cascading mechanism by a case study of **Friendster**, fitting approximated time series of active users through the spread of user departures predicted by the k-core decomposition. We also showed how a generalised version of the k-core decomposition enables the empirical measurement of resilience, and presented the first empirical analysis of social resilience in OSNs. To this end, we provided an empirical study of social resilience across five influential OSN, including successful ones like **Facebook** and unsuccessful ones like **Friendster**. We have shown that the hypothesis of a power-law degree distribution cannot be accepted for any of these communities, discarding the epidemic properties of complex networks as a possible explanation for large-scale cascades of user departures. Our k-core analysis overcomes this limitation, quantifying social resilience as a collective phenomenon using the CCDF of node coreness. We found that the topologies of two successful sites, **LiveJournal** and **Facebook**, are less resilient than the unsuccessful **Friendster** and **Orkut**. This indicates that the environmental condition of an OSN play a major role for its success. Thus, we conclude that the topology of the social network alone cannot explain the stories of success and failure of the studied OSN, and it is necessary to focus future empirical analysis in measuring these costs and benefits. Additionally, we found very high maximum coreness numbers for most of the OSN we studied. The existence of these super-connected cores indicates that information can be spread efficiently through these OSN [134].

Our analysis is focused on the macroscopic resilience of OSN, but additional research is necessary to complete our findings. Microscopic data on user activity and churn can provide estimators for the benefits and costs of each network, to further validate the work presented here. Furthermore, the generalised k-core can be applied when user decisions are more complex than just staying or leaving the network, for example introducing heterogeneity of benefits or weights in the social links.

Another important open question is the role of directionality in the social network. More specifically, how can we measure resilience when asymmetric relations are allowed? The benefits of users of these networks would represent both the reputation of a user and the amount of information it receives from its neighbourhood. This is the topic of the next chapter.

Chapter 6

Cost of participation increases resilience of Online Social Networks

Summary

We follow up on the study of OSN resilience by considering directionality in the link relations. To this end, we focus on reputation-based OSNs, such as Twitter, as prime examples of communities where asymmetric relations are the norm. We argue that social resilience in these OSNs can be recast in terms of average user reputation, as an indication for information quality and user engagement. Moreover, individual reputation is seen as recursively dependent on the reputation of one's followers. Building on these foundations, we then extend the rational model of user behaviour from the previous chapter by incorporating rudimentary network dynamics. Our new model of user reputation is thus simple, yet able to reproduce generic features of real-world OSNs such as core-periphery structure and (direct and indirect) reciprocity. Most importantly, the model demonstrates that introducing cost of usage in an OSN can increase its resilience by promoting core of users who mutually re-enforce their reputation, while driving non-productive members out.

Based on Schweitzer F., Mavrodiev P., Seufert A., Garcia D. *Modeling User Reputation in Online Social Networks: How Reciprocity impacts Resilience*, manuscript pending submission. F.S and D.G wrote an intermediate version of the manuscript. A.S. did the main computer simulations. P.M. did all technical analysis and considerably re-wrote many parts of the manuscript to make it suitable for this thesis.

6.1 Introduction

Previously, we saw resilience of an online social network (OSN) as the ability to withstand detrimental changes in its user-base. These changes may be caused by an external shock (e.g. technological change), which increases the cost of participation to some users and triggers their departure. These leaving users, in turn, decrease the benefits of their connections, who now have fewer friends and may decide to leave too, resulting in departure cascades. Formally, we quantified resilience as the size of the community which finds it beneficial to remain engaged in the OSN, given certain costs of participation. This approach relied on the fact the benefits from being connected to another user are symmetric, as is the case for bilateral friendship relations¹. The topic of this chapter is to extend our investigations on the resilience of OSNs by considering communities with directional links, i.e. asymmetric relations in which the benefits of two mutually connected users are qualitatively different. In particular, we introduce the notion of *cost* of usage and find that the cost which maximises resilience is unexpectedly non-zero.

A well-known example for an OSN with directionality and asymmetrical links is Twitter, where a user chooses to follow other users with the purpose of receiving information from them. One might argue that the benefit a user receives from being part of such OSN consists of the value of this information, generally in the form of relevant content, e.g. links, pictures, opinions. On the other hand, users also benefit when the content they produce is valued and disseminated through the community. As evidenced by the modern rapid pace of information production and spreading in an OSN, the benefit received through content consumption diminishes, and the new scarce resource of online communities becomes attention [117].

Indeed, the growth of the *market value* of attention has been astonishing [194, 240]. A sizeable industry of “follower services”, e.g. `boostlikes.com` and `followfly.com`, has emerged with the sole value proposition of selling likes and followers. Particularly conspicuous is `followfly.com`’s slogan,

In order to indicate professionalism and attract new users to your page, your social media profile needs to be backed by loyal followers. Reinforce your status as a go to brand by supplementing your account with Instagram likes and boost your popularity.

¹This is only approximately true, as one could argue that mutual friends in an OSN may value the connection differently. Nevertheless, such complexity can be accounted for in the approach from the previous chapter by adjusting the benefit function, e.g. considering weighted links [79]

Websites like these, use click-farms in developing countries that employ thousands of people to mimic attention in a given online community. These artificial or “fake” likes and followers do not correspond to genuine interest in the information produced by a given user. Nevertheless the astonishing growth of their popularity indicates that even factitious interest is perceived as an indicator of individual influence or reputation. In fact, the spread of these services has been so pervasive that it has started to harm legitimate businesses by overcrowding their content². Even OSN providers, themselves, offer the possibility to attract attention by paying for promoting one’s content³. All this is an indication that the attention economy is as real in the online world as it is in the offline world [59].

Arguably, a key factor in the struggle to capture attention is user reputation. Reputation increases the influence of one’s content, as measured by the likelihood that this content will be acknowledged, consumed and re-distributed in the community. This is important not only for private individuals, but also for commercial entities which benefit directly from delivering content to millions of users. Individual reputation is, therefore, a key component in the benefit of participation. Measuring and analysing individual reputation in an OSN can be done in a variety of ways. For the case of Twitter, reputation cannot be mapped simply to amount of followers, as ranking users according to this amount significantly differs from ranking according to the popularity of one’s tweets, as measured by re-tweeting behaviour [148]. It seems that reputation, and thus ability to influence information spread and attract attention, cannot be captured by simple metrics, such as number of followers, but depends on a combination of the topic of the produced content and the ability to attract followers who, themselves, have high reputation [41].

There are still many open questions about the influence of user reputation on the success of an OSN, and how reputation mechanisms affect user rewards. From the perspective of OSN providers individual reputation is a reliable proxy for the quality of produced information and allows filtering out the destructive effect of spammers [85]. It is well known that boosting individual reputation by rewarding users for valuable content, e.g. with virtual “badges”, is a powerful incentive to encourage quality contributions [6]. An example of a successful OSN where reputation incentives have brought success is **Instagram**, a photo-sharing mobile application, which grew to 200 million users in its four years of existence [23, 74]. In contrast, other sites like **Foursquare** have changed their focus to recommendation, as reputation did not provide the desired user engagement [118].

In this chapter, we contribute to the study of how user reputation affects OSNs success.

²see [215] and references therein

³Every **Facebook** page and Twitter post can be promoted for a fee, by being displayed to targeted audience. There are controversial indications that Facebook, itself, uses click farms for such promotions [274].

We propose a model which applies the concepts of resilience and departure cascades from Chapter 5 to reputation-based OSNs. The model extends previous notions of reputation, based on degree and traditional centrality metrics [85], by including directionality, cost of usage, and decay of reputation over time requiring users to be involved. Interestingly, the model predicts that the optimal resilience of such OSNs is attained for a non-zero cost of usage. Section 6.2 introduces the model and its analytical analysis. In Section 6.3 we present the results of computer simulations and the emergence of optimal cost of usage. Discussion and concluding remarks are provided in Section 6.4.

6.2 A reputation model

Based on the above discussion, our reputation model accounts for the following two basic ideas. First, it explicitly includes the costs and benefits of users of an OSN. We assume that the benefit of users to join, and to stay in the OSN, results from the *relative* reputation that they receive from being connected to others (see Section 6.2.1).

Second, directionality is taken into account by calculating reputation as a combination of the number of users who unilaterally connect to an individual, i.e. *follow* him, and the reputation of one's own followers. This is quite common in different ranking schemes. For example, **PageRank** calculates the rank of a website as a function of the number and ranks of the sites linking to it. Similarly, the vulnerability of financial institutions can be calculated dependent on the number and vulnerability of the counterparties connected to them. In more general terms, such dynamics follows (hetero)catalytic models e.g. from chemistry and biology, where the concentration of a particular (chemical) species depends on the concentration of those species that produce, or feed it [217]. In [122] and [123], this dynamics was combined with a network dynamics that runs on a different time scale. At each time step (measured in network time), the system was perturbed by an *extremal dynamics* where the least performing node in the network was replaced by a new node that randomly rewires itself back to the system. This model was already analysed in detail [228] and extended to cover other phenomena, such as strategic link formation between economic agents [141].

We build on this modelling approach by extending it in the context of OSNs in a number of important points. First, we introduce cost of usage, which users need to account for in their decisions to participate. These include, among others, the efforts to login, the effort of learning how to efficiently interact with the user interface or to adapt if the interface

changes⁴. Even if most users are willing to disregard such efforts, OSN administrators would still like to know how users would respond, if monetary costs, e.g. monthly membership fee, are introduced at some point. Will members remain loyal? What would be the impact of less active users leaving compared to core users leaving? Intuitively, one would assume that the higher the cost, the higher the likelihood of users leaving, the larger the resulting cascades, and the lower the resilience. Interestingly, this is not the case as we show later. Note that by introducing cost of usage, we depart from the extremal dynamics because *many* more users can decide to leave dependent on how the cost of usage compares to their benefit.

Second, we focus on the effects that cascades of leaving users have on the OSN by investigating the resulting structural features of the network. And third we provide an analytical and a numerical investigation that links the cost of usage with the resilience of the OSN.

Recall that in Chapter 5 we formalised the qualitative notion of resilience via the k-core decomposition. In our model, we define resilience as the long-term average benefit of all users in the OSN. Intuitively, the more satisfied users are in being part of an OSN, the more the community is able to preserve its members after a shock. Note that defined in this way, resilience is not associated with a fixed, stationary stable state, but with the dynamic capabilities of the OSN to limit the number of users leaving in a cascade, and to quickly recover from such cascades.

6.2.1 Costs and benefits

Analogous to Chapter 5, we pose the same rational choice framework on users who ponder their participation in an OSN. Simply, users leave when their cost of staying in the network exceed, at a given point in time, their benefits of being members. This can be expressed by the dynamics

$$s_i(t) = \Theta [b_i(t) - c_i(t)]. \quad (6.1)$$

Here, $s_i(t)$ characterises the current state of user i at time t as a binary variable: $s_i = 1$ means that the user is part of the network and $s_i = 0$ means that at time t the user leaves. $\Theta[z]$ is the Heavyside function which, returns 1 if $z \geq 0$ and 0 if $z < 0$. Thus, the current state of user i depends on the difference between his benefits $b_i(t)$ and costs $c_i(t)$.

In the following, we assume that the main component in determining the benefits b_i is the user reputation, $X_i(t)$, received from interacting with other users. By ignoring the

⁴As I am writing this chapter, Facebook has just introduced, yet another change in their interface by re-arranging the news feed and the information bar.

component received from consuming content, we focus on the importance of individual influence and ability to capture attention. There are many OSNs in which reputation is a key reward mechanism that encourages participation: in **Twitter** users obtain reputation from the number of their followers, in product review communities like **Amazon** or **Doyoo** users earn their reputation from the **likes** of other users, in knowledge-creation communities like **Wikipedia** and **Stack Overflow** reputation-based incentives (e.g. badges) are key in driving contributions.

Such a reputation measure can be explicitly displayed on the site, like the *RG score* of **Researchgate** or the **Reddit** score, which increases with the reputation of followers and the feedback of the community. On the other hand, user reputation can be implicit and not part of a user profile, but can still be perceived through the activity of other users. Examples of this implicit reputation are *re-tweets* in **Twitter** and *likes* in **Facebook**.

We can express the reputation of a user by the following dynamics

$$\frac{d}{dt}X_i(t) = \sum_{j=1}^N a_{ij}(t)X_j(t) - \phi X_i(t). \quad (6.2)$$

The coefficients a_{ij} 's are elements of the adjacency matrix of the OSN, \mathbf{A} (see Figure 6.1). They represent the link between users j and i in the OSN at time t . These are unweighted, but directed links, because it makes a difference whether user j follows user i , or the other way around. $a_{ij}(t) = 1$ if there is a link from j to i , i.e. j is called a *follower* of i , and $a_{ij}(t) = 0$ otherwise. Since it is unrealistic for a user to follow himself, we set $a_{ii}(t) = 0$ for all t .

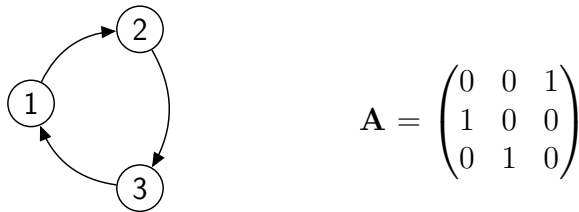


Figure 6.1: An example of a small, directed network (left) and its adjacency matrix \mathbf{A} (right). The network consists of three users, each with one follower.

Note that in the sum of Eq. 6.2, the reputation of a user, i , does not depend only on the sheer number of users that follow i , but also on their reputation. In our model we keep the total number of users, N , constant. This implies that if a certain number of users leave the OSN, they are replaced by the same number of new users joining. This is, arguably, an oversimplification in including *entry and exit* dynamics. While more realistic approaches

could certainly provide additional insights, it suffices for our purposes, as we exclusively focus on the reputation dynamics and resilience.

The second term on the right-hand side of Eq. 6.2 takes into account the efforts required to keep a certain level of reputation. Reputation that is not actively maintained will fade out in the course of time, i.e. users who are not able to attract followers with sufficient impact automatically lose reputation. This is captured by the exponential decay rate ϕ . The larger the ϕ , the higher the effort to maintain one's reputation or equivalently the faster reputation decays for the same amount of effort. Considering that users face the same functional tasks of participation (e.g. logging in, consuming and producing content, etc.), we approximate their costs, $c_i(t)$, as a constant τ equal for all users. Note that there is no cost involved in maintaining individual links. According to Eq. 6.2 and dependent on the specific social network, the reputation of users can grow to large numbers or become infinitely small, but never zero.

Finally, borrowing from social comparison theory [71], we pose that users would pay more attention to their *relative* reputation compared to others rather than to the absolute value in Eq. 6.2. Social comparison theory suggests that people form their attitudes and behaviours by comparing themselves to others. This psychological tendency is likely to play a role in online communities as well [70, 106]. By the various ways in which **Facebook** profiles can be polished to others, e.g. career success, education, family status, travel destinations, etc., users consciously measure their own life achievements against those of their friends. There are already reports that raise awareness to the growing negative social comparison in **Facebook** that increases the number of users suffering from depressions [70]. Ironically, **Facebook** started precisely as a website for comparing physical attractiveness among college students.

We, thus, define the relative reputation as $X_i(t)$ re-scaled by the user reputation in the OSN, $X_{\max}(t)$. In this way, different networks become comparable.

With these considerations, the benefits and costs in our model are now fully specified

$$b_i(t) = \frac{X_i(t)}{X_{\max}(t)} \in (0, 1) ; \quad c_i(t) = \tau \in [0, 1). \quad (6.3)$$

Note that in contrast to Eq. 6.3, it is common in the general class of heterocatalytic models to express the relative reputation in terms of the sum of the individual reputations, $\sum X_i(t)$

$$x_i(t) = \frac{X_i(t)}{\sum_j X_j(t)}. \quad (6.4)$$

In the context of OSNs, however, this is less realistic, as it requires that at least the total reputation is public knowledge. The most reputable individual, on the other hand, is often visible in top ranking lists that almost all reputation-based OSNs provide. Despite the conceptual difference, a solution to Eq. 6.3 can be mapped directly to a corresponding solution to Eq. 6.4, by normalising x_i with respect to x_{\max} as follows

$$\frac{x_i(t)}{x_{\max}(t)} = \frac{X_i(t)/\sum_j X_j(t)}{X_{\max}(t)/\sum_j X_j(t)} = \frac{X_i(t)}{X_{\max}(t)} = b_i(t) \quad (6.5)$$

In Appendix 12.1, we further show that an equilibrium solution to Eq. 6.2 is also an equilibrium for $b_i(t)$ and $x_i(t)$ up to a scaling factor. This means that the entry/exit dynamics in Section 6.2.3 is invariant to the particular way in which users evaluate their relative reputation.

According to Eq. 6.1, users leave the OSN at time t if their *relative reputation* is lower than the fixed threshold τ . Their links $a_{ij}(t)$ are then set to zero, which according to Eq. 6.2 reduces the reputation of other users j at the next time step. This can lead to cascades of users leaving the OSN at consecutive time steps. The main contribution of our model is to understand how a deterioration of user motivation to stay engaged, due to an increase of the fixed costs τ , will impact the resilience of the OSN. Therefore, as a next step, in Section 6.2.2 we first investigate how the structural properties of the underlying social network affect the reputation dynamics, before turning our attention to the entry/exit dynamics in Section 6.2.3.

6.2.2 Quasistationary Equilibrium

Consider a *fixed* social network. Expressing the reputation dynamics of $b_i(t)$ from Eq. 6.3 yields (see Appendix 12.1)

$$\frac{db_i}{dt} = \sum_{j=1}^N a_{ij} b_j(t) - b_i(t) \sum_{j=1}^N a_{mj} b_j(t), \quad (6.6)$$

where m is the index of the individual with highest absolute reputation X_{\max} . The first term describes the reputation boost that individual i obtains from all his followers. The second term is a scaling factor and represents the reputation decay with strength equal to the total boost in reputation that user m receives.

Eigenvalues and eigenvectors. The set of Eqs. 6.6 forms a linear dynamical system of coupled first-order differential equations with the initial conditions described by the vector $\mathbf{b}(0) = \{b_1(0), b_2(0), \dots, b_N(0)\}$. The (constant in time) a_{ij} forms the elements of an adjacency matrix \mathbf{A} of size $N \times N$ (see Figure 6.2). At equilibrium we require $\dot{b}_i = 0$, hence

$$\sum_{j=1}^N a_{ij} b_j(t) = b_i(t) \sum_{j=1}^N a_{mj} b_j(t). \quad (6.7)$$

In matrix form, this equation becomes

$$\mathbf{A}\mathbf{b}(t) = \mathbf{b}(t) \sum_{j=1}^N a_{mj} b_j(t). \quad (6.8)$$

If \mathbf{X}^λ is an eigenvector of \mathbf{A} with a corresponding eigenvalue λ , then re-scaling \mathbf{X}^λ will also produce an eigenvector, i.e. $\mathbf{b}^\lambda = \mathbf{X}^\lambda / X_{\max}^\lambda$ is a solution to Eq. 6.8. In this case the scaling factor $\sum_j a_{mj} b_j(t)$ gives the m^{th} component, $(\lambda b^\lambda)_m$, of the vector $\lambda \mathbf{b}^\lambda$. Since $(\lambda b^\lambda)_m = X_m^\lambda / X_{\max}^\lambda = 1$, it follows that $\sum_j a_{mj} b_j(t) = \lambda$.

Further, we know that \mathbf{A} is a real, non-negative matrix, hence the *Perron-Frobenius Theorem* tells us that the largest eigenvalue of \mathbf{A} is real and positive. We denote this eigenvalue as λ_1 and refer to it as the Perron-Frobenius eigenvalue [19]. The theorem also guarantees that the corresponding Perron-Frobenius eigenvector $\mathbf{b}^{\lambda_1} = \{b_1^{\lambda_1}, b_2^{\lambda_1}, \dots, b_N^{\lambda_1}\}$, contains only real and non-negative entries, which define the relative reputation values of each user in equilibrium⁵.

Let us illustrate this by an instructive example of the rather small network shown in Figure 6.2.

The characteristic polynomial determining the eigenvalues of the corresponding adjacency matrix \mathbf{A} given in Figure 6.2 is

$$-\lambda^5 + \lambda^3 + \lambda^2 = 0 \quad (6.9)$$

and the largest eigenvalue is $\lambda_1 = 1.32$. The corresponding eigenvector gives us the absolute reputation: $\mathbf{X}^{\lambda_1} = \{2.32, 1.75, 1.32, 1.32, 1\}$. Rescaling this eigenvector by $1/X_{\max}^{\lambda_1}$ gives

⁵The Perron-Frobenius eigenvector represents the unique asymptotically stable attractor of this dynamics. To see this, imagine our initial vector $\mathbf{b}(0)$ as a linear combination of all eigenvectors of \mathbf{A} . Then Eq. 6.6 tells us that for large times t , the component of $\mathbf{b}(0)$ corresponding to the largest eigenvalue will dominate all others as $\mathbf{b}(t) = e^{\lambda_1 t} \mathbf{b}^{\lambda_1}$ where \mathbf{b}^{λ_1} is the Perron-Frobenius eigenvector.

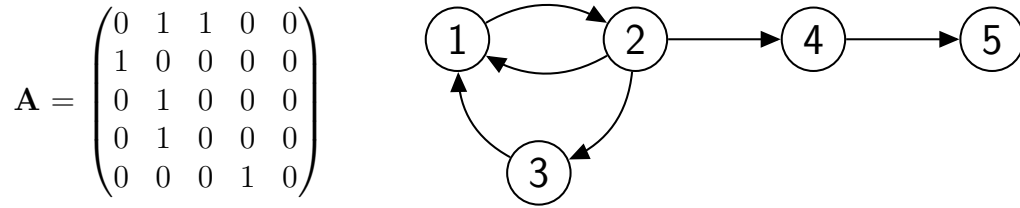


Figure 6.2: An example of simple OSN (**right**) that displays a core-periphery structure (see main text) and the corresponding adjacency matrix **A** (**left**)

the relative reputation in equilibrium: $\mathbf{b}^{\lambda_1} = \{1, 0.75, 0.57, 0.57, 0.43\}$.

Core-periphery structure. To facilitate the interpretation of these relative reputation values, let us take a look at the corresponding network structure. This toy network already shows a topological feature typical for many social networks, known as a *core-periphery* structure [37, 40, 53, 111, 113, 147, 259]. Following the seminal work by Borgatti and Everett [29, 68], the core is defined as a cohesive subgroup (e.g. a clique, n-clique, n-club or n-clan, k-plex) and the periphery is everything else. Accordingly and accounting for the directionality in reputation-based OSNs, we define the core, Q , as the largest strongly-connected component (SCC). Each node in the SCC is reachable by all other nodes in the SCC. The periphery consists of all nodes that do not belong to the core.

It is known that the Perron-Frobenius eigenvalue of a directed OSN (λ_1) is equal to the Perron-Frobenius eigenvalue of its largest SCC and can be used to characterise the structure of the core Q (see Appendix 12.2). In Figure 6.2, the core is the SCC formed by users 1, 2, 3, who mutually boost each others' reputations. The periphery consists of users 4 and 5 who only benefit from being connected to the core. The periphery usually contains simple *chains* of users ($4 \rightarrow 5$) which emanate from the core, in this case from user 2. The core itself does not contain simple chains, but *cycles*, i.e. *closed* directed chains that involve $n \geq 2$ users [123]. In the example, we observe two cycles, $1 \rightarrow 2 \rightarrow 1$ and $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$, i.e. users can be part of more than one cycle. Moreover, all users have one follower each, except for user 1 who has two followers, thus his reputation can be expected to be higher than that of the others. Note that even though users 4 and 5 have the same number of followers, their reputations are different, as user 4 has a more reputable follower belonging to the core.

More formally, a *cycle* is defined as a subgraph in which there is a closed path from every node of the subgraph back to itself. Cycles and structures of interlocking cycles represent *irreducible subgraphs*. The core Q must always contain at least one cycle for it to be a strongly connected component. In Appendix 12.2, we show the dependency between the Perron-Frobenius eigenvalue and the length and number of the cycles in the core.

Direct and indirect reciprocity. The shortest cycle, $n = 2$, expresses *direct reciprocity*. In Figure 6.2 users 1 and 2 mutually follow each other and boost each other's reputations. For $n \geq 3$, we have *indirect reciprocity*. In the same figure, user 2 contributes to the reputation of user 3 who, however, does not reciprocate. Instead, user 3 follows user 1, and it is only through user 1 that the loop is closed and user 2 receives a reputation boost.

Direct reciprocity is very common in OSNs, e.g. in **Twitter** or **Google+** it is seen as good practice to link back to someone that has chosen to follow you or to have you as his/her friend. Likewise, **likes**, **+1**, or shared posts often take direct reciprocity into account. Compared to this, indirect reciprocity is more difficult to detect. To boost interaction along a chain of followers, and to hopefully close the loop, some OSNs like **Google+** or **Researchgate**, indicate for each follower the number of additional users that the focal user and the follower both have in common. This may increase the likelihood of creating shortcuts and also shorter cycles. Both direct and indirect reciprocity are mechanisms that increase connectivity in an OSN, and promote the creation of larger and more stable cores.

Different OSNs show varying reciprocity patterns. Previous research has reported high reciprocity for **Yahoo!** [360] and **Flickr** [147], but **Twitter** seems to have a much lower reciprocity [148]. Empirical studies have also shown that users have increased likelihood of creating shortcuts that shorten the cycles of the network [263].

Length of simple chains. At equilibrium we can plug in the eigenvector \mathbf{b}^{λ_1} corresponding to λ_1 into Eq. 6.6 to obtain

$$\sum_j a_{ij} b_j^{\lambda_1} = \lambda_1 b_i^{\lambda_1}, \quad (6.10)$$

or conversely

$$b_i^{\lambda_1} = \frac{1}{\lambda_1} \sum_j a_{ij} b_j^{\lambda_1}. \quad (6.11)$$

This means that, in the long run, the reputation b_i of user i is equal to the sum of the reputations of all users j that follow i , *attenuated* by a factor $1/\lambda_1$.

Eq. 6.11 allows us to draw some conclusions about the maximum length of simple chains involving peripheral users. In our toy example from Figure 6.2, the reputation of users 4 and 5 are related to the attenuation factor in Eq. 6.11 such that $b_4 = b_2/\lambda_1$, $b_5 = b_2/\lambda_1^2$, and in general $b_n = b_2/\lambda_1^{n-1}$ for a chain of length n . If we require a simple chain to be

exactly of length n , $b_n > \tau$ and $b_{n+1} \leq \tau$ must hold. In other words, the n^{th} peripheral user finds it beneficial to stay while the $(n + 1)^{\text{th}}$ leaves. Hence, we obtain for n

$$n \in \left[\frac{\ln(b_2/\tau)}{\ln \lambda_1}, \frac{\ln(b_2/\tau)}{\ln \lambda_1} + 1 \right), \quad (6.12)$$

and since we require an integer value

$$n = \left\lceil \frac{\ln(b_2/\tau)}{\ln \lambda_1} \right\rceil. \quad (6.13)$$

The maximum length of a simple chain in a core-periphery network thus depends on the cost level τ , the connectedness within the core Q expressed by the largest eigenvalue λ_1 , and the relative reputation of the core user who connects the core to the chain (in our example user 2).

Unstable cores. Imposing the condition $n > 0$ in Eq. 6.13 requires that $\lambda_1 > 1$ and $b_2 > \tau$, which holds only if the OSN contains *cycles*. Without these cycles, $\lambda_1 = 0$, i.e. the core-periphery structure breaks down. The condition, $b_2 > \tau$, requires that the core user 2, who connects to the simple chain, needs to obtain a net gain from participating in the OSN. Otherwise, due to the attenuation factor in Eq. 6.11, none of the users in the chain would have an incentive to stay in the network.

The special case of $\lambda_1 = 1$ represents an important exception. Network structures like these are characterised with a core that consists of a minimally connected cycle, that is a cycle that has *exactly* as many edges as number of nodes in it [124]. Referring to Figure 6.3, users 1, 2, and 3 form such a core.

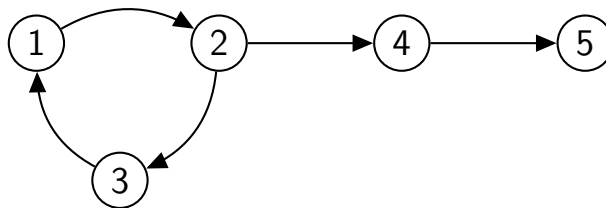


Figure 6.3: An example of a minimally connected core and a peripheral chain of theoretically unbounded size.

The *absolute* population dynamics in Eq. 6.2 for this simple core, assuming $\phi = 0$, is given by

$$X_i(t) \stackrel{t \rightarrow \infty}{\sim} e^{\lambda_1 t}, \quad \text{for } i \in [1, 3] \quad (6.14)$$

Hence the Perron-Frobenius eigenvalue can be interpreted as the “reputation growth rate” of the core. Since $\lambda_1 = 1$, the core users are not connected densely enough to re-enforce their reputation above their contribution to the periphery. In this case, and provided that $b_2 > \tau$, Eq. 6.13 tells us that the length of the simple chains is no longer bounded. Consequently all users in the peripheral chain would have the same relative reputation as the core. More importantly, however, network structures like these are very unstable, because core users have a reputation comparable to the periphery. In fact, peripheral users can often obtain a higher reputation from multiple connections to the core, which reduces the benefit of the core users and increases their likelihood of leaving. Different from peripheral users, the leave of core users considerably affects other core users that are part of the same cycle. This starts cascades of users leaving and thus destroys the core. Without the core, the periphery would not be able to sustain its reputation and would break down as well.

Number and length of cycles. Unstable core-periphery structures can be avoided as long as interlocking cycles appear in the core. These contain users involved in multiple cycles from which they receive a much higher reputation and increase the benefit, i.e. the relative reputation, also for others. Both the number of cycles in a network and their length have an impact on the largest eigenvalue λ_1 as illustrated in Appendix 12.2. In general, we can conclude that λ_1 *increases* with the *number of cycles*, but *decreases* with the *length* of the cycle, keeping everything else fixed. The number of cycles in the core of the network further depends on the average density m (average number of connections per user), a parameter discussed in the next section when we introduce entry/exit dynamics.

6.2.3 Network dynamics

In the previous section, we explained that the relative reputation dynamics of Eq. 6.6, for a *fixed network*, converges to an equilibrium state in which the relative reputations of users given by the eigenvector corresponding to λ_1 . This convergence time defines the time scale for the reputation dynamics. Dependent on the stationary reputation values that define their benefits in relation to their costs τ , users can decide to leave the OSN, Eq. 6.1, and will be replaced by new users joining the OSN. As described in Section 6.2.1, we assume that $N_{\text{exit}}(T) = N_{\text{entry}}(T)$, to keep N constant. T is the time at which entry and exit happens. We assume that the time scale for entry and exit, i.e. for changing the network structure, can be separated from the much shorter time scale of the reputation dynamics. In other words, users make their decision based on the quasi-stationary benefit, which can only change after the network has changed.

If a user leaves the OSN, all his incoming and outgoing links are set to zero. The newcomer will establish the same number of incoming and outgoing links *on average* but, assuming that he has no knowledge of other users, these links will be distributed at random. This mechanism is even simpler than the well-known model of preferential attachment in which followers connect with higher probability to more reputable users. We also assume that each user follows on average m other users. More precisely, if there is a constant probability p that a new user links to any of the $(N - 1)$ other users, then $m = p(N - 1)$ is a constant related to the average density of a random OSN.

Our major interest is in the role of the cost τ that, according to Eqs. 6.1 and 6.3 defines the level at which users will leave the OSN, measured in terms of relative reputation. The latter is between zero and one, so τ gives the fractional benefit that has to be reached to stay in the OSN. In our computer simulations, we will vary this level from zero to 0.5 to study the impact of increasing costs.

$\tau=0$ would imply no costs. To still allow for a network dynamics in this case, we apply the usual extremal dynamics, i.e. the user with the lowest relative reputation is removed and a new user enters. In case several users have the lowest reputation, we choose one of them randomly. The other limiting case $\tau = 1$ would imply that all N users will leave and be replaced by a completely new cohort. Then, the network at every time T starts as a new random network with no chance to evolve. Hence, small or intermediate values of τ would be most appropriate.

6.3 Results of computer simulations

6.3.1 Resilience

Recall that we defined *resilience* as the *long-term average benefit* in an OSN. As benefits are measured in terms of relative reputations, resilience becomes the long-term average user reputation in the network. This quantity can be seen as a determinant for the success and attractiveness of an OSN, as high average reputation would signal the high quality of the information being disseminated through the network. Calculated in this way, resilience conforms to our intuitive notion that a resilient OSN is able to withstand changes. If most of the users are highly-reputable, the cost of usage must increase considerably to cause a collapse. In addition, the size of the departure cascades will be limited as only peripheral users will be affected for moderate cost levels.

With this in mind, we denote the long-term average reputation as $\langle b \rangle$, and obtain the

following expressions

$$\bar{b}(T) = \frac{1}{N} \sum_{i=1}^N b_i(T) = \frac{1}{N} \sum_{i=1}^N \frac{X_i^{\lambda_1}(T)}{X_{\max}^{\lambda_1}(T)}; \quad \langle b \rangle = \frac{1}{R} \sum_{r=1}^R \bar{b}_r(T_{\max}). \quad (6.15)$$

$\bar{b}(T)$ refers to the population average at a given time T , which can considerably fluctuate because of the stochastic influences in changing the network structure at every network time step T . Therefore, the long-term average benefit $\langle b \rangle$ is a system average taken over a large number of independent simulations R . Each simulation was run for $T_{\max} = 12000$ network time steps. This is long enough for the network to undergo various cycles of growth and crashes. The results of our simulations are shown in Figure 6.4 for various values of the costs τ and the average number of links m .

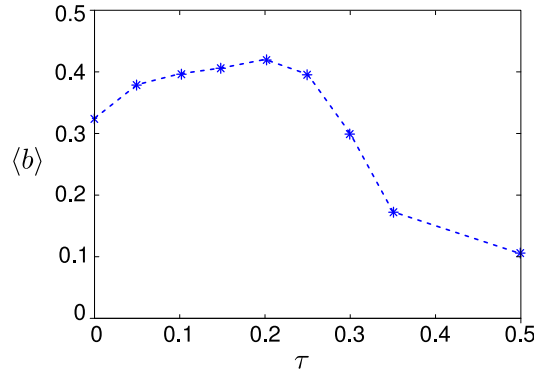


Figure 6.4: Long-term average benefit $\langle b \rangle$, Eq. 6.15 for varying costs τ . Other parameters $m = 0.25$, $N = 100$.

The most interesting observation shown on the left side of Figure 6.4 is the maximum average benefit $\langle b \rangle_{\max}$ for a non-zero cost at $\tau_{\max}=0.2$. Counterintuitively, this result implies that adding a cost for participating in an OSN maximises the resilience in the system. Appendices 12.3 and 12.4 describe precisely how this optimal value emerges. Figure 12.1 also provides additional computer simulations to verify the finding. Here, we summarise the explanations.

Recall from Section 6.2.2 that resilience depends on the existence of a distinct core-periphery structure. Without a differentiated core in which core users re-enforce their reputation, the overall reputation in the network will necessarily vanish. Once a core is established, the entry/exit dynamics from Section 6.2.3 puts pressure on its existence that is determined by the cost τ . In general, any cost level will trigger cascades of leaving users. As long as only peripheral users are affected these cascades will not destroy the

whole network. The core will remain intact and the resilience of the OSN will actually increase due to the removal of low-reputation users from the periphery.

Irrevocably, the core will be destroyed as costs increase, however an OSN should be able to recover quickly as to not affect the long-term satisfaction of its users. Hence, not only the life time of the core (i.e. the core-periphery structure), but also the time to (re)establish this structure matter. As Figure 12.1 shows, the value $\tau_{\max}=0.2$ optimises the ratio between these two time spans. Importantly, as Appendix 12.4 shows, the network is not considerably disturbed at this optimal cost level. This shows that the maximum average benefit is not concentrated in a minority of core users, at the expense of a large and sparsely connected periphery.

The lifetime of the core-periphery structure is mostly challenged by peripheral users who attract followers, without reciprocally contributing to the reputation of others. A higher cost $0 < \tau < \tau_{\max}$ prevents this “behaviour” to some extent, as peripheral users, who have in general less reputation than the core, are removed. However, increasing the cost further, $\tau_{\max} < \tau < 1$, would likely affect the highly reputable core users as well, and destroy the core, consequently the depending periphery. Interestingly, the recovery time of the core-periphery structure becomes shorter if τ increases. The reason is that higher cost means more new users enter the OSN, and their combined re-linking back to the network increases the chances of establishing reciprocal relations, that is a core. But, again, if the randomness associated with this process becomes too high, favourable structures may be destroyed quickly. The small, but considerable cost $\tau_{\max}=0.2$ is able to balance these counteracting processes.

To better understand the improvement of resilience with costs, we discuss some further properties of the OSN in the next sections.

6.3.2 Core size and largest eigenvalue

In Section 6.2.2 we mentioned that an increase in cost τ affects the structure of the OSN. In particular, the length of both simple chains and of cycles of followers will be reduced (Eq. 6.13). This results in a decreasing size Q of the core built by users that belong to one or more cycles. Figure 6.5 shows the histogram of the core sizes, $P(Q)$ for two different cost levels, a reference $\tau=0$ and $\tau=0.25$. To allow for a real comparison of the different simulations, we have taken into account only those realisations where the core-periphery structure encompasses the whole network (i.e. one connected component, and no isolated users, or groups of users), and each user has at least one follower. One recognises that with increasing cost, the distribution becomes more right-skewed, with its maximum shifted to

smaller values of Q . This is to be expected as higher costs makes larger re-enforcing cycles in the core less likely to occur due to random chance. Actual snapshots of this network are shown in Figure 12.2.

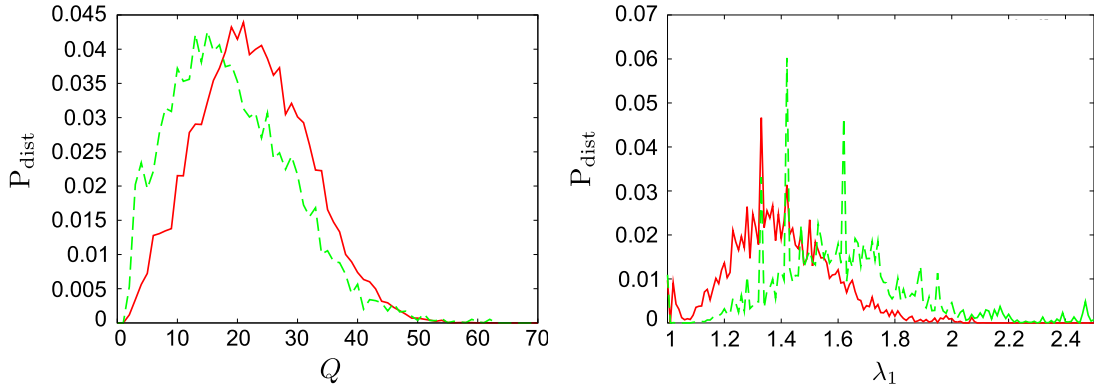


Figure 6.5: **Left:** Distribution $P(Q)$ of core sizes Q . **Right:** Distribution $P(\lambda_1)$ of the largest eigenvalue λ_1 . (red) $\tau = 0$, (green) $\tau = 0.25$. Other parameters $N = 100$, $m = 0.25$, $T = 10^6$ time steps. Figure from [227] used with permission.

The second structural insight comes from the histogram of Perron-Frobenius eigenvalue, $P(\lambda_1)$ shown in the right-hand side of Figure 6.5. Compared to the case with no costs, we observe a considerably broader distribution, with the maximum shifted to larger values of λ_1 . Recall that in Section 6.2.2 we concluded that λ_1 increases with the number of cycles in the core. Since higher costs decrease the size of the core, an increase of λ_1 can only mean that the core, while being smaller, is better connected. This makes sense, as it is the only way for a self-organised core to survive harsher environments. Comparing the snapshots in Figure 12.2 confirms this conclusion.

We remind that the condition in Eq. 6.13 refers only to *simple* chains, that is, chains in which each user has exactly one follower. If users have several followers, their benefit-cost difference can overcome the condition in Eq. 6.13 even if their position in the chain would not, otherwise, allow it. Such users automatically boost the reputation of all other users down the chain. As the cost τ increases, having more than one follower becomes crucial in particular for peripheral users to stay. Having more than one follower, on the other hand, also increases the chance of creating new cycles, which in turn increases λ_1 . This again feeds back both on the core and the periphery of the OSN, increasing the pressure towards a more compact core and shorter chains. Hence, it is in fact the relation with λ_1 that facilitates the role of the cost τ in shaping the network.

6.4 Conclusion

Introducing costs of usage can be a powerful tool to increase the resilience of an OSN by preventing it from “dying from success”. This phenomenon occurs, when an OSN attracts many users so quickly that the information filtering mechanisms fail to cope with all the content users produce, and the resulting information overload drives users away. A possible way to avoid this is including socially-aware recommender systems [261] and filtering mechanisms.

We pose an alternative: the costs of using an OSN can be tuned to encourage quality above quantity, driving the community towards a higher benefit for its users, and thus higher resilience, but not necessarily larger user base. Our agent-based model, though simplistic, captures several generic features of such OSNs:

Core-periphery structure. Most OSNs are characterised by a core of active users that are closely linked to each other, and a periphery of less active users that are loosely connected to the core. Our model is able to reproduce such a structure, but also to explain its origin, as a combination of direct and indirect reciprocity. Indeed, the various linking mechanisms that the most popular OSNs implement (e.g. friendship suggestions based on number of common friends) indicate that OSN providers recognise that reciprocity drives the creation of more stable and resilient structure.

Resilience. The entry/exit dynamics in Section 6.2.3 ensures that the structure of the OSN never converges to an equilibrium. Because user reputation depends also on the number of followers, the decision of some users to leave the OSN, due to a bad cost-benefit evaluation, will decrease the reputation of others as well. This is the basic mechanism which creates cascades of users leaving the OSN. We propose that an OSN is *resilient* if the majority of users with a considerable reputation *stays*, despite the fact that a large number of users may leave at any time step and will be replaced by newcomers. In this way, resilience conforms to our intuitive notion that a resilient OSN is able to withstand changes. If most of the users are highly-reputable, the cost of usage must increase considerably to cause a collapse.

Our main focus was on the relationship between resilience and the *cost of usage* τ . Increasing τ represents a pressure on the users to leave, because it widens their cost-to-benefit trade-off. In order to estimate the impact of increasing costs, we used the long-term benefit averaged over the whole OSN, $\langle b \rangle$, as a resilience measure. Intuition would suggest that $\langle b \rangle$ monotonously decreases with increasing costs up to a point where the whole OSN

collapses. Interestingly, this picture does not hold for comparably small cost values. On the contrary, a small cost up to $\tau_{\max}=0.2$ is beneficial, as it *increases* the long-term average benefit, compared to a reference case without any costs. In other words, a small cost level expels those users who were never able to attract any followers.

We also investigated the precise way in which the cost level changes the structure of the network and affects resilience. First the optimal cost level balances two dynamic effects: (i) the *emergence* of a core of users that, by means of direct or indirect reciprocity, maintain a relatively high reputation, and (ii) the *maintenance* of this core by the users. Resilience does not simply mean that this core stays alive, it also means that a new core is quickly established when the old one crashes. This way, the average reputation of most users remains high *over time*.

Second, we discovered that the Perron-Frobenius eigenvalue, λ_1 , of the adjacency matrix of the network can be used to determine the structure of the core itself. The eigenvalue increases with the number of cycles in the core, and decreases with the size of the core, keeping the number of cycles fixed. On the other hand, λ_1 also increases with τ . Hence, increasing the cost reduces the size of the core, but forces core users to increase their connectivity in order to keep a positive net benefit.

As concluding remarks, we emphasise that the above conclusions are obtained from a model that, as every modelling approach, only captures a part of reality. Our model certainly does not include all features of a real OSN, and even simplifies those that are included. For example, we ignored the component of the benefit that users obtain from consuming content, and only focused on the contributions from the ability to capture attention, i.e. reputation. Though justified, this simplification certainly leaves out other interesting insights. Another example is our entry/exit dynamics. Arguably, new users do not form connections completely at random, but may decide based on the reputation of existing users (in a multiplicative process akin to preferential attachment) or based on existing offline relationships.

In choosing the right amount of complexity we were driven by one main consideration – “what is the minimum amount of model complexity needed to study the interplay between cost of usage and resilience?”. This is an important question, because there is always a trade-off between complexity and tractability. It is surprising to find out that we do not need complex assumptions about user engagement, user entry or benefit function. Already accounting for user reputation, as recursively dependent on the reputation of one’s followers, and including a rudimentary network dynamics allows us to demonstrate that there exists an optimal cost of usage for an OSN, which maximises resilience. That this result arises from a simple model speaks only about its fundamental nature.

Part II

Decisions in Animal Societies

“For instance, on the planet Earth, man had always assumed that he was more intelligent than dolphins because he had achieved so much — the wheel, New York, wars and so on — whilst all the dolphins had ever done was muck about in the water having a good time. But conversely, the dolphins had always believed that they were far more intelligent than man — for precisely the same reasons.

The last ever dolphin message was misinterpreted as a surprisingly sophisticated attempt to do a double-backwards-somersault through a hoop whilst whistling the 'Star Spangled Banner', but in fact the message was this: So long and thanks for all the fish.”

DOUGLAS ADAMS

The Hitchhiker's Guide to the Galaxy (1979)

Chapter 7

Mitigating negative consequences

Summary

In this last chapter, we introduce the coordination problem associated with communal roosting in Bechstein bats. Communal roosting provides evolutionary benefits, such as vital thermoregulation and increased protection from predators, however, it is conditional on individuals coordinating their roosting decisions. This is not a trivial task, as bats possess limited and often conflicting information on the roosting preferences of others. Empirical studies have observed that information transfer about suitable roosts is a likely mechanism for achieving collective coordination. We follow these works by investigating the actual recruitment patterns underlying the information transfer to infer the individual-level rules governing it. To this end we analyse unique datasets on bats movements in and out of experimental roosts and show a robust way of inferring recruitment events. Next, in a novel application of social network theory, we test various recruitment mechanisms to find the simplest one that reproduces the observed recruitment patterns. Finally, this approach allows us to isolate individuals who seemingly display more complex behaviour and can, thus, serve as a guide for future experimental studies.

Based on Mavrodiev, P., Fleischmann D., Kerth, G., Schweitzer F., *Information transfer in Bechstein's bats: a network perspective* (working title), advanced-stage manuscript. D.F. and G.K provided the datasets on animal behaviour. P.M. conducted all data processing and analysis, as well as the theoretical investigations on recruitment mechanisms. P.M. and D.F work jointly on writing the manuscript in an article format suited for a biologically oriented journal.

7.1 Introduction

The first part of this thesis was largely concerned with designing interaction mechanisms that govern collective decisions in human groups. In doing so, we were able to shed light on some of the unintended outcomes of collective decisions regarding the role of payoff-related information on the emergence of cooperation, the effects of herding and social influence on the wisdom of crowds, and the impact of user departures on the resilience of online communities.

In this second part, we focus on collective decisions in animal systems, in which individual interactions are hard to design from the ground up. Indeed, by observing animal group decision-making in natural habitats, we are limited in our ability to influence individual behaviour, and instead try to *infer* the mechanisms underlying this behaviour. We will be studying two colonies of Bechstein bats facing a coordination problem critical to their survival. The nature of the problem stems from the limited and conflicting information that individuals have on the knowledge and preferences of their colony mates. Failure to coordinate bears a highly negative and unintended collective outcome – that of group splitting – which endangers the survival of the whole colony. In the rest of the chapter, we follow up on empirical work dealing with this problem by modelling the coordination process with the aim of eliciting the inter-individual mechanisms behind it.

7.2 Coordination problem in Bechstein bats

Study animals. The Bechstein’s bat (*Myotis bechsteinii*) is a medium sized Vespertilionid bat species that lives up to 21 years in nature. The females breed communally in maternity colonies of 10 to 45 individuals that are established from the beginning of April to the end of August (males are solitary) [130]. During the summer months, colony members switch between up to 50 communal roosts in tree cavities and bat boxes, spread over a forest area of 30-50 ha, on an almost daily basis [130].

Communal roosting under limited information. A distinguishing characteristic of Bechstein bat colonies is their fission-fusion dynamics [130]. The fission phase takes place during the night, when the colony splits into individuals who forage alone or in small groups. From an evolutionary perspective, splitting is an adaptation to resource scarcity allowing decreased competition for unpredictable and patchy food resources [55]. During the day, individuals assemble back, or “fuse”, into two to six groups to roost. Communal roosting is believed to provide evolutionary benefits, in terms of social thermoregulation

and increased protection from predators [199]. It is also proposed that roosting in groups enables the exchange of information about vital resources, such as the location of food patches or novel roosts [132, 207]. In this sense, the communal roost represents a type of information center, in which individuals mutually benefit from exchanging information¹.

During foraging, individual bats explore their habitat and accumulate private information and preferences about novel roosts. For example, upon discovering a roost, the presence of other animals is perceived as negative, while warmth – as positive. Additionally, some individuals are intrinsically more active in exploring their habitat [73], and thus become better informed about the location of suitable roosts, while others remain relatively uninformed. As a result, bats have limited information of the knowledge or roosting preferences of others. If individuals relied only on their private information in choosing a roost, they run the risk of roosting either alone or in a too small of a group to benefit from communal roosting. The latter poses a real danger to the survival of the individual, as it deprives her of crucial thermoregulation. Therefore, for a colony to be sustainable, the ability to coordinate heterogeneous preferences, under limited information, is vital for mitigating the detrimental consequences of colony fission during roosting.

Field experiments have shown that to achieve this critical coordination, Bechstein bats engage in active information transfer during the fission phase [132, 267]. Individuals, who have discovered the locations of suitable roosts through independent exploration, transfer their knowledge to conspecifics by recruiting and leading them to these locations. The recruiter and the recruitee, thus form a leading-following pair. Information transfer in such a way provides both individual and group benefits. First, by advertising preferred roosts to others, an individual bat increases the likelihood of communally roosting with conspecifics, who liked these roosts, and therefore reduces the risk of roosting alone. Second, the combined information transfer of all individuals generates communal knowledge of a large set of roosts. Since Bechstein bats switch communal roosts frequently [133] (e.g. to avoid predators or parasites, or to maintain social bonds), communal knowledge of day roosts represents an important resource for the colony.

Despite the importance of information transfer for generating communal knowledge and preventing colony fission, the individual-level mechanisms underlying the recruitment process in Bechstein bats are largely unknown. There are novel empirical studies in this direction that have already found the absence of reciprocity and kin-selection [73, 132, 154, 190], however, the full complexity of the inter-individual recruitment process is still unclear. For

¹Note that we do not claim that individual benefits are symmetric. Empirical works have shown that colony members do not contribute equally to disseminating information about food sources [73, 132]. More active contributors are proposed to benefit in other ways, such as profit from foraging in larger groups and obtaining preferential roosting positions [207].

example, even though individuals are capable of individual recognition [210], do they employ it for recruitment? Or is there path dependence in forming leading-following pairs? Do leaders prefer to associate more with other leaders? Moreover, are there individuals with significant importance for spreading information through the colony? In this chapter, we use social network theory to address these questions. We analyse information transfer in Bechstein bats on unprecedented fine-grained level. For the first time, we construct leading-following networks that allow us to focus on individual behaviour and to *test* inter-individual mechanisms governing the recruitment process. The theoretical insights we provide can, thus, serve as a guide for future field work.

7.3 Mechanisms for information transfer

7.3.1 Data and methods

The work done in this chapter was conducted in close collaboration with the research group of “Applied Zoology and Nature Conservation” in University of Greifswald, Greifswald, Germany². The analysis in this chapter was predicated on raw data on bats’ behaviour (introduced below) that were collected and provided by the German research group [16, 73, 154, 190, 206].

Study site and roost monitoring. From 2007 to 2011, we studied two colonies (BS and GB2) of Bechstein bats within their home ranges located in the Guttenberg forest near Würzburg, Germany (Figure 7.1, left). The study period in each year was between the beginning of May and end of September. In that time, the home ranges were equipped with man-made, experimental bat boxes (Figure 7.1, right). These boxes were to serve as day roosts, similar to natural roosts in trees (e.g branch cavities, and woodpecker or rot cavities), in which the Bechstein bats sleep during the day. The difference from natural roosts, however, is that the experimental boxes were equipped with special reading devices that recorded bats’ visits. Since 1996, all bats in both colonies have been individually marked with individual microchips (PIT-tags) in their first year of life. Each microchip is programmed with a unique 10-digit ID that can be identified and recorded by the reading devices in the experimental boxes. In this way, every time a bat passes the entrance of an experimental box, her unique ID would be read out by the reading device with minimal disturbance to the individual.

²<http://www.mnf.uni-greifswald.de/institute/fr-biologie/institute-und-forschung/zool-institut-museum/angewandte-zoologie-und-naturschutz.html>



Figure 7.1: Left: Study site. Right: Experimental box

Once installed, the locations of the experimental boxes are unknown to the bats until they discover them through random exploration. In the beginning of the study period in each year, the experimental boxes were relocated within the home ranges to prevent memory effects from the previous year, and thus had to be discovered anew. Importantly, not all experimental boxes were discovered by the colony in a given year. Moreover, not all discovered and visited experimental boxes were used as day roosts. However, once an experimental box was used as a day roost by more than two bats, it was assumed that the group-decision process for that box was completed. The box was subsequently moved to a new location within the home range, and used as an experimental roost again.

Our datasets, thus, consist of the yearly raw recordings of the reading devices from all experimental boxes for each of the two colonies. A small sample dataset is shown below in Listing 7.1

```

1. Ser.-Nr.: [0401]
2. 02.06;00:50:25;00065db1f6;OK
3. 02.06;00:50:25;fffffffffff;OK
4. 02.06;01:00:47;00068e1ac4;OK
5. 02.06;01:00:49;fffffffffff;OK
6. 02.06;01:00:51;00068e1ac4;OK
7. 02.06;01:00:52;fffffffffff;OK
8. 02.06;01:00:52;00068e1ac4;OK

```

Listing 7.1: An excerpt from the recordings of an experimental box for the GB2 colony in year 2008. Line numbers serve as a visual guide only and are not part of the data.

Each line corresponds to one reading³, i.e. one activation of the reading device by a

³Except for the 1st line, which contains the serial number of the reading device; it is ignored.

visiting bat. Columns are separated by semicolon. The first column shows the date of the reading (in this case June 2nd), the second column indicates the time of the recording in 24-hour format, the third column contains the unique 10-digit ID of each bat⁴, and the last column is a status message. Table 7.1 shows a summary of the total number of readings and the number of installed, discovered and occupied experimental roosts, for each colony throughout the years.

Colony	Year	colony size	#readings	#boxes installed	#boxes discovered	#boxes occupied
GB2	2007	31	1002	17	11	4
	2008	34	4243	32	32	25
	2009	21	1273	21	16	6
	2010	44	878	17	12	3
	2011	16	1929	18	18	6
BS	2007	16	5600	25	20	12
	2009	17	9102	32	28	16
	2010	19	2169	23	19	7
	2011	7	2016	20	13	9

Table 7.1: Data summary

To obtain the identify of the individuals who roost in the experimental boxes, each box was checked daily between 8am and 11am for bat presence [16]. Since the entrances of the boxes face the ground (Figure 7.1, right), roosting bats can be seen by visual inspection using a flashlight. The individual PIT tags were then scanned by a mobile reading device.

Defining and inferring leading following events. We refine the nomenclature used by [132] to denote the information status, i.e. the knowledge, that individuals possess about the location of experimental boxes. An individual is said to be *naïve* at time t_1 regarding a given box, if she has not been recorded by the reading device in that box at all times $t < t_1$. Similarly, an individual is considered *experienced* at time t_2 regarding a given box, if she has been recorded in that box for any previous time $t < t_2$.

Out of the raw recordings for a given year and colony, we can identify *patterns* that correspond to certain events. Some pertinent patterns are (i) a “discovery” event defined as the first registered recording in an experimental box by a naïve bat; the bat is, appropriately, termed the box’s discoverer, and the box is considered discovered, (ii) an “exploration” event, which is the visit of a naïve bat in an already discovered box, and (iii) a “re-visit”

⁴Lines with `ffffffffff` contain irrelevant information produced as a peculiarity of the reading device. This information is ignored in the analysis and in all subsequent illustrations.

event registered when experienced individuals independently visit a discovered box [73]. These examples constitute *personal information gathering*, as individuals acquire, or reinforce, information about experimental boxes, such as their location or perceived quality, independently. In contrast, when such information is transferred socially, i.e. from experienced to naïve individuals, we refer to *social information gathering*, and the relevant pattern is a “leading-following” (L/F) event. Following [132], we define a leading-following (L/F) event to a box as the *joint* visit of a naïve and an experienced individual. The experienced bat is called the *leader*, and the naïve bat is the *follower* in the event. We denote an L/F event by $B \Rightarrow A$, where B is the follower, and A is the leader.

In our analysis, we associate with each L/F event (i) the experimental box in which it was detected, and (ii) the times at which the leader and the follower were recorded by the reading device in the box. Note that, it is not necessary for the leader to enter the box before the follower. Often it is the latter who is registered first. In case the leader and the follower were recorded multiple times, we take those times that minimise the difference between their appearances in the dataset (see Listing 7.2 and associated explanation). Finally, we refer to the **time difference** of an L/F event as the absolute difference between the recording times of the leader and the follower.

The actual inference of L/F events from the definition above depends on three parameters. The first parameter is the maximum allowed time difference between consecutive recordings of a leader and a follower, regardless of order. We refer to it as `lf_delay`. The `lf_delay` is important in determining which patterns constitute a joint visit of two individuals, as bats do not enter a box immediately upon arriving: females returning at night to a day roost usually encircle it several times before entering [132]. The `lf_delay` limit controls the sheer number of events we detect, since the higher the limit, the more likely it is to find an experienced and a naïve individual recorded within `lf_delay` of each other. In the limit of `lf_delay` $\rightarrow \infty$, we would detect the maximum number of leading-following events, many of which would be false positives, as bats recorded days apart would still be assumed to have “jointly” arrived at a box.

The second parameter represents the *minimum* time a follower in an L/F event needs to potentially become a leader, i.e. the time needed to find, recruit, and lead other followers. We denote it as `turnaround_time`. The importance of this parameter becomes apparent in Listing 7.2, which shows a frequently occurring recording pattern.

Assume that, for this box, individual `00065db1f6` is experienced at time `01:00:00` (line 1), individual `00068e1ac4` is naïve at `01:00:20` (line 2), and individual `00065ded81` is naïve at `01:01:10` (line 4). Taking `lf_delay` = 3 minutes (which is a good rule-of-thumb [132]) we can deduce from the first two lines that, according to the definition of an L/F event,

1. 02.06;01:00:00;00065db1f6;OK
2. 02.06;01:00:20;00068e1ac4;OK
3. 02.06;01:01:00;00068e1ac4;OK
4. 02.06;01:01:01;00065ded81;OK

Listing 7.2: A simplified example of how `turnaround_time` affects the inference of L/F events.

individual 00068e1ac4 followed individual 00065db1f6 to that box, i.e. $00068e1ac4 \Rightarrow 00065db1f6$. More precisely, we infer an L/F event to this box with the leader recorded at 01:00:00 and the follower at 01:00:20. The time difference of this event is 20 seconds.

Let us further assume that 00068e1ac4 liked the box she was just led to, and in turn would like to show it to other individuals. Her second recording in this dataset is on the third line – 40 seconds after her first appearance as a follower. Assuming that `turnaround_time` < 40 seconds, then 00068e1ac4 would have had enough time to fly around her home range, meet other individuals, recruit and ultimately lead them back to this box. In this example, she led individual 00065ded81 who appeared within a time of `lf_delay` from her, i.e. we would deduce the L/F event $00065ded81 \Rightarrow 00068e1ac4$. In addition, however, we see that 00065db1f6 and 00065ded81 appear within `lf_delay` of each other, hence we must also form the L/F pair $00065ded81 \Rightarrow 00065db1f6$. Evidently, this last L/F event contradicts $00065ded81 \Rightarrow 00068e1ac4$.

The issue is that, in reality, the 40-second delay between the two readings of 00068e1ac4 is not due to her having led another individual to the box. Instead, it is highly probable that, after having followed 00065db1f6, she was simply encircling the box for 40 seconds, and triggered the reading device a second time upon re-entry. The proper distinction between actual recruitment and such behavioural variability is the role of the parameter `turnaround_time`. In the toy example from Listing 7.2, a more realistic interpretation is that 00065db1f6 led both 00068e1ac4 and 00065ded81, i.e. we would only infer two L/F events. Note that since 00068e1ac4 appears twice, we associate the time of her first recording (01:00:20) with the L/F event $00068e1ac4 \Rightarrow 00065db1f6$, since it minimises the time difference to the recording of the leader.

The third parameter is the hour in the morning, on the day of a box occupation, after which subsequent recordings from this box are ignored. The necessity to ignore some recordings comes from the need to distinguish between genuine information exchange about suitable roosts (in terms of leading-following) and “pre-occupation” behaviour. Several hours before the occupation of a given box, experienced individuals who have decided to roost there, fly around the box and emit echolocation calls that serve to attract naïve

bats to the same box [188]. It has been suggested that this broadcasted social information is used by naïve bats (especially juveniles) to learn the location of suitable roosts from experienced conspecifics [129]. The result is that occupation is preceded by a growing group of individuals (experienced and naïve) flying around, or *swarming*, the roost for several hours. As it takes time to respond to the echolocation calls, the readings of naïve and experienced individuals in our datasets will tend to exhibit relatively large time distance. As a result, additional L/F events will be identified with time differences skewed toward the allowable limit of `lf_delay` (see Appendix 13.1 for illustration). These L/F events do not constitute genuine recruitment, in the sense that naïve individuals were led to a roost, but rather reflect the swarming phenomenon. Therefore, we define the parameter `occupation_deadline` as the temporal deadline on the day of a box occupation, after which subsequent readings in this box are attributed to swarming, and thus ignored.

Selecting parameter values. As illustrated above, each of the three parameters affects the inference of L/F events differently. Therefore, it is important to choose proper values that allow us to identify an adequate number of genuine leading-following events for statistical analysis.

Empirical research in the field of group-decision making in Bechtein bats has suggested 3 minutes for `lf_delay` and 3am for `occupation_deadline` as a reasonable rule of thumb [132]. We build upon these heuristics by comparing the distributions of time differences of all L/F events, fixing `lf_delay` and varying the other two within a reasonable range (see Figure 7.2). To generate sufficient sample sizes for the comparison, the dataset we chose to analyse was the GB2 colony in 2008 (Table 7.1). The reason is that, in 2008, the colony had the highest number of discovered and occupied boxes, the second largest colony size, and a large amount of raw readings. Therefore, we expected to identify the largest number of L/F events, which indeed proved to be the case (see Table 7.5)

Note that any combination of the three parameters is a 3-tuple, which generates a set of L/F time differences from all identified L/F events in the dataset. An example is presented in Figure 7.2, where we show histograms of the L/F time differences for `lf_delay = turnaround_time = 3` minutes, and `occupation_deadline = 2am` (left) and `occupation_deadline = 3am` (right).

Figure 7.2 also illustrates why we focus on the distributions of L/F time differences to select the values of the three parameters. As there is no objective method⁵ to quantify the behaviour underlying each of the parameters, we argue that L/F time differences best

⁵Objective, as in best reflection of reality. Indeed, one cannot “ask” a bat how much time she needs for recruitment or how far away she travels from a follower.

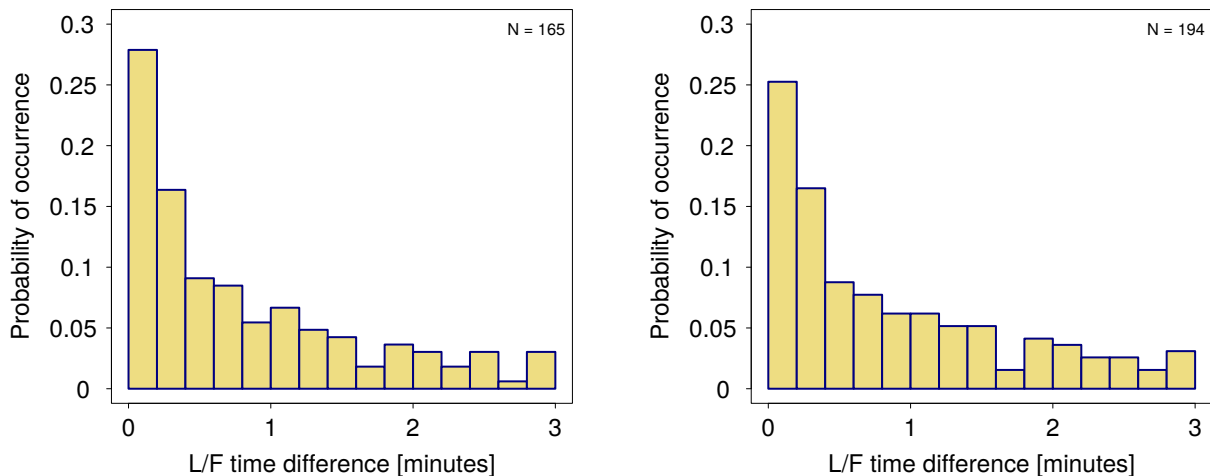


Figure 7.2: L/F time differences for the GB2 colony in 2008. Histograms show the absolute differences between the times at which the leader and the follower were recorded in all identified L/F events. Parameters: `turnaround_time = lf_delay = 3` minutes (both plots), `occupation_deadline = 2am` (left) and `occupation_deadline = 3am` (right). Insets indicates the total number of identified L/F events.

capture the effect that varying the parameters has on the L/F events we identify. For example, a visual inspection of Figure 7.2 hints that increasing `occupation_deadline` from 2am to 3am, does not change the time difference distributions. This implies that swarming has not yet set in (otherwise, we would expect quantitatively more events with larger time difference), and the additional L/F events on the right-hand side are genuine. Consequently, we would prefer `occupation_deadline=3am`, as it increases our sample size.

Table 7.2 formalizes this argument. Here, `lf_delay` is fixed at 3 minutes, while `occupation_deadline` is varied in $\{2\text{am}, 3\text{am}, 5\text{am}, 8\text{am}\}$, and `turnaround_time` – in $\{2, 3, 5, 7, 9\}$ minutes. For each value of `turnaround_time` (rows in the table), we compare the time difference distributions ($\mathcal{X}_i/\mathcal{Y}_i$) between all possible pairs of `occupation_deadline`. The comparison is done via a bootstrapped Wilcoxon rank-sum test on the null hypothesis that the two distributions are the same, against the two-sided alternative \mathcal{H}_1 , and the one-sided alternative \mathcal{H}_2 that $\mathcal{X}_i < \mathcal{Y}_i$. Each table cell shows the p-value for the two-sided and one-sided test, respectively.

As an example, fixing `turnaround_time = 2` minutes, we see that the distribution of L/F time differences for `occupation_deadline` at 2am is not statistically different from the distribution with `occupation_deadline` at 3am (p-value = 0.602). This is an indication that the nature of the identified L/F events is invariant to the later deadline, hence it is unlikely that we have inadvertently included swarming effects. Further inspection of the

turnaround_time	occupation_deadline pairs $\mathcal{X}_i/\mathcal{Y}_i$					
	2am/3am	2am/5am	2am/8am	3am/5am	3am/8am	5am/8am
2	0.602/0.301	0.243/0.122	0.009/0.005	0.527/0.264	0.033/0.016	0.104/0.052
3	0.477/0.238	0.1 /0.05	0.003/0.001	0.357/0.179	0.02 /0.01	0.13/0.065
5	0.639/0.319	0.134/0.067	0.002/0.001	0.295/0.147	0.005/0.002	0.06/ 0.03
7	0.676/0.338	0.106/0.053	0.001/0	0.216/0.108	0.002/0.001	0.048/0.024
9	0.765/0.382	0.137/0.069	0.001/0.001	0.22 /0.11	0.002/0.001	0.045/0.023

Table 7.2: GB 2 colony in 2008. Wilcox rank-sum test with 10^3 bootstraps and `lf_delay=3` minutes. Table cells are formatted as p_1/p_2 where p_1 and p_2 are the p-values for the hypotheses \mathcal{H}_1 and \mathcal{H}_2 respectively (see main text).

table reveals that qualitative changes in L/F time differences occur when `occupation_deadline=8am`, but not for the other pair-wise comparisons. The one-sided test indicates the type of these changes, namely that L/F events inferred up to 8am on the day of occupation, tend to have larger time differences compared to earlier occupation deadlines. This is in line with the reasoning in Appendix 13.1 and implies the presence of swarming effects. Therefore, `occupation_deadline=8am` is likely too late.

Moreover, this conclusion holds when varying `turnaround_time`, as well. The impact of this parameter on the L/F time differences seems to be small, in the range considered, except for values smaller than 5 minutes and comparing `occupation_deadline = 5am` vs. `occupation_deadline = 8am`. In these cases, too many events with small time differences are identified, which conceals the swarming events. The effect of `turnaround_time` is primarily on the number of identified L/F events, as assuming larger recruitment delays excludes events where the leader found a follower relatively quickly (Table 7.4).

Based on these arguments, for a fixed `lf_delay=3` minutes, we would choose `turnaround_time=3` minutes and `occupation_deadline=5am` on the day of occupation. This gives us an optimal trade-off between the number of inferred L/F events, and the interference due to swarming.

In Table 7.3 we apply the same comparison procedure, but this time we fix `lf_delay=5` minutes. Again, `occupation_deadline=8am` produced consistently larger time differences that are not present when comparing all other `occupation_deadline` pairs. Additionally, the effect of `turnaround_time` is again small. Considering that higher `lf_delay` further increases our sample of identified L/F events (Table 7.4), we fix `lf_delay=5` minutes.

7.3.2 Leading-following networks

Social network theory. The network concept is an abstraction that has its fundamental roots in the mathematical field of graph theory. A graph is a construct used to model

turnaround_time	occupation_deadline pairs					
	2am/3am	2am/5am	2am/8am	3am/5am	3am/8am	5am/8am
2	0.725/0.362	0.522/0.261	0.005/0.003	0.782/0.391	0.011/0.006	0.012/0.006
3	0.619/0.31	0.349/0.175	0.006/0.003	0.671/0.335	0.019/0.009	0.03/0.015
5	0.457/0.229	0.135/0.068	0/0	0.47/0.235	0.004/0.002	0.018/0.009
7	0.457/0.228	0.094/ 0.047	0/0	0.36/0.18	0.002/0.001	0.015/0.008
9	0.514/0.257	0.085/ 0.043	0/0	0.29/0.145	0.001/0	0.012/0.006

Table 7.3: GB 2 colony in 2008 with lf_delay=5 minutes.

occupation_deadline	lf_delay = 3					lf_delay = 5				
	turnaround_time					turnaround_time				
	2	3	5	7	9	2	3	5	7	9
2am	173	165	158	155	154	211	201	185	181	181
3am	202	194	184	181	178	245	235	221	217	206
5am	274	269	249	248	234	329	321	298	297	290
8am	354	349	326	325	321	456	440	411	410	405

Table 7.4: Number of identified L/F events for the GB2 colony in 2008 with different values of the three parameters.

pairwise relations between objects, and as such offers a powerful abstraction to real-world social structures. Early works on graph theory date as far back as the eighteenth century and had a predominantly mathematical focus [20]. It was not until the first half of the twentieth century that major interdisciplinary progress in psychology [181], anthropology [166, 201] and sociology [192] advanced the idea of human social interactions constituting networks that can be modeled and understood in the mathematical framework of graph theory. What became known as *social network theory* [262] quickly gained footing and spurred prolific research in the study of social organisation. The network paradigm revealed some of the most well-known regularities in modern human societies, such as the “small world effect” [177, 232, 254] and the “strength of weak ties” [92].

Even more importantly, social network theory has transcended the human domain and has become widely accepted as an important conceptual framework for studying social interactions in animal groups [57, 143, 195, 265]. As social structures in vertebrate animal systems are founded on behavioural interactions among individuals [266], the same methods can be applied for studying social organisation in these systems as well. Recent research has revealed novel structural insights in different taxa, such as identifying mechanisms for community formation and evolution in dolphins [162, 163], predicting male reproductive success in the long-tailed manakin based on their network connectivity [171], detecting the “small world” phenomenon in guppies [58], and revealing the association choices in zebra and onager societies [245]. In bats, social network theory has unveiled the presence of long-term social relationships despite the fission-fusion dynamics, thereby imparting novel insights on the relation between cognitive abilities and social complexity

[131].

The attractiveness of applying social network theory in studying animal systems is in the level of abstraction (individuals become nodes, and their interactions become links) it provides, that allows us to quantitatively analyse social organisation at *all* levels (individual, group, community, population, etc.) across a wide range of interactions (recruitment, friendship, conflict, communication, etc.) [144]. In this way, we can understand complex behavioural patterns as the emerging result of individual-level interaction mechanisms that, in the ideal case, make biological sense. Here, we apply this network paradigm to analyse the patterns of leading-following behaviour, in the hope of gaining insights on the individual importance in spreading information and the mechanisms through which leading-following occurs.

A network is a collection of *nodes* and *connections* (also called *links* or *edges*) among these nodes. In our case, an L/F network is a network in which nodes represent individual bats and connections among nodes represent L/F events. Following the procedure above, we identified all L/F events in our data, and constructed directed, aggregated leading-following *networks* for each dataset. More specifically, a directed link from node *A* to node *B*, denoted as $A \Rightarrow B$, means that individual *A* followed individual *B*. Figure 7.3 shows the L/F network for the GB2 colony in 2008.

In Table 7.5, we show basic characteristics of the L/F networks in all datasets.

Colony	Year	#nodes	#L/F events	network density	#WCC	#SCC	size of largest SCC
	2007	31	60	0.07	4	23	9
	2008	34	262	0.23	1	2	33
GB2	2009	21	33	0.08	2	19	3
	2010	44	142	0.08	1	22	14
	2011	16	86	0.35	1	2	15
	2007	16	169	0.7	1	1	16
BS	2009	17	201	0.74	1	1	17
	2010	19	148	0.43	1	3	17
	2011	7	26	0.62	1	1	7

Table 7.5: Basic structural properties of the leading-following networks from the GB2 and BS colonies. Shown are number of nodes (bats), number of identified L/F events (edges), network density, number of weakly connected components, number of strongly connected components, and the size of the largest strongly connected component. Rows in cyan are the dataset we consider for further analysis (see main text).

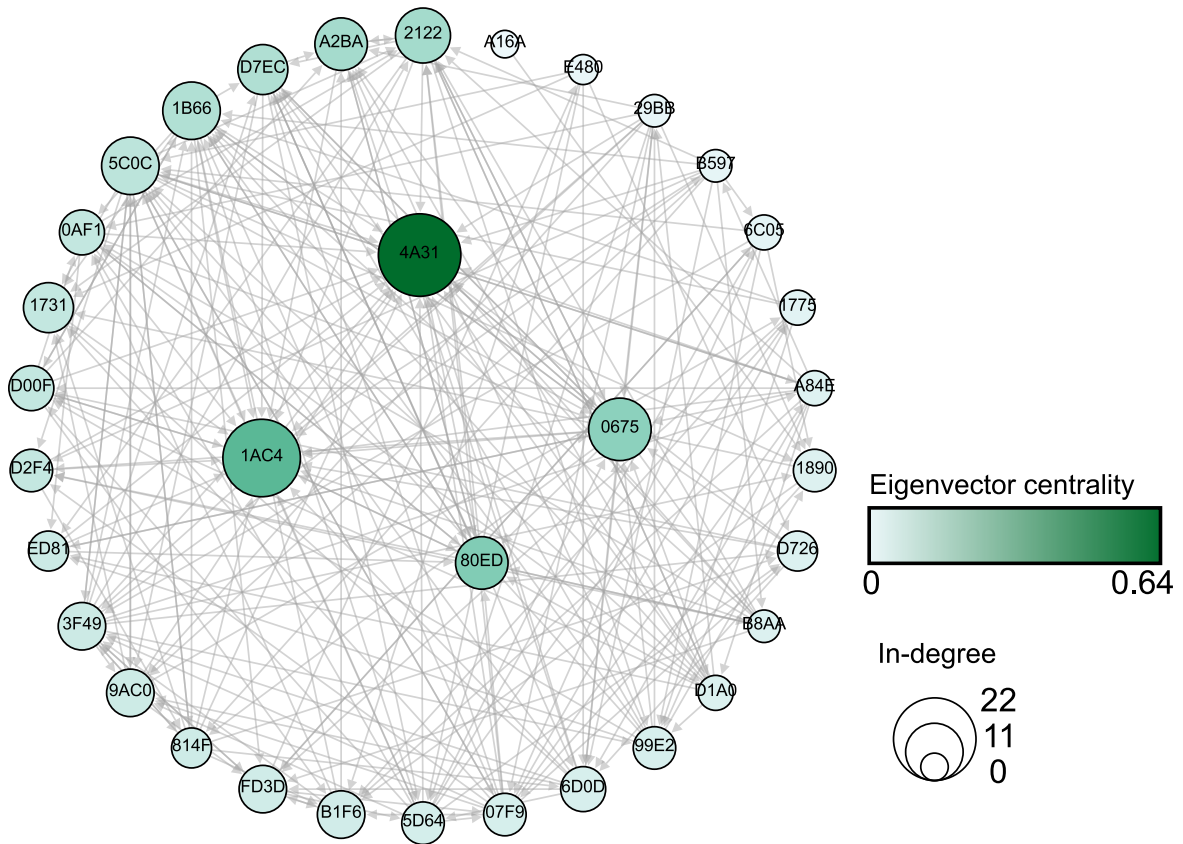


Figure 7.3: Aggregated leading-following network for the GB2 colony in 2008. Nodes represent individual bats, and directed edges represent following behaviour. Node colors indicate eigenvector centrality (see main text), whereas node sizes indicate in-degree centrality. The four (arbitrary) individuals with highest eigenvector centrality are shown in the middle. Note that for the sake of illustration, edges show only *unique* L/F events, i.e. leading-following between the same leader and follower, but to different roosts, are omitted. Total number of unique L/F events is 262, while the total number of L/F events, including multiple leading-following between the same individuals, is 321 (Table 7.3).

Defining importance. In social network theory, the importance, or influence, of a node within a network is referred to as *centrality*. There are various centrality measures in use, and each makes certain implicit assumptions about the manner in which information flows through the network. Choosing a centrality measure is, thus, context-dependent and determined by the underlying process that takes place in the network [28]. An improperly selected centrality metric, can lead to either losing the ability to interpret the measure correctly, or to deducing wrong answers (akin to applying linear regression in non-linear relations).

In our case, an appropriate centrality measure would have to reflect what we mean by in-

dividual influence in spreading information about novel roosts. If influence is best proxied by the total amount of roosts that a given bat made known to the colony, then a suitable centrality measure is *in-degree* centrality. This quantity represents the total number of first-degree followers an individual had during the whole study period, i.e. it is the number of L/F events in which an individual participated as a leader, including those events in which she led the same follower several times to different roosts.

In-degree centrality measures the total leading “volume”, that is *direct* influence, without considering how the information, distributed by a leader to her followers, propagates further through the colony. If we wish to include such *indirect* effects to our notion of importance, one popular centrality measure is the *eigenvector* centrality. Eigenvector centrality is defined as the principal eigenvector of the adjacency matrix defining the network. The defining equation of an eigenvector is

$$\lambda \vec{c} = A \vec{c}$$

where A is the adjacency matrix of the network, λ is a constant (the eigenvalue) and \vec{c} is the eigenvector (see Appendix 13.2 for a more detailed example). The elements of the eigenvector are the centralities of each node. According to this equation, a node has high eigenvector centrality if it is adjacent to nodes that themselves have high centralities. In other words, a bat leading a few bats, who themselves lead a lot, can be more important than a bat leading many who never lead. The observant reader may already spot the analogy with the benefit calculation in our resilience model from Chapter 6. In a similar way, an influential individual, with high eigenvector score, is followed by others, who are themselves influential.

We can see in-degree and eigenvector centrality as the two ends of a continuum, with one end measuring exclusively direct influence, and the other measuring all possible indirect ways, in which information can flow from one individual to all the rest. A possible disadvantage to eigenvector centrality, however, is that its recursive definition may not apply entirely, as far as information transfer is concerned. In Figure 7.4, we have plotted the relative frequency, aggregated over all datasets, of observing L/F events *chain* of a certain length, through which information about a *fixed* roost is spread. For example, two L/F events, $A \Rightarrow B$ and $C \Rightarrow A$, constitute a chain of length two (in addition to forming two separate chains of length one), provided both were to the same roost. In other words B spread the information to A , and A , in turn, disseminated it further to C . Therefore, B ought to obtain direct importance from having led A , but also indirect contribution, for were it not to her, A would not have recruited C ⁶.

⁶This is not entirely true. It is possible, though unknowable, that A would have found the roost by

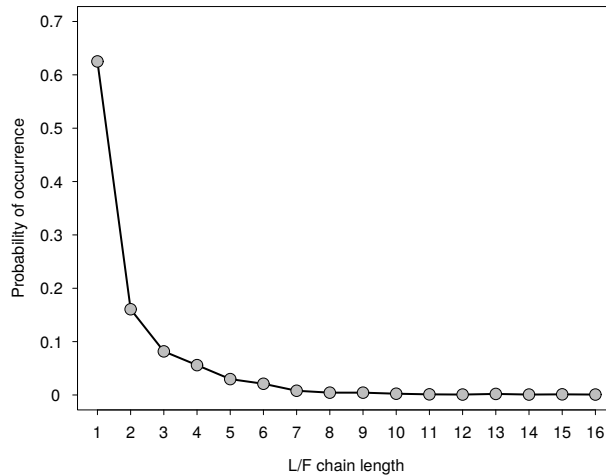


Figure 7.4: Probability distribution of the lengths of L/F event chains, calculated over all nine datasets.

The figure shows two observations. First, chains longer than 16 did not occur in any of the datasets we have. More importantly, event chains of length up to two constitute about 80% of all lengths observed, and the probability of longer chains decreases drastically. Since we construct aggregated L/F networks (i.e. a link represents leading-following, disregarding the target box), the longer L/F chains we observe, as in Figure 7.3, likely do not represent information spreading for the same roost. Eigenvector centrality, however, considers all chains equally, and an individual at the beginning of an L/F chain would be attributed more importance, the longer the chain is. This does not reflect her information reach, as it is quite likely that beyond length two, the target roost of the L/F events further down the chain, changes. It may, however, reflect her influence, in terms of some individual quality (i.e., dominance), which causes individuals with many followers to follow her.

To address this issue with limited information spread, we define a new metric – *second-degree centrality* – which calculates centrality as the sum of the total number of followers of the focal individual, and the sum of the in-degrees of these followers weighted by a factor α (in that sense the followers of one’s followers are her second-degree followers). This reflects our observation that chains of length up to two constitute the majority in all datasets. Second-degree centrality is a simplified version of Katz centrality [128], which accounts for all higher degree followers, with decreasing weight factors. In our analysis, we will use all three centrality measures. Figure 7.5 illustrates how conclusions about individual importance are affected by the chosen measure.

her own exploration, or that she “forgot” the information obtained from B , and re-visited the box before leading C . This becomes more likely with the length of the event chains we consider.

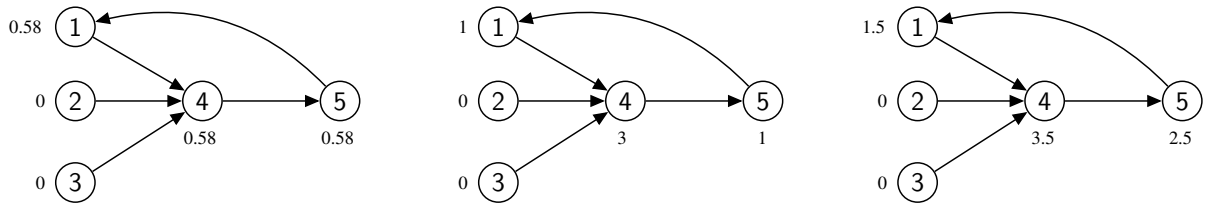


Figure 7.5: Differences between the three candidate centrality measures. The centralities for each measure are indicated next to each node. **(Left)** *Eigenvector centrality*. Since individuals 2 and 3 have no followers, they are attributed zero influence, and thus contribute nothing to the influence of their leader, individual 4. In turn, 1, 4, and 5, each have one follower of non-zero importance, hence they have the same eigenvector scores. **(Center)** *Indegree centrality*. Here, only direct influence is measured. Individual 4 is most influential, as she spread information to three different individuals. Individuals 1, and 5 with one follower each, have still equal importance. **(Right)** *Second-degree centrality* with $\alpha = 0.5$. Individual 4 has a higher centrality than her in-degree score, as we account for the indirect contribution of individual 1 ($3 + 0.5 \times 1 = 3.5$). However, 5 is now more important than 1, because 4 contributes to 5 indirectly ($1 + 0.5 \times 3 = 2.5$).

Calculating importance To calculate the importance of each individual, we use only the cyan-coloured datasets in Table 7.5, as they provide the most reliable sample sizes of detected L/F events for a statistical analysis. In particular, the L/F networks constructed from these datasets, are weakly connected, which means that all individuals participated in L/F events. Moreover, judging from the size of the largest strongly connected component, the vast majority of individuals participated as both leaders and followers. The high density of these networks, further indicates their high connectivity. In contrast, the network for the GB2 colony in 2009, consisted of 21 individuals and contained 19 strongly connected components of which the largest one had a size of three. This implies that the majority of the individuals were either exclusive followers or were isolated nodes (i.e. did not partake in leading-following activity).

Next, an important question to address is whether the individual importance we calculate, regardless of the centrality measure, is significant in the statistical sense. In other words, can similar centrality scores arise in networks where leading-following behaviour is governed by some random process? This approach is not new. Instead of building biologically- or stochastically-based micro-foundations for explaining observed L/F patterns, we rather ask whether observed network properties can be reproduced by simple network-generation processes that only match some empirical constraints, but are otherwise fully random. If that is the case, then we may conclude that such regularities (e.g. centrality sequences) are not that “interesting”, as no alternative, more structural, model would pass any test discriminating against the random counterpart. Conversely, if observed regularities can-

not be reproduced by the null random model, we argue that more complex, structural mechanisms are responsible for what we observe. Null random-network models, therefore, serve as a sieve that can help us reveal interesting patterns in real networks.

As there is no ready-made random process for our case, we investigate the following five null models of randomly re-wiring the empirical L/F networks. Each model adds various structural properties, extracted from the real networks, to define “randomness” with increasing level of complexity. To test the presence of significantly important individuals, we compare individual centrality scores computed from the empirical L/F networks to those generated by each model.

- M1. **Random rewiring.** Each directed link in the L/F network is re-connected to two randomly chosen nodes. This is equivalent to randomly choosing two bats for an L/F event, while keeping the total number of L/F events the same as in the original network. This is the simplest model that destroys all structure present in the network, and therefore makes no assumptions about the behaviour of bats – each individual is equally likely to become a leader or a follower, independent of any idiosyncratic characteristics.
- M2. **Random rewiring with individual activity.** Here, the re-wiring procedure is almost the same as in the first model, the only difference being that the probability of a bat to be selected as a leader is proportionate to her *activity*. Activity of an individual is the total number of revisits, explorations, discoveries and leading-followings (either as a leader or as a follower) that she participated in. The probability of being chosen as a leader is the normalised activity with respect to the total activity of all individuals in the study period. According to this model, L/F events are still formed randomly, however, more active bats have a higher likelihood to be followed, as their frequent flights render them more “visible” to potential followers.
- M3. **Preserve following behaviour.** For each link in the network, we keep the follower, but randomly choose a new leader. The underlying assumption is that the observed frequency of following represents a unique following “budget”, or tendency (e.g. by individual characteristics, such as age, weight, etc.), that each bat is endowed with. This model addresses the question of whether individuals, as important as we observe, can arise if leaders are chosen randomly (i.e. identity of leaders is irrelevant), but following behaviour is restricted empirically.
- M4. **Preserve leading behaviour.** We apply the same idea as in the third model, but instead, fix the number of leading events to the empirically observed. Therefore, in an L/F pair, the identity of the follower is irrelevant, as she is randomly assigned, and

the individual leading volume is constrained. Importantly, the association between a leader and a follower is still random, i.e. bats do not have special preference to lead another individual. Should this model *fail* to reproduce empirical centrality scores, it would imply the opposite – leaders do in fact associate non-randomly in an L/F pair, e.g. due to kin selection or reciprocity among long-term associated partners.

M5. Preserve following behaviour with individual activity. We build on M3. by incorporating activity (see M2) in the probability of an individual to be chosen as a leader.

Results. In the following Figures 7.6, 7.7, 7.8, 7.9, 7.10 and 7.11, we present the results of calculating the centrality scores, according to the three centrality measures, for each highlighted dataset from Table 7.5, and compare them to those generated by the five random models. More specifically, for each random model, we re-wired the corresponding empirical network accordingly, 10^5 times, and recorded the individual centrality scores generated in each re-wired network. In this way, we obtained a distribution of model-generated centralities per individual. Blue bands in the figures denote the 95% innermost range of these distributions, and point symbols indicate the centralities computed from the empirical L/F networks. The in-degree and second-degree centrality scores are normalised with respect to the largest centrality score (empirical or model-generated) in the corresponding model. Eigenvector centralities are normalised by default. Note that, since model M4 assumes fixed leading behaviour (i.e. in-degree), it cannot be evaluated with in-degree centrality, as re-wiring will produce no variation in individual in-degrees.

Furthermore, Table 7.6 shows a comparison between the three centrality measures across the six models, in terms of how well each describes the empirical centrality scores. For each distribution of model-generated centralities⁷, we estimated its Gaussian kernel density, and evaluated the density at the empirical centrality score for the corresponding individual. We used this “density score” as an indication of the extent, to which the importance of that individual is reproduced by the given model and centrality measure. The sum of all density scores is then the total density score for this model and centrality measure. Therefore, we account not only for the presence of empirical centralities in the 95% ranges of the centrality distributions, but also for their likelihood within these distributions.

⁷Recall that every individual has such distribution for each model and centrality measure.

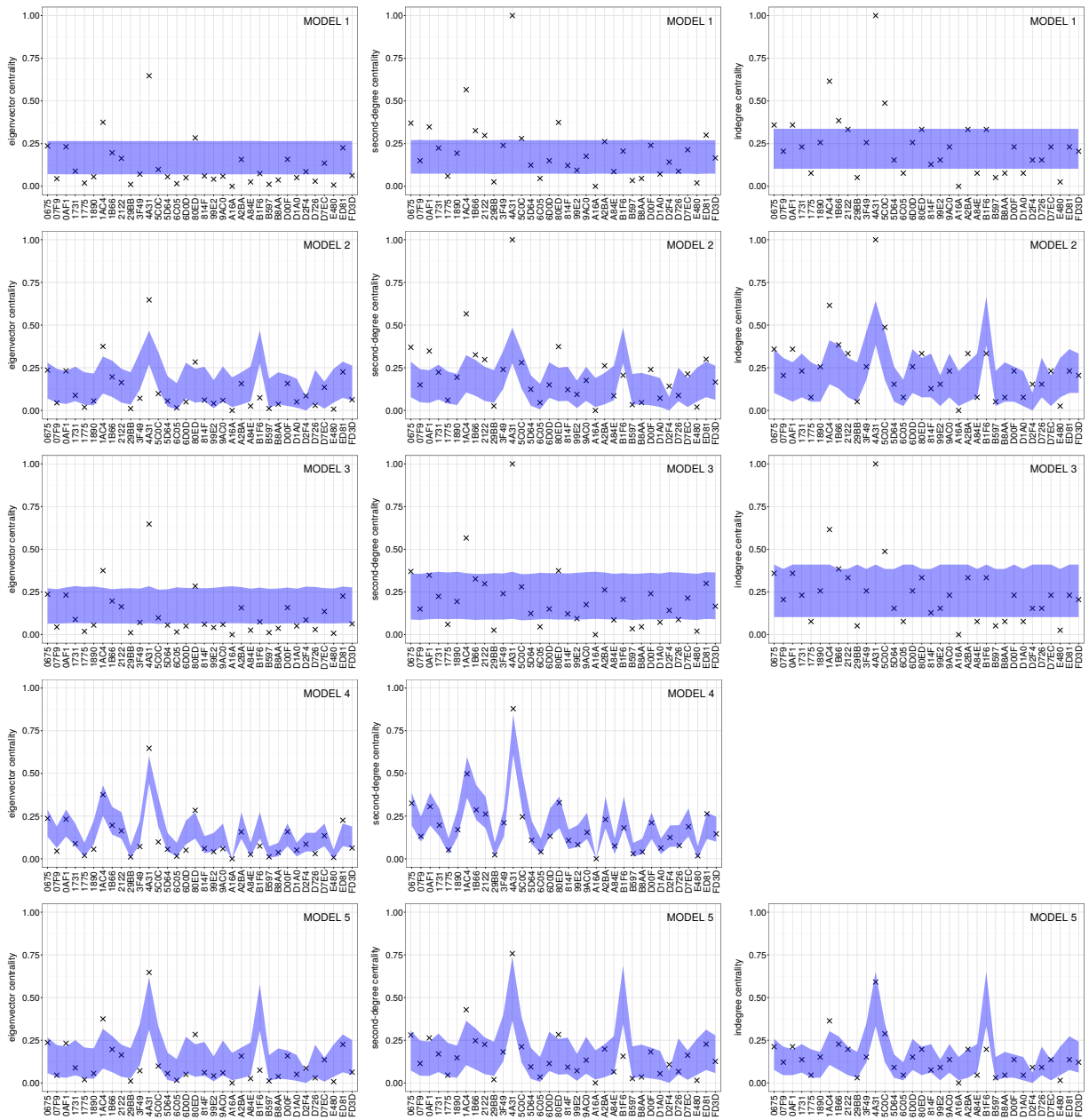


Figure 7.6: Model-generated vs. empirical centrality scores for colony GB2 in year 2008.

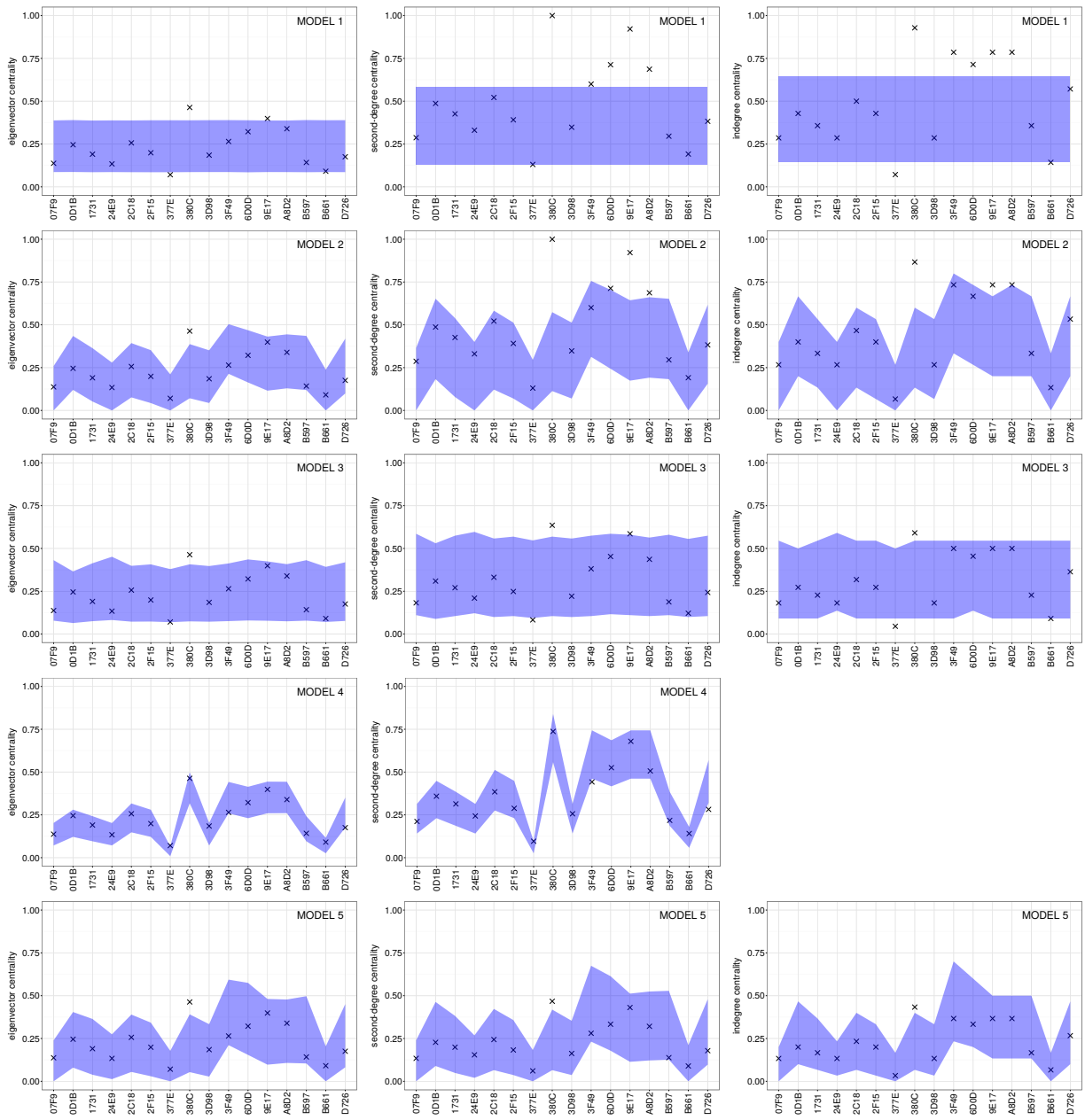


Figure 7.7: Model-generated vs. empirical centrality scores for colony GB2 in year 2011.

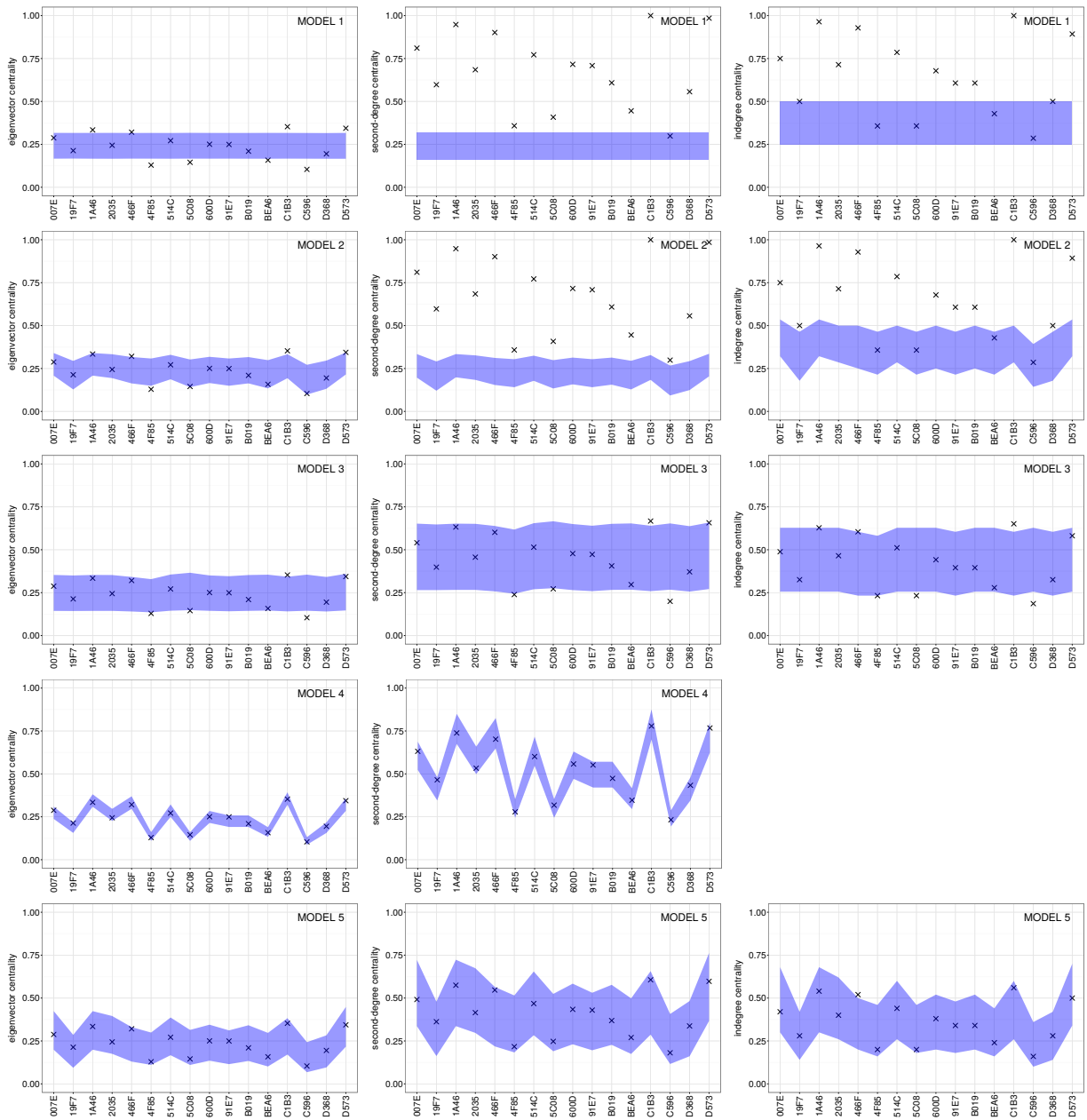


Figure 7.8: Model-generated vs. empirical centrality scores for colony BS in year 2007.

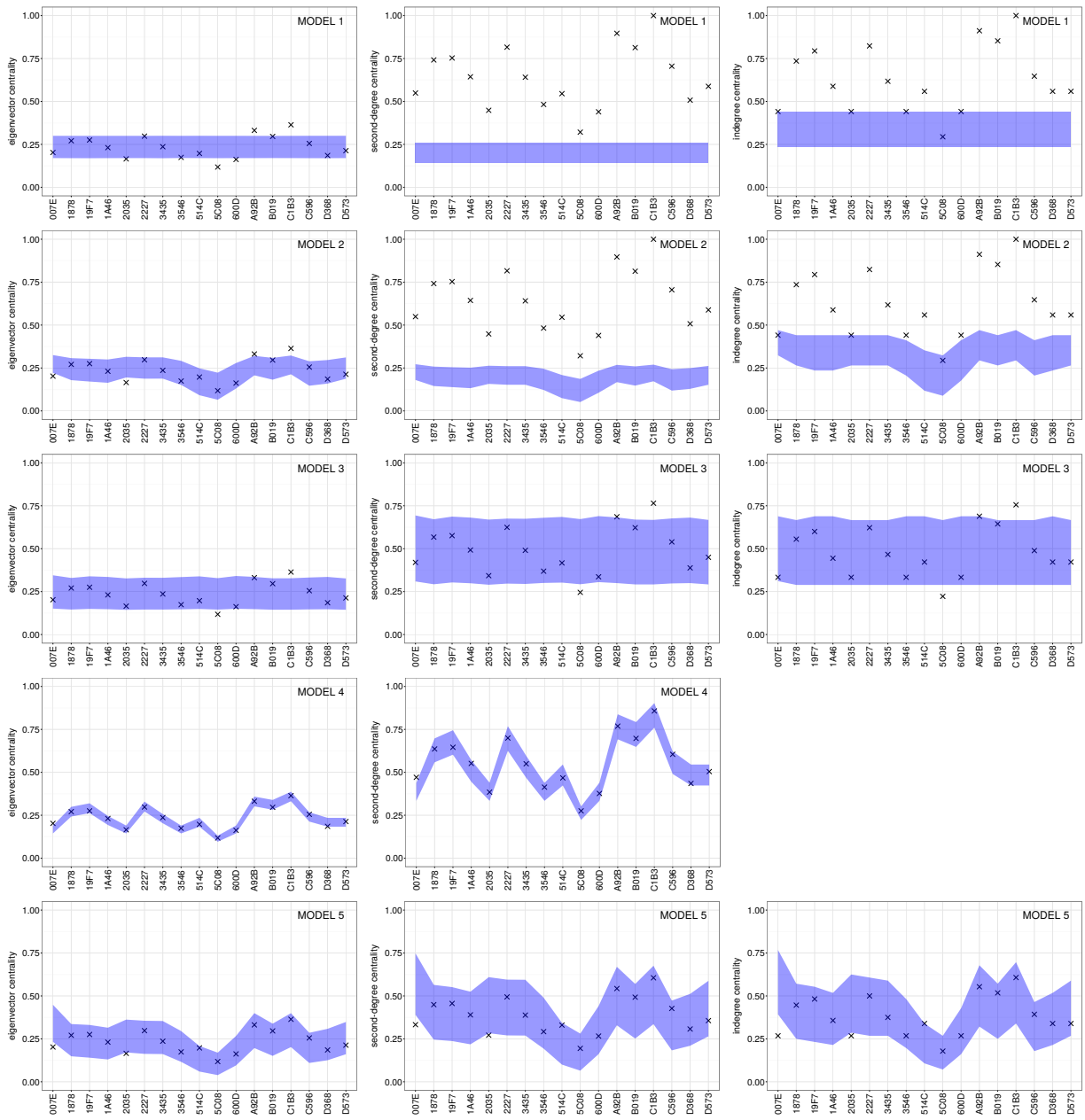


Figure 7.9: Model-generated vs. empirical centrality scores for colony BS in year 2009.

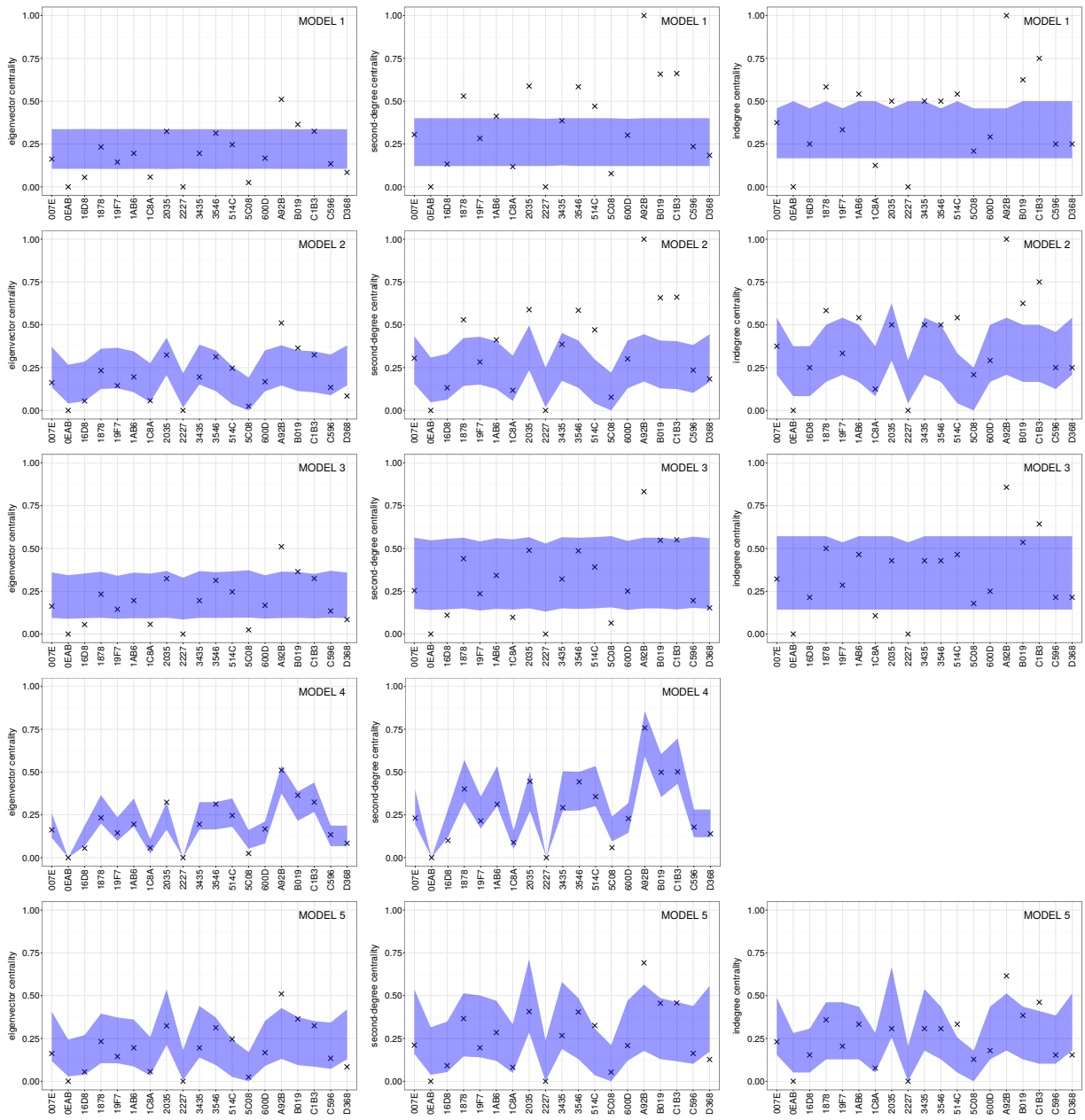


Figure 7.10: Model-generated vs. empirical centrality scores for colony BS in year 2010.

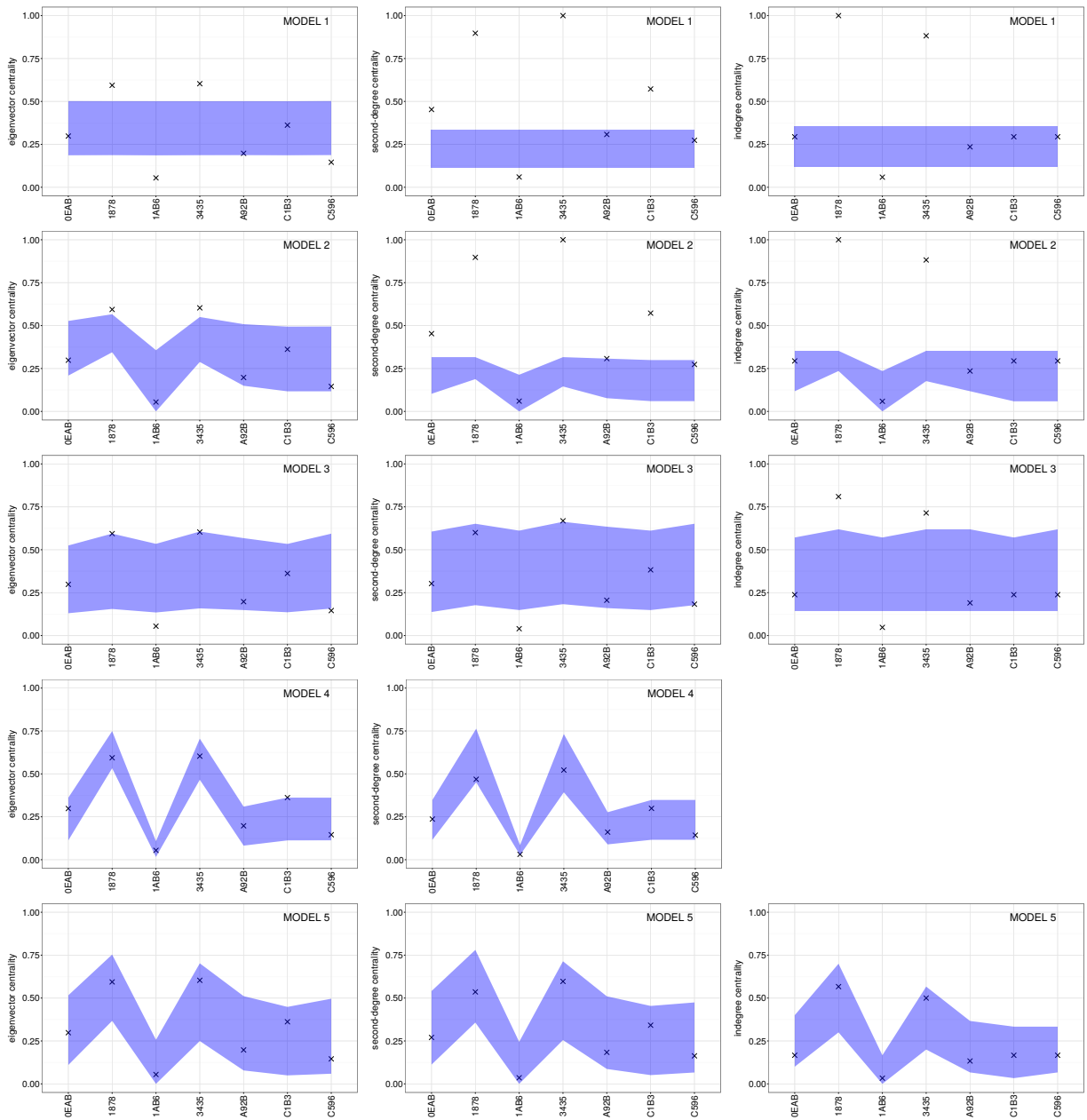


Figure 7.11: Model-generated vs. empirical centrality scores for colony GB2 in year 2011.

dataset	measure	Model				
		M1	M2	M3	M4	M5
GB2 2008	eigenvector	65.82	87.04	67.59	262.62	103.52
	in-degree	80.62	102.59	88.13	NA	161.27
	second-degree	80.78	106.79	89.42	333.37	122.68
GB2 2011	eigenvector	44.76	61.84	45.26	110.02	57.74
	in-degree	28.26	58.49	32.46	NA	98.16
	second-degree	26.92	29.54	32.38	83.3	58.16
BS 2007	eigenvector	62.52	87.28	66.15	293.37	88.25
	in-degree	26.33	27.81	32.6	NA	59.05
	second-degree	4.42	0.28	33.35	131.8	52
BS 2009	eigenvector	81	93.53	78.1	385.78	92.5
	in-degree	14.96	20.69	35.63	NA	54.55
	second-degree	0	0	37.95	152.78	55.7
BS 2010	eigenvector	44.72	47.88	45.7	107.14	52.81
	in-degree	27.66	36.64	31.6	NA	50.25
	second-degree	26.97	32.34	28.68	85.44	48.88
BS 2011	eigenvector	9.07	11.4	10.34	39.13	21.2
	in-degree	21.4	22.77	8.36	NA	30
	second-degree	8.03	9.75	9.27	38.63	20.56

Table 7.6: Comparison between the five random models and three centrality measures, for each analysed dataset. Table cells indicate the total density score (see main text) of each model for the given centrality measure.

Is there evidence for individual preferences in L/F events? We use model M4 to isolate individual preferences, e.g. kin selection, in forming L/F pairs. Due to its stringent assumptions, this null model is not meant as a test for the presence of significantly important individuals in spreading information. Indeed, by quantifying importance based on leading behaviour, model M4 explicitly assumes, to a large extent, the very quantity we are trying to reproduce.

Instead, we focus on the repercussions, should this model be inadequate in explaining an individual’s centrality. M4 assumes that the identity of the follower is inconsequential, and that the association between a leader and a follower in an L/F pair is random. Therefore, if we find individuals whose empirical centrality scores are significantly different from the model-predicted score, it must be because they have associated non-randomly in some L/F events. Figure 7.12 illustrates the idea.

On the left, we show a toy-example of a six-individual L/F network, in which individuals 1,2 and 3 show a strong preference to associate with one another in reciprocating recruit-

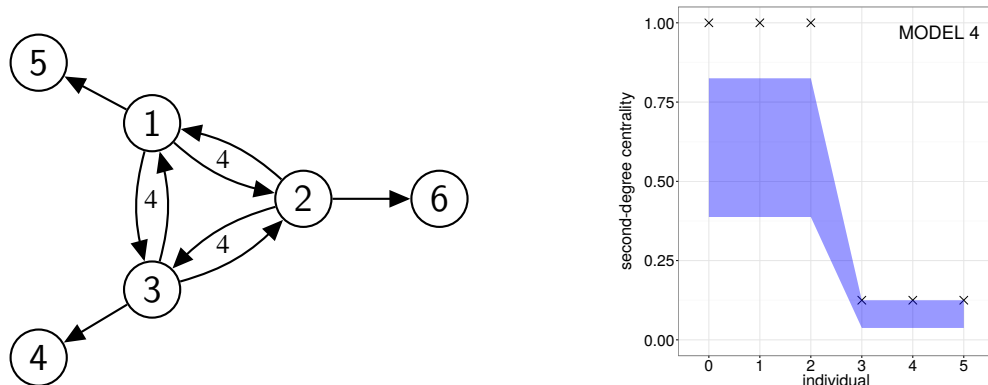


Figure 7.12: An example network which displays strong leading-following associations between a core of three individuals. Each member of the core has followed and led the other two 4 times. The plot to the right displays empirical and (normalised) model-generated second-degree centrality scores.

ment patterns. To the right, we see that their expected second-degree centrality scores based on model M4 are never as high as observed. By assuming random association in L/F pairs, we have effectively destroyed the underlying association structure. However, we note that failure to detect an individual with a significantly different centrality than expected, does not indicate evidence for the absence of individual preference. It only implies that such preference is not needed in explaining the formation of L/F pairs, after controlling for heterogeneity in leading tendencies.

In all datasets, we find the centrality scores, generated by model M4, closely match those extracted from the data, regardless of the centrality measure. In fact, this model has the highest overall density score. We, thus, conclude that there is no evidence that bats recruit others preferentially, as long as observed leading propensity is respected. This general result is in line with empirical works in Bechstein bats, which have found no preference based on kin or reciprocity for leading-following behaviour [132].

To what extent can activity explain leading behaviour? One conclusion we can draw from the presented results, which notably holds across all studied datasets and centrality measures, is that the simplest model, M1, is the least successful in reproducing the empirical centrality scores. Qualitatively, this is evidenced by the number of individuals falling outside the 95% range in all figures. Quantitatively, the model has the worst density score in four of the six datasets (GB 2008, GB2 2011, BS 2007, and BS 2010) and second-worst in the remaining two. This is a clear indication that L/F events are not formed entirely by chance, and that individuals are not equally likely to be leaders or followers.

Is individual activity, alone, the missing ingredient? Judging from Figures 7.6, 7.7 and

Table 7.6, model M2 seems to offer an improvement, over model M1, for the GB2 colony. In the BS colony, M2 is only marginally better than M1 in terms of density score. However, in all datasets, there is still a considerable number of individuals whose centrality cannot be reproduced by an activity-only model.

Model M2 also shows a striking contrast between the centrality scores from the eigenvector, and the other two centrality measures, for the BS colony in 2007, 2009, and 2011 (Figures 7.8, 7.9, and 7.11). The figures illustrate that based on activity alone, we should expect lower, and more evenly distributed, in-degree and second-degree scores than observed in reality. The contrast is due to the indirect influence that is captured only by the eigenvector calculation, and given disproportionate weight in dense networks. Compared to BS in 2010 and to the other colony, the three BS datasets are considerably smaller (approximately half the size of the GB2 colony in all observation years) and denser, and exhibit higher average activity which is distributed almost uniformly among the colony members. On one hand, since the normalised activity represents the probability of being selected as a leader, model M2 predicts that leading behaviour, and thus in-degree and second-degree centrality scores, will be similarly distributed. However, in the real networks, observed leading tendencies are heterogeneous, as evidenced by the individual in-degrees (Figure 7.13). This means that other factors that contribute to one's activity, such as number of independent explorations, conceal this heterogeneity to the effect that the model-generated in-degree and second-degree centrality scores are negatively biased.

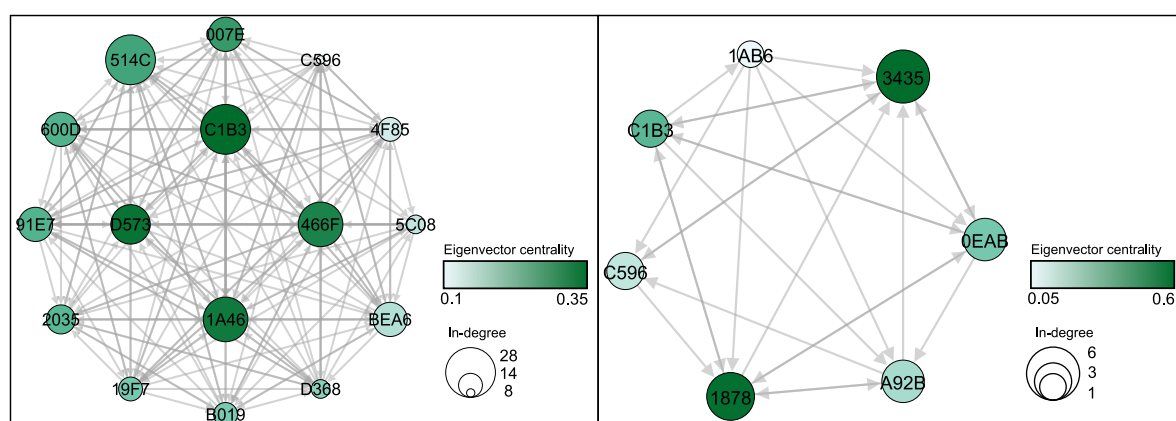


Figure 7.13: Aggregated leading-following network for the BS colony in 2007 (left) and 2011 (right).

On the other hand, due to the small size and high density of these three networks, indirect influence will be accentuated and more prevalent. The size of the largest connected component in Table 7.5 and Figure 7.13 illustrate that there are numerous closed L/F event chains that involve the majority of the individuals, and as shown in Chapter 6, individuals

in these chains will mutually re-enforce their eigenvector centralities. The end result is that eigenvector scores will be similar across the whole network.

In model M3, we ignore activity and assume empirically constrained following behaviour coupled with random leader selection. Quantitatively, the model is not much better than the previous two; in fact it is next to worst for all considered datasets. Even though, the confidence bands in some datasets (GB2 2011, BS 2007, BG 2009, and BS 2011) indicate that the centrality scores of most individuals are reproduced, the distributions of model-generated centralities display high variance, which betrays poor predicting power and translates to the low density score. Models M2 and M3 shows us that taken separately, activity and fixed following budget do not reproduce well individual importance. The conclusions from these two models are also sensitive to the centrality metric used, and the underlying network topology.

The next logical step is combining both factors, which is done in the last model M5. Remarkably, M5 is able to reproduce with high-fidelity individual importance in all datasets across the three centrality measures, and attains the highest overall density score⁸. In three of the six datasets (Figures 7.8, 7.9 and 7.11), model-generated centralities are not significantly different than empirical ones for all individuals. This holds also in the remaining three datasets, with a few notable exceptions. These are four individuals – **1AC4** and **B1F6** (GB2 2008), **380C** (GB2 2011) and **A92B** (BS 2010) – whose centrality scores are considerably outside the 95% confidence range.

Based on the performance of model M5, we conclude that its underlying mechanism is able to reproduce, to a large extent, the observed L/F patterns. In particular, we propose that individuals possess a natural propensity to acquire information socially (i.e. to be followers), and randomly choose to follow those who appear most often available. An individual's leading behaviour and resulting influence in spreading information about suitable roosts is, therefore, determined by her flight activity while exploring her home range.

We note that the value of this explanation is two-fold. First, it suggests the most parsimonious recruitment mechanism, with sufficient explanatory power, that holds across diverse datasets and centrality measures. Second, it serves to focus research attention by isolating those cases, in which a more complex mechanism may be at play. This is particularly useful in biological systems, in which high individual diversity makes it harder to identify common determinants of observed behaviour.

In our case, three of the four notable individuals mentioned above are significantly more influential than expected from M5. The mismatch means that they led more than ex-

⁸Excluding M4, which, as we explained, is not to be used for explaining importance

pected from their flight activity alone. Since activity also includes behaviour unrelated to leading-following and is normalised to the activity of others, it could be that these individuals are “specialised” in recruitment without engaging in other forms of activity. Therefore, relative to others, their overall activity may be lower, which decreases their probability of leading. As a result, the model-generated centrality scores would necessarily be lower than observed. Another reason for the mismatch may have to do with an individual’s recruitment “efficiency”. That is the number of followers that a leader recruits, on average. A consistent tendency to recruit several followers per L/F event, can result in centrality scores (especially in-degree) larger than expected from a random mechanism, in which such tendency is absent. A particularly interesting exception is individual **B1F6** in the GB2 2008 dataset. Her recruitment activity, as a leader in L/F events, is markedly lower than her overall flight activity. In fact, most of her recording in the dataset came from discovery, exploration, and revisit events. The model, then, would predict a higher expected centrality than observed in the real network. This hints to a potential preference for personal information gathering over participation in information spreading.

7.4 Conclusion

In this chapter we complemented our investigations on unintended outcomes by studying an animal system in which mitigation of negative collective consequences is critical for its survival; in this sense we argue that the negative can be seen as unintended, even though the notion of conscious intent cannot be fully attributed to an animal, in the same way it can be for a human. We introduced the fission-fusion society of the Bechstein’s bat and the coordination problem of communal roosting that a bat colony faces. What is interesting in such an animal system is that the colony has naturally evolved mechanisms to avoid colony splitting during communal roosting. Therefore, by applying the complex system perspective, our goal was to infer these individual rules underlying the collective coordination.

From collaboration with field biologists we obtained novel datasets from two colonies, recording bat movements in and out of experimental roosts. These recordings could, then, be used to infer specific events, corresponding to personal or social information gathering. The latter is defined in terms of leading-following (L/F) events which are characterised by a leading individual who recruits a follower and leads her to a particular roost. This represents *information transfer*, which was already observed and proposed as a way to facilitate communal roosting in bats [132, 267], however, the inter-individual rules governing recruitment are largely unknown.

Here, we first proposed a rigorous method to analyse the recording patterns from the data, and infer L/F events reliably. We argued that the time differences of L/F events can be used to calibrate the three parameters underlying our method, which adds a valuable systematic treatment to the analysis of such recording data by means of algorithms. In the main text and also in Appendix 13.1, we showed that comparing (statistically) the distributions of L/F time differences for different parameter values can filter out the effects of swarming behaviour, which is the main source of “noise” in inferring genuine L/F events.

Continuing with the analysis, our main result consisted of using social network theory to construct aggregated leading-following networks from all datasets. The network perspective allowed us to focus on the individual role in the recruitment process, and in particular on individual importance (quantified by three network centrality measures) in information transfer. In particular, we created five different network generation processes and compared the individual centralities produced by each process to the centralities observed in the empirical networks. In doing so, we could statistically test which generation mechanism best reproduces the empirical centrality scores for the largest number of individuals. We found that the simplest mechanism that achieves this posits random associations in leading-following pairs, together with natural tendency for social information gathering regarding potential roosts (i.e. following behaviour). Moreover, the observed leading behaviour could be explained, to a large extent, by individual flight activity. In other words, more active individuals tend to get followed more by others, simply because of the higher likelihood of being observed by potential followers. We note that the random association in our mechanism corroborates existing findings on the lack of kin-selection and reciprocity in the recruitment process [132], but generalises these findings to any kind of association between a leader and a follower. Finally, the network approach allowed us to identify four individuals whose importance for information transfer cannot be explained by flight activity alone. Rather than being conclusive evidence against our proposed mechanism as a whole, we see this as an indication for a more complex individual behaviour (e.g. specialisation in recruitment, higher recruitment efficiency or explicit preference for personal information gathering) that, otherwise, would have been difficult to single out.

We believe that our results can guide future empirical studies in two ways. First, our mechanistic explanation for recruitment needs to be confirmed by field work. We need experiments that explicitly test whether recruitment is passive (as we postulated, since followers simply follow randomly the individuals who are around) or active in the sense that leaders intentionally attract the attention of their potential followers by e.g. acoustic signals or specific aerial displays. Second, isolating the “missing” behavioural complexity to *specific* individuals can ease the design of experiments targeted to these individuals. For example, relating demographic, health or genetic characteristics to displayed inconsisten-

cies with flight activity may reveal the fundamental reasons underlying their behavioural variability.

Chapter 8

Conclusions

8.1 Summary

This thesis set out with the ambitious task to provide an alternative perspective on understanding unexpected consequences of collective decisions. In doing so, we adopted a complex systems perspective and presented a broad array of decision-making scenarios, from human to animal societies, all of which can be approached from this view. Our main proposition advocated the recognition of the inherent complexity in systems composed of a large number of interacting individuals, whether they are atoms, ants in a colony, or neurons firing in a human brain. These, so called complex systems, exhibit positive *and* negative feedback processes, which cause a highly non-linear response to interventions that cannot be predicted a-priori by classical reductionist methods.

In Part I, we focused on contemporary decision-making scenarios in humans. This part began with a theoretical investigation on promoting cooperation in social dilemmas. By a standard game-theoretical approach, we illustrated that depriving a fraction of individuals from access to payoff-related information, thus making them uninformed, can induce herding behaviour that, under the right conditions, leads to large-scale cooperation. Moreover, the right conditions can be established in a relatively cost-efficient way. We showed that providing cooperative incentives only when an individual's local neighbourhood contains an equal number of cooperators and defectors, is enough to tilt the scale in favour of cooperation. Less information can, thus, boost cooperation.

In Chapter 3, we followed up this theoretical insight with a more general (empirical and theoretical) investigation of herding and imitation in the context of the wisdom of crowds phenomenon. First, we provided experimental evidence that social influence can under-

mine the phenomenon, even when monetary incentives are designed to prevent that. In a decision-making task, subjects had to repeatedly guess the answer to quantitative questions, while being exposed to different types of information about others' guesses. The result was that already trace amounts of social influence, in terms of exposure to an average opinion, can induce herding behaviour so that individuals collectively drift in the wrong direction and lose diversity at the same time. In the same chapter, we modelled individual behaviour in this experiment with the help of a simple agent-based model. The only two model ingredients – strength of perceived social influence and individual conviction – were sufficient to reproduce the observed “social influence” and “range reduction” effects. More importantly, the model suggested that it is the crowd's initial accuracy and diversity that ultimately determines the polarity of the social influence effect.

This assumption was tested in Chapter 4, where our main concern was the design of interaction mechanisms that can restore the wisdom of crowds phenomenon in the absence of individual expertise. In the experimental study we conducted, it was shown that by making participants aware of their relative performance (i.e. ranks) *and* by introducing competition, we can effectively design a crowd that converges remarkably close to the correct solution, while maintaining considerable diversity. Moreover, since no subject was able to consistently outperform the others, we argue that the displayed collective accuracy truly comes as unintended from an individual's perspective. Finally, we concluded the chapter by demonstrating the effect of the crowd's initial accuracy and diversity on its convergence to the correct solution. The speed of convergence can be seen as an indication of the polarity of the social influence effect from Chapter 3. Despite the fact that positive or negative connotation cannot be *directly* attributed to convergence, we argue that, in practice, slowly approaching a solution, even the correct one, can often be perceived as less desirable, hence negative, especially under time constraints. In our experiment, we found that initial diversity separates crowds that start as equally accurate on average – the more diverse crowd quickly becomes more accurate and outperforms the less diverse one. Diversity also helps initially less wise crowds (in terms of collective accuracy) to “catch up” with wiser, but homogeneous groups.

The concluding two chapters of the first part dealt with collective effects of individual behaviour on the resilience of online communities. In Chapter 5, we presented the first empirical analysis of social resilience in OSNs. With the help of a theoretical model of user behaviour, we were able to (i) define social resilience as a compounded result of individual actions and (ii) use the k-core decomposition to measure and compare the resilience of several OSNs using the CCDF of node coreness. In our theoretical framework, the collapse of an online community is due to large-scale, *uncoordinated*, cascades of users leaving. These cascades can be caused by a number of external shocks, e.g. technological

modifications and competitor dynamics, which get amplified by the network structure and propagate through the OSN. We verified the cascading mechanism by first rejecting the epidemic properties of scale-free complex networks as a possible explanation, and second by applying it to a case study of **Friendster**. Our k-core analysis showed that the topologies of two successful OSNs, **Livejournal** and **Facebook**, are less resilient than the unsuccessful **Friendster** and **Orkut**, which indicates that the topology alone is not enough to explain success and failure.

Finally, Chapter 6 proposed a mechanism that counter-intuitively increases resilience in reputation-based OSNs. Using the same assumption of rational user behaviour as in Chapter 5, we showed how the unintended cascades of users leaving can be managed in a way that maximises resilience. In our model, resilience is quantified as the average long-term reputation of all users. We were able to reproduce empirically observed network features in reputation-based OSNs, such as core-periphery structure, and direct and indirect reciprocity, but most importantly, to demonstrate that introducing *cost of usage* increases resilience. We argued that the optimal (non-zero) cost is a mechanism that balances two dynamic forces. First, the cost of usage filters out peripheral users with low reputation who, in the early stages of fast community growth, overwhelm existing filtering mechanisms and distribute low-quality content. This allows a core of well-connected users to emerge, which by means of direct and indirect reciprocity maintain high reputation, and thus overall resilience. Second, the optimal cost minimises the time to establish a new core by providing the right amount of influx of new users.

Moving away from the design of inter-individual interactions, in Part II we investigated decision-making scenarios, in which inference of interaction rules was the best we can do. We studied two colonies of Bechtein bats which face unintended collective effects on a daily basis. The colony relies on communal roosting to survive, yet individuals must make roosting decisions under limited information, hence, regularly, risk colony splitting. Solving this coordination problem is a fascinating evolved ability of this system and, in the spirit of complex systems, we focused on the inter-individual rules that make this possible. We followed up on existing empirical work that showed the importance of information transfer via leading-following relationships for achieving coordination. Analysing unique datasets on bat activity in experimental boxes, we applied for the first time a large-scale network analysis on leading-following behaviour in Bechtein bats. From the constructed leading-following networks, we were able to test various rules underlying the association in leading-following pairs, consequently underlying the coordination process in general. We showed that a parsimonious recruitment mechanism able to reproduce the observed individual importance in spreading information, is that bats possess a natural tendency to follow others (i.e. to gather information socially) and randomly choose whom to follow.

As a result, leaders are selected based on their flight activity as activity determines their “availability” to potential followers. With this, we confirmed previous analyses on the lack of kin-selection and reciprocity in the recruitment process. However, in addition, we demonstrated that solving the complex coordination problem can be achieved by following simple individual rules.

Finally, our modelling methodology allowed us to identify systemically important individuals with respect to information transfer. We found four individuals whose flight activity was not sufficient to reproduce their observed leading behaviour. As we argued, this could indicate the presence of individual specialisation in recruitment or higher recruitment “efficiency” for the first three individuals, and potential preference for personal information gathering over information transfer for the fourth one. One way or another, we believe that these results can guide future field work by designing novel ways to precisely test the simple recruitment mechanism above, and to investigate idiosyncrasies not explained by flight activity alone.

8.2 Scientific contributions and future research

Original scientific contribution. The work presented in this dissertation led to the writing of several manuscripts and peer-reviewed articles in journals and conferences. An article based on the role of information and social herding for promoting cooperation from Chapter 2.1 was published in a complex systems journal [224]. An abridged form of the agent-based model in Chapter 3 was presented at a conference on collective intelligence [170], while the full model is pending submission. The work on ranking and competition from Chapter 4 has been adapted to an article format and is to be submitted shortly after the completion of this thesis. Our model on social resilience in OSNs has been digitally archived [82] and presented to an ACM conference [80]. This work also received considerable media attention (see Appendix 11.1). The theoretical model of resilience in reputation-based OSNs is in an article format and is pending submission upon completion of this thesis. Finally, the work on Bechstein’s bats has been presented in the form of two scientific talks given in the group of our German collaborators. The application of network theory to studying individual-level recruitment mechanisms was met with great enthusiasm and interest. So was the robust methodology in calibrating the inference of meaning from empirical data and the streamlined analysis via the help of computer algorithms.

Relevance for sociology. Sociology has always faced the issue of complexity for studying society implies, by definition, studying complexity [34, 161]. In the nineteenth century,

it was commonly accepted to study complexity from the perspective of *evolutionism*, i.e. the view that societies *intrinsically* develop along a time-line from simpler to more complex forms of existence. Indeed that seemed to be the case in those days when the advent of industrialisation and mass production, the rural/urban shift, the growth of the welfare state and class conflicts to name a few, greatly increased the complexity of western societies as seen by e.g. the growth in division of labour or bureaucratic institutions [38]. It was common for evolutionism to model societies as systems that have a life of their own. This systemic perspective resulted in famous intellectual insights – Marx and Engels studied economic systems and the class structures they produce, Durkheim focused on cultural systems and Comte investigated the evolutionary stages through which societies passed. However, understanding how complex social structures and institutions on the systemic level can result from the actions of individual actors, who are after all the fundamental constituents of social systems, remained problematic [38].

An alternative view that is slowly making its way into sociological research is that of complexity as inherent characteristic of any social system [38, 213]. It is this view that we have adopted in this dissertation in our pursuit of understanding unintended consequences as emerging from individual interactions. Our main contribution to sociology, thus, revolves around demonstrating how simple interaction rules can lead to unexpected collective effects, such as global cooperation, wisdom of crowds and collapse of an online community. The interaction mechanisms we presented have intentionally been made much simpler than they could potentially be. For example, the decision of a user to join/remain or leave an OSN may account for factors such as existing friends already in the network, information overload, quality of consumed content, or any other number of external factors. However, including all these parameters brings the risk of overfitting and making the model no less tractable than the original scenario it was intended to characterise. As a result, one may easily lose track of which parameter is precisely responsible for the collective behaviour we observe. Instead, we emphasise that agent-based models bring *qualitative* understanding of the individual-level mechanisms by which collective phenomena arise. Because the individual agents are greatly simplified, the complexity of the emerging behaviour at the level of the population may only result from their interactions, rather than be due to inherent individual complexity. It is this emphasis on understanding rather than exact prediction that we believe sociology can benefit from in studying social complexity in general.

Relevance for behavioural biology. The leading-following networks we constructed provided evidence for the potency of the network approach in studying individual role in biological systems. While social network theory has already been applied in animal systems (see the review in Chapter 7), we applied it for the first time in studying information

transfer about suitable roosts. In this respect, we hope that the methodology we employed in inferring the interaction rules underlying recruitment and in identifying systemically important individuals will stimulate the development of novel experimental set-ups that can employ the same network perspective on studying other evolutionary research questions as well. Moreover, our results were obtained by an automated and rigorous method of extracting and calibrating meaningful patterns from inherently noisy behavioural data. We hope that this serves as a demonstration that, given higher quality individual data, systematic methods can provide even more considerable insights.

Future research. In addition to the results presented in this dissertation, we see potential for further investigation, both empirical and theoretical, on several of the topics we addressed. First, an experimental study on the effects of restricting payoff-related information on the global level of cooperation may be carried out. It would be interesting to test the extent to which uninformed individuals actually imitate their local neighbourhood, provided the right non-linear herding incentives. Second, the empirical results on rank and competition can be studied theoretically with the help of an agent-based model. In simplest terms, the model shall only include the strength of an individual's response to his current rank. Should this model, after calibration, succeed in reproducing the crowd's collective accuracy and diversity, it would imply that individual behaviour in such a setting may be surprisingly simple on average. Otherwise, in the spirit of agent-based modelling, we would need to increase the complexity of individual response to rank information. Third, additional research is needed on our social resilience work. In particular, regarding Chapter 5, additional microscopic data on user activity and churn can provide estimators for the benefits and costs of each network to further validate that the k-core decomposition applies in real scenarios as well. The work presented in the same chapter is theoretically limited to the study of monotonously increasing, convex objective functions of benefit versus active neighborhood. While empirical studies support this assumption [15, 270], it is possible to imagine a scenario where information overload decreases the net benefit of users with very large neighbourhoods, creating nonlinearities where the generalized k-core is not a stable solution. Furthermore, the generalised k-core can be applied when user decisions are more complex than just staying or leaving the network, for example introducing heterogeneity of benefits or weights in the social links. Fourth, with respect to the coordination problem in Bechstein bats, we see the potential to link information transfer about suitable roosts with actual roost occupation. We can employ the same individual-level approach and pose the question of how individual bats decide to occupy a roost, and how these decisions are affected by previous participation in the recruitment process. For example, one can imagine an agent-based model where each bat receives a certain utility

from occupying a given roost. This utility may depend on personal preference for this roost and belief about others' preferences. These preferences, in turn, can be proxied by the observed leading-following behaviour in the colony.

Bibliography

- [1] Achlioptas, D.; Clauset, A.; Kempe, D.; Moore, C. (2005). On the bias of traceroute sampling. In: *STOC '05*.
- [2] Achrekar, H.; Gandhe, A.; Lazarus, R.; Yu, S.-H.; Liu, B. (2011). Predicting Flu Trends using Twitter data. In: *Computer Communications Workshops (INFOCOM WKSHPS)*. pp. 702–707.
- [3] Ahn, Y.-Y.; Han, S.; Kwak, H.; Moon, S.; Jeong, H. (2007). Analysis of topological characteristics of huge online social networking services. In: *WWW '07*.
- [4] Ainslie, G. (2001). *Breakdown of will*. Cambridge University Press.
- [5] Ammitzboell, K. (2007). Unintended consequences of peace operations on the host economy from a people's perspective. In: C. Aoi; C. de Coning; R. Thakur (eds.), *Unintended consequences of peacekeeping operations*, United Nations University Press, vol. 1 of *Lecture Notes in Computer Science*. pp. 69–90.
- [6] Anderson, A.; Huttenlocher, D.; Kleinberg, J.; Leskovec, J. (2013). Steering user behavior with badges. In: *Proceedings of the 22nd international conference on World Wide Web*. International World Wide Web Conferences Steering Committee, pp. 95–106.
- [7] Andreoni, J.; Miller, J. H. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *Economic Journal* **103**(418), 570–85.
- [8] Anscombe, G. (1958). Modern Moral Philosophy. *Philosophy* **33**(124), 1–16.
- [9] Armstrong, J. S. (2001). *Principles of forecasting: a handbook for researchers and practitioners*, vol. 30. Springer.
- [10] Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgements. *Groups, Leadership and Men. S* , 226–236.

- [11] Axelrod, R. (1980). Effective choice in the prisoner's dilemma. *Journal of Conflict Resolution* **24(1)**, 3–25.
- [12] Axelrod, R. (2006). *The evolution of cooperation: revised edition*. Basic books.
- [13] Axelrod, R.; Dion, D. (1988). The further evolution of cooperation. *Science* **242**, 1385–1390.
- [14] Axelrod, R.; Iliev, R. (2014). Timing of cyber conflict. *Proceedings of the National Academy of Sciences* .
- [15] Backstrom, L.; Huttenlocher, D.; Kleinberg, J.; Lan, X. (2006). Group formation in large social networks. In: *KDD '06*.
- [16] Baigger, A. (2012). *The effect of social sub-structuring on information transfer and group decision-making in female Bechstein's bats*. Master's thesis, University of Würzburg.
- [17] Ball, J. (1973). The zoo hypothesis. *Icarus* **19(3)**, 347–349.
- [18] Benevenuto, F.; Rodrigues, T.; Cha, M.; Almeida, V. (2009). Characterizing User Behavior in Online Social Networks. In: *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference*. IMC '09, New York, NY, USA: ACM, pp. 49–62.
- [19] Berman, A.; Plemmons, R. J. (1987). *Nonnegative Matrices in the Mathematical Sciences*. Society for Industrial and Applied Mathematics, 360 pp.
- [20] Biggs, N.; Lloyd, K. E.; Wilson, R. (1999). *Graph Theory 1736-1936*. Oxford University Press, 240 pp.
- [21] Bikhchandani, S.; Hirshleifer, D.; Welch, I. (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy* **100(5)**, 992–1026.
- [22] Bikhchandani, S.; Sharma, S. (2000). Herd Behavior in Financial Markets. *IMF Staff Papers* **47(3)**, 279–300.
- [23] Blog, I. (2014). Instagram Today: 200 Million Strong. <http://blog.instagram.com/post/80721172292/200m>. Accessed: 2014-04-15.
- [24] Blume, L. E. (1993). The Statistical Mechanics of Strategic Interaction. *Games and Economic Behavior* **5(3)**, 387–424.

- [25] Bon, G. L. (1895). *Psychologie Des Foules*. Livres Généraux, 132 pp.
- [26] Bonabeau, E. (2009). Decisions 2.0: The Power of Collective Intelligence. *MIT Sloan Management Review* **50**.
- [27] Boortz, C.; Jurkatis, S.; Kremer, S.; Nautz, D. (2013). *Herding in financial markets: Bridging the gap between theory and evidence*. Tech. rep., Humboldt University, Collaborative Research Center 649.
- [28] Borgatti, S. (2005). Centrality and network flow. *Social Networks* **27(1)**, 55 – 71.
- [29] Borgatti, S.; Everett, M. (2000). Models of core/periphery structures. *Social Networks* **21(4)**, 375–395.
- [30] Boudon, R. (1981). *The Logic of Social Action: An Introduction to Sociological Analysis*. Routledge, 1 edn., 208 pp.
- [31] Boudon, R. (1989). Why do Social Scientists Tend to See the World as Over-Ordered? *Philosophica* **44(2)**, 15–31.
- [32] Brenner, D. J.; Doll, R.; Goodhead, D. T.; Hall, E. J.; Land, C. E.; Little, J. B.; Lubin, J. H.; Preston, D. L.; Preston, R. J.; Puskin, J. S.; Ron, E.; Sachs, R. K.; Samet, J. M.; Setlow, R. B.; Zaider, M. (2003). Cancer risks attributable to low doses of ionizing radiation: Assessing what we really know. *Proceedings of the National Academy of Sciences* **100(24)**, 13761–13766.
- [33] Brown, J. J.; Hildreth, R.; Ford, S. E. (2005). When the World Is Not Your Oyster. *Science* **309(5732)**, 244.
- [34] Buckley, W. F. (1998). *Society—a Complex Adaptive System: Essays in Social Theory*, vol. 9. Taylor & Francis.
- [35] Buessler, K. O.; Jayne, S. R.; Fisher, N. S.; Rypina, I. I.; Baumann, H.; Baumann, Z.; Breier, C. F.; Douglass, E. M.; George, J.; Macdonald, A. M.; Miyamoto, H.; Nishikawa, J.; Pike, S. M.; Yoshida, S. (2012). Fukushima-derived radionuclides in the ocean and biota off Japan. *Proceedings of the National Academy of Sciences of the United States of America* **109(16)**, 5984–8.
- [36] Cable, D. M.; Shane, S. (1997). A prisoner’s dilemma approach to entrepreneur-venture capitalist relationships. *Academy of Management Review* **22(1)**, 142–176.

- [37] Capocci, A.; Servedio, V. D. P.; Colaioni, F.; Buriol, L. S.; Donato, D.; Leonardi, S.; Caldarelli, G. (2006). Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia. *Phys. Rev. E* **74**, 036116.
- [38] Castellani, B.; Hafferty, F. W. (2009). *Sociology and complexity science: a new field of inquiry*. Springer.
- [39] Castellano, C.; Fortunato, S.; Loreto, V. (2009). Statistical Physics of Social Dynamics. *Reviews of Modern Physics* **81(2)**, 591–646.
- [40] Cattani, G.; Ferriani, S. (2008). A Core/Periphery Perspective on Individual Creative Performance: Social Networks and Cinematic Achievements in the Hollywood Film Industry. *Organization Science* **19(6)**, 824–844.
- [41] Cha, M.; Haddadi, H.; Benevenuto, F.; Gummadi, P. K. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy. *ICWSM* **10**, 10–17.
- [42] Chun, H.; Kwak, H.; Eom, Y.-H.; Ahn, Y.-Y.; Moon, S.; Jeong, H. (2008). Comparison of online social relations in volume vs interaction. In: *IMC '08*.
- [43] Clauset, A.; Shalizi, C. R.; Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. *SIAM Review* **51(4)**, 661.
- [44] Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* **5(4)**, 559–583.
- [45] Clements, K. C.; Stephens, D. W. (1995). Testing models of non-kin cooperation: mutualism and the Prisoner's Dilemma. *Animal Behaviour* **50(2)**, 527–535.
- [46] Coleman, J. (1987). Microfoundations and macrosocial behavior. In: J. Alexander; B. Giesen; R. Münch; S. N.J (eds.), *The Micro-Macro Link*, University of California Press. pp. 153–173.
- [47] Coleman, J. S. (1978). *Introduction to Mathematical Sociology*. MacMillan Pub Co.
- [48] Coleman, J. S.; Fararo, T. J. (1993). *Rational choice theory*. Sage Publications.
- [49] Coleman, S.; Blumler, J. G. (2011). The Wisdom of Which Crowd? On the Pathology of a Listening Government. *The Political Quarterly* **82(3)**, 355–364.
- [50] Cont, R.; Bouchaud, J.-P. (2000). Herd Behavior and Aggregate Fluctuations in Financial Markets. *Macroeconomic Dynamics* **4**.

- [51] Corfman, K. P.; Lehmann, D. R. (1994). The prisoner's dilemma and the role of information in setting advertising budgets. *Journal of Advertising* **23(2)**, 35–48.
- [52] Cornelius, W. A. (2001). Death at the Border: Efficacy and Unintended Consequences of US Immigration Control Policy. *Population and Development Review* **27(4)**, 661–685.
- [53] Corten, R. (2012). Composition and Structure of a Large Online Social Network in the Netherlands. *PLoS ONE* **7(4)**, e34760.
- [54] Cosmelli, D.; Thompson, E. (2010). Embodiment or envatment? Reflections on the bodily basis of consciousness. In: J. Stewart; O. Gapenne; E. Di Paolo (eds.), *Enaction: Toward a New Paradigm for Cognitive Science.*, The MIT Press, Cambridge, MA, 2010. pp. 361–385.
- [55] Couzin, I.; Laidre, M. (2009). Fission–fusion populations. *Current biology* **19(15)**, R633–R635.
- [56] Couzin, I. D.; Ioannou, C. C.; Demirel, G.; Gross, T.; Torney, C. J.; Hartnett, A.; Conradt, L.; Levin, S. A.; Leonard, N. E. (2011). Uninformed Individuals Promote Democratic Consensus in Animal Groups. *Science* **334(6062)**, 1578–1580.
- [57] Croft, D.; James, R.; Krause, J. (2008). *Exploring animal social networks*. Princeton University Press.
- [58] Croft, D. P.; Krause, J.; James, R. (2004). Social networks in the guppy (*Poecilia reticulata*). *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271(Suppl 6)**, S516–S519.
- [59] Davenport, T. H.; Beck, J. C. (2001). *The attention economy: Understanding the new currency of business*. Harvard Business Press.
- [60] Dawid, A.; DeGroot, M.; Mortera, J.; Cooke, R.; French, S.; Genest, C.; Schervish, M.; Lindley, D.; McConway, K.; Winkler, R. (1995). Coherent combination of experts' opinions. *TEST* **4**, 263–313.
- [61] Dawkins, R. (1990). *The Selfish Gene*. Oxford University Press, USA, 368 pp.
- [62] Devenow, A.; Welch, I. (1996). Rational herding in financial economics. *European Economic Review* **40(3–5)**, 603–615.

- [63] DiNardo, J.; Lemieux, T. (2001). Alcohol, marijuana, and American youth: the unintended consequences of government regulation. *Journal of Health Economics* **20(6)**, 991 – 1010.
- [64] Echeburúa, E.; de Corral, P. (2009). [Addiction to new technologies and to online social networking in young people: A new challenge]. *Adicciones* **22(2)**, 91–95.
- [65] Einhorn, H. J.; Hogarth, R. M.; Klempner, E. (1977). Quality of group judgment. *Psychological Bulletin* **84(1)**, 158.
- [66] Endreß, M. (2010). Unvorhergesehene Effekte – altes Thema, neue Probleme? In: G. Albert; R. Greshoff; R. Schützeichel (eds.), *Dimensionen und Konzeptionen von Sozialität*, VS Verlag für Sozialwissenschaften. pp. 13–32.
- [67] Engel, K. (2005). Mitigating global climate change in the United States: a regional approach. *NYU Envtl. LJ* **14**, 54.
- [68] Everett, M.; Borgatti, S. (2000). Peripheries of cohesive subsets. *Social Networks* **21(4)**, 397 – 407.
- [69] Fehr, E.; Fischbacher, U. (2003). The nature of human altruism. *Nature* **425(6960)**, 785–791.
- [70] Feinstein, B. A.; Hershenberg, R.; Bhatia, V.; Latack, J. A.; Meuwly, N.; Davila, J. (2013). Negative social comparison on Facebook and depressive symptoms: Rumination as a mechanism. *Psychology of Popular Media Culture* **2(3)**, 161.
- [71] Festinger, L. (1954). A theory of social comparison processes. *Human relations* **7(2)**, 117–140.
- [72] Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.
- [73] Fleischmann, D. (2008). *Roosting behaviour, information transfer, and group decision making in a maternity colony of Bechstein's bats*. Master's thesis, University of Würzburg.
- [74] Forbes (2012). Facebook Buys Instagram For \$1 Billion. Smart Arbitrage. <http://www.forbes.com/sites/bruceupbin/2012/04/09/facebook-buys-instagram-for-1-billion-wheres-the-revenue/>. Accessed: 2014-04-15.
- [75] Galam, S. (2008). Sociophysics: A Review Of Galam Models. *International Journal of Modern Physics C* **19(03)**, 409–440.

- [76] Gale, D.; Shapley, L. (1962). College admissions and the stability of marriage. *American Mathematical Monthly* , 9–15.
- [77] Galton, F. (1907). Vox Populi. *Nature* **75(1949)**, 450–451.
- [78] Garas, A.; Garcia, D.; Skowron, M.; Schweitzer, F. (2012). Emotional persistence in online chatting communities. *Scientific Reports* **2**, 402.
- [79] Garas, A.; Schweitzer, F.; Havlin, S. (2012). A k-shell decomposition method for weighted networks. *New Journal of Physics* **14**, 083030.
- [80] Garcia, D. (2013). Social Resilience in Online Communities: The Autopsy of Friendster. In: *Conference in Online Social Networks*.
- [81] Garcia, D.; Garas, A.; Schweitzer, F. (2012). Positive words carry less information than negative words. *EPJ Data Science* **1**, 3.
- [82] Garcia, D.; Mavrodiev, P.; Schweitzer, F. (2013). Social Resilience in Online Communities: The Autopsy of Friendster. *arXiv arXiv:1302.6109*.
- [83] Garcia, D.; Weber, I.; Garimella, R. V. K. (2014). Gender Asymmetries in Reality and Fiction : The Bechdel Test of Social Media. In: *International AAAI Conference on Weblogs and Social Media*.
- [84] de Gardelle, V.; Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences* **108(32)**, 13341–13346.
- [85] Gayo-Avello, D. (2013). Nepotistic relationships in Twitter and their impact on rank prestige algorithms. *Information Processing & Management* **49(6)**, 1250 – 1280.
- [86] Genest, C.; Zidek, J. V. (1986). Combining Probability Distributions: A Critique and an Annotated Bibliography. *Statistical Science* **1(1)**, 114–135.
- [87] Giles, M. (Jan 28th, 2010). A world of connections - A special report on social networking. *The Economist* .
- [88] Gjoka, M.; Kurant, M.; Butts, C. T.; Markopoulou, A. (). Walking in Facebook: A Case Study of Unbiased Sampling of OSNs. In: *INFOCOM '10*.
- [89] Golbeck, J. (2009). Trust and Nuanced Profile Similarity in Online Social Networks. *ACM Trans. Web* **3(4)**, 12:1–12:33.
- [90] Golub, B.; Jackson, M. O. (2010). Naïve Learning in Social Networks and the Wisdom of Crowds. *American Economic Journal: Microeconomics* **2(1)**, 112–49.

- [91] Gordon, K. (1924). Group Judgments in the Field of Lifted Weights. *Journal of Experimental Psychology* **7(5)**, 398–400.
- [92] Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology* **78(6)**, 1.
- [93] Griffith, T. L.; Fuller, M. A.; Northcraft, G. B. (1998). Facilitator Influence in Group Support Systems: Intended and Unintended Effects. *Information Systems Research* **9(1)**, 20–36.
- [94] Groeber, P.; Lorenz, J.; Schweitzer, F. (2013). Dissonance minimization as a micro-foundation of social influence in models of opinion formation”. *Journal of Mathematical Sociology* .
- [95] Groeber, P.; Rauhut, H. (2009). Does ignorance promote norm compliance? *Computational and Mathematical Organization Theory* **16(1)**, 1–28.
- [96] Groeber, P.; Schweitzer, F.; Press, K. (2009). How Groups Can Foster Consensus: The Case of Local Cultures. *Journal of Artificial Societies & Social Simulation* **12(2)**.
- [97] Hardin, G. (1968). The Tragedy of the Commons. *Science* **162(3859)**, 1243–1248.
- [98] Harkins, A. (unpublished). Network Games with Perfect Complements. Warwick University Draft.
- [99] Harri, O.-K. (2008). Network Analysis and Crowds of People as Sources of New Organisational Knowledge. In: A. Koohang; K. Harman; J. Britz (eds.), *Knowledge Management: Theoretical Foundation*, Informing Science Press. pp. 173–189.
- [100] Hauert, C.; De Monte, S.; Hofbauer, J.; Sigmund, K. (2002). Volunteering as Red Queen mechanism for cooperation in public goods games. *Science (New York, N.Y.)* **296(5570)**, 1129–32.
- [101] Hauert, Christoph (2001). Fundamental clusters in spatial 2x2 games. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **268(1468)**, 761–769.
- [102] Hegselmann, R.; Krause, U. (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation* **5(3)**.
- [103] Hegselmann, R.; Krause, U. (2005). Opinion Dynamics Driven by Various Ways of Averaging. *Computational Economics* **25**, 381–405.

- [104] Helbing, D.; Johansson, A.; Al-Abideen, H. (2007). Dynamics of crowd disasters: An empirical study. *Physical Review E* **75(4)**, 046109.
- [105] Helbing, D.; Yu, W. (2008). Migration as a mechanism to promote cooperation. *ACS - Advances in Complex Systems* **11**, 641–652.
- [106] Henderson, J. (2012). Is Social Media Destroying Your Self-Esteem? <http://www.forbes.com/sites/jmaureenhenderson/2012/07/11/is-social-media-destroying-your-self-esteem/>. Accessed: 2014-04-15.
- [107] Herzog, S. M.; Hertwig, R. (2009). The Wisdom of Many in One Mind Improving Individual Judgments With Dialectical Bootstrapping. *Psychological Science* **20(2)**, 231–237.
- [108] Highfield, R. (1996). *Frontiers of complexity: The search for order in a chaotic world*. Ballantine Books.
- [109] Hirshleifer, J.; Carlos, J.; Coll, M. (1988). What Strategies Can Support the Evolutionary Emergence of Cooperation? *The Journal of Conflict Resolution* **32(2)**, 367–398.
- [110] Hodas, N.; Lerman, K. (2014). The Simple Rules of Social Contagion. *Scientific Reports* **4(4343)**, 402.
- [111] Hojman, D.; Szeidl, A. (2008). Core and periphery in networks. *Journal of Economic Theory* **139(1)**, 295 – 309.
- [112] Holley, R. A.; Liggett, T. M. (1975). Ergodic Theorems for Weakly Interacting Infinite Systems and the Voter Model. *The Annals of Probability* **3(4)**, 643–663.
- [113] Holme, P. (2005). Core-periphery organization of complex networks. *Phys. Rev. E* **72**, 046111.
- [114] Hølst, J. A.; Kacperski, K.; Schweitzer, F. (2001). Social impact models of opinion dynamics. In: *In: Annual Reviews of Computational Physics*. pp. 253–272.
- [115] Hommes, C.; Sonnemans, J.; Tuinstra, J.; Van de Velden, H. (2005). Coordination of expectations in asset pricing experiments. *Review of Financial Studies* **18(3)**, 955–980.
- [116] Hong, L.; Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of the United States of America* **101(46)**, 16385–9.

- [117] Huberman, B. A.; Romero, D. M.; Wu, F. (2009). Crowdsourcing, attention and productivity. *Journal of Information Science* **35(6)**, 758–765.
- [118] Insider, B. (213). Foursquare CEO Dennis Crowley Responds To All Of His Company’s ”Haters”. <http://www.businessinsider.com/foursquare-ceo-dennis-crowley-responds-to-all-of-his-companys-haters-2013-4>. Accessed: 2014-04-15.
- [119] Issacharoff, S. (2008). Democracy and collective decision making. *International Journal of Constitutional Law* **6(2)**, 231–266.
- [120] Jackson, C. (2012). Instagram loses focus of its reputation. <http://www.trackur.com/instagrams-loses-focus-of-its-reputation>. Accessed: 2014-04-15.
- [121] Jackson, J. B. C.; Kirby, M. X.; Berger, W. H.; Bjorndal, K. A.; Botsford, L. W.; Bourque, B. J.; Bradbury, R. H.; Cooke, R.; Erlandson, J.; Estes, J. A.; Hughes, T. P.; Kidwell, S.; Lange, C. B.; Lenihan, H. S.; Pandolfi, J. M.; Peterson, C. H.; Steneck, R. S.; Tegner, M. J.; Warner, R. R. (2001). Historical Overfishing and the Recent Collapse of Coastal Ecosystems. *Science* **293(5530)**, 629–637.
- [122] Jain, S.; Krishna, S. (1998). Autocatalytic sets and the growth of complexity in an evolutionary model. *Physical Review Letters* **81(25)**, 5684–5687.
- [123] Jain, S.; Krishna, S. (2002). Crashes, recoveries, and ”core shifts” in a model of evolving networks. *Phys. Rev. E* **65**, 026103.
- [124] Jain, S.; Krishna, S. (2005). Graph theory and the evolution of autocatalytic networks. In: *Handbook of Graphs and Networks*. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 355–395.
- [125] Jervis, R. (1998). *System Effects: Complexity in Political and Social Life*. Princeton University Press, 328 pp.
- [126] Kahan, D. M. (1997). Social Influence, Social Meaning, and Deterrence. *Virginia Law Review* **83(2)**, 349–395.
- [127] Kairam, S. R.; Wang, D. J.; Leskovec, J. (2012). The life and death of online groups. In: *WSDM ’12*.
- [128] Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika* **18(1)**, 39–43.

- [129] Kerth, G.; Kiefer, A.; Trappmann, C.; Weishaar, M. (2003). High gene diversity at swarming sites suggest hot spots for gene flow in the endangered Bechstein’s bat. *Conservation Genetics* **4**(4), 491–499.
- [130] Kerth, G.; König, B. (1999). Fission, fusion and nonrandom associations in female Bechstein’s bats (*Myotis bechsteinii*). *Behaviour* **136**(9), 1187–1202.
- [131] Kerth, G.; Perony, N.; Schweitzer, F. (2011). Bats are able to maintain long-term social relationships despite the high fission-fusion dynamics of their groups. *Proceedings of the Royal Society B: Biological Sciences* .
- [132] Kerth, G.; Reckardt, K. (2003). Information transfer about roosts in female Bechstein’s bats: an experimental field study. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**(1514), 511–515.
- [133] Kerth, G.; Weissmann, K.; König, B. (2001). Day roost selection in female Bechstein’s bats (*Myotis bechsteinii*): a field experiment to determine the influence of roost temperature. *Oecologia* **126**(1), 1–9.
- [134] Kitsak, M.; Gallos, L. K.; Havlin, S.; Liljeros, F.; Muchnik, L.; Stanley, H. E.; Makse, H. A. (2010). Identification of influential spreaders in complex networks. *Nature Physics* **6**(11), 888–893.
- [135] Kittur, A.; H., C. E. (2007). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In: *25th Annual ACM Conference on Human Factors in Computing Systems (CHI 2007)*.
- [136] Kittur, A.; Kraut, R. E. (2008). Harnessing the wisdom of crowds in wikipedia: quality through coordination. In: *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. CSCW ’08, New York, NY, USA: ACM, pp. 37–46.
- [137] Klugman, S. F. (1945). Group judgments for familiar and unfamiliar materials. *The Journal of General Psychology* **32**(1), 103–110.
- [138] Klugman, S. F. (1947). Group and individual judgments for anticipated events. *The Journal of Social Psychology* **26**(1), 21–28.
- [139] Kogan, D. (2011). How Facebook Saved My Son’s Life. <http://www.slate.com/id/2297933/pagenum/a11/> [Last accessed: August 7th, 2011].
- [140] König, M.; Battiston, S.; Schweitzer, F. (2009). Modeling evolving innovation networks. In: A. Pyka; A. Scharnhorst (eds.), *Innovation networks: New approaches in modelling and analyzing*, Berlin: Springer. pp. 189–269.

- [141] König, M.; Battiston, S.; Schweitzer, F. (2009). Modeling Evolving Innovation Networks. In: A. Pyka; A. Scharnhorst (eds.), *Innovation Networks*, Understanding Complex Systems, Springer Berlin Heidelberg. pp. 189–269.
- [142] Kraskov, A.; Stögbauer, H.; Grassberger, P. (2004). Estimating mutual information. *Phys. Rev. E* **69**.
- [143] Krause, J.; Croft, D.; James, R. (2007). Social network theory in the behavioural sciences: potential applications. *Behavioral Ecology and Sociobiology* **62(1)**, 15–27.
- [144] Krause, J.; Lusseau, D.; James, R. (2009). Animal social networks: an introduction. *Behavioral Ecology and Sociobiology* **63(7)**, 967–973.
- [145] Krause, J.; Ruxton, G. D. (2002). *Living in Groups*. Oxford University Press, 210 pp.
- [146] Kuhlman, M.; Marshello, A. (1975). Individual differences in game motivation as moderators of preprogrammed strategy effects in prisoner’s dilemma. *Journal of personality and social psychology* **32(5)**, 922.
- [147] Kumar, R.; Novak, J.; Tomkins, A. (2010). Structure and Evolution of Online Social Networks. In: P. S. Yu; J. Han; C. Faloutsos (eds.), *Link Mining: Models, Algorithms, and Applications*, Springer New York. pp. 337–357.
- [148] Kwak, H.; Lee, C.; Park, H.; Moon, S. (2010). What is Twitter, a social network or a news media? In: *Proceedings of the 19th international conference on World wide web*. ACM, pp. 591–600.
- [149] Laine, M. S. S.; Ercal, G. (2011). User Groups in Social Networks: An Experimental Study on YouTube. In: *HICSS ’11*.
- [150] Landis, J. D.; Hamm, N. T.; Renshaw, C. E.; Dade, W. B.; Magilligan, F. J.; Gartner, J. D. (2012). Surficial redistribution of fallout 131iodine in a small temperate catchment. *Proceedings of the National Academy of Sciences* **109(11)**, 4064–4069.
- [151] Langevoort, Donald (2000). The Human Nature of Corporate Boards: Law, Norms and the Unintended Consequences of Independence and Accountability. *The Georgetown Law Journal* **89(797)**, 797–832.
- [152] Laughlin, R. B. (2006). *A Different Universe: Reinventing Physics from the Bottom Down*. Basic Books, 272 pp.

- [153] Lee, M. D.; Zhang, S.; Shi, J. (2011). The wisdom of the crowd playing The Price Is Right. *Memory & Cognition* **39(5)**, 914–23.
- [154] Leinert, V. (2010). *Fission-fusion dynamics and group decision making in female Bechstein's bats*. Master's thesis, University of Osnabrück.
- [155] Leskovec, J.; Horvitz, E. (2008). Planetary-scale views on a large instant-messaging network. In: *WWW '08*.
- [156] Linstone, H. A.; Turoff, M.; *et al.* (1975). *The Delphi method: Techniques and applications*. Addison-Wesley Publishing Company, Advanced Book Program Boston, MA.
- [157] Lorenz, J.; Rauhut, H.; Schweitzer, F.; Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences* **108(22)**, 9020–9025.
- [158] Lorenz, J.; Rauhut, H.; Schweitzer, F.; Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences* .
- [159] Lorge, I.; Fox, D.; Davitz, J.; Brenner, M. (1958). A survey of studies contrasting the quality of group performance and individual performance, 1920-1957. *Psychological Bulletin* **55(6)**, 337.
- [160] Lorge, I.; Fox, D.; Davitz, J.; Brenner, M. (1958). A survey of studies contrasting the quality of group performance and individual performance, 1920-1957. *Psychological Bulletin* **55(6)**, 337–372.
- [161] Luhmann, N. (1995). *Social systems*. Stanford University Press.
- [162] Lusseau, D. (2003). The emergent properties of a dolphin social network. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270(Suppl 2)**, S186–S188.
- [163] Lusseau, D.; Newman, M. (2004). Identifying the role that animals play in their social networks. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271(Suppl 6)**, S477–S481.
- [164] Mackay, C. (2003). *Extraordinary Popular Delusions*. Dover Publications, 112 pp.

- [165] Mage, D.; Ozolins, G.; Peterson, P.; Webster, A.; Orthofer, R.; Vandeweerd, V.; Gwynne, M. (1996). Urban air pollution in megacities of the world. *Atmospheric Environment* **30(5)**, 681 – 686.
- [166] Malinowski, B. (1913). *The family among the Australian aborigines; a sociological study*. University of London Press, 352 pp.
- [167] Mannes, A. E. (2009). Are We Wise About the Wisdom of Crowds? The Use of Group Judgments in Belief Revision. *Management Science* **55(8)**, 1267–1279.
- [168] Mardia, K.; Jupp, P. (1999). *Directional Statistics*. Wiley.
- [169] Mavrodiev, P.; Tessone, C. J.; Schweitzer, F. (2013). Quantifying the effects of social influence. *Scientific reports* **3**.
- [170] Mavrodiev, P.; Tessone, J. C.; Schweitzer, F. (2012). Effects of Social Influence on the Wisdom of Crowds. In: *Collective Intelligence conference, 2012*.
- [171] McDonald, D. (2007). Predicting fate from early connectivity in a social network. *Proceedings of the National Academy of Sciences* **104(26)**, 10910–10914.
- [172] McFarland, D.; Tinbergen, N. (1981). *The Oxford companion to animal behaviour*. Oxford University Press Oxford.
- [173] McKelvey, R. D.; Palfrey, T. R. (1995). Quantal Response Equilibria for Normal Form Games. *Games and Economic Behavior* **10(1)**, 6–38.
- [174] Merton, R. (1936). The Unanticipated Consequences of Purposive Social Action. *American Sociological Review* **1(6)**, 894–904.
- [175] Mica, A.; Peisert, A.; Winczorek, J. (eds.) (2012). *Sociology and the Unintended: Robert Merton Revisited (Polish Studies in Culture, Nations and Politics)*. Peter Lang International Academic Publishers, 387 pp.
- [176] Miguel, M. S.; Eguíluz, V. M.; Toral, R.; Klemm, K. (2005). Binary and multivariate stochastic models of consensus formation. *Computing in Science & Engineering* **7(6)**, 67–73.
- [177] Milgram, S. (1967). The small world problem. *Psychology today* **2(1)**, 60–67.
- [178] Milinski, M.; Sommerfeld, R. D.; Krambeck, H.-J.; Reed, F. A.; Marotzke, J. (2008). The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proceedings of the National Academy of Sciences* **105(7)**, 2291–2294.

- [179] Mislove, A.; Marcon, M.; Gummadi, K. P.; Druschel, P.; Bhattacharjee, B. (2007). Measurement and analysis of online social networks. In: *IMC '07*.
- [180] Moorhead, G.; Ference, R.; Neck, C. P. (1991). Group Decision Fiascos Continue: Space Shuttle Challenger and a Revised Groupthink Framework. *Human Relations* **44(6)**, 539–550.
- [181] Moreno, J. L. (1934). *Who shall survive?: A new approach to the problem of human interrelations*. Nervous and Mental Disease Publishing.
- [182] Moussaïd, M.; Kämmer, J. E.; Analytis, P. P.; Neth, H. (2013). Social Influence and the Collective Dynamics of Opinion Formation. *PLoS ONE* **8(11)**, e78433.
- [183] Murr, A. E. (2011). "Wisdom of crowds"? A decentralised election forecasting model that uses citizens' local expectations. *Electoral Studies* **30(4)**, 771–783.
- [184] Nowak, M. (2006). Five rules for the evolution of cooperation. *Science* **314(5805)**, 1560–1563.
- [185] Olson, M.; Olson, M. (2009). *The logic of collective action: public goods and the theory of groups*, vol. 124. Harvard University Press.
- [186] Onnela, J.-P.; Reed-Tsochas, F. (2010). Spontaneous emergence of social influence in online systems. *Proceedings of the National Academy of Sciences* **107(43)**, 18375–18380.
- [187] Opp, K.-D. (2011). Modeling Micro-Macro Relationships: Problems and Solutions. *The Journal of Mathematical Sociology* **35(1-3)**, 209–234.
- [188] O'Shea, T.; Vaughan, T. (1977). Nocturnal and seasonal activities of the pallid bat, *Antrozous pallidus*. *Journal of Mammalogy* , 269–284.
- [189] Page, S. E. (2007). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press.
- [190] Parchem, M. (2009). *Group decisions in a fission-fusion society: An experimental field study in a large Bechstein's bat colony*. Master's thesis, University of Würzburg.
- [191] Parker, W. D.; Prechter, R. R. (2005). Herding : An Interdisciplinary Integrative Review From A Socionomic Perspective. *International Conference on Cognitive Economics* , 11.

- [192] Parsons, T. (1937). *Structure of Social Action.: A Study in Social Theory with Special Reference to a Group of Recent European Writers*. McGraw-Hill.
- [193] Pastor-Satorras, R.; Vespignani, A. (2002). Epidemic dynamics in finite size scale-free networks. *Physical Review E* **65(3)**, 1–4.
- [194] Perlroth, N. (2013). Fake twitter followers become multimillion-dollar business. <http://bits.blogs.nytimes.com/2013/04/05/fake-twitter-followers-becomes-multimillion-dollar-business/>.
- [195] Pinter-Wollman, N.; Hobson, E. A.; Smith, J. E.; Edelman, A. J.; Shizuka, D.; de Silva, S.; Waters, J. S.; Prager, S. D.; Sasaki, T.; Wittemyer, G.; *et al.* (2013). The dynamics of animal social networks: analytical, conceptual, and theoretical advances. *Behavioral Ecology* **25**, 242–255.
- [196] Pollay, W. R. (1986). The Distorted Mirror: Reflections on the Unintended Consequences of Advertising. *Journal of Marketing* **50**, 18–36.
- [197] Portes, A. (2000). The Hidden Abode: Sociology as Analysis of the Unexpected. *American Sociological Review* **65(1)**, 1–18.
- [198] Poundstone, W. (2011). *Prisoner's dilemma*. Random House LLC.
- [199] Pretzlaff, I.; Kerth, G.; Dausmann, K. H. (2010). Communally breeding bats use physiological and behavioural adjustments to optimise daily energy expenditure. *Naturwissenschaften* **97(4)**, 353–363.
- [200] Raafat, R. M.; Chater, N.; Frith, C. (2009). Herding in humans. *Trends in Cognitive Sciences* **13(10)**, 420 – 428.
- [201] Radcliffe-Brown, A. R. (1940). On social structure. *Journal of the Anthropological Institute of Great Britain and Ireland* , 1–12.
- [202] Rapoport, A.; Chammah, A. M. (1965). *Prisoner's dilemma: a study in conflict and cooperation*. University of Michigan Press, 258 pp.
- [203] Raub, W.; Buskens, V.; Van Assen, M. A. L. M. (2011). Micro-Macro Links and Microfoundations in Sociology. *The Journal of Mathematical Sociology* **35(1-3)**, 1–25.
- [204] Rauhut, H.; Lorenz, J. (2011). The wisdom of crowds in one mind: How individuals can simulate the knowledge of diverse societies to reach better decisions. *Journal of Mathematical Psychology* **55(2)**, 191–197.

- [205] Ray, R. (2006). Prediction Markets and the Financial "Wisdom of Crowds". *Journal of Behavioral Finance* **7(1)**, 2–4.
- [206] Reuter, M. (2011). *Group decision-making in female Bechstein's bats: Mechanisms of the choice of day roosts in a fission-fusion society*. Master's thesis, University of Würzburg.
- [207] Richner, H.; Heeb, P. (1996). Communal life: honest signaling and the recruitment center hypothesis. *Behavioral Ecology* **7(1)**, 115–118.
- [208] Riker, W. (1982). *Liberalism Against Populism: A Confrontation Between the Theory of Social Democracy and the Theory of Social Choice*. San Francisco, CA: WH Freeman.
- [209] Roca, C.; Cuesta, J.; Sánchez, A. (2009). Effect of spatial structure on the evolution of cooperation. *Physical Review E* **80(4)**, 046106.
- [210] Safi, K.; Kerth, G. (2003). Secretions of the interaural gland contain information about individuality and colony membership in the Bechstein's bat. *Animal Behaviour* **65(2)**, 363–369.
- [211] Sansone, C.; Morf, C. C.; Panter, A. T. (2004). *The Sage Handbook of Methods in Social Psychology*. SAGE, 119 pp.
- [212] Savage, N. (2011). Twitter As Medium and Message. *Commun. ACM* **54(3)**, 18–20.
- [213] Sawyer, R. K. (2005). *Social Emergence: Societies As Complex Systems*. Cambridge University Press, 276 pp.
- [214] Schelling, T. C. (1971). Dynamic models of segregation. *The Journal of Mathematical Sociology* **1(2)**, 143–186.
- [215] Schneider, J. (2014). Likes or lies? How perfectly honest businesses can be overrun by Facebook spammers. <http://thenextweb.com/facebook/2014/01/23/likes-lies-perfectly-honest-businesses-can-overrun-facebook-spammers/>.
- [216] Schroeder, D. A. (1995). *Social Dilemmas: Perspectives on Individuals and Groups*. Praeger, 240 pp.
- [217] Schweitzer, F. (1997). Active brownian particles: Artificial agents in physics. In: L. Schimansky-Geier; T. Pöschel (eds.), *Stochastic Dynamics*, Springer Berlin / Heidelberg, vol. 484 of *Lecture Notes in Physics*. pp. 358–371.

- [218] Schweitzer, F. (2007). *Brownian Agents and Active Particles*. Springer Series in Synergetics, Springer Berlin Heidelberg.
- [219] Schweitzer, F.; Behera, L. (2009). Nonlinear voter models: The transition from invasion to coexistence. *European Physical Journal B* **67(3)**, 301–318.
- [220] Schweitzer, F.; Behera, L. (2012). Optimal migration promotes the outbreak of cooperation in heterogeneous populations. *ACS - Advances in Complex Systems* **15(Supp 1)**, 1250059.
- [221] Schweitzer, F.; Behera, L.; Mühlenbein, H. (2002). Evolution of Cooperation in a Spatial Prisoner’s Dilemma. *ACS - Advances in Complex Systems* **5(2-3)**, 269–299.
- [222] Schweitzer, F.; Holyst, J. A. (2000). Modelling collective opinion formation by means of active Brownian particles. *The European Physical Journal B* **15(4)**, 723–732.
- [223] Schweitzer, F.; Mach, R.; Mühlenbein, H. (2005). Agents with heterogeneous strategies interacting in a spatial IPD. In: T. Lux; S. Reitz; E. Samanidou (eds.), *Nonlinear Dynamics and Heterogenous Interacting Agents*, Springer, vol. 550 of *Lecture Notes in Economics and Mathematical Systems*. pp. 87–102.
- [224] Schweitzer, F.; Mavrodiev, P.; Tessone, J. C. (2013). How can social herding enhance cooperation? *Advances in Complex Systems* **16(04n05)**, 1350017.
- [225] Schweitzer, F.; Zimmermann, J.; Mühlenbein, H. (2002). Coordination of decisions in a spatial agent model. *Physica A* **303(1-2)**, 189–216.
- [226] Seidman, S. B. (1983). Network structure and minimum degree. *Social Networks* **5(3)**, 269–287.
- [227] Seufert, A. (2004). *Dynamics of Complex Adaptive Systems*. Master’s thesis, Technical University Berlin.
- [228] Seufert, A. M.; Schweitzer, F. (2007). Aggregate dynamics in an evolutionary network model. *International Journal of Modern Physics C* **18(10)**, 1659–1674.
- [229] Skoulakis, G. (2008). A recursive formula for computing central moments of a multivariate lognormal distribution. *The American Statistician* **62(2)**.
- [230] Smith, A. (1790). *The Theory of Moral Sentiments*. ΜεταLibri, 6 edn., 387 pp.
- [231] Sokolski, H. D. (2004). *Getting MAD: nuclear mutual assured destruction, its origins and practice*. DIANE Publishing.

- [232] de Sola Pool, I.; Kochen, M. (1979). Contacts and influence. *Social networks* **1(1)**, 5–51.
- [233] Solomon, M. (2006). Groupthink versus The Wisdom of Crowds: The Social Epistemology of Deliberation and Dissent. *The Southern Journal of Philosophy* **44(S1)**, 28–42.
- [234] Sood, V.; Redner, S. (2005). Voter Model on Heterogeneous Graphs. *Phys. Rev. Lett.* **94**, 178701.
- [235] Sorroos, M. (1994). Global change, environmental security, and the prisoner’s dilemma. *Journal of Peace Research* **31(3)**, 317–332.
- [236] Stark, H.-U.; Tessone, C. J.; Schweitzer, F. (2008). Decelerating microdynamics can accelerate macrodynamics in the voter model. *Physical Review Letters* **101**, 018701.
- [237] Stark, H.-U.; Tessone, C. J.; Schweitzer, F. (2008). Slower is faster: Fostering consensus formation by heterogeneous inertia. *ACS - Advances in Complex Systems* **11**, 551–563.
- [238] Steinfield, C.; Ellison, N. B.; Lampe, C. (2008). Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology* **29(6)**, 434 – 445.
- [239] Stephens, D. W.; McLinn, C. M.; Stevens, J. R. (2002). Discounting and reciprocity in an iterated prisoner’s dilemma. *Science* **298(5601)**, 2216–2218.
- [240] Stringhini, G.; Wang, G.; Egele, M.; Kruegel, C.; Vigna, G.; Zheng, H.; Zhao, B. Y. (2013). Follow the Green: Growth and Dynamics in Twitter Follower Markets. In: *Proceedings of the 2013 Conference on Internet Measurement Conference*. IMC ’13, New York, NY, USA: ACM, pp. 163–176.
- [241] Stroop, J. R. (1932). Is the judgment of the group better than that of the average member of the group? *Journal of experimental Psychology* **15(5)**, 550.
- [242] Stumpf, M. P. H.; Porter, M. A. (2012). Critical truths about power laws. *Science* **335(6069)**, 665–6.
- [243] Suchecki, K.; Eguíluz, V. M.; Miguel, M. S. (2005). Conservation laws for the voter model in complex networks. *EPL (Europhysics Letters)* **69(2)**, 228.

- [244] Sueur, C.; King, A. J.; Conradt, L.; Kerth, G.; Lusseau, D.; Mettke-Hofmann, C.; Schaffner, C. M.; Williams, L.; Zinner, D.; Aureli, F. (2011). Collective decision-making and fission-fusion dynamics: a conceptual framework. *Oikos* .
- [245] Sundaresan, S.; Fischhoff, I.; Dushoff, J.; Rubenstein, D. (2007). Network metrics reveal differences in social organization between two fission-fusion species, Grevy’s zebra and onager. *Oecologia* **151(1)**, 140–149.
- [246] Surowiecki, J. (2005). *The Wisdom of Crowds*. Anchor, 336 pp.
- [247] Szabó, G.; Fath, G. (2007). Evolutionary games on graphs. *Physics Reports* **446(4-6)**, 97–216.
- [248] Sznajd-Weron, K.; Sznajd, J. (2000). Opinion evolution in closed community. *Int. J. Mod. Phys. C* **11(6)**.
- [249] Szolnoki, A.; Perc, M.; Szabó, G.; Stark, H. (2009). Impact of aging on the evolution of cooperation in the spatial prisoner’s dilemma game. *Physical Review E* **80(2)**, 021901.
- [250] Tenner, E. (1997). *Why Things Bite Back: Technology and the Revenge of Unintended Consequences*. Vintage, 448 pp.
- [251] Tessone, C. J.; Sánchez, A.; Schweitzer, F. (2013). Diversity-induced resonance in the response to social norms. *Phys. Rev. E* **87**, 022803.
- [252] Tessone, C. J.; Toral, R.; Amengual, P.; Wio, H. S.; San Miguel, M. (2004). Neighborhood models of minority opinion spreading. *The European Physical Journal B* **39(4)**, 535–544.
- [253] Thurstone, L. (1994). A Law of Comparative Judgement. *Psychological Review* **101(2)**, 226–270.
- [254] Travers, J.; Milgram, S. (1969). An experimental study of the small world problem. *Sociometry* **32(4)**, 425–443.
- [255] Turner, M.; Pratkanis, A. (1998). Twenty-Five Years of Groupthink Theory and Research: Lessons from the Evaluation of a Theory. *Organizational Behavior and Human Decision Processes* **73(2/3)**, 105–15.
- [256] Turner, P. E.; Chao, L. (1999). Prisoner’s dilemma in an RNA virus. *Nature* **398(6726)**, 441–443.

- [257] Van Lange, P. A. M.; Joireman, J. A. (2008). How We Can Promote Behavior That Serves All of Us in the Future. *Social Issues and Policy Review* **2(1)**, 127–157.
- [258] Vazquez, F.; González-Avella, J. C.; Equiluz, V.; Miguel, M. S. (2009). Collective Phenomena in Complex Social Networks. In: V. In; P. Longhini; A. Palacios (eds.), *Applications of Nonlinear Dynamics Model and Design of Complex Systems*, Springer Verlag. pp. 189–199.
- [259] Vitali, S.; Glattfelder, J. B.; Battiston, S. (2011). The Network of Global Corporate Control. *PLoS ONE* **6(10)**, e25995.
- [260] Vul, E.; Pashler, H. (2008). Measuring the crowd within probabilistic representations within individuals. *Psychological Science* **19(7)**, 645–647.
- [261] Walter, F. E.; Battiston, S.; Schweitzer, F. (2008). A model of a trust-based recommendation system on a social network. *Autonomous Agents and Multi-Agent Systems* **16(1)**, 57–74.
- [262] Wasserman, S.; Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press, 825 pp.
- [263] Weng, L.; Ratkiewicz, J.; Perra, N.; Gonçalves, B.; Castillo, C.; Bonchi, F.; Schifanella, R.; Menczer, F.; Flammini, A. (2013). The Role of Information Diffusion in the Evolution of Social Networks. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '13, New York, NY, USA: ACM, pp. 356–364.
- [264] Werner, G. M.; Dyer, M. G. (1993). Evolution of herding behavior in artificial animals. *From Animals to Animats* **2**, 393–399.
- [265] Wey, T.; Blumstein, D. T.; Shen, W.; Jordán, F. (2008). Social network analysis of animal behaviour: a promising tool for the study of sociality. *Animal Behaviour* **75(2)**, 333–344.
- [266] Whitehead, H. (2008). *Analyzing animal societies: quantitative methods for vertebrate social analysis*. University of Chicago Press.
- [267] Wilkinson, G. S. (1992). Information transfer at evening bat colonies. *Animal Behaviour* **44, Part 3(0)**, 501–518.
- [268] Winkler, R. L.; Makridakis, S. (1983). The Combination of Forecasts. *Journal of the Royal Statistical Society. Series A (General)* **146(2)**, 150–157.

- [269] Wu, F.; Huberman, B. A.; Adamic, L. A.; Tyler, J. R. (2004). Information flow in social groups. *Physica A: Statistical Mechanics and its Applications* **337**(1–2), 327 – 335.
- [270] Wu, S.; Das Sarma, A.; Fabrikant, A.; Lattanzi, S.; Tomkins, A. (2013). Arrival and departure dynamics in social networks. In: *WSDM '13*.
- [271] Yang, J.; Leskovec, J. (2012). Defining and evaluating network communities based on ground-truth. In: *MDS '12*.
- [272] Yaniv, I.; Milyavsky, M. (2005). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes* **103**(1), 104–120.
- [273] Ye, S.; Wu, S. (2010). Measuring Message Propagation and Social Influence on Twitter.com. In: L. Bolc; M. Makowski; A. Wierzbicki (eds.), *Social Informatics*, Springer Berlin Heidelberg, vol. 6430 of *Lecture Notes in Computer Science*. pp. 216–231.
- [274] Youtube (2014). Facebook Fraud. <https://www.youtube.com/watch?v=oVfHeWTKjag>. Accessed: 2014-04-15.
- [275] Zimmermann, M. G.; Eguíluz, V. M.; San Miguel, M. (2004). Coevolution of dynamical states and interactions in dynamic networks. *Phys. Rev. E* **69**, 065102.
- [276] Zingerle, A. (1998). The Unanticipated Consequences of Action: Sociological and Ethical Aspects. In: C. Mongardini; S. Tabboni (eds.), *Robert K. Merton & Contemporary Sociology*, Transaction Publishers New Brunswick (USA) and London (UK). 1 edn., pp. 177–186.

Appendix

Chapter 9

9.1 Derivation of $\langle \ln x(t) \rangle$

In this appendix we show how to derive a closed-form solution of the collective error \mathcal{E} from Eq. 3.5. As argued in the main text, the dynamics of \mathcal{E} depend on the mean of the log-transformed estimates $\langle \ln x(t) \rangle$.

First, let us express $x_i(t)$ and $x_i(0)$ as the sum of their means plus a small deviation

$$\begin{aligned} x_i(t) &= \langle x(t) \rangle + \delta_i(t) & \langle \delta(t) \rangle &= 0, t \geq 0 \\ x_i(0) &= \langle x(0) \rangle + \sigma_i(0) & \langle \sigma(0) \rangle &= 0. \end{aligned} \quad (9.1)$$

Note that $x_i(0) \stackrel{d}{\neq} x_i(t)$, because $x_i(0)$ come from a log-normal distribution, but the subsequent $x_i(t)$ follow 3.12. In general, these would not be identically distributed¹, therefore require individual fluctuation terms δ and σ .

By plugging in Eq. 9.1 into the dynamics of $\langle \ln x(t) \rangle$ in Eq. 3.17, we obtain

$$\begin{aligned} \frac{d \langle \ln x(t) \rangle}{dt} &= \alpha \left\langle \underbrace{\frac{1}{1 + \frac{\delta(t)}{\langle x(t-1) \rangle}}}_{\text{Taylor expansion around 0}} \right\rangle - \alpha + \beta \left\langle \frac{x(0)}{x(t)} \right\rangle - \beta + \frac{D}{\sqrt{N}} \left\langle \frac{\xi(t)}{x(t)} \right\rangle = \\ &= \alpha \left\langle \sum_{n=0}^{\infty} \left(-\frac{\delta(t)}{\langle x(t-1) \rangle} \right)^n \right\rangle + \beta \left\langle \frac{\langle x(0) \rangle \left(1 + \frac{\sigma(0)}{\langle x(0) \rangle} \right)}{\langle x(t-1) \rangle \left(1 + \frac{\delta(t)}{\langle x(t-1) \rangle} \right)} \right\rangle - \alpha - \beta + \frac{D}{\sqrt{N}} \left\langle \frac{\xi(t)}{x(t)} \right\rangle \end{aligned}$$

¹A trivial exception is the no information regime ($\alpha = 0$). The ensemble average, as given by 3.14 and 3.15, would be approximately constant, in which case estimates would fluctuate around their starting values. As a result $x_i(t)$ would be equal in distributions for all $t \geq 0$

$$\begin{aligned}
&= \alpha \left\langle \sum_{n=0}^{\infty} \frac{(-1)^n}{\langle x(t-1) \rangle^n} \langle \delta(t)^n \rangle \right\rangle + \beta \frac{\langle x(0) \rangle}{\langle x(t-1) \rangle} \left\langle \left(1 + \frac{\sigma(0)}{\langle x(0) \rangle} \right) \cdot \underbrace{\left(1 + \frac{\delta(t)}{\langle x(t-1) \rangle} \right)^{-1}}_{\text{Taylor expansion around 0}} \right\rangle - \\
&\quad - \alpha - \beta + \frac{D}{\sqrt{N}} \left\langle \frac{\xi(t)}{x(t)} \right\rangle = \\
&= \alpha \sum_{n=0}^{\infty} \frac{(-1)^n}{\langle x(t-1) \rangle^n} \langle \delta(t)^n \rangle + \beta \frac{\langle x(0) \rangle}{\langle x(t-1) \rangle} \left\langle \left(1 + \frac{\sigma(0)}{\langle x(0) \rangle} \right) \cdot \sum_{n=0}^{\infty} \frac{(-1)^n}{\langle x(t-1) \rangle^n} \langle \delta(t)^n \rangle \right\rangle - \\
&\quad - \alpha - \beta + \frac{D}{\sqrt{N}} \left\langle \frac{\xi(t)}{x(t)} \right\rangle = \\
&= \alpha \sum_{n=0}^{\infty} \frac{(-1)^n}{\langle x(t-1) \rangle^n} \langle \delta(t)^n \rangle + \beta \frac{\langle x(0) \rangle}{\langle x(t-1) \rangle} \left[\sum_{n=0}^{\infty} \frac{(-1)^n}{\langle x(t-1) \rangle^n} \langle \delta(t)^n \rangle + \sum_{n=0}^{\infty} \frac{(-1)^n \langle \sigma(0) \delta(t)^n \rangle}{\langle x(0) \rangle \langle x(t-1) \rangle^n} \right] - \\
&\quad - \alpha - \beta + \frac{D}{\sqrt{N}} \left\langle \frac{\xi(t)}{x(t)} \right\rangle. \tag{9.2}
\end{aligned}$$

Appendix 9.2 shows that general form of $\langle \sigma(0)^m \delta(t)^n \rangle$ is given by

$$\langle \sigma(0)^m \delta(t)^n \rangle = \frac{\langle \sigma(0)^{m+n} \rangle}{(\alpha + \beta)^n} [\beta + \alpha e^{-(\alpha+\beta)t}]^n, \quad m, n \geq 0 \tag{9.3}$$

From here we can express $\langle \delta(t)^n \rangle$ and $\langle \sigma(0) \delta(t)^n \rangle$

$$\langle \delta(t)^n \rangle = \frac{\langle \sigma(0)^n \rangle}{(\alpha + \beta)^n} [\beta + \alpha e^{-(\alpha+\beta)t}]^n. \tag{9.4}$$

$$\langle \sigma(0) \delta(t)^n \rangle = \frac{\langle \sigma(0)^{n+1} \rangle}{(\alpha + \beta)^n} [\beta + \alpha e^{-(\alpha+\beta)t}]^n. \tag{9.5}$$

Hence

$$\begin{aligned}
\frac{d \langle \ln x(t) \rangle}{dt} &= (\alpha + \beta) \sum_{n=0}^{\infty} \frac{(-1)^n}{\langle x(0) \rangle^n} \frac{\langle \sigma(0)^n \rangle}{(\alpha + \beta)^n} [\beta + \alpha e^{-(\alpha+\beta)t}]^n + \\
&\quad + \beta \sum_{n=0}^{\infty} \frac{(-1)^n}{\langle x(0) \rangle^{n+1}} \frac{\langle \sigma(0)^{n+1} \rangle}{(\alpha + \beta)^n} [\beta + \alpha e^{-(\alpha+\beta)t}]^n - \\
&\quad - (\alpha + \beta) + \frac{D}{\sqrt{N}} \frac{\langle \xi(t) \rangle}{\langle x(t) \rangle}, \tag{9.6}
\end{aligned}$$

where we have used that $\langle x(t) \rangle = \langle x(0) \rangle$ for large t . By transforming the second term, we obtain

$$\begin{aligned} \frac{d \langle \ln x(t) \rangle}{dt} &= (\alpha + \beta) \sum_{n=0}^{\infty} \frac{(-1)^n \langle \sigma(0)^n \rangle}{\langle x(0) \rangle^n (\alpha + \beta)^n} [\beta + \alpha e^{-(\alpha+\beta)t}]^n - \\ &\quad - \frac{\beta(\alpha + \beta)}{\beta + \alpha e^{-(\alpha+\beta)t}} \sum_{n=0}^{\infty} \frac{(-1)^{n+1} \langle \sigma(0)^{n+1} \rangle}{\langle x(0) \rangle^{n+1} (\alpha + \beta)^{n+1}} [\beta + \alpha e^{-(\alpha+\beta)t}]^{n+1} - \\ &\quad - (\alpha + \beta) + \frac{D}{\sqrt{N}} \frac{\langle \xi(t) \rangle}{\langle x(t) \rangle}. \end{aligned}$$

Note that $\langle \sigma(0) \rangle = 0$, because by definition $\langle \sigma(0) \rangle$ is the first central moment of the initial distribution $x(0)$. We use this to re-index the sums above

$$\begin{aligned} \frac{d \langle \ln x(t) \rangle}{dt} &= (\alpha + \beta) \left[1 + \sum_{n=1}^{\infty} \frac{(-1)^n \langle \sigma(0)^n \rangle}{\langle x(0) \rangle^n (\alpha + \beta)^n} [\beta + \alpha e^{-(\alpha+\beta)t}]^n \right] - \\ &\quad - \frac{\beta(\alpha + \beta)}{\beta + \alpha e^{-(\alpha+\beta)t}} \sum_{n=1}^{\infty} \frac{(-1)^n \langle \sigma(0)^n \rangle}{\langle x(0) \rangle^n (\alpha + \beta)^n} [\beta + \alpha e^{-(\alpha+\beta)t}]^n - \\ &\quad - (\alpha + \beta) + \frac{D}{\sqrt{N}} \frac{\langle \xi(t) \rangle}{\langle x(t) \rangle}. \end{aligned} \tag{9.7}$$

Therefore the dynamics can finally be written as

$$\frac{d \langle \ln x(t) \rangle}{dt} = (\alpha + \beta) \left(1 - \frac{\beta}{\beta + \alpha e^{-(\alpha+\beta)t}} \right) \sum_{n=1}^{\infty} \frac{(-1)^n \langle \sigma(0)^n \rangle}{\langle x(0) \rangle^n (\alpha + \beta)^n} [\beta + \alpha e^{-(\alpha+\beta)t}]^n + \frac{D}{\sqrt{N}} \frac{\langle \xi(t) \rangle}{\langle x(0) \rangle}. \tag{9.8}$$

The solution is given by integrating the time-dependent exponential term

$$\frac{1}{\alpha + \beta} \int \frac{d \langle \ln x(t) \rangle}{dt} dt = \sum_{n=1}^{\infty} \frac{(-1)^n \langle \sigma(0)^n \rangle}{\langle x(0) \rangle^n (\alpha + \beta)^n} \underbrace{\int [\beta + \alpha e^{-(\alpha+\beta)t}]^n dt}_{:= \mathcal{A}(n,t)}$$

$$\begin{aligned}
& -\beta \sum_{n=1}^{\infty} \frac{(-1)^n \langle \sigma(0)^n \rangle}{\langle x(0) \rangle^n (\alpha + \beta)^n} \int [\beta + \alpha e^{-(\alpha+\beta)t}]^{n-1} dt + \\
& + \frac{D}{\langle x(0) \rangle \sqrt{N}} \int_0^t e^{\beta(s-t)} \langle \xi(s) \rangle ds.
\end{aligned}$$

$\mathcal{A}(n, t)$ can be expanded with the Binomial theorem

$$\begin{aligned}
\mathcal{A}(n, t) &= \beta^n t + C + \sum_{k=1}^n \binom{n}{k} \beta^{n-k} \alpha^k \int e^{-(\alpha+\beta)tk} dt = \\
&= \beta^n t + C - \sum_{k=1}^n \binom{n}{k} \beta^{n-k} \alpha^k \left[\frac{1}{(\alpha + \beta)k} e^{-(\alpha+\beta)kt} - C \right] = \\
&= \beta^n t + C [1 + (\alpha + \beta)^n - \beta^n] - \frac{1}{\alpha + \beta} \sum_{k=1}^n \frac{1}{k} \binom{n}{k} \beta^{n-k} \alpha^k e^{-(\alpha+\beta)kt}. \quad (9.9)
\end{aligned}$$

Now we can write $\mathcal{B}(n, t) := \mathcal{A}(n, t) - \beta \mathcal{A}(n-1, t)$ and express $\langle \ln x(t) \rangle$ as

$$\langle \ln x(t) \rangle = \sum_{n=1}^{\infty} \frac{(-1)^n \langle \sigma(0)^n \rangle}{\langle x(0) \rangle^n (\alpha + \beta)^{n-1}} \mathcal{B}(n, t) + \frac{D(\alpha + \beta)}{\langle x(0) \rangle \sqrt{N}} \int_0^t e^{\beta(s-t)} \langle \xi(s) \rangle ds. \quad (9.10)$$

$\mathcal{B}(n, t)$ equals

$$\begin{aligned}
\mathcal{B}(n, t) &= \beta^n t + C [1 + (\alpha + \beta)^n - \beta^n] - \frac{1}{\alpha + \beta} \sum_{k=1}^n \frac{1}{k} \binom{n}{k} \beta^{n-k} \alpha^k e^{-(\alpha+\beta)kt} - \\
&- \beta^n t - \beta C [1 + (\alpha + \beta)^{n-1} - \beta^{n-1}] + \frac{\beta}{\alpha + \beta} \sum_{k=1}^{n-1} \frac{1}{k} \binom{n-1}{k} \beta^{n-1-k} \alpha^k e^{-(\alpha+\beta)kt}. \quad (9.11)
\end{aligned}$$

Using the property of the binomial coefficient that $\binom{n}{k} = \binom{n}{k-1} + \binom{n-1}{k-1}$ we write further

$$\mathcal{B}(n, t) = C [1 - \beta + \alpha(\alpha + \beta)^{n-1}] - \frac{\alpha^n}{n(\alpha + \beta)} e^{-(\alpha+\beta)nt} - \frac{1}{\alpha + \beta} \sum_{k=1}^{n-1} \frac{1}{k} \binom{n-1}{k-1} \beta^{n-k} \alpha^k e^{-(\alpha+\beta)kt}. \quad (9.12)$$

At $t = 0$, $\mathcal{B}(n, t)$ is given by

$$\mathcal{B}(n, 0) = C[1 - \beta + \alpha(\alpha + \beta)^{n-1}] - \frac{\alpha^n}{n(\alpha + \beta)} - \frac{1}{\alpha + \beta} \sum_{k=1}^{n-1} \frac{1}{k} \binom{n-1}{k-1} \beta^{n-k} \alpha^k. \quad (9.13)$$

The first term involves the integration constant, C , and can be determined from the initial condition $\langle x(0) \rangle$

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{(-1)^n}{\langle x(0) \rangle^n} \frac{\langle \sigma(0)^n \rangle}{(\alpha + \beta)^{n-1}} \left(C[1 - \beta + \alpha(\alpha + \beta)^{n-1}] \right) &= \langle \ln x(0) \rangle + \\ &+ \sum_{n=1}^{\infty} \frac{(-1)^n}{\langle x(0) \rangle^n} \frac{\alpha^n \langle \sigma(0)^n \rangle}{n(\alpha + \beta)^n} + \\ &+ \sum_{n=1}^{\infty} \frac{(-1)^n}{\langle x(0) \rangle^n} \frac{\langle \sigma(0)^n \rangle}{(\alpha + \beta)^n} \sum_{k=1}^{n-1} \frac{1}{k} \binom{n-1}{k-1} \beta^{n-k} \alpha^k, \end{aligned} \quad (9.14)$$

where we have also zeroed the stochastic term, which is valid for large N . Hence finally

$$\begin{aligned} \langle \ln x(t) \rangle &= \langle \ln x(0) \rangle + \\ &+ \sum_{n=1}^{\infty} \frac{(-1)^n}{\langle x(0) \rangle^n} \frac{\alpha^n \langle \sigma(0)^n \rangle}{n(\alpha + \beta)^n} + \\ &+ \sum_{n=1}^{\infty} \frac{(-1)^n}{\langle x(0) \rangle^n} \frac{\langle \sigma(0)^n \rangle}{(\alpha + \beta)^n} \sum_{k=1}^{n-1} \frac{1}{k} \binom{n-1}{k-1} \beta^{n-k} \alpha^k - \\ &- \sum_{n=1}^{\infty} \frac{(-1)^n}{\langle x(0) \rangle^n} \frac{\alpha^n \langle \sigma(0)^n \rangle}{n(\alpha + \beta)^n} e^{-(\alpha+\beta)nt} + \\ &- \sum_{n=1}^{\infty} \frac{(-1)^n}{\langle x(0) \rangle^n} \frac{\langle \sigma(0)^n \rangle}{(\alpha + \beta)^n} \sum_{k=1}^{n-1} \frac{1}{k} \binom{n-1}{k-1} \beta^{n-k} \alpha^k e^{-(\alpha+\beta)kt}, \end{aligned} \quad (9.15)$$

or more compactly

$$\begin{aligned}
\langle \ln x(t) \rangle &= \langle \ln x(0) \rangle + \\
&+ \sum_{n=1}^{\infty} \frac{(-1)^n \alpha^n \langle \sigma(0)^n \rangle}{\langle x(0) \rangle^n n(\alpha + \beta)^n} (1 - e^{-(\alpha+\beta)nt}) + \\
&+ \sum_{n=1}^{\infty} \frac{(-1)^n \langle \sigma(0)^n \rangle}{\langle x(0) \rangle^n (\alpha + \beta)^n} \sum_{k=1}^{n-1} \frac{1}{k} \binom{n-1}{k-1} \beta^{n-k} \alpha^k (1 - e^{-(\alpha+\beta)kt}).
\end{aligned}$$

Simplifying the last term yields

$$\begin{aligned}
\langle \ln x(t) \rangle &= \langle \ln x(0) \rangle + \\
&+ \sum_{n=1}^{\infty} \frac{(-1)^n \alpha^n \langle \sigma(0)^n \rangle}{\langle x(0) \rangle^n n(\alpha + \beta)^n} (1 - e^{-(\alpha+\beta)nt}) + \\
&+ \sum_{n=1}^{\infty} \frac{(-1)^n \langle \sigma(0)^n \rangle}{\langle x(0) \rangle^n n(\alpha + \beta)^n} [\alpha^n (e^{-(\alpha+\beta)nt} - 1) + (\alpha + \beta)^n - (\beta + \alpha e^{-(\alpha+\beta)t})^n],
\end{aligned}$$

hence

$$\begin{aligned}
\langle \ln x(t) \rangle &= \langle \ln x(0) \rangle + \\
&+ \sum_{n=1}^{\infty} \frac{(-1)^n \alpha^n \langle \sigma(0)^n \rangle}{\langle x(0) \rangle^n n(\alpha + \beta)^n} [(1 - \alpha^n)(1 - e^{-(\alpha+\beta)nt}) + (\alpha + \beta)^n - (\beta + \alpha e^{-(\alpha+\beta)t})^n].
\end{aligned} \tag{9.16}$$

9.2 Derivation of $\langle \sigma(0)^m \delta(t)^n \rangle$

First, we simplify notation by letting

$$\langle \sigma(0)^m \delta(t)^n \rangle = A_{m,n}(t), \quad \text{for } m, n, t \geq 0$$

Note that $A_{m,n}(0) = \langle \sigma(0)^{m+n} \rangle$ since $\delta(0) = \sigma(0)$ by definition.

The dynamics of $A_{m,n}(t)$ can be expressed as

$$\frac{d}{dt}A_{m,n}(t) = \left\langle \sigma(0)^m n \delta(t)^{n-1} \frac{d}{dt} \delta(t) \right\rangle = \left\langle \sigma(0)^m n \delta(t)^{n-1} \frac{d}{dt} [x(t) - \langle x(t) \rangle] \right\rangle.$$

Plugging in Eqs. 3.12 and 3.14, and ignoring the noise terms yields

$$\begin{aligned} \frac{d}{dt}A_{m,n}(t) &= \left\langle n\sigma(0)^m \delta(t)^{n-1} \left[\alpha(\langle x(t) \rangle - x(t)) + \beta(x(0) - \langle x(0) \rangle) + \langle x(t) \rangle - \langle x(0) \rangle \right] \right\rangle = \\ &= \left\langle n\sigma(0)^m \delta(t)^{n-1} \left[-\alpha\delta(t) + \beta\sigma(0) - \beta\delta(t) \right] \right\rangle = \\ &= -n(\alpha + \beta) \langle \sigma(0)^m \delta(t)^n \rangle + n\beta \langle \sigma(0)^{m+1} \delta(t)^{n-1} \rangle = \\ &= -n(\alpha + \beta)A_{m,n}(t) + n\beta A_{m+1,n-1}(t). \end{aligned} \quad (9.17)$$

Further, assume that the general form of $A_{m,n}(t)$ is

$$C_{m,n} \cdot f(t)^n, \quad (9.18)$$

where $C_{m,n}$ are coefficients that depend only on m and n , and $f(t)$ is an arbitrary function that depends only on t . We now plug this ansatz in Eq. 9.17

$$nC_{m,n}f(t)^{n-1} \frac{d}{dt}f(t) = -n(\alpha + \beta)C_{m,n}f(t)^n + n\beta C_{m+1,n-1}f(t)^{n-1}.$$

Dividing both sides by $f(t)^{n-1}$ yields a non-homogeneous first-order differential equation for $f(t)$

$$C_{m,n} \frac{d}{dt}f(t) = -(\alpha + \beta)C_{m,n}f(t) + \beta C_{m+1,n-1}.$$

The general solution is the sum of the solution to the homogeneous equation and a particular solution to the non-homogeneous. The latter, in turn, can be obtained by the method of undetermined coefficients to finally yield

$$f(t) = Ke^{-(\alpha+\beta)t} + \frac{\beta}{\alpha + \beta} \cdot \frac{C_{m+1,n-1}}{C_{m,n}}.$$

The constant K is given by $A_{m,n}(0) = \langle \sigma(0)^{m+n} \rangle$, therefore

$$f(t) = \left[\frac{\langle \sigma(0)^{m+n} \rangle}{C_{m,n}} \right]^{1/n} \cdot e^{-(\alpha+\beta)t} + \frac{\beta}{\alpha + \beta} \cdot \frac{C_{m+1,n-1}}{C_{m,n}} \left[1 - e^{-(\alpha+\beta)t} \right].$$

As $f(t)$ depends only on t , we require the terms involving n and m to reduce to constants, i.e.

$$\left[\frac{\langle \sigma(0)^{m+n} \rangle}{C_{m,n}} \right]^{1/n} = B,$$

therefore

$$C_{m,n} = \frac{\langle \sigma(0)^{m+n} \rangle}{B^n}, \quad (9.19)$$

for some constant B . It easily follows that

$$\frac{C_{m+1,n-1}}{C_{m,n}} = B,$$

hence, $f(t)$ is given by

$$f(t) = B \left[\frac{\alpha}{\alpha + \beta} e^{-(\alpha+\beta)t} + \frac{\beta}{\alpha + \beta} \right]. \quad (9.20)$$

Finally by combining Eqs. 9.19 and 9.20 into Eq. 9.18, B cancels out and we are left with the solution

$$A_{m,n}(t) = \frac{\langle \sigma(0)^{m+n} \rangle}{(\alpha + \beta)^n} \left[\alpha e^{-(\alpha+\beta)t} + \beta \right]^n. \quad (9.21)$$

Chapter 10

10.1 Results

In agreement with our aggregate analysis in the main text, Figure 10.1 illustrates that overall all ranks betray considerable deviation in the first round. This indicates that individuals naturally do not trust their randomly assigned initial positions. However, immediately afterwards, lower-ranks start to consistently reduce their exploration range. This is not true for those ranked 16 and higher, whose frustration of being among the last widens their exploration range to span virtually the whole circle in all ten rounds.

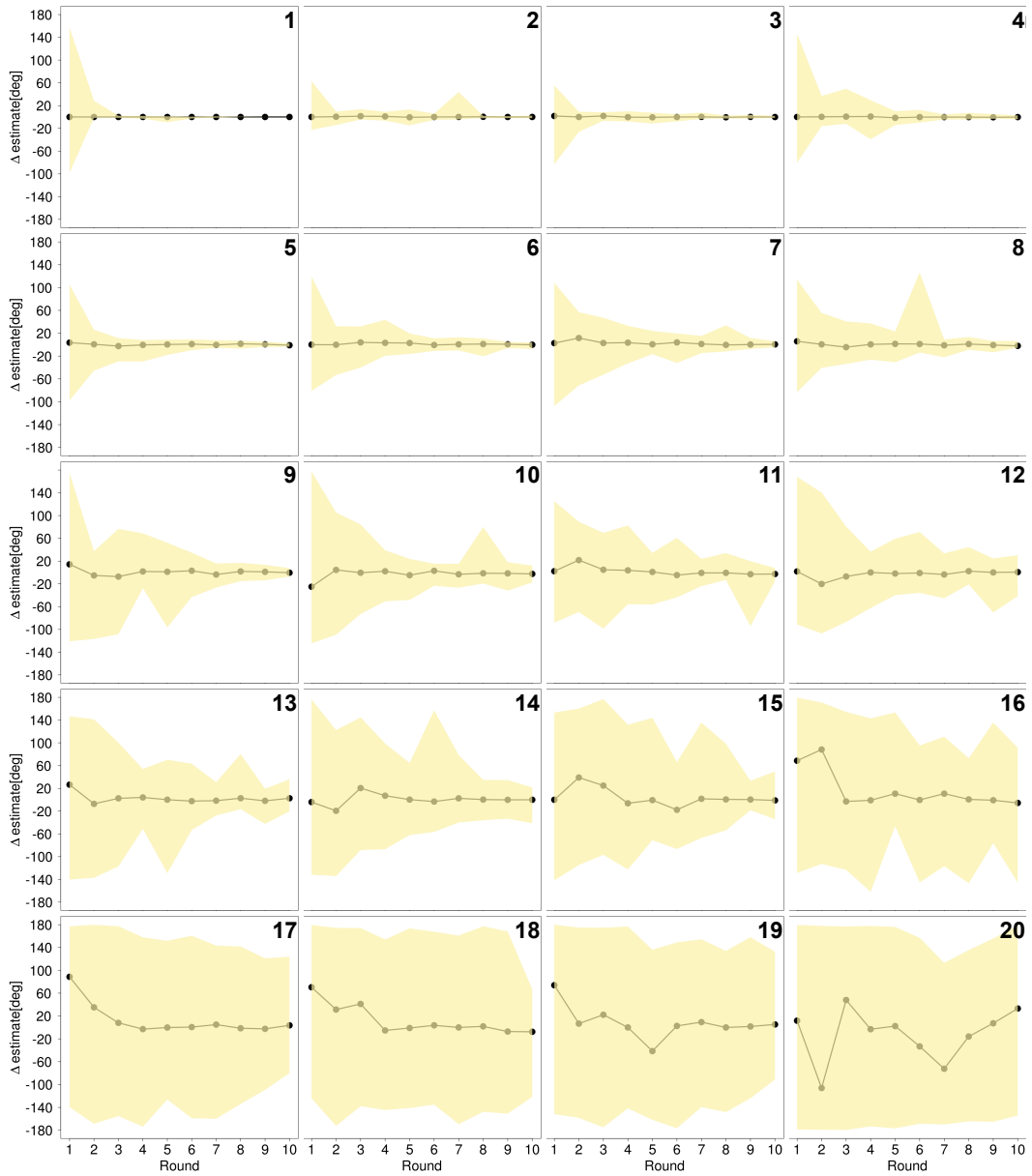


Figure 10.1: Estimates deviation during the progression of a stage, dependent on rank. Black point symbols indicate the median estimate deviation for a particular rank in a given round. Coloured bands represent the inter-quartile range. Ranks are denoted in the top-right corner of each plot.

Chapter 11

11.1 Media Coverage

The study was first uploaded onto the arXiv platform¹ on February 22nd, 2013. It was quickly picked up by MIT Technology Review in their *Emerging Technology From arXiv* section, and subsequently by Wired Enterprise. After these popular science and technology websites reported on our work, a long wave of interest followed by a variety of social media – Slate.com, The Connectivist, Herald Sun and GIZMODO – to name a few. An exhaustive list can be found in the media section of the Chair of Systems Design website². Toward the end of September 2013, the television network 3sat featured our work in their science series *Nano*³, which was an opportunity to provide our direct account of the relevance and lessons of the study.

It has to be noted that many of the media reports saw our work as an opportunity to attract viewership by speculating about the fate of Facebook. A case in point is the title of the Slate.com article: *How Will Facebook Die?*. As we explained in the *Nano* series, we propose a mechanism for OSNs collapse, based on propagation of exiting behaviour. This mechanism does suggest some measures that OSNs administrators can take to ensure that external shocks propagate as little as possible; examples are encouraging members to connect to other well-connected members, or implementing changes gradually in a way that does not disturb users with high coreness. However, we cannot predict how and when user exodus will be triggered, especially when the complete network structure is generally unavailable for research? Though we propose that technological change is a likely trigger, other factors, such as availability of substitute OSNs, play a role too. In fact, an OSN might disappear through other mechanisms as well. For example, in 2012 Instagram – a popular photo-sharing mobile application – faced a threat of a massive and *coordinated*

¹An electronic archive of scientific preprints, see <http://www.arxiv.org/>

²<http://www.sg.ethz.ch/research/response-media/>

³Online at: <http://www.3sat.de/page/?source=/nano/technik/172240/index.html>

user boycott after changes in their privacy policy allowed Instagram owners to freely sell users' photos [120]. Such abuse of user data or privacy violations could destroy a community, by driving away a sizable part of it at once.

Below, we provide an example of two notable media articles.

11.1.1 Researchers conduct “autopsy” of social network Friendster

A team of systems designers conducted an “autopsy” of social networking site Friendster by analysing several online communities.

Researchers at the Swiss Federal Institute of Technology examined Friendster, Livejournal, Facebook, Orkut and Myspace in a bid to identify what makes a social network survive or decline, and what makes them capable of withstanding changes.

As the paper on the topic explains, “changes may cause users to leave, which may trigger further leaves of others who lost connection to their friends. This may lead to cascades of users leaving.”

Friendster was founded in 2002 and at its peak had more than 100 million users. In 2009, having undergone a comprehensive redesign and suffered some technical problems, the site started haemorrhaging users, and was eventually closed down in 2011 and reopened as an online gaming portal.

The study seeks to find out what went wrong and takes the form of a “digital autopsy” on Friendster using data collected at the time by the Internet Archive.

The team – led by David Garcia and co-authored by Pavlin Mavrodiev and Frank Schweitzer – defines the social resilience of an online community as “the ability of the community to withstand external stresses and disturbances as a result of environmental changes”, particularly the user interface of the social network.

They found that when the time and effort (the costs) associated with being a member of a social network outweigh the benefits, then a decline in users becomes likely. If one person leaves, their friends become more likely to leave and as more people leave, this can lead to a cascading collapse in membership.

Each network has some resistance to this decline, depending on how many friends each users has. If a user has a thousand friends, they will hardly notice when a couple leave. But if a user has three friends and one leaves then they are much more likely to leave themselves.

So if a large proportion of people who use the network have a small number of friends, it can be highly vulnerable to mass exodus.

The team used “k-core analysis” to identify the fraction of the network in which all users have at least a certain number (k) of friends. This fraction – the k-core distribution – was analysed for each of the five aforementioned networks.

Being vulnerable to mass exodus does not mean a network automatically fails. In order for that to happen, the cost-to-benefit ration must drop to a point where individual members choose to leave. So the combination of a vulnerable k-core and a low cost-to-benefit ratio create a recipe for disaster.

Just before Friendster collapsed, the cost-to-benefit ratio fell significantly due to the changes to the user-interface combined with technical issues. “This measure can be seen as a precursor of the later collapse of the community,” says the study. This was the ultimate cause of death.

Interestingly, the study found that the topologies of Livejournal and Facebook are less resilient than the unsuccessful networks Friendster and Orkut. “This indicates that the environmental condition of an online social network plays a major role in its success. Thus, we conclude that the topology of the social network alone cannot explain the stories.”

The report flags up a comedy video made by The Onion, which sees fictitious “internet archaeologists” commenting on the decay of Friendster.

“While proposed as a satire of the speed of internet culture, this video illustrates the opportunities that a failed social network offers for research. The users of such a community leave traces that allow us to investigate its failure. In this sense, we can name our work as Internet Archaeology, because we analyse non-written traces of a disappeared society, aiming at understanding the way it worked and the reasons for its demise.”

By Liar Clark, Wired UK, 27 February 2013⁴

11.1.2 The Friendster Autopsy: How a Social Network Dies

What kills a social network? A group of internet archeologists have picked over the digital bones of Friendster - the pioneering social networking site that drowned in Facebook’s wake - and we now have a clearer picture of its epic collapse.

Friendster was once the hottest thing in social networking. Google wanted to buy it for \$30 million back in 2003, but - burdened by technical glitches and a more nimble competitor in

⁴Online at: <http://www.wired.co.uk/news/archive/2013-02/27/autopsy-of-friendster>

Facebook - it was pretty much dead in the U.S. by 2006. That said, it trudged along for a few more years, helped by a relatively strong following in southeast Asia. Then, around 2009, a site redesign crushed it.

It ended up being a kind of “controlled demolition,” with weakly connected chains of friends quickly disintegrating, says David Garcia, a professor with the Swiss Federal Institute of Technology and one of the authors of a recent paper analyzing Friendster’s demise.

Just before Friendster relaunched itself as a gaming site in 2011, the Internet Archive crawled the dead network, grabbing a snapshot. Garcia and his fellow researchers used that snapshot of that controlled demolition as the basis for their research, which they describe as both a work of internet archeology and an autopsy.

What they found was that by 2009, Friendster still had tens of millions of users, but the bonds linking the network were not particularly strong. Many of the users were not connected to a lot of other members, and the people they had befriended came with just a handful of their own connections. So they ended up being so loosely affiliated with the network, that the burden of dealing with a new user interface just was not worth it.

“First the users in the outer cores start to leave, lowering the benefits of inner cores, cascading through the network towards the core users, and thus unraveling,” Garcia told us during an online chat.

The researchers describe heart of successful networks in terms of what that they call K-cores. These are subset of users who not only have a lot of friends, but they have “resilience and social influence,” Garcia says. As these K-cores disintegrated, the whole Friendster thing fell apart.

If there is a lesson to be learned from the data, it is that it takes more than a lot of users to build a viable social network. They need to have strong connections too. So Facebook should be looking at the types of connections it users have and encourage them to connect to other strongly connected users, Garcia says. In other words, strong networks are made up of strongly-linked people, not of stragglers.

By Robert Mcmillan, Wired Enterprise, 27 February 2013⁵

⁵Online at: <http://www.wired.com/2013/02/friendster-autopsy/>

Chapter 12

12.1 Absolute and relative reputation

In this Appendix, we show that the equilibrium *absolute* reputations given by a stationary solution to Eq. 6.2 are a scaled representation of the equilibrium *relative* reputations \mathbf{x} and \mathbf{b} .

Let us rewrite Eq. 6.2 in matrix form

$$\frac{d\mathbf{X}(t)}{dt} = \mathbf{A}\mathbf{X}(t) - \phi\mathbf{X}(t). \quad (12.1)$$

where \mathbf{I} is the identity matrix. It is clear that an equilibrium exists only if ϕ is an eigenvalue of \mathbf{A} . In this case, the corresponding eigenvector gives the equilibrium absolute reputation. Additionally, as ϕ represents a decay rate, and $X_i(t) \geq 0$ for all i , we require that the matrix \mathbf{A} has at least one positive eigenvalue, and that its corresponding eigenvector does not have negative components. As \mathbf{A} is a non-negative real matrix, we obtain these conditions from the Perron-Frobenius theorem (see main text). However, we are still bound by a limited set of values for the decay rate.

To avoid this last restriction, a common transformation is to normalise the absolute reputation, $X_i(t)$, with respect to the total reputation in the system. By doing so, the free parameter ϕ disappears, provided it is the same for all users. More importantly, using such relative reputation is more plausible in an OSN where individuals tend to evaluate their benefits with respect to others, rather than to an arbitrary absolute scale.

Let $x_i(t) = X_i(t) / \sum_j X_j(t)$. Then we can express the dynamics of $x_i(t)$ as follows

$$\begin{aligned}
\frac{dx_i(t)}{dt} &= \frac{\frac{dX_i(t)}{dt} \left(\sum_j X_j(t) \right)}{\left(\sum_j X_j(t) \right)^2} - \frac{X_i(t) \frac{d \left(\sum_j X_j(t) \right)}{dt}}{\left(\sum_j X_j(t) \right)^2} = \\
&= \frac{dX_i(t)/dt}{\sum_j X_j(t)} - x_i(t) \frac{d \sum_j X_j(t)/dt}{\sum_j X_j(t)} = \\
&= \frac{\sum_j a_{ij} X_j(t) - \phi X_i(t)}{\sum_j X_j(t)} - x_i(t) \frac{\sum_j \sum_k a_{jk} X_k(t) - \phi X_j(t)}{\sum_j X_j(t)} = \\
&= \sum_j a_{ij} x_j(t) - \phi x_i(t) + \phi x_i(t) - x_i(t) \sum_j \sum_k a_{jk} x_k(t) = \\
&= \sum_j a_{ij} x_j(t) - x_i(t) \sum_j \sum_k a_{jk} x_k(t). \tag{12.2}
\end{aligned}$$

In matrix form Eq. 12.2 becomes

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) - \mathbf{x}(t) \sum_j \sum_k a_{jk} x_k(t). \tag{12.3}$$

The equilibrium solutions to Eq. 12.3 are given by the eigenvectors of \mathbf{A} . If \mathbf{X}^λ is one such eigenvector with a corresponding eigenvalue λ , then the rescaled vector $\mathbf{x}^\lambda = \mathbf{X}^\lambda / \sum_i X_i^\lambda$ is also an eigenvector, and the constant, $\sum_j \sum_k a_{jk} x_k(t)$, equals λ .

However, defining relative reputation in this way, presumes that at least the total reputation in the system is common knowledge. This is unlikely to be the case in any OSN. For this reason, we normalise $X_i(t)$ with respect to the individual with maximum absolute reputation $X_{\max}(t)$.

Let $b_i(t) = X_i(t)/X_{\max}(t)$. The dynamics of $b_i(t)$ becomes

$$\begin{aligned}
\frac{db_i(t)}{dt} &= \frac{\frac{dX_i(t)}{dt} X_{\max}(t)}{X_{\max}(t)^2} - \frac{X_i(t) \frac{dX_{\max}(t)}{dt}}{X_{\max}(t)^2} = \\
&= \frac{dX_i(t)}{dt} \frac{1}{X_{\max}(t)} - b_i(t) \frac{1}{X_{\max}(t)} \frac{dX_{\max}(t)}{dt} =
\end{aligned}$$

$$\begin{aligned}
&= \frac{\sum_j a_{ij} X_j(t)}{X_{\max}(t)} - \phi \frac{X_i(t)}{X_{\max}(t)} - b_i(t) \frac{1}{X_{\max}(t)} \left(\sum_j a_{mj} X_j(t) - \phi X_{\max}(t) \right) \\
&= \sum_j a_{ij} b_j(t) - \phi b_i(t) + \phi b_i(t) - b_i(t) \sum_j a_{mj} b_j(t) = \\
&= \sum_j a_{ij} b_j(t) - b_i(t) \sum_j a_{mj} b_j(t), \tag{12.4}
\end{aligned}$$

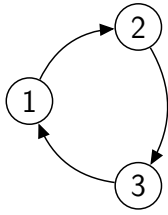
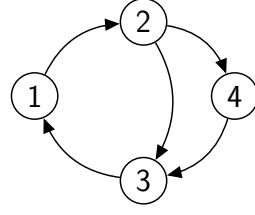
where m is the index of the individual with the highest relative reputation $x_{\max}(t)$.

In a similar manner, a rescaled eigenvector of \mathbf{A} , $\mathbf{b}^\lambda = \mathbf{X}^\lambda / X_{\max}^\lambda$, is an equilibrium solution to Eq. 12.4. In this case the constant, $\sum_j a_{mj} b_j(t)$ again equals λ .

To summarise, an eigenvector \mathbf{X}^λ of \mathbf{A} is a solution to both Eqs. 12.3 and 12.4. This shows that the two ways of defining relative information are in fact equivalent, up to a scaling factor.

12.2 Impact of cycles on λ_1

To further illustrate the impact of the network structure, in particular the number and the length of cycles, on the largest eigenvalue of the adjacency matrix, we present some didactical examples in Table 12.1.

Adjacency matrix	Corresponding network
$ \mathbf{A} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} $ $ \lambda_1 = 1 $ $ \mathbf{b}^{\lambda_1} = (1, 1, 1) $	
$ \mathbf{A} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} $ $ \lambda_1 = 1.22 $ $ \mathbf{b}^{\lambda_1} = (0.819, 0.671, 1, 0.55) $	

$A = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$	
$\lambda_1=1.36904$ $\mathbf{b}^{\lambda_1}=(0.73, 0.533, 1, 0.835, 0.61)$	
$A = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$	
$\lambda_1=1.194$ $\mathbf{b}^{\lambda_1}=(0.838, 0.702, 1, 0.588, 0.492)$	
continues on the next page	
Table 12.1 – continued from the previous page	
$A = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$	
$\lambda_1=1.325$ $\mathbf{b}^{\lambda_1}=(0.755, 0.57, 1, 0.755, 0.57, 0.43)$	

Table 12.1: Examples of how the network structure affects the adjacency matrix

The example in the first row has the largest eigenvalue of $\lambda_1 = 1$, and the corresponding eigenvector gives the same relative reputation for every user. A possible next step in the evolution of the core structure is shown in the next row, where a new user 4 joins the network. User 2 follows this newcomer, who in turn links itself back to user 3. The core is now composed of two cycles of length 3 and 4. As a consequence, λ_1 jumps up from 1 to a new value 1.22. Hence, the addition of a cycle to the core increases λ_1 . With this, the reputation of the users in the core is no longer homogeneous. Because of his two

followers, user 3 has the highest relative reputation, whereas user 4, being the last in the cycle starting with user 3 has the lowest reputation.

If a new user 5 joins the core structure, adding yet another cycle, as shown in the third row, λ_1 increase further, and the reputation vector is re-shuffled. User 3 is still dominant, but now user 4 is able to boost his own reputation thanks to user 5. This leaves user 2 as the loser in this new configuration.

To illustrate the effect of cycle sizes on the core structure, imagine that user 5 joined the network in a different way. Instead of creating a short cycle with user 4, he might have extended an existing one, as shown in the fourth row. The core now consists of a two cycles, $1 \rightarrow 2 \rightarrow 3$, and $1 \rightarrow 2 \rightarrow 4 \rightarrow 5 \rightarrow 3$. Yet, user 5 now *reduces* the value of λ_1 to 1.194, down from 1.22 before. User 3 is still dominant, but user 5, being the last in the cycle starting at user 3, has the lowest population. This core structure would be unfeasible in a system with a cost level $\tau=0.5$, since user 5 would decide to leave.

Let us go back to the sustainable core structure in row 3. Imagine that the small $4 \rightarrow 5$ cycle expanded after a new user 6 joined as a follower of user 4, and being followed by user 5 (last row). Again, the effect of increasing the length of an existing cycle was to reduce λ_1 .

To summarize these observations we conclude that adding a cycle to an existing core, while keeping all else equal, increases the value of λ_1 . On the other hand, if the number of cycles is kept constant, then *increasing* the length of a cycle *decreases* the value of λ_1 .

12.3 Cascades of users leaving the OSN

Table 12.2 shows consecutive steps without cost ($\tau_1=0$, left column) and with optimal cost ($\tau_2=0.2$, right column) to illustrate how cascades of users leaving the OSN emerge. Note that the STEPS \mathcal{T}_1 to *mathcal{T}_4* refer to points in the simulation when a certain structure of interest has emerged. In this sense, STEP is used didactically. For example, STEP \mathcal{T}_1 refers to a network time step (see Section 6.2.3) when a core-periphery structure emerges with the same reputation of the core and peripheral node. The *actual* network time steps, T , at which the displayed structures occur, are different for both cost levels and are not displayed.

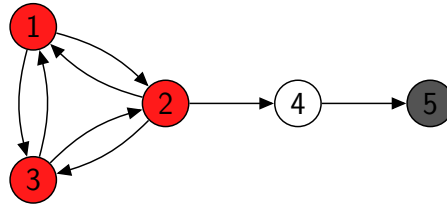
We start with an “initial” configuration at STEP \mathcal{T}_0 of Table 12.2, that has emerged after a certain time measured in network time steps. It has a core-periphery structure (users 1, 2, 3 belong to the core coloured in red, and users 4, 5 to the periphery). Users coloured in grey are posed to leave the OSN at the given time step. In the initial configuration user

5 has the lowest relative reputation of 0.24 and will leave the network, because Section 6.2.3 imposes extremal dynamics when $b_i > \tau$, for all i . In the next network time step, it will be replaced by another user 5 with a different connections to the network. Because of the tightly connected core, most dropout events during subsequent time steps (not shown) will likely affect only the periphery, replacing users 4 and 5. However, due to the random rewiring of m incoming and outgoing links for each new user, configurations like the ones shown at a later STEP \mathcal{T}_1 of Table 12.2 can appear. In fact, the consecutive steps \mathcal{T}_1 - \mathcal{T}_4 are drawn from real simulations of this small OSN in Appendix 12.4.

Interestingly, in STEP \mathcal{T}_1 the peripheral users have managed to attract a number of incoming links. This may not happen immediately at their entry, but we assume that, within one network time step T , after a short relaxation time users become known to at least part of the network and then are able to receive incoming links. Because of this link structure, the peripheral users can receive the same relative reputation as the core users. I.e. all users become susceptible to dropout in the same manner. In the example, user 1 (left column) and user 2 (right column) will leave because, among users with the equal reputation, one is chosen at random. Their dropout leads to a smaller core and, consequently, to a lower relative reputation of core users which becomes even lower than the one of peripheral users. This in turn motivates more core users to leave, as shown at STEP \mathcal{T}_2 . In the right column, user 1, which was a follower of the previous user 2, leaves now. In the left column, a new user 1 has entered the OSN, but has failed to establish links to other users, or to attract other links. Hence, this user is also determined to leave at the current time step. In the consecutive STEP \mathcal{T}_3 , the cascade of core users leaving further erodes the structure of the core and hence also affects the peripheral users, as shown in STEP \mathcal{T}_4 . In conclusion, after the core was affected by the cascades in STEP \mathcal{T}_1 , the OSN could no longer remain resilient and its core breaks down.

As the snapshots of Table 12.2 come from real computer simulations, we can also tell that it took a considerable time before, because of random rewiring, a new core of users emerged. Precisely, the recovery time was $T_{\text{recovery}}=492$ in the case of $\tau_1=0$, but only $T_{\text{recovery}}=297$ in the case of $\tau_2=0.2$. As this example indicates, the emergence of cascades of users leaving is largely independent of the precise cost level. However the time for the OSN to *recover* and the *average life time* of the core significantly depends on it. This is quantitatively analysed in Appendix 12.4.

STEP \mathcal{T}_0



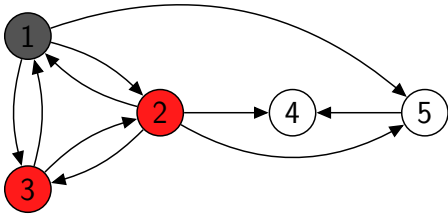
$$\mathbf{b}^{\lambda_1} = (1, 1, 1, 0.49, 0.24)$$

$$\langle b \rangle = 0.74$$

$\tau_1 = 0$

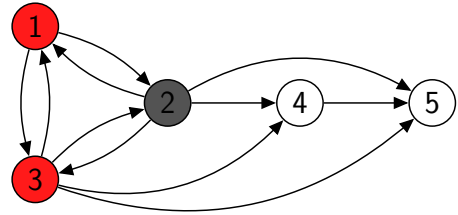
$\tau_2 = 0.2$

STEP \mathcal{T}_1



$$\mathbf{b}^{\lambda_1} = (1, 1, 1, 1, 1)$$

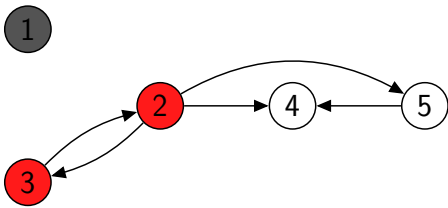
$$\langle b \rangle = 1$$



$$\mathbf{b}^{\lambda_1} = (1, 1, 1, 1, 1)$$

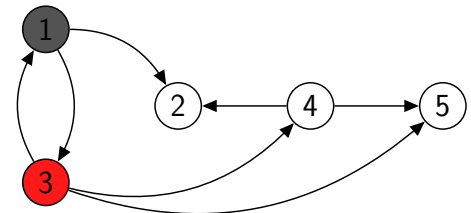
$$\langle b \rangle = 1$$

STEP \mathcal{T}_2



$$\mathbf{b}^{\lambda_1} = (0, 0.5, 0.5, 1, 0.5)$$

$$\langle b \rangle = 0.5$$



$$\mathbf{b}^{\lambda_1} = (0.5, 1, 0.5, 0.5, 1)$$

$$\langle b \rangle = 0.7$$

continues on the next page

Table 12.2 – continued from the previous page

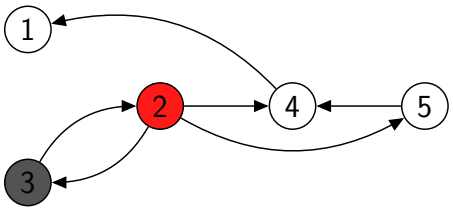
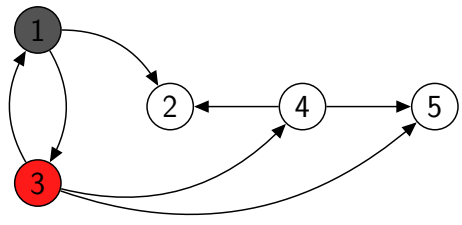
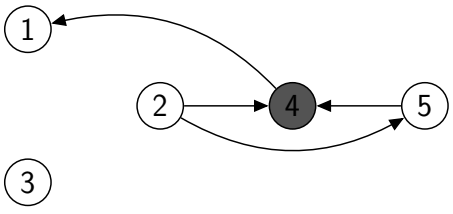
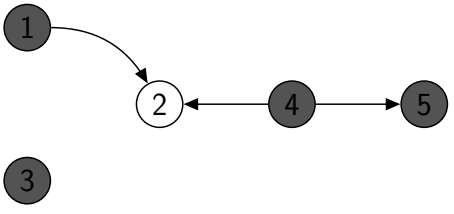
STEP \mathcal{T}_3	
	
$\mathbf{b}^{\lambda_1} = (1, 0.5, 0.5, 1, 0.5)$ $\langle b \rangle = 0.7$	$\mathbf{b}^{\lambda_1} = (0.5, 1, 0.5, 0.5, 1)$ $\langle b \rangle = 0.7$
STEP \mathcal{T}_4	
	
$\mathbf{b}^{\lambda_1} = (1, 0, 0, 0, 0)$ $\langle b \rangle = 0.2$ $T_{\text{recovery}} = 492$	$\mathbf{b}^{\lambda_1} = (0, 1, 0, 0, 0)$ $\langle b \rangle = 0.2$ $T_{\text{recovery}} = 297$

Table 12.2: Examples of emerging cascades of users leaving obtained from the computer simulations in Appendix 12.4. The top row shows the initial configuration of a sample OSN. Subsequent rows show consecutive steps of the evolution of the OSN for two different cost levels: (left column) $\tau_1 = 0$, (right column) $\tau_2 = 0.2$. Red colors indicates core users, whereas users leaving at the given time step are shown in grey. Peripheral users are in white. The relative reputation \mathbf{b}^{λ_1} of users and the average reputation $\langle b \rangle$ (Eq. 6.15), are given for each configuration. Other parameters: $m = 0.25$, $N = 5$.

12.4 Optimal cost level

To illustrate how the cost, τ , affects the average reputation in the network, we simulated the dynamics of a small network of five nodes over $T = 10^4$ network time steps. The left part of Figure 12.1, which should be compared with Figure 6.4, shows again the average reputation of the OSN. We observe that for our toy network the maximum average benefit is attained for a nonzero cost. More precisely, we confirm that the optimal cost is $\tau=0.2$,

as it was for the much larger network simulated in Fig. 6.4, so this cannot be attributed to a size effect. Furthermore, the network is not considerably disturbed at this maximum cost level, as the average number of new users is about 40% of the network size. This shows that the maximum average benefit is not concentrated in a minority of core users, at the expense of a large and sparsely connected periphery.

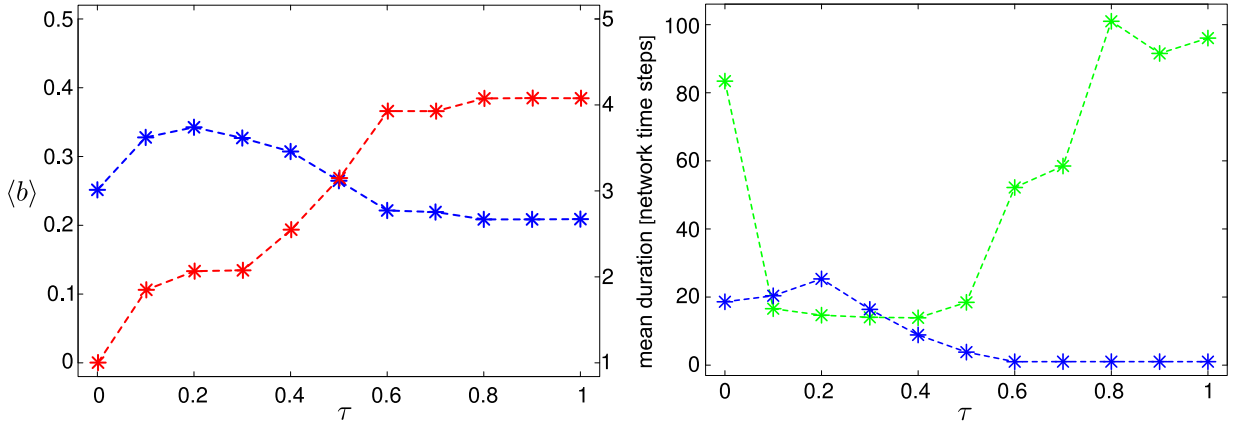


Figure 12.1: **Left:** Long term average benefit $\langle b \rangle$ (Eq. 6.15) (blue), and average number of rewired users over the whole simulation (red, right y-axis). **Right:** average lifetime of a core, $\langle \Omega_Q \rangle$ (blue), and average time to establish a core, $\langle \Pi_Q \rangle$ (green), both measured in network time T , as a function of the cost τ . Wilcoxon-rank test shows that $\langle \Omega_Q \rangle_{\tau=0.2}$ is significantly different from $\langle \Omega_Q \rangle_{\tau=0.1}$ (p-value = 0.02) and $\langle \Omega_Q \rangle_{\tau=0.3}$ (p-value = 10^{-8}). Other parameters: $m = 0.25$.

The right part of Figure 12.1 explains why this particular value of τ appears to be the optimal one. Here, we have plotted two different time spans, the average lifetime of a core, $\langle \Omega_Q \rangle$, and the average time to establish a core, $\langle \Pi_Q \rangle$, both measured in network time T , as a function of the cost τ . With the exception of $\tau=0$, which only serves as a reference case, the time $\langle \Pi_Q \rangle$ to establish a core in the OSN monotonously increases with τ until it reaches a saturation value at high levels of $\tau \rightarrow 1$. As explained in Section 6.2.3, this increase is due to the increasing number of users that leave the OSN at each time step T . If τ gets too large, a considerable fraction of the OSN is replaced, this way also destroying favorable reciprocal social links. For large τ the average life time of the core, $\langle \Omega_Q \rangle$, decreases for the very same reasons. For small τ , however, we observe a non-monotonous behavior – the life time increases with τ until it reaches a maximum at $\tau=0.2$.

Therefore, the optimal cost level balances two different dynamic effects: (i) the *emergence* of a core of users that, by means of direct or indirect reciprocity, maintain a relatively high reputation, and (ii) the *maintainence* of this core by the users. *Resilience* does not simply mean that the core stays alive, it also means that a new core is quickly established

when the old one crashes. This way, the average reputation of most of the users remain high *over time*. Cascades of users leaving cannot be prevented as shown in Appendix 12.3. The question is how fast the OSN is able to cope with it.

12.5 Network structure

Here we present three typical snapshots of the network structure for different values of the cost τ . In all three simulations, the parameter that determines the network density is fixed to $m=0.25$ and the system size is $N=100$. The most immediate differences in these three snapshots are (i) the core size, which strongly decreases with increasing cost τ and (ii) the maximum out-degree which increases with increasing cost τ . These findings are in agreement with our theoretical discussions in Section 6.2.2 about changes in the largest eigenvalues λ_1 dependent on the network structure.

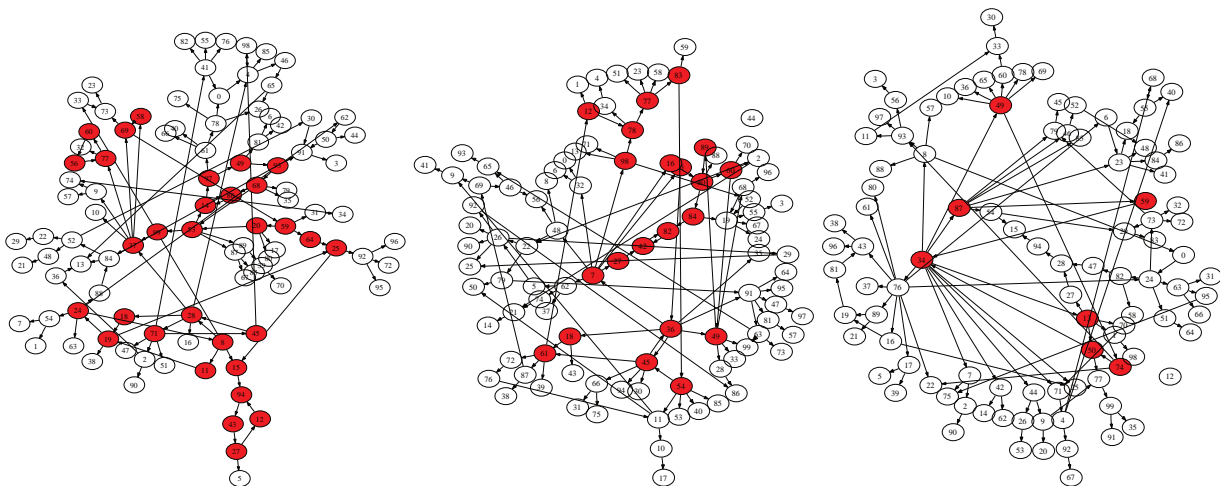


Figure 12.2: **Left:** $\tau=0.0$. $\lambda_1=1.133$, number of users in the core is 31. The largest out-degree is 5. In total, there are 3 users (71, 28, and 24) with that number. **Middle:** $\tau=0.2$. $\lambda_1=1.47$, number of users in the core is 21. The largest out-degree is 7. User 7 has 7 outgoing links, user 36 has 6, and at least five users have 5 (users 49, 91, 26, 54, and 45). **Right:** $\tau=0.3$. λ_1 is 1.40, number of users in the core is 7. The largest out-degree is 10. Two users have out-degree 10, (users 34 and 76), one has out-degree 7 (user 49) and one has out-degree 6 (user 87).

Chapter 13

13.1 Indications of Swarming

Below we illustrate how swarming affects L/F time differences. In particular, the time differences, in presence of swarming, are skewed towards the `lf_delay` limit.

Listing 13.1 shows a typical swarming pattern from a representative experimental box close to 6am on the day of the box occupation. Here, five experienced and five naïve individuals were recorded in the time of about 10 minutes.

```
1. NAIVE      ;05:52:28;00068E1731
2. EXPERIENCED;05:53:52;00060F6D0D
3. NAIVE      ;05:54:41;00064407F9
4. EXPERIENCED;05:55:02;0005FDFD3D
5. EXPERIENCED;05:56:12;00065EA84E
6. NAIVE      ;05:57:22;00060D6C05
7. EXPERIENCED;05:59:03;0006979AC0
8. NAIVE      ;06:00:31;0006011890
9. EXPERIENCED;06:02:22;00064407F9
10. NAIVE     ;06:02:32;0006011890
```

Listing 13.1: An excerpt from the *processed* recordings of an experimental box for the GB2 colony in 2008. Each line contains the individual information status (see Section 7.3.1), recording time, and unique identification number, in that order, separated by semi-colon.

The recording pattern shows an interesting peculiarity. The readings of no two individuals appear within less than a minute of each other, as one would, otherwise, expect from an experienced leader and a naïve follower that arrive at the box together. Table 13.1 contains all L/F events identified from the sample with `lf_delay` fixed at 5 minutes. Indeed, the mean time difference of these events is 2.4 minutes, and the minimum is strictly above 1 minute. This characteristic is more consistent with swarming behaviour, in which a few experienced individuals attract naïve conspecifics by circling around the roost and emitting echolocation calls.

Follower	Leader	L/F time difference	L/F times
00060F6D0D	00068E1731	1.4	05:53:52 / 05:52:28
00064407F9	00068E1731	2.21667	05:54:41 / 05:52:28
0005FDFD3D	00068E1731	2.56667	05:55:02 / 05:52:28
00065EA84E	00068E1731	3.73333	05:56:12 / 05:52:28
00060F6D0D	00060D6C05	3.5	05:53:52 / 05:57:22
00064407F9	00060D6C05	2.68333	05:54:41 / 05:57:22
0005FDFD3D	00060D6C05	2.33333	05:55:02 / 05:57:22
00065EA84E	00060D6C05	1.16667	05:56:12 / 05:57:22
00065EA84E	0006011890	4.31667	05:56:12 / 06:00:31
0006979ACO	00060D6C05	1.68333	05:59:03 / 05:57:22
0006979ACO	0006011890	1.46667	05:59:03 / 06:00:31
00064407F9	0006011890	1.85	06:02:22 / 06:00:31

Table 13.1: The L/F events corresponding to the recording pattern in Listing 13.1. Parameters `lf_delay=5` minutes and `occupation_deadline=8am`. `turnaround_time` does not affect this example, as no naïve individual appears again as a leader.

Since experienced and naïve individuals are not grouped together as in genuine leading-following, it takes time for a naïve individual to respond to the calls and fly to the roost. As a result, most L/F events identified in this way would tend to have larger time differences closer to the allowable limit of `lf_delay`. We use precisely this observation when fine-tuning the `occupation_deadline` parameter.

As a comparison, consider the following recording pattern from another box close to 2am on the day of its occupation.

1. NAIVE ;01:55:21;00064380ED	22. NAIVE ;01:58:45;000697D00F
2. NAIVE ;01:55:25;00065DB1F6	23. NAIVE ;01:58:46;000697D00F
3. NAIVE ;01:55:26;00065DB1F6	24. NAIVE ;01:58:47;000697D00F
4. NAIVE ;01:55:36;000697D00F	25. EXPERIENCED;01:58:48;0005FE0AF1
5. NAIVE ;01:55:49;00065DB1F6	26. EXPERIENCED;01:58:49;0005FE0AF1
6. NAIVE ;01:55:50;00065DB1F6	27. EXPERIENCED;01:58:50;0005FE0AF1
7. NAIVE ;01:55:51;00065DB1F6	28. EXPERIENCED;01:58:52;0005FE0AF1
8. NAIVE ;01:55:52;00065DB1F6	29. NAIVE ;01:58:53;000697D00F
9. NAIVE ;01:55:55;00065DB1F6	30. EXPERIENCED;01:58:54;0005FE0AF1
10. NAIVE ;01:57:12;00065DB1F6	31. EXPERIENCED;01:58:54;00068E1B66
11. NAIVE ;01:57:14;00065DB1F6	32. NAIVE ;01:58:54;000697D00F
12. NAIVE ;01:57:21;000697D00F	33. EXPERIENCED;01:58:55;0005FE0AF1
13. NAIVE ;01:57:22;000697D00F	34. EXPERIENCED;01:58:56;0005FE0AF1
14. NAIVE ;01:57:23;000697D00F	35. EXPERIENCED;01:58:56;00068E1B66
15. NAIVE ;01:57:25;000697D00F	36. EXPERIENCED;01:58:57;00068E1B66
16. NAIVE ;01:57:27;000697D00F	37. EXPERIENCED;01:58:58;00068E1B66
17. NAIVE ;01:58:30;000697D00F	38. EXPERIENCED;01:58:59;00068E1B66
18. NAIVE ;01:58:31;000697D00F	39. EXPERIENCED;01:59:00;00068E1B66
19. NAIVE ;01:58:32;000697D00F	40. NAIVE ;01:59:09;00065DB1F6
20. NAIVE ;01:58:35;000697D00F	41. NAIVE ;01:59:10;00065DB1F6
21. NAIVE ;01:58:36;000697D00F	42. EXPERIENCED;02:00:02;000697A2BA

Listing 13.2: An excerpt from the *processed* recordings of an experimental box for the GB2 colony in 2008. The pattern is formatted as in Listing 13.1.

The L/F events corresponding to this pattern are shown in Table 13.2. The mean time difference is 1.5 minutes and the minimum is zero, as individual arrivals exceeded the time resolution of the reading device.

Follower	Leader	L/F time difference	L/F times
0005FE0AF1	000697D00F	0	01:58:54 / 01:58:54
00068E1B66	000697D00F	0	01:58:54 / 01:58:54
0005FE0AF1	00064380ED	3.45	01:58:48 / 01:55:21
00068E1B66	00064380ED	3.55	01:58:54 / 01:55:21
000697A2BA	00064380ED	4.68333	02:00:02 / 01:55:21
0005FE0AF1	00065DB1F6	0.216667	01:58:56 / 01:59:09
00068E1B66	00065DB1F6	0.15	01:59:00 / 01:59:09
000697A2BA	00065DB1F6	0.866667	02:00:02 / 01:59:10
000697A2BA	000697D00F	1.13333	02:00:02 / 01:58:54

Table 13.2: The L/F events corresponding to Listing 13.1. Parameters as in Listing 13.1.

Consistent with leading-following behaviour, most L/F events have low time difference, indicating that an experienced individual did appear close together with a designated follower. As for the couple of events with large time differences, they are most likely due to naïve individuals remaining at the entrance of the box, thereby triggering the reading device repetitively, than to swarming. As seen from Listing 13.1, three such individuals, 00065DB1F6, 000697D00F and 00068E1B66, generate long recording sequences that prevent other followers from examining the box upon arrival, thereby forming an L/F event with large time difference to the leader.

13.2 Eigenvector centrality

Consider the following simple leading-following network,

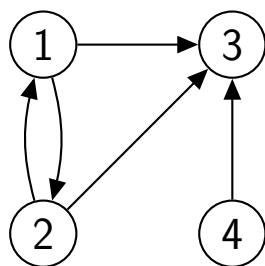


Figure 13.1: Example of a leading-following network with 4 nodes (bats) and 5 leading-following events (edges)

An alternative representation of this network is through its so called *adjacency* matrix, A , which indicates which two nodes are adjacent to each other.

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The elements $a_{i,j}$ (i , and j index rows and columns, respectively) in this matrix are 1 if a directed link exists between nodes j and i . In other words, $a_{i,j} = 1$ if j followed i . Otherwise, $a_{i,j} = 0$. For example the first column $a_{i,1}$ gives all nodes that node 1 follows. We see that $a_{2,1} = a_{3,1} = 1$, so 1 has followed both 2 and 3.

The main idea behind eigenvector centrality is that the centrality of a node i , c_i , is proportionate to the sum of the centralities of all nodes who follow it. Staying with node 1, its centrality is the sum of the centralities of nodes 2 and 3, i.e. $c_1 = \frac{1}{\lambda}c_2 + \frac{1}{\lambda}c_3$ or $\lambda.c_1 = c_2 + c_3$ for some proportionality constant λ . In this way, we can express the centralities of all nodes and write them as a system of equations:

$$\begin{aligned} \lambda.c_1 &= 0.c_1 + 1.c_2 + 1.c_3 + 0.c_4 \\ \lambda.c_2 &= 1.c_1 + 0.c_2 + 0.c_3 + 0.c_4 \\ \lambda.c_3 &= 1.c_1 + 1.c_2 + 0.c_3 + 1.c_4 \\ \lambda.c_4 &= 0.c_1 + 0.c_2 + 0.c_3 + 0.c_4 \end{aligned}$$

In matrix form the above system can be rewritten as:

$$\lambda \cdot \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix}$$

or in vector notation:

$$\lambda.\vec{c} = A.\vec{c}.$$

This is the familiar eigenvector problem. We need to find a vector \vec{c} such that upon applying matrix A to it, the result is a scaled version of \vec{c} with a scaling factor λ . The unknown vector \vec{c} is called an eigenvector of the matrix A , and λ is referred to as the eigenvalue, which corresponds to that eigenvector.

Solving the system of equations yields,

$$\vec{c} = \{0.408, 0.408, 0.816, 0\}, \quad \lambda = 1$$

Therefore node 3 is most central since it is followed by everyone. Nodes 1 and 2 follow

each other so they boost their own centrality, and node 4 is not followed by anyone so its centrality is 0.

Pavlin Mavrodiev

Curriculum Vitæ

Education

- Mar'11–Apr'14 **PhD**, *ETH Zurich*, Zurich.
- Chair of Systems Design
 - Data-driven mathematical modelling of social systems
 - understand the emergence of collective phenomena
 - analysis of large (online) datasets from human and animal societies
 - development of agent-based models to qualitatively and quantitatively reproduce observed collective behaviour
 - Thesis – “*Decisions and Their Unintended Consequences*”
- Feb'09–Dec'10 **M.Sc.**, *ETH Zurich*, Zurich.
- M.Sc. in Management, Technology and Economics
 - Top 10% of graduating class
 - Thesis – “*Enhancement of Enterprise Service Architecture*”. Grade: 5.75 out of 6.0
- Sep'03–Jun'06 **B.Sc.**, *Jacobs University Bremen*, Bremen, Germany.
- B.Sc. in Electrical Engineering and Computer Science
 - Thesis – “*Test of Dynamic Random Access Memories for Retention Time Fails*”. Grade: 1.33 out of 1.0
- Sep'98–Jun'03 **High School**, *Mathematics High School*, Burgas, Bulgaria.
- Graduated with honors

Experience

Industry

- May'10–Nov'10 **Intern/Master student**, *SAP Research*, Zurich.
- Developed and implemented a framework for advanced decision-making in distributed systems of modular business software.
- Jun'09–Sep'09 **Intern**, *SAP Research*, Zurich.
- Analysis of state-of-the-art academic and industrial efforts in designing and implementing Service Oriented Architecture and Event Driven Architecture applications. An existing framework within SAP was improved as a result of this internship.
Contact: Dr. Oliver Kasten (oliver.kasten@sap.com)
- Jun'06–Feb'09 **Development Engineer and Team Leader**, *Qimonda*, Munich, Germany.
- Head of a three-person, cross-functional team responsible for redundancy optimization. Redundancy is additional memory cells built on a memory chip to take over the functions of defect cells.
 - Maintained and improved a software product (Java) to emulate memory testers (used to test the functionality of memory chips).
- Jun'05–Aug'05 **Intern**, *Infineon Technologies*, Munich, Germany.
- Implemented a software solution (C++) to automatically configure memory testers.

- Jun'04–Aug'04 **Intern**, *Siltronic*, Burghausen, Germany.
- Implemented an internal knowledge management database.
- Teaching**
- 2011–2012 **Lecturer**, *ETH Zurich*, Zurich.
- “*Collective Dynamics of Firms*” – a quantitative course that uses industrial datasets to develop agent-based models of the collective properties of firms, e.g. firm size distribution and growth rates.
- 2004–2006 **Teaching Assistant**, *Jacobs University Bremen*, Bremen, Germany.
- Various teaching assistant positions for introductory and advanced courses in communication technologies, electronics, and electrical engineering.

Technical Skills

- Data analysis Processing and exploratory data analysis of structured and semi-structured data sets, e.g. user behaviour in online social networks. **Tools:** R, Python, Hadoop.
- Modelling Identifying and formulating hypothesis for statistical inference. Development of econometric or agent-based models that explain statistical regularities in data. Methods from statistics, statistical physics, machine learning, and complex network theory. **Tools:** R, C/C++, Python.
- Visualization Novel visualization techniques to enhance understanding of complex relationships in large datasets. **Tools:** D3.js
- Programming **Regular:** C/C++, R, Java, Python. **Other:** HTML, Emacs, Vensim, Stata, Matlab, Mathematica, ...
- Databases MySQL, PostgreSQL.

Languages

- | | | |
|-----------|-----------------|---------------------------|
| Bulgarian | Fluent | <i>My native language</i> |
| English | Fluent | <i>Written and spoken</i> |
| German | Advanced | <i>Written and spoken</i> |

Awards

TIMES, European business case study tournament in corporate strategy

- member of the team that won the local qualification round in Zurich
- member of the runner-up team in the Lisbon semi-finals

Jacobs University Bremen

- university grant of €16500 covering half of the study costs

ETH Zurich

- high performance scholarship for my last M.Sc. semester

Bulgarian 2001 team competition in informatics

- 2nd place

Publications

Journal

- 2013 **Mavrodiev P.**, Tessone, Claudio J., Schweitzer F., *Quantifying the effects of social influence*, Scientific Reports, Vol. 3
- 2013 Schweitzer F., **Mavrodiev P.**, Claudio J., *How can social herding enhance cooperation*, Advances in Complex Systems, Vol. 16
- 2013 Tasca P., **Mavrodiev P.**, Schweitzer F., *Quantifying the Impact of Leveraging and Diversification on Systemic Risk*, Journal of Financial Stability (Submitted)
- 2014 Garcia, D., Tessone, Claudio J., **Mavrodiev, P.**, Perony, N., *The digital traces of bubbles: Feedback cycles between socio-economic signals in the Bitcoin economy*, Journal of the Royal Society Interface (submitted)

Proceedings

- 2012 **Mavrodiev, P.**, Tessone, Claudio J., Schweitzer, F., *Effects of social influence on the wisdom of crowds*, Proceedings of Collective Intelligence conference 2012

Working papers

- 2014 **Mavrodiev, P.**, Tessone, Claudio J., Schweitzer, F., *Designing wise crowds: the importance of ranking and competition*.
- 2014 **Mavrodiev, P.**, Tessone, Claudio J., Schweitzer, F., *The effects of social influence on the wisdom of crowds*.
- 2014 **Mavrodiev, P.**, Fleischmann D., Kerth, G., Schweitzer F., *Information transfer in Bechstein's bats: a network perspective* (working title).

