Chapter 26

Modeling collective emotions in online social systems

David Garcia, Antonios Garas, and Frank Schweitzer
ETH Zürich

Every day, millions of Internet users leave online traces that are publicly accessible. Data about forum comments, video downloads, or product reviews provide a valuable insight into human online behavior. The retrieval of datasets of unprecedented size may eventually also allow testing of hypotheses or validation of theories that have been developed in the social sciences, for example, about preferences, social influence (Lorenz, 2009; Onnela & Reed-Tsochas, 2010), trust, and cooperation (Walter, Battiston, Yildirim, & Schweitzer, 2011). This recent scientific development has led to the emerging field of *computational social science* (Lazer et al., 2009) which combines methods and tools from different technical and social disciplines. Also, psychology can benefit from this development by getting access to data without designing expensive experimental setups of limited size. For example, the analysis of *Twitter messages* allows studying the influence of the circadian cycles on human mood (Golder & Macy, 2011), and sentiment analysis of large-scale datasets reveals patterns of emotional expression predicted by earlier theories (Garcia, Garas, & Schweitzer, 2012).

A special feature of online communities is the frequent occurrence of collective emotions, which are not so easily observable in offline interaction. Spontaneously, large numbers of users share similar emotional states, due to their ability to reach many other users in a quick and often anonymous way. Such collective emotions can result from exogenous as well as from endogenous causes. For example, external events, such as a natural disaster, are able to trigger the online expression of emotions of millions of users. But collective emotions can also be created within online communities, for example, in forum discussions (Chmiel et al., 2011).

Our aim is to study emergent collective emotions, as their dynamics and preconditions can be studied based on the textual expressions of users in online communities. Online data allows us to precisely measure how and when collective emotional states emerge, but the analysis of this spontaneous behavior cannot be simply reduced to the activity of single users. Instead, these collective states should be understood as emergent phenomena resulting from the interaction of a large number of individuals. In our approach, we try to relate the statistical regularities observed in online communities to the interactions between users. The distinction between the micro level of individual users and the macro level at which their collective behavior can be observed is one of the specific features of the theory of complex systems. Over the last 40 years, methods and tools from computer science, statistical physics, and applied mathematics have been utilized to address this micro–macro link and to predict the collective dynamics of a system from individual interactions of many system elements or agents. Agent-based modeling provides a useful approach to understand collective phenomena by studying the rules of the agents involved. In particular, some agent-based models in social psychology (Kuppens, Oravecz, & Tuerlinckx, 2010) support the need to apply these models to understand human behavior.

We present a general framework to model the emergence of collective emotions based on the emotional expressions of individuals and their interactions. In our framework, individuals are represented by Brownian agents (Schweitzer, 2003), which allows the use of established methods from statistical physics to derive information about the collective dynamics of the agents. The emotional state of an agent is characterized by its valence, measuring the emotions' pleasantness or unpleasantness, and its arousal, for example, the level of activity associated with the emotion. This framework follows the dimensional theory of core affect (Russell, 1980), which represents the short-lived, intense features of emotional life. While more complex models and additional dimensions can be taken into account, core affect is closely linked to emotional expression, for example, written text (posts) in the case of online communication. Such posts are transmitted over the Internet to other agents who may read them and react to them in an emotional way. Modeling such kinds of emotional feedback between distant users via the exchange of emotional messages under different circumstances is one of the main purposes of the framework developed.

Our framework is specific enough to allow analytical results to predict simulation outcomes (Schweitzer & Garcia, 2010), but general enough to cover a wide range of online emotional interactions. The main feedback loops of this framework, as sketched in Fig. 26.1, are comprised of two orthogonal layers: an internal layer describing the agent (shown horizontally) and an external layer describing the communication process

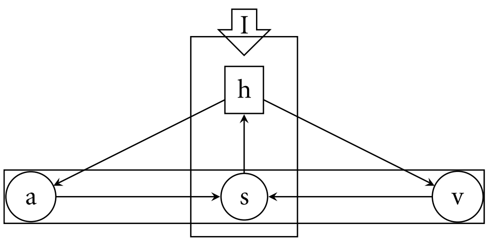

Fig. 26.1 Causation among the components of the general model. Reproduced from The European Physical Journal B—Condensed Matter and Complex Systems, 77(4), 2010, pp. 533–45, An agent-based model of collective emotions in online communities, F. Schweitzer. With kind permission from Springer Science and Business Media.

(shown vertically). In the internal layer, the arousal a and the valence v of an agent determine its emotional expression s, which reaches the external layer by contributing to the communication field h. The latter has its independent dynamics and can, in addition to contributions from other agents, also consider input from external sources, I. The causality loop is closed by considering that both valence and arousal of an agent are affected by the communication field.

Since we are interested in modeling the emotional dynamics of Internet communities, the general framework can be easily adjusted to consider the particularities of various online platforms such as user expression limitations, external influence on users, or communication in networks as opposed to broadcast. In this chapter, we will provide different examples of how to specify our modeling framework to cope with different online communities.

The concept of Brownian agents

Brownian agents are described by a set of k different state variables $u_i^{(k)}$, where the index $i = 1, \ldots, N$ refers to each individual agent i. Each of these variables could be *external* if they can be observed in experimental data, or *internal* if they can only be indirectly concluded from the observable data. Each of these state variables can be time dependent due to interaction with the agent's environment, or due to internal dynamics that do not require external influence. In a general way, we can formalize the dynamics of each state variable u_i as a superposition of two influences of different nature:

$$\frac{du_i^{(k)}}{dt} = f_i^{(k)} + F_i^{\text{stoch}} \tag{1}$$

This formulation is based on the principle of causality, as the change of any variable u is produced by some causes which are listed on the right hand side of the equation. In the case of Brownian agents, these causes are assumed to be described by a superposition of deterministic and stochastic influences.

The stochastic term F_i^{stoch} models all the influences on the variables that are not observable on the time and length scale of the available data. This stochastic influence does not drive the dynamics of the agent state in any particular direction, and it is commonly, but not necessarily, modeled by white noise. Furthermore, the strength of the stochastic influences might be different among agents, depending on local parameters of the agents, as in Schweitzer (2003).

The deterministic term $f_i^{(k)}$ represents all the specified influences that change the corresponding state variable $u_i^{(k)}$. For example, non-linear interactions with other agents can be modeled by a function that depends on the state variables of any set of agents, which can also include agent i itself. $f_i^{(k)}$ can also describe the agent's response to the available information, which is the case for our modeling framework. It can depend on external conditions, such as the influence of mass media in online communities. Additionally, $f_i^{(k)}$ can reflect the *eigendynamics* of the agent, which are the changes in the variables $u_i^{(k)}$ not caused by any influence external to the agent. Examples of eigendynamics are saturation or exhaustion, common in the modeling of human behavior (Lorenz, 2009). In order to design a multi-agent system, we have to define the agent's state variables, $u_i^{(k)}$, and the dynamics of their change, $f_i^{(k)}$, specifying the interaction among agents. These dynamics are defined at the level of the individual agent and not at the collective level, in a way that the macroscopic dynamics emerge from the interaction of a large amount of agents, just as collective emotions emerge in online communities.

Emotional states and their internal dynamics

Following the bidimensional representation of core affect (Russell, 1980), we quantify the emotional state of an agent through the variables of valence $\upsilon_i(t)$, and arousal $a_i(t)$. Different emotional states can be mapped to different points in this bidimensional space. For example, "happy" is an emotional state with high valence and arousal, while "satisfied" has positive valence but negative arousal. Other states like "depressed" have negative valence and arousal, and "angry" has positive arousal and negative valence. These two variables are known to capture most of the information of emotional experience (Russell, 1980), so we will define the state of the agent as $e_i(t) = \{\upsilon_i(t), a_i(t)\}$. Note that valence and arousal are internal variables. They can only be indirectly observed, for example, through physiological measurements or individual reports.

In the absence of any external excitation, any emotional state should relax to an equilibrium state. This assumption is supported by empirical studies that show how emotional states exponentially decay in a stochastic manner (Kuppens et al., 2010). This relaxation, $e_i(t) \to 0$, implies $\upsilon_i(t) \to 0$, $a_i(t) \to 0$. Thus, in accordance with Eq. (1) we define the dynamics of the Brownian agent as follows:

$$\begin{aligned} \dot{\upsilon}_i &= -\gamma_{\upsilon i}\, \upsilon_i(t) + F_\upsilon + A_{\upsilon i}\, \xi_\upsilon(t) \\ \dot{a}_i &= -\gamma_{ai}\, a_i(t) + F_a + A_{ai}\, \xi_a(t) \end{aligned} \tag{2}$$

The first terms on the right-hand side of the equations describe the exponential relaxation of valence and arousal toward the equilibrium state. The parameters $\gamma_{\upsilon i}$ and γ_{ai} define the time scales of this relaxation, which can be different for valence and arousal and across agents. The second terms describe the deterministic influences as explained below, and the third terms model the stochastic influences. $\xi_\upsilon(t)$, $\xi_a(t)$ are random numbers drawn from a given distribution of white noise, for example, they have zero mean and no temporal correlations. The strengths of the stochastic components are quantified by $A_{\upsilon i}$ and A_{ai}, which can also vary across agents.

The deterministic influences on the emotional state of the agent are described by the functions F_υ, F_a. They depend on specific assumptions applicable to online collective emotions, in particular the agents' interaction, access to information, or their response to the media. They can also depend on the emotions of other agents, as for example, empathy (Preston & de Waal, 2002) would drive the emotional state of an agent toward the one perceived from others. In the following sections, we extend the description of the agent by defining the actions an agent can take, and by specifying the forms of these functions.

Emotional expression

When the deterministic and stochastic influences become negligible, Eq. (2) defines the stationary state $e_i(t) \to 0$. If the influences are large, for example, if information with emotional content becomes available to the agent, there should be excited emotional sates. This excited state is not externally observable unless the agent decides to communicate, creating a message or posting a comment in a discussion. Consequently, our assumption for the expression of emotions is that the agent expresses its valence through the externally observable variable $s_i(t)$ if its arousal exceeds a certain individual threshold, $\mathcal{T}_i$:

$$s_i(t + \Delta t) = \mathrm{sign}(\upsilon_i(t))\Theta[a_i(t) - \mathcal{T}_i] \tag{3}$$

Where $\Theta[x]$ is the Heaviside's function which is one only if $x \geq 0$ and zero otherwise. If $\Theta[x] = 1$, we assume that the agent is not able to perfectly communicate its valence state, for example, the exact value of $\upsilon_i(t)$, and its expression is simplified to the sign of its valence, $r_i(t) = \mathrm{sign}(\upsilon_i(t))$. Of course this is only an assumption, and we can easily change it to allow perfect valence communication. Thus, it is essential to assess its validity through the analysis of real data. This way the communication process receives a coarse-grained representation of the valence of individual agents, which can be adjusted to the accuracy of the data analysis techniques available.

An additional assumption is that the agent might not be able to immediately express its emotions if the arousal hits the threshold at a particular time t. This expression might be delayed for a certain time Δt, as the agent might not have immediate access to communication media. Empirical studies investigated these kind of waiting time distributions of human communication (Crane, Schweitzer, & Sornette, 2010; Garas, Garcia, Skowron, & Schweitzer, 2012; Rybski, Buldyrev, Havlin, Liljeros, & Makse, 2009), letting us assume $P(\Delta t) \propto \Delta t^{-\alpha}$, where α should be empirically determined.

Communication of emotions in online communities

After describing the dynamics of emotional states and emotional expression, we need to specify how this emotional expression is communicated to the other agents. In line with previous models of social interaction (Schweitzer & Holyst, 2000), we assume that every positive and negative expression is stored in a communication field $h_\pm(t)$ with a component for positive communication $h_+(t)$, and another component for negative information $h_-(t)$. This variable essentially stores the "amount" of available comments of a certain emotional content at a given moment in time. We propose the following equation for the dynamics of the field:

$$\dot{h}_\pm = -\gamma_\pm h_\pm(t) + cn_\pm(t) + I_\pm(t) \tag{4}$$

where each agent contribution $s_i(t)$ increases the corresponding field component by a fixed amount c at the exact time the expression occurred. This parameter c represents the impact of the information created by the agent to the information field, defining a time scale.

The variable $n_{\pm}(t)$ shows the total number of agents contributing positive or negative emotional expression at time t. These expressions are in general time dependent, for example, they lose importance as they become older, usually due to the creation of new information in the community. This is represented by the exponential decay present in the first term of the right-hand side of Eq. (4), which is parametrized through $\gamma_{\pm}$. In addition, externally produced positive or negative emotional content might change the communication field, as, for example, news can have a great impact in the overall emotional state of an online community. We model this mechanism through the agent-independent term $I_{\pm}(t)$, which can be modeled as a stochastic input, or used to analyze the reactions of the model to external stimuli.

To finish the description of our framework, we need to specify how the available information influences the state of the individual agents, which is covered by the functions F_{υ} and F_a.

Feedback of communication into emotional states

The aim of our model is to reproduce the emergence of a collective emotion, assuming that it cannot be understood as a simple superposition of individual emotional states. Our assumption is that the emotional expression of an agent may change the emotional state of a number of other agents, either directly or indirectly. For this influence we can only use hypotheses and test them in computer simulations in order to investigate various possible scenarios. Additionally, these can also be empirically tested when individual users are exposed to emotional content in experiments (Kappas, Tsankova, Theunis, & Kuester, 2011).

In the communication field of our model, there are two components for positive, $h_+(t)$, and negative, $h_-(t)$, emotional information. Depending on the state of an agent, it might be affected by these different kinds of information in different ways. For example, if we assume that agents with negative (positive) valence mostly respond to negative (positive) emotional information, we can specify:

$$\mathrm{F}_{\upsilon} \propto \frac{r_i}{2}\left\{(1+r_i)\, f\left[h_+(t)\right] - (1-r_i)\, f\left[h_-(t)\right]\right\} \tag{5}$$

where $r_i(t) = \mathrm{sign}(\upsilon_i(t))$ and $f[h_{\pm}(t)]$ are functions depending either on h_+ or on h_- only. An alternative scenario would be that agents pay attention to the prevalence of positive or negative emotional content independently of their valence. In that case, we may assume:

$$\mathrm{F}_{\upsilon} \propto g\left[h_+(t) - h_-(t)\right] \tag{6}$$

where g is a function of the difference between the two components of the information field. Other combinations might be tested as well.

A general assumption for Eq. (5) is that valence increases with the respective information perceived by the agent. The strength of this influence should depend on the emotional state of the agent, often in a non-linear manner. For example, if an agent is happy (sad), it may become happier (more sad) if receiving information about happy (sad) agents or events. A general formulation for this kind of dynamics has the form:

$$F_{\upsilon}\left[h_{\pm}(t),\upsilon_i(t)\right]=h_{\pm}(t)\sum_{k=0}^{n}b_k\upsilon^k(t) \tag{7}$$

where the key assumption is that the coefficients b_k are the same for any value of the valence.

Dynamics of arousal

As already explained, the arousal measures the degree in which the emotion encourages or discourages activity. It becomes important when it reaches a threshold $\mathcal{T}_i$, which is assumed to be the precondition for emotional expression (Rime, 2009). Emotional expression should have some impact on the arousal, and we assume that the arousal is lowered after producing a message, or set back to the ground state in the most simple case. This means that the dynamics of arousal should be divided into two parts: one applying before the arousal reaches the threshold, and one at the exact moment when it is reached. Hence, we define the dynamics of the arousal $a_i(t)$ as:

$$\dot{a}_i=\dot{\bar{a}}_i(t)\Theta\left[\mathcal{T}_i-a_i(t)\right]-a_i(t)\Theta\left[a_i(t)-\mathcal{T}_i\right] \tag{8}$$

As long as $x = \mathcal{T}_i - a_i(t)>1$, and the arousal dynamics are defined by $\dot{\bar{a}}_i(t)$ as in Eq. (2). Once the threshold is reached, $x\leq 0$, $\Theta[x]=0$ and $\Theta[-x]=1$, deterministically resetting the arousal back to zero.

To conclude the dynamics of arousal, we must specify the function F_a, which applies when the arousal is below the threshold. The arousal was designed to be an orthogonal variable to valence, measuring the activity level of an emotion. It is reasonable to assume that agents respond to all the emotional content available in the community, for example, the sum of both field components, in a way that depends on their own arousal in a non-linear manner, regardless of the valence dimension. Following the same general point of view as for the case of valence, we propose the following non-linear dependence:

$$F_a \propto \left[h_+(t)+h_-(t)\right]\sum_{k=0}^{n}d_k a^k(t) \tag{9}$$

Alternatively, we may argue that agents pay attention to the information only as long as their arousal is positive because negative arousals are associated with states of inactivity (tired, sleepy, depressed, bored). In this case, it is reasonable to assume, for example, that the impact of information increases linearly with the activity level:

$$F_a \propto \left[h_+(t)+h_-(t)\right]\, a(t)\Theta\left[a(t)\right] \tag{10}$$

This description defines a complete framework to design agent-based models of collective emotions in online communities. Simulation and statistical analysis of the properties of these models can explain the reasons for the emergence of collective emotional states from the online interaction of large amounts of users. In the following sections we present two instances of applications of our framework to emotional interaction in product reviews communities and chat room discussions.

A model for emotional product reviews

As an application of our modeling framework we study the emotional interaction in product review communities. Product review communities provide their users with the means to overcome information barriers typically present in traditional media and marketing by gathering independent information generated by other users.

This process of information exchange is subject to emotional interaction, changing the impact of a review depending on subjective factors like opinions and emotions.

The influence of emotions in product reviews has been subject of previous research (Dellarocas & Narayan, 2006), but the analysis of reviews was done in an individual manner. Collective emotions regarding products are a fairly new point of view that is gaining importance in order to predict and to optimize product acceptance. Datasets on user-generated product reviews provide a valuable resource to study how certain products become famous, or "beloved" by means of the Internet. In this section, we present our agent-based model which is able to reproduce two properties of emotions in product reviews: the time response of the community to the release of the product, and the typical distribution of emotional expression through the review text. We compare simulations of our model with a large dataset from *Amazon.com*, with emotional information extracted by sentiment detection tools.

Modeling emotional interaction through product reviews

The structure of this model is the same as explained in Fig. 26.1, where the emotional state of the agents is composed of valence and arousal and is influenced by a collective information field using specific assumptions about this kind of communication (Garcia & Schweitzer, 2011). In our model, we focus on the discussion at the product level, ignoring relations between products. This means that the communication between agents always refers to the reviewed product. It is a particular property of a product that every user is allowed to review it only once. We introduce this constraint in the arousal dynamics. Specifically, after an agent's arousal reaches its threshold T_i, the threshold is reset to a value of ∞, preventing the agent from making a second review on the same product. We assume that the initial values of these thresholds are heterogeneous among agents, sampled from a normal distribution with mean μ and standard deviation σ.

For this application, we assume that the arousal dynamics depends on the sum of both components of the field (h_+ and h_-), as formalized in Eq. (9). For this case, the polynomial function of Eq. (9) goes up to degree 2, modeling a quadratic dependence on the agent's own arousal. Our simulation results (Schweitzer & Garcia, 2010) show that this form of arousal dynamics is able to produce spontaneous emergence and disappearance of collective emotional states. For the valence dynamics, we assume that the influence of the information field in the agent's valence F_υ depends on the previous value of the agent's valence. This means that previous negative experiences of the product lead to a tendency to pay less attention to the positive expression of other agents. On the other hand, agents with positive experiences will be more influenced by positive emotional information than by negative information. We can formalize this asymmetry of agent perception through an exponential function with a cubic decay, as explained in Garcia and Schweitzer (2011).

Writing reviews is heavily influenced by preferences of the users and their relation to the properties of the product. In our model, user preferences are included as an agent internal variable u_i constant in time. The heterogeneity on these preferences is captured by sampling u_i from a uniform distribution in the interval [0, 1]. This way we do not assume any kind of general preference toward a particular value, as preferences simply determine what is subjectively preferred and not what is better or worse. Product properties are represented in the same scale as user preferences, as described by a parameter $q \in [0,1]$.

It is a common assumption in product review communities that a reviewer has previously purchased or experienced the reviewed product. In our model, this experience determines the initial value of the valence as the difference between the agent's preference u_i and the product property q. If a product is at perfect match with a user's preference $|u_i - q| = 0$ (Walter, Battiston, & Schweitzer, 2007), the agent starts with a maximum initial valence ($\upsilon_i(0) = 1$). If the product happens to be the complete opposite to the agent's expectations, the value of the difference between both would be maximum and the agent's valence ($\upsilon_i(0) = -1$).

According to our framework, the value of an agent's expression s_i is determined by its valence υ_i. We assume that agent expressions influence the field more the more emotional they are. As product reviews are fairly long texts compared to other kinds of online communication, sentiment analysis techniques are able to provide values for different degrees of emotionality. A review might contain only factual information and not influence the emotions of a reader, but it could also contain mild or extreme emotional content. The value of an agent's expression s_i ranges from −5 to 5, proportional to the value of its valence when creating the review.

Comparison of simulations with reviews data

Our model for emotions in product reviews aims at reproducing collective properties of emotional expression toward products. Our dataset contains more than 1.7 million reviews from *Amazon.com* for more than 16,000 products. Each review has been processed with SentiStrength (Thelwall & Kappas, Chapter 25, this volume; Thelwall et al., 2010), a sentiment analysis tool that gives values of positive and negative emotions in a text in a scale from 1 to 5. Statistical analysis of this dataset (Garcia & Schweitzer, 2011) showed the existence of two patterns of the reaction of the community to the release of a product. Furthermore, emotional expression regarding products followed distributions of a characteristic shape, which our model should reproduce.

Given a particular set of values for the parameters of our model, the initial value of the communication field determines the type of collective dynamics of a simulation. This way the model is able to reproduce the different scenarios we found in the real data, which correspond to reviews resulting from mass media versus word-of-mouth influence. The first row of Fig. 26.2 shows the time series of emotional expression in two model simulations. The second row shows two example time series for two products of our *Amazon.com* dataset. The left column shows the case when there is a strong input to the field at the beginning of the simulation. This initial impulse, simulating marketing campaigns, forces the dynamics of the community into a vastly decaying single spike. An example of this kind of scenarios in real products is *Harry Potter and the Deathly Hallows*, which was subject to a large amount of media attention around its release due to the fact that it is part of series of already successful books.

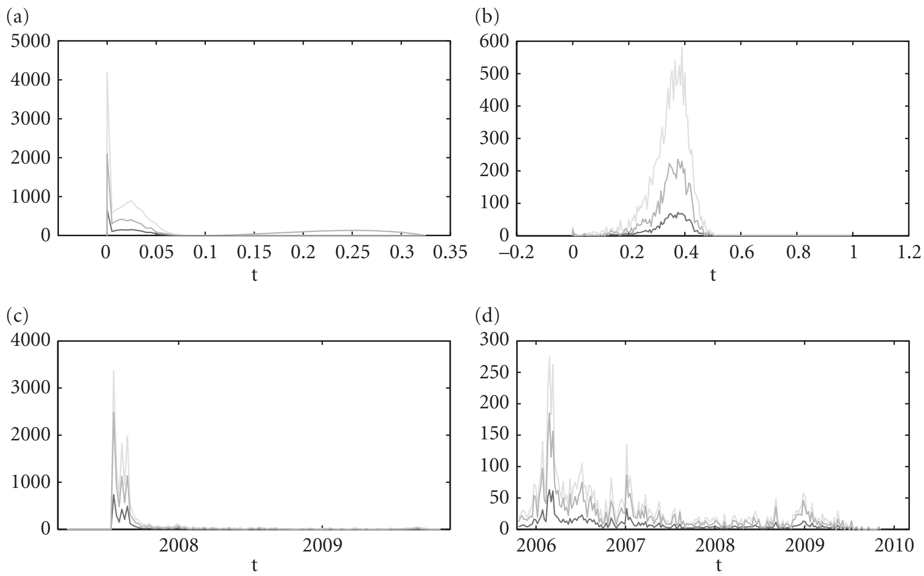

Fig. 26.2 Amount of ratings (black), total positive expression (light gray) and total negative expression (dark gray) for the simulated time. Rate of reviews and emotions for a strong media impulse (a) and when the emotions spread through the community (b). Weekly statistics for *Harry Potter and the Deathly Hallows* (c) and *Marley and Me* (d). © 2011 IEEE. Reprinted, with permission, from IEEE Proceedings, Emotions in Product Reviews—Empirics and Models, Garcia, D., & Schweitzer, F., pp. 483–488. See also Plate 1.

The right column of Fig. 26.2 shows the alternative case of a slower increase of the activity in the community. The simulated time series shows that, in the absence of initial information, the model can build up endogenous cascades of reviews. This kind of dynamics requires a variance of the threshold distribution large enough to trigger some agents that lead the activity in early stages of the simulation. This behavior of our model can be compared with the real example of the book *Marley and Me*, as shown in the bottom right panel. Compared to other books, *Marley and Me* was not subject of an important marketing campaign, and the increase in review activity was due to word of mouth effects among readers. This slower growth also leads to a slower decay compared to the left column of Fig. 26.2, which is also present in our simulations. These results are similar in a qualitative sense, but many questions are still open for future analysis of the dynamics of the model. For example, the model might be able to reproduce different forms of the growth and decay of reviewing behavior, which in turn might be more similar to the ones showed for *Marley and Me*.

The valence dynamics of this model were designed to reproduce different patterns of positive and negative emotional expression in product reviews. The dark gray bars in Fig. 26.3 show a typical histogram of emotional expression in our Amazon.com dataset. In general, the distribution of negative emotions is more uniformly distributed than the expression of positive emotions, which usually have a large bias toward the maximum value. The light gray bars in Fig. 26.3 show the histogram of emotional expression from our simulations. The similarity between both histograms shows how we are able

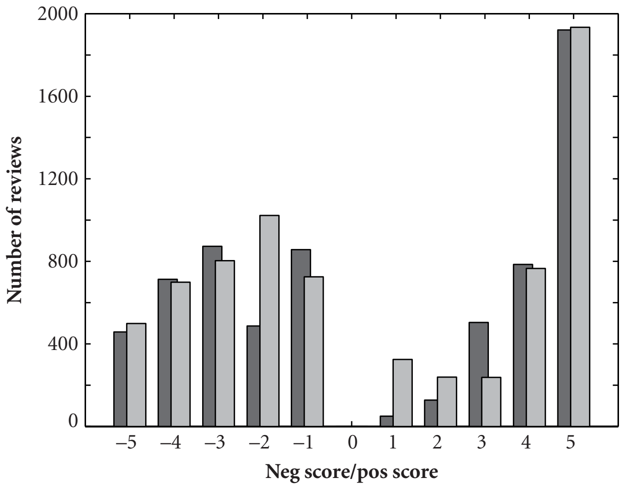

Fig. 26.3 Comparison between the emotional distribution of the reviews for "Harry Potter" (dark gray) and the simulation results (light gray). © 2011 IEEE. Reprinted, with permission, from IEEE Proceedings, Emotions in Product Reviews—Empirics and Models, Garcia, D., & Schweitzer, F., pp. 483–488. See also Plate 2.

to reproduce the distribution of emotional expression in product reviews, given certain parameter values.

To conclude, Fig. 26.3 shows that the outcome of our model has macroscopic properties similar to real world data on product reviews. Our model provides a phenomenological explanation based on psychological principles, linking the microscopic interaction between agents with the macroscopic behavior we observed in our *Amazon.com* dataset. In particular, the different time responses and emotion distributions of the community have the same qualitative properties in model simulations and real data. Within our framework, further explorations of the relation between model and data are possible. For example, each product can be mapped to a set of parameter values that reproduce the collective properties of the community reaction. This would provide a measure of the impact of product properties and marketing in the psychometric space of the customers.

A model for emotions in chat rooms

The second application of our modeling framework aims to provide insights on the nature of human communication in real-time online discussions, for example, chat rooms. Online communication like that in chat rooms has recently received much attention from the scientific community (Garas et al., 2012; Rybski et al., 2009). As a result, many statistical regularities of our communication patterns are revealed, such as the power-law nature of the waiting time distribution $P(\tau)$, where τ is the elapsed time between two consecutive actions of the same user. Such regularities should be, and are, considered in the design of our model. For example, instead of being driven by the arousal dynamics the level of activity is sampled from the real inter-activity time distribution $P(\tau)$: $\sim \tau^{-1.54}$, as reported by Garas and colleagues (2012).

Using our framework, the valence dynamics should follow Eq. (2) and is composed by a superposition of stochastic and deterministic influences:

$$\dot{\upsilon}_i = -\gamma_\upsilon \upsilon_i + b\left(h_+ - h_-\right)\upsilon + A_\upsilon \xi_i \tag{11}$$

The exponential decay of the valence is determined by γ_υ and the influence of the information fields is modeled through $b(h_+ - h_-)\upsilon$. The parameter b quantifies the valence change per time unit due to the discussion of emotional content, which depends on the balance between positive, h_+, and negative, h_-, components of the field. This differs from the previous assumption of Eq. (5) used for the modeling of product review communities, but is more appropriate to capture communication in chat rooms. Chat discussions are usually very fast real-time interactions that display a limited amount of messages to the users, unlike fora in which large amounts of messages can be accessed at any time. In this model, we aim to reproduce plausible chat room interactions, in which users are just able to read a smaller amount of messages created in a short time.

As mentioned before, agents create messages with time intervals sampled from the empirical inter-event distribution. When posting a message, the variable s_i of the agent is set to a value that depends on its valence υ_i. As chat messages are usually very short, we cannot assume the existence of very rich emotional content like in the case of product reviews, but just some emotional orientation as positive, negative, or neutral messages. We formalize the expression of valence polarity as:

$$s_i \begin{cases} -1 & \text{if} \quad \upsilon_i < V_- \\ +1 & \text{if} \quad \upsilon_i > V_+ \\ 0 & \quad \text{otherwise} \end{cases} \tag{12}$$

where the thresholds V_- and V_+ represent the limit values that determine the emotional content of the agent's expression. These thresholds do not need to be symmetric around zero because, as we discuss in the "Measuring the baseline of emotional expression" section, human expression is systematically positively biased (Garcia et al., 2012). If humans communicate in the presence of social norms that encourage positive expression, thresholds should satisfy $|V_+| < |V_-|$.

In this application of our framework, the communication field is formulated exactly as in Eq. (4), for example, it increases by a fixed amount c when an agent expresses its emotions. By analyzing the parameter space of the model, we are able to identify parameter values that reproduce observable patterns of real human communication. Garas and colleagues (2012) have shown that there is a striking emotional persistence in online human communication. This emotional persistence can be reproduced by simulated conversations between agents chatting. It is interesting to note that we only assumed that the inter-activity time τ of agents follows a power-law distribution, to obtain an inter-event distribution for the time lapse between consecutive messages that has the same form as obtained from the real data, but with faster dynamics on a short time range.

Measuring the baseline of emotional expression

The assumptions of our modeling framework are formulated such that they are empirically testable. In this section, we present our results of the empirical analysis of emotional words, which support the assumption of asymmetric thresholds of emotional expression in our chat room model, as discussed in the previous section. This modeling assumption corresponds to the previously formulated Pollyanna hypothesis (Boucher & Osgood, 1969), which asserts that there is a bias toward the use of positive words.

We have tested this hypothesis in the case of online textual communication. Specifically, we analyzed the patterns of online usage of words contained in three established lexica. These lexica contain estimations of the *valence* contents of emotional words in three of the most used languages on the Internet, namely English (Bradley & Lang, 1999), German (Vo, Conrad, Kuchinke, Urton, & Hofmann, 2009), and Spanish (Redondo, Fraga, Padron, & Comesana, 2007). We estimate the word frequency using Google's *N*-gram dataset (Brants

& Franz, 2009), one of the largest datasets available on Internet word usage. Combining these two datasets, we are able to verify the existence of a positive bias in online written expression.

For the three lexica, the valence distributions have mean values very close to zero, as shown in the upper panel of Fig. 26.4. The picture is different when these distributions are rescaled by the frequency of appearance of each word in the Internet. The mean valence of a word chosen at random from online text is considerably larger than zero for all three languages. This has a particular importance for quantitative analysis of the emotions in written text, as the "emotional reference point" is not at zero, but at considerably higher valence values (about 0.3).

The existence of this bias in emotional expression influences the communication process between individuals. For example, the information content of a word is closely related to its frequency of usage. While we are unable to estimate information content perfectly, the

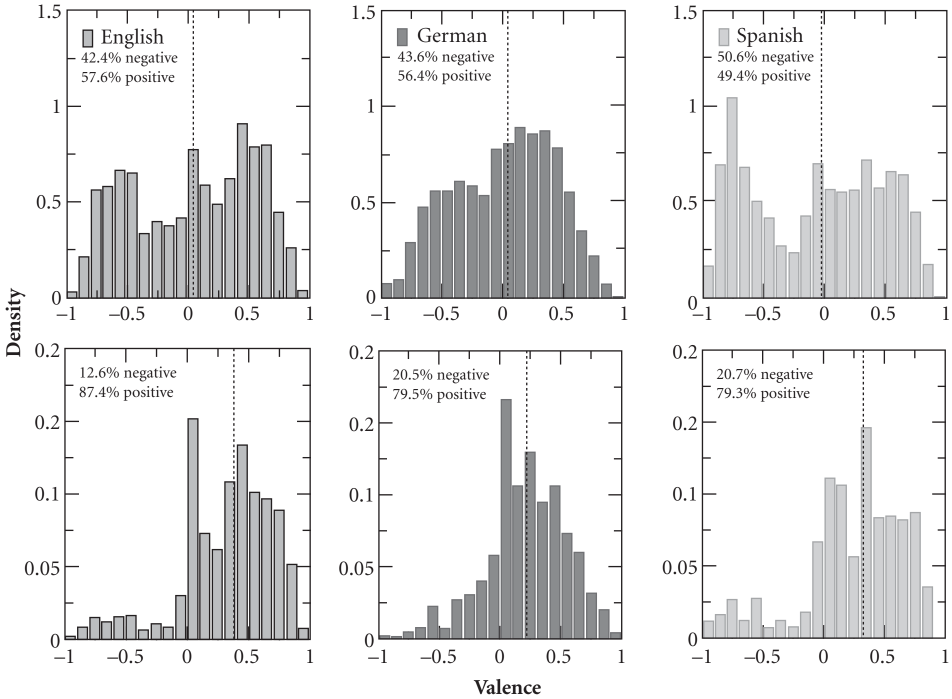

Fig. 26.4 (Upper panel) Distributions of reported valence values for words in English (left panel), German (middle panel), and Spanish (right), normalized by the size of the lexica. (Lower panel) Normalized distributions of reported valence values weighted by the frequency of word usage, obtained from the same lexica. Average valence (median) 0.314 (0.375) for English, 0.200 (0.216) for German, and 0.238 (0.325) for Spanish. The dashed lines indicate the median. Inset numbers: ratio of positive and negative areas in the corresponding distributions. Reproduced from EPJ Data Science, 1(1), 2012, pp. 3, Positive words carry less information than negative words, David Garcia. With kind permission from Springer Science and Business Media. See also Plate 3.

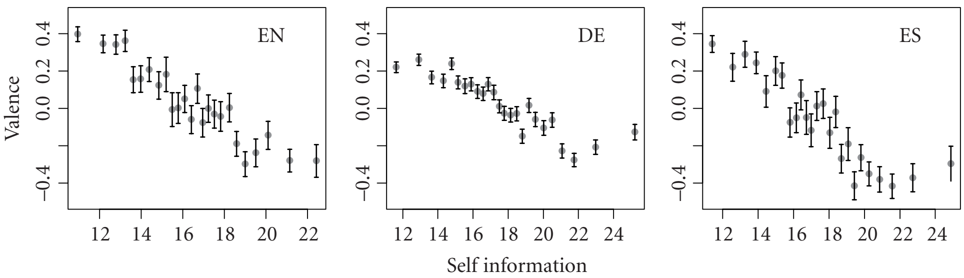

Fig. 26.5 Relation between self-information and valence. Average valence is shown for bins that contain 5% of the data, with error bars showing the standard error. For all the three languages, valence clearly decreases with the self-information of the word, i.e., positive words carry less information than negative words. Reproduced from EPJ Data Science, 1(1), 2012, pp. 3, Positive words carry less information than negative words, David Garcia. With kind permission from Springer Science and Business Media. See also Plate 4.

frequencies of individual words and word sequences let us provide valid empirical estimations. Thus, we estimated the information content through self-information (Garcia et al., 2012), which is defined as $I(w) = -\log P(w)$. Looking at the relation between word valence and information content, we found a consistent pattern in all three languages: the information content of a word decreases with the valence it expresses. Fig. 26.5 shows the clear negative trend in information content, which is supported by correlation measures (correlation coefficients between −0.3 and −0.4). Our analysis extends information measures taking into account the co-occurrence of words up to distance 4, finding similar results.

These results are consistent with previous studies in social psychology, which support that the expression of positive emotions encourages communication and strengthens social bonds (Rimé, 2009). This provides an evolutionary advantage to communities where communication shows a positive bias, as it increases prosocial behavior and collaboration.

Social sharing of emotional content

In our modeling framework, online communication is affected by an information field h that is modified by agents' emotional expression. A message created by an agent might trigger new messages from other agents, depending on the emotional content of the message and the emotional states of the involved agents. We tested this kind of feedback with data from the online microblogging site Twitter, where users create and share short messages called *tweets*. In particular, we tested the influence of emotional content in the social sharing of tweets, by combining sentiment analysis with data on *retweets*, for example, redistributed messages which were previously created by other users.

We analyzed a large dataset composed of more than 30 million tweets processed with SentiStrength in a similar fashion as the product reviews in the previous section (Pfitzner, Garas, & Schweitzer, 2012). Due to the very short nature of tweets, we combined the output of the sentiment classification to generate two values: emotional polarity from the

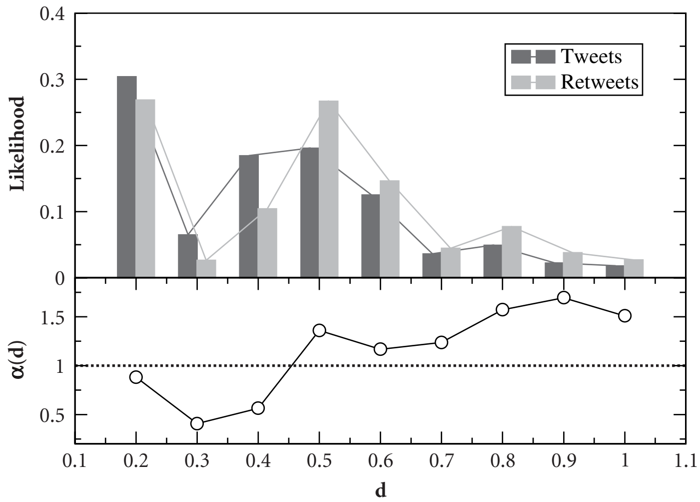

Fig. 26.6 Emotional divergence of tweets and retweets. (Top) Likelihood of tweets and retweets to have emotional divergence *(d)*. (Bottom) Likelihood ratio. See also Plate 5. Adapted from Pfitzner, R., Garas, A., & Schweitzer, F., Emotional Divergence Influences Information Spreading in Twitter, Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, p. 545, figure 2, Copyright (c) 2012, Association for the Advancement of Artificial Intelligence (www.aaai.org), with permission.

comparison of positive and negative content, and emotional divergence d, or strength of the overall emotions expressed through the tweet. We calculate the divergence of a tweet as $d = (p - n)/10$, where p and n are the positive and negative values given by SentiStrength.

Our statistical analysis revealed that most of the emotional tweets are positive, in line with the positive bias discussed in the "Measuring the baseline of emotional expression" section. In addition, the ratios of positive, negative and neutral content are very similar for tweets and retweets. This result suggests that, for the case of Twitter, there is not a clear unbalance in social sharing according to the sign of the valence expressed through a message. However, when looking into the values of divergence between tweets and retweets, the pattern is different. Retweets consistently contain stronger emotional content, although it is not biased toward positive or negative. In particular, the ratio between retweet and tweet likelihoods α increases significantly when emotional divergence is above 0.5, as shown in Fig. 26.6. As an illustrative example, our estimations show that a tweet with $d = 0.9$ has a 14% chance of being shared, whereas a tweet with $d = 0.3$ only 3%.

The interpretation of these results is consistent with the concept that *emotions elicit social sharing* (Rime, 2009), showing that humans tend to share more experiences in which stronger emotions are involved. This analysis provides useful insights that should be taken into account for future models of emotional spread in social networking sites, as a message with strong emotional content is substantially more likely to spread in a community like Twitter.

Conclusion

To summarize, our modeling framework provides the means to understand and predict the emergence of collective emotional states, based on the interaction between individual agents. Its analytical tractability allows us to find conditions in which these states appear and disappear, leading us to the formulation of testable hypothesis of emotion dynamics. We tested some of these hypotheses against datasets of online origin, lending support to the existence of asymmetries in emotional expression. Instances of our models have been proven successful in reproducing collective behavior in product review communities and chat rooms, but our framework has also been used to define an agent-based model for emotional behavior in social networking sites (Šuvakov, Garcia, Schweitzer, & Tadić, 2012), and in virtual human platforms where three-dimensional avatars show facial emotional expression (Ahn, Gobron, Garcia, Silvestre, & Thalmann, 2012). Future applications aim at applying our framework to other types of online communication, such as forum discussions, open source communities, and dialog systems (Rank, Skowron & Garcia, 2013).

Acknowledgments

The research leading to the results discussed in this chapter has received funding from the European Community's Seventh Framework Programme FP7-ICT-2008-3 under grant agreement no 231323 (CYBER-EMOTIONS).

References

Ahn, J., **Gobron, S.**, **Garcia, D.**, **Silvestre, Q.**, **Thalmann, D.**, **& Bulic, R.** (2012). An NVC emotional model for conversational virtual humans in a 3D chatting environment. In *Proceedings of the Articulated Motion and Deformable Objects. 7th International Conference, AMDO 2012* (pp. 47–57). Port d'Andratx, Mallorca, Spain, July 11–13, 2012.

Boucher, J., & Osgood, C. E. (1969). The Pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior*, **8**(1), 1–8.

Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Stimuli, Instruction Manual and Affective Ratings. Technical Report C-1*. Gainesville, FL: The Center for Research in Psychophysiology.

Brants, T., & Franz, A. (2009). *Web 1T 5-gram, 10 European Languages Version 1*. Philadelphia, PA: Linguistic Data Consortium.

Chmiel, A., **Sienkiewicz, J.**, **Thelwall, M.**, **Paltoglou, G.**, **Buckley, K.**, **Kappas, A.**, **& Hołyst, J.** (2011). Collective emotions online and their influence on community life. *PLoS ONE* **6**(7): e22207.

Crane, R., **Schweitzer, F.**, **& Sornette, D.** (2010). Power law signature of media exposure in human response waiting time distributions. *Physical Review E*, **81**(5), 1–6.

Dellarocas, C., & Narayan, R. (2006). A statistical measure of a population's propensity to engage in post-purchase online word-of-mouth. *Statistical Science*, **21**(2), 277–285.

Garas, A., **Garcia, D.**, **Skowron, M.**, **& Schweitzer, F.** (2012). Emotional persistence in online chatting communities. *Scientific Reports*, **2**(402).

Garcia, D., **Garas, A.**, **& Schweitzer, F.** (2012). Positive words carry less information than negative words. *EPJ Data Science*, **1**(1), 3.

Garcia, D., & Schweitzer, F. (2011). Emotions in product reviews—empirics and models. *Proceedings of SocialCom 2011*, 483–488.