

Diss. ETH No. 20699

Modeling collective emotions in online communities

A dissertation submitted to the
ETH ZURICH

for the degree of
Doctor of Sciences

presented by
DAVID GARCÍA BECERRA
M. Sc. ETH Zurich, Computer Science
born July 16, 1985
citizen of Spain

accepted on the recommendation of
Prof. Dr. Dr. Frank Schweitzer, examiner
Prof. Dr. Johan Bollen, co-examiner

2012

Acknowledgments

This thesis is the fruit of three years of joy, effort, and satisfaction that would have been impossible without the support of all those who walked with me along the beautiful but winding paths of Life and Science.

I am very grateful to Frank Schweitzer, who provided the invaluable guidance that allowed me to complete this work, and who shaped my professional and scientific character with each discussion we had. I would like to thank my scientific collaborators, starting with Antonios Garas, whose constant advice was often what I needed to go on. I thank all my colleagues from the Chair of Systems Design, Dorian Tanase, Claudio Tessone, Marcelo S. Zanetti, Ingo Scholtes, Markus Geipel, and all those who provided the best environment for my research. Among the members of the group, I am especially indebted with the student assistants Adiya Abisheva and Petar Tsankov, whose assistance was of great help.

I would like to thank my collaborators in the Cyberemotions project, Arvid Kappas, Mathias Theunis, and Dennis Küster, for teaching me everything I know about psychology. I thank Marcin Skowron, Stéphane Gobron, Ahn Junghyun, and Mike Thelwall for being exemplar fellow computer scientists. And I also thank the physicists Bosilika Tadić, Marija Mitrović, and Janusz Hołyst, whose constructive criticism was always of great help.

But all my efforts would have been completely fruitless without the support of my closest ones. I would like to thank my parents Tomas and Pilar, and my sister Cristina, because everything I have achieved in these early years of my life is thanks to their unconditional love. I cannot forget thanking all my friends, Mehmet, Tizoc, Emre, Olga, Altug, Matteo, Giorgos, Gonzalo, Alberto, and all those I cannot include here and helped me over these three years. And last but not least I thank Brigitte, the best source of inspiration and the strongest support in the toughest moments. Words cannot express the gratitude I feel towards you.

They have all being the best support to complete three years of effort, and a constant source of inspiration behind my contribution. I am not only grateful to them for encouraging me in this endeavor, but also for their hope and trust in my future academic career.

Contents

Table of contents	i
Abstract	ix
Summary	xiii
Introduction	1
Collective emotions on the Internet	3
Analyzing emotions	6
Quantitative analysis of emotional interaction	11
Our approach to collective emotions	16
I Modeling and empirics of collective emotions	19
1 A framework for modeling collective emotions in online communities	21
1.1 A unified approach to collective emotions	22
1.2 An agent-based model of emotions	23
1.2.1 The concept of Brownian agents	23
1.2.2 Emotional states	24
1.2.3 Emotional actions	25
1.2.4 Communicating emotions	27
1.2.5 Emotional feedback	28

1.2.6	Arousal and threshold	29
1.3	Emergence of collective emotions	30
1.3.1	Valence dynamics	30
1.3.2	Arousal dynamics	33
1.3.3	Expression of emotions	35
1.4	Computer simulations of collective emotions	37
1.5	Discussion and applications of our framework	40
2	Collective emotions in product reviews communities	43
2.1	Collective emotions of consumers	44
2.2	The role of emotions in the reviewer paradox	45
2.3	A dataset of emotions in product reviews	47
2.3.1	Valence in product reviews	49
2.3.2	Collective emotional polarity towards products	51
2.3.3	Reviewer activity and emotional expression	54
2.4	User activity in a cultural market	57
2.4.1	Product inequality in featured media	58
2.4.2	Influence among reviewers	61
2.4.3	Word of mouth versus mass media	63
2.4.4	Product success and emotions	66
2.5	Extended analysis of review emotions	68
2.5.1	Positive and negative emotions in product reviews	68
2.5.2	Positivity and negativity towards products	70
2.6	An agent-based model of emotions in product reviews	71
2.6.1	Arousal dynamics	71
2.6.2	Valence dynamics	72
2.6.3	Intensity of emotional expression	73
2.6.4	Simulation of emotions in product reviews	74

2.7	Conclusions from review emotions	76
3	Modeling emotions in chatrooms	79
3.1	The question of online communication	80
3.2	Data from Internet Relay Chats	83
3.3	User activity patterns	84
3.3.1	Individual activity	84
3.3.2	Collective discussion activity	86
3.3.3	Long range interactions	87
3.4	Patterns of emotional expression	89
3.4.1	Measuring emotional persistence	89
3.4.2	Individual emotions	91
3.4.3	Emotions at the discussion level	92
3.5	Modeling persistence in chatrooms	95
3.5.1	An agent-based model for chat communities	95
3.5.2	Simulation of emotions in a chatroom	98
3.5.3	Computational analysis of the model	99
3.5.4	Model limitations	102
3.6	Discussion	104
II	Individual dynamics of emotional agents	107
4	Dynamics of individual emotions	109
4.1	Empirical study of emotion dynamics	110
4.2	Experimental setups	112
4.2.1	Measurements of emotions	114
4.3	Physiology of online interaction	117
4.3.1	The Zygomaticus peak	117

4.3.2	Response to pictures	119
4.3.3	Response to threads	119
4.3.4	Physiological memory of thread experience	122
4.4	Emotion eigendynamics	125
4.4.1	The emotion relaxation equation	125
4.4.2	Eigendynamics parameter estimation	127
4.5	Emotion dynamics under perception of online content	129
4.5.1	Emotional states after reading forum discussions	129
4.5.2	Changes in emotional states due to reading forum discussions	131
4.5.3	Parameter estimation for agent-based models	134
4.6	Antecedents and consequences of expression	137
4.7	Conclusions	141
5	The positive bias of emotional expression	143
5.1	The role of emotional expression in social interaction	144
5.2	Frequency of emotional words	146
5.3	Word valence and information content	149
5.4	Relation between valence and self-information	149
5.4.1	Context-based information measures	152
5.4.2	Additional analysis of valence, length and self-information	154
5.5	Discussion	156
6	Discussion and conclusions	159
	Contributions in a nutshell	160
	Applications of the framework for collective emotions	165
	Emotions in online political campaigns	170
	Relevance in other fields	171

Appendix	177
A Supplementary Material to Chapter 2	179
B Supplementary Material to Chapter 3	181
C Supplementary Material to Chapter 4	183
D Response in the media	187
List of figures	194
List of tables	195
Bibliography	197

Abstract

Abstract in English

Collective emotions in online social systems emerge through the interaction of thousands of Internet users. We analyze their digital traces, i.e. the messages posted in chatrooms, social networking sites, or product reviews communities, by means of statistical methods and sentiment detection, to reveal statistical regularities, or patterns, in their online activity and emotional expression. But collective emotions do not trivially follow from the accumulation of user expressions, and their analysis requires an approach that can tackle their complexity.

We study collective emotions in online communities through a unified approach that combines statistical analysis of digital traces, agent-based modeling of emotional interaction, and empirical analysis of individual emotion dynamics. We provide a modeling framework that links collective emotional states with the interactions between individual users. This framework allows us to define our models according to psychological theories, aiming to reproduce the statistical regularities found in online data. We show that, under testable assumptions, agent-based models within this framework can reproduce states of collective emotional polarization, which can be predicted from analytical results.

We illustrate our approach through the analysis and modeling of two very different online communities, namely product reviews communities and chatrooms. First, we study emotions in the product reviews of [Amazon.com](https://www.amazon.com), showing the effect of mass media and user interaction through emotional expression. An agent-based model within our framework reproduces different types of collective responses to products, and the distributions of positive and negative emotions. We follow by analyzing a large dataset of daily logs from IRC chatrooms, studying the patterns of emotional interaction in real-time online discussions. Our analysis reveals that the activity patterns of these chatrooms do not differ from other

means of communication, and that there is emotional persistence at the discussion level. We designed and simulated an agent-based model that reproduces the emergence of this persistence from a combination of emotion dynamics, and the empirical activity patterns found in the data.

The dynamics of emotions of our models are supported by results on the analysis of data from psychological experiments, including physiological signals and subjective emotion reports. We combine data from four independent experimental setups to analyze the changes in emotions during online interaction, validating some of the hypotheses formulated in our modeling framework. We also develop a technique to study changes in emotions through physiological signals, advancing towards new experiments of emotions in online communities. Furthermore, we find new insights on dynamics of emotions and their influence in online behavior, which can improve future agent-based models of emotions.

Finally, we studied some properties of individual emotions through large datasets of emotional expression. We empirically tested the existence of a positive bias in the usage of emotional words, finding support for this hypothesis on Internet data. By using information measures that take into account word context, we found that, as a consequence, positive words carry less information than negative words.

Kurzfassung auf Deutsch

Kollektive Emotionen in Online Social Systems entstehen durch Interaktionen, welche zwischen tausenden Internetnutzern stattfinden. Mittels statistischer Methoden und Sentimentanalyse analysieren wir digitale Spuren wie z.B. Nachrichten in Chats, Social Networking Sites und Communities zur Produktbewertung, um statistische Regularitäten und Muster in der Onlineaktivität und dem emotionalen Ausdruck von Internetnutzern aufzudecken. Da kollektive Emotionen jedoch nicht einfach Folge der Akkumulation des emotionalen Ausdrucks einzelner Nutzer sind, erfordert ihre Analyse einen Ansatz, welcher ihrer Komplexität gerecht wird.

Daher untersuchen wir kollektive Emotionen in Online Communities mittels einer integrierten Herangehensweise welche die statistische Analyse digitaler Spuren und eine agentenbasierte Modellierung emotionaler Interaktionen mit der empirischen Analyse individueller emotionaler Dynamik kombiniert. Wir stellen ein Modellierungs-Framework bereit, welches kollektive emotionale Zustände mit den Interaktionen zwischen einzelnen

Nutzern verbindet. Unser Framework erlaubt es uns, Modelle auf Grundlage psychologischer Theorien zu definieren. Diese Modelle sollen dann dazu genutzt werden, um in Online Daten gefundene statistische Regelmäßigkeiten zu reproduzieren. Wir zeigen, dass agentenbasierte Modelle innerhalb dieses Frameworks unter gewissen, überprüfbaren Voraussetzungen Zustände kollektiver emotionaler Polarisierung reproduzieren können, welche auf Basis einer analytischen Herangehensweise vorhersagbar sind.

Wir verdeutlichen unsere Herangehensweise durch die Analyse und Modellierung zweier sehr unterschiedlicher Online Communities, nämlich Communities zur Produktbewertung und Chatrooms. Zunächst untersuchen wir Emotionen in Produktbewertungen auf [Amazon.com](https://www.amazon.com) und zeigen den Einfluss von Massenmedien und der Interaktion zwischen Nutzern durch emotionalen Ausdruck. Ein agentenbasiertes Modell innerhalb unseres Frameworks erlaubt die Reproduktion verschiedener Typen kollektiver Rückmeldungen zu Produkten sowie die Verteilungen positiver und negativer Emotionen. Daraufhin analysieren wir eine große Zahl täglicher Protokolle von IRC Chaträumen und untersuchen die Muster emotionaler Interaktion in Echtzeit-Onlinediskussionen. Unsere Analyse zeigt, dass die Aktivitätsmuster dieser Chaträume sich nicht von anderen Kommunikationsformen unterscheiden. Darüber hinaus zeigt unsere Studie die Existenz emotionaler Persistenz. Daraufhin entwickeln wir ein agentenbasiertes Modell welches die Entstehung dieser Persistenz mittels Kombination von Emotionsdynamik und empirischen Aktivitätsmustern reproduziert.

Die Dynamik von Emotionen in unseren Modellen werden durch Ergebnisse psychologischer Experimente unterstützt, welche sowohl physiologische Signale wie auch subjektive Berichte von Emotionen erfassen. Wir kombinieren Daten von vier unterschiedlichen Experimenten, um Emotionsänderungen während Onlineinteraktionen zu analysieren. Hierdurch validieren wir einige der in unserem Modellierungs-Framework formulierten Hypothesen. Darüber hinaus entwickeln wir eine Methode die es erlaubt, Emotionsänderungen durch physiologische Signale zu untersuchen. Hiermit liefern wir die Grundlage zur Durchführung neuartiger Experimente zu Emotionen in Online Communities. Des Weiteren gewinnen wir neue Erkenntnisse über die Dynamik von Emotionen sowie ihren Einfluss auf das Verhalten von Internutzern, welche eine Verbesserung zukünftiger agentenbasierter Modelle versprechen. Schließlich untersuchen wir einige Eigenschaften individueller Emotionen durch große Datensätze emotionaler Ausdrücke. In empirischer Art und Weise testen wir die Existenz eines positiven Bias in der Nutzung emotionaler Worte in Internet Daten und finden Hinweise, welche eine entsprechende Hypothese stützen. Als Konsequenz stellen wir durch die Nutzung informationstheoretischer Maße welche den Kontext von Worten berücksichtigen fest, dass der Informationsgehalt positiver Worte geringer ist als der negativer Worte.

Summary

For each chapter of the dissertation, a detailed summary is presented on the first page.

Introduction We start by motivating the reasons to study collective emotions, and provide an overview of previous research and theories relevant for the topic. We explain our approach to collective emotions in online communities, and how we combine different disciplines.

Part I: Modeling and empirics of collective emotions

A framework for modeling collective emotions in online communities We introduce our agent-based modeling framework, defining agent dynamics based on psychological states and social interaction. We show how models within this framework can reproduce collective states of emotional polarization.

Collective emotions in product reviews communities We study the role of emotions in product reviews communities, testing the influence of mass media and user interaction. We provide an agent-based model for this kind of online communities, and show simulation results that reproduce patterns of emotional expression in product reviews.

Modeling emotions in chatrooms We explore the effect of anonymity in real-time chatrooms, analyzing the activity patterns and emotional expression of public discussions. Our agent-based model reproduces the emergence of emotional persistence in these chat communities.

Part II: Individual dynamics of emotional agents

Dynamics of individual emotions We test the assumptions of our models against data from four independent psychological studies. We perform parameter estimations for the dynamics of emotions under online interaction, and provide results to improve agent-based models of emotions.

The positive bias of emotional expression The combination of datasets of word emotional content and word frequency from the whole Internet lead us to find robust biases of emotional expression. As a result, we measure the relation between the emotional content of words and the amount of information they contain.

Discussion and conclusions The results and contributions of this work are outlined in this chapter. We mention the impact and external applications of our framework, and comment on future research.

Appendix

Some supplementary figures are provided in the appendix, which should be consulted by the readers that are interested in applying the models and results explained in this thesis.

Introduction

“It has been said that man is a rational animal. All my life I have been searching for evidence which could support this.”

Bertrand Russell, *Unpopular Essays*, 1950

The argument of rationality as the fundamental essence of humanity can be traced back to Seneca, who in his letters to Lucilius proposed: *“And what quality is best in man? It is reason; by virtue of reason he surpasses the animals, and is surpassed only by the gods”*. For centuries, the definition of humans was conceived as a dialectic differentiation from animals, but this paradigm changed during the XX century, when humans managed to build the perfect rational agent: the computer. The question of humanity, often phrased as “what makes us different from animals?” was reformulated as “what makes us different from machines?”. As an answer to this question, the old argument of emotions gained weight as the basis of humanity. But the role of emotions in human life is very different from the role of reason, as Plato explained:

“It’s not at all uncommon to find a person’s desires compelling him to go against his reason, and to see him cursing himself and venting his passion on the source of the compulsion within him. It’s as if there were two warring factions, with passion fighting on the side of reason. But I’m sure you won’t claim that you had ever, in yourself or in anyone else, met a case of passion siding with his desires against the rational mind, when the rational mind prohibits resistance.”

Plato, *The Republic*

This quotation highlights the asymmetry in human will between emotions and reason, which consists of some momentum that drives us to follow the dictates of our emotions,

fighting them in order to follow our reason. The main conflict between them is due to the transient nature of emotions, which often can be incompatible with the long-term vision of our reason. It seems rather ethical to assume that emotions are a necessary condition for a human to be considered as such, changing the assumption of pure rationality as *bounded rationality* (Simon, 1997), or subject to certain limitations within the self.

But in spite of their differences, there is an intersection between emotion and logic: human language. Our emotions, our deepest and irrational feelings, can only be connected to our reason by the symbolic space provided by our ability to communicate. Emotions alone would not have led to the creation of human cultures, or civilization, bringing us back to the real difference from animals:

“There seems to be no substance to the view that human language is simply a more complex instance of something to be found elsewhere in the animal world. This poses a problem for the biologist, since, if true, it is an example of true emergence- the appearance of a qualitatively different phenomenon at a specific stage of complexity of organization. In fact, the process by which the human mind achieved its present state of complexity and its particular form of innate organization are a total mystery... It is perfectly safe to attribute this development to natural selection, so long as we realize that there is no substance to this assertion, that it amounts to nothing more than a belief that there is some naturalistic explanation for these phenomena.”

Noam Chomsky, *Language and Mind*, 1998

The phase transition that separates the complexity of animal languages from that of human languages would be one of the essential, and still greatly unexplained, features of humanity. Under this view, most human action a person can do is the expression of emotions. The social sharing of emotional experiences, the empathy of people towards the suffering of others, the creative conflict that improves a situation through the expression of anger and frustration... All these instances of emotional interaction would be examples of what turns our society into a *human* society.

The complex emotional interactions between individuals can lead to the emergence of collective states, in which large amounts of humans can share an emotion. Such collective emotions are of extreme importance to human societies, as they are one of the sources of the *ethos*, i.e. just as individual emotional experiences shape the ethics of a person, collective emotions reshape the ethics shared by a whole society (Alexander *et al.*, 2004). Ideology, as a collective construction on individual subjectivity (Zizek, 1989), makes emotions one

of the key components in understanding human societies. But, how can we explain the emergence of emotions at the collective level from interactions between individuals? This question has been addressed in the field of social psychology (Bar-Tal, 2006), calling for an interdisciplinary approach to understand this process. In this thesis, we want to show our steps towards the understanding of collective emotions, in particular in online communities, by a combination of theories and techniques from psychology, physics, and computer science.

Collective emotions on the Internet

According to (Bar-Tal, 2006), collective emotions are shared by large numbers of individuals, in contrast to group-based emotions that are felt by individuals as a result of their membership to a certain group or society. The former concept suggests that group members may share the same emotions for a number of different reasons, whereas the latter refers to emotions that individuals experience as a result of identifying with their fellow group members. In this general sense, collective emotions are not restricted to online communities. Instead they can emerge in any social context. The aim of our research, however, is to understand their dynamics in *online* communities.

Examples of collective emotions in online communities can be observed en masse on the Internet. One particular example was the large amount of emotional discussions which followed the death of Michael Jackson, other examples are the "memes" and heated discussions in anonymous fora like 4chan.org. For the case of the death of Michael Jackson, we can say that an external event (the death of a celebrity) triggered a large amount of emotional expressions (messages of grief in social networks). This was a large-scale emotional event, with the simple explanation that lots of users simultaneously shared similar emotions. But not all the collective emotions have such a simple explanation, and they can emerge from complex interactions among the members of an online community. An illustrative example of this kind of *complex* collective emotions is the events following a particular video in the videoblog of the Youtube user "boxxybabe", uploaded in early 2009¹. That seemingly uncontroversial video was subject to a set of discussions in various fora, in which users divided into lovers and haters of the character shown in the video, known as *Boxxy*. The emotional expression of the involved users quickly escalated to a collective emotion that spanned few different online communities, and that drove an amount of traffic to boxxybabe's channel that made it the most subscribed channel in Youtube during January 2009. The attention to this channel dropped quickly, but traces of this

¹<http://knowyourmeme.com/memes/boxxy>

polarizing collective emotion are still found in online media, specially through the usage of images like the ones shown in Figure 1.



Figure 1: Set of images shared online due to the collective emotion generated around *Boxxy*, popularly regarded as a *meme* (Leskovec *et al.*, 2009). These images were created by anonymous users of various online communities.

These collective emotions in online communities follow a very similar scheme: Users which have subscribed to social networking sites or to blogs or discussion fora, become enraged or excited about a particular event (like the performance of a beloved soccer team in a world competition) or a personal (good or bad) experience. Importantly, these individual feelings are then shared with other users, i.e. they are *communicated* by means of online media, most likely by writing a personal statement. Obviously, users do not transmit an emotion, instead they communicate a piece of information, which may trigger an emotional reaction in participants reading it. Dependent on such an impact, other users may decide to involve themselves in such an emotional communication, e.g. by sharing the feeling or opposing to it. Under certain circumstances, we may observe mutual communication in a small group of users, but there are also scenarios where many users express their feeling once, in a sequence, or where many users repeatedly fire the discussions by emotional statements. These discussions show the existence of emergent collective states in which the users share their emotions, rather than an aggregate of the emotions of the community. The discussions do not necessarily have to be centered around just one kind of emotion. In many cases, we see the emergence of two collective emotions, a 'positive' and a 'negative'

one, which may coexist or 'fight' each other, as it was for the example of *Booxy* mentioned above. These collective states usually only have a finite life time, i.e. they disappear, but could come back when triggered by a new event or post.

The cyberspace does not have emotions, but individuals that interact online can share emotions. Here, certain conditions of individual interactions apply that are not present for offline communities. Online communities react on other time scales (not necessarily faster, but often with a time shift), they act on different stimuli (there are hardly any offline assemblies that share emotions about a Youtube video), they have different thresholds to express emotions, and they do it in a very different manner, namely by writing a text. Collective emotions are fostered by Internet communication because of (a) the fast information distribution, and (b) the anonymity of users in the Internet, which often seduces people to drift away from established norms and show a salient private personality. In fact, empirical studies (Sassenberg and Boos, 2003) which compared the attitude change in virtual and face to face interactions, have demonstrated that human behavior and social norms are affected by Internet interaction.

The fundamental differences between online and offline interaction make the former an idoneous ground for the emergence of collective emotions. Granovetter's theory of weak ties (Granovetter, 1983), tested in offline and online social interaction, highlights the importance of social connections with less interaction strength but larger in amount, as the medium for fast information spreading. Compared to strong ties in the schema of Figure 2, weak ties are very present in online communication (Szell and Thurner, 2010), as the cost of maintaining online social links is much less than their offline counterpart.

Online communication also allows the existence of a weaker form of interaction, present between users that do not even know each other's names. These weaker ties can still lead to non-negligible influence between users, but they eventually allow large sets of users to interact in a very fast way. For example, a set of users that never interacted before can engage in a heated discussion in a forum in which they are either anonymous, or their identities are irrelevant. This kind of interaction can take place many times within an online community, and is not defined by explicit social ties, but by user interests and conditions of online activity. Online social media serves as a technology to overcome the cognitive limitations that lead to the absence of such numerous and pervasive weak ties in offline interaction, but their *de facto* anonymity might imply changes in user behavior. For example, users can tend to exaggerate or use stronger language to attract the attention of others, while they would refrain from doing so in offline interaction.

We remark that the Internet is indeed shaping the phenomena in mind, it is not a mere interface for monitoring 'real' social interaction. While we agree that there are certain

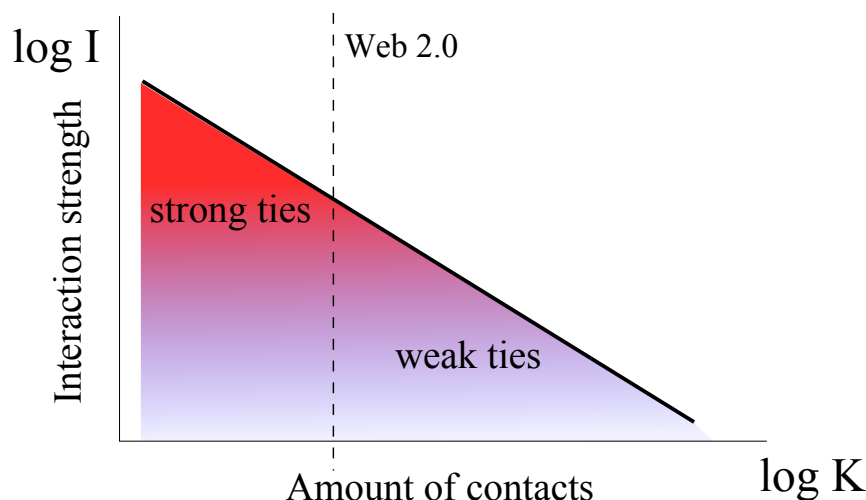


Figure 2: Schema of the types of social ties. Vertical axis represents interaction strength, and horizontal axis the possible amount of contacts associated with such strength. Both are represented in an approximated logarithmic scale.

commonalities (mostly based on social herding and amplification), there are also substantial differences in communication. For the very same reason, we argue that real phenomena like mass hysteria can indeed be seen as instances of collective emotions, however, data and models of online interaction might not be valid in that case.

Analyzing emotions

Emotions are one of the most relevant aspects of anyone's life, often escaping to the explanations of our own reason. Introspection is often not an illustrative experiment, as our perception of an emotion eliciting situation is doomed to be biased by our own subjectivity. The same event can evoke very different emotional responses, leading to heterogeneous understanding of the same phenomenon. Our own subjectivity covers emotions with an aura of mysticism that hides them from our own rationality, historically leaving them far from the necessary quantization of the scientific method.

Humanities and social sciences were the first disciplines with the ability to explore the previously dark field of human emotions, providing many theories and questions about our emotional experiences. Such theories have been very difficult to prove, due to the difficulty to gather experimental data on human behavior. Empirical data on humans is rather scarce and limited compared with the data usually present in other disciplines like physics, chemistry, or biology. For example, a physicist is free to extract any feasible amount of information from a subatomic particle, but a psychologist faces great limitations

and ethical issues when testing with human subjects. It would be difficult (and expensive) to find a sufficient amount of experiment participants that agree on being monitored 24 hours a day, with complete loss of privacy. Here, we outline how our research is framed in a wider context of emotion research in psychology, sociology, and computer science.

Representation of emotions

Emotional life is inherently complicated, being very difficult to study simultaneously at all levels. The interaction between personality, cognition, subjective states, and other individuals makes it unfeasible to have an empirically testable theory of the whole phenomenon that is affect. For this reason, psychology has divided emotional experience in different levels, which allows the analysis of certain features of emotional life. From the fastest to the slowest in time scale, these are:

- **Reflex reactions.** Human emotions serve as very fast reactions to react to changes in the environment, such as threats or opportunities. Many of these reactions have a direct equivalent in neural activity, such as a fear reaction due to pain.
- **Core affect,** defined as psychological states of high relevance for the individual that are not targeted and have a physiological manifestation. These states relax quickly in time at the expense of large amounts of energy of the organism, and often do not imply the full awareness of the individual.
- **Mood and feelings.** At a slower time scale, emotions can interact with human cognition to create sustained states of mood, or targeted perceptions associated with events, known as feelings. Mood and feelings have a much more complicated nature than core affect, and are usually represented and studied in a different manner.
- **Personality traits.** Lifelong conditions of human behavior that are heterogeneous among individuals. These can be related to emotions, as in the case of traumas, phobias, or expression patterns. Their change evolves through life and is usually related to psychological development or relevant events.

In this work, we fix our level of analysis and modeling to the phenomenon of *core affect* (Russell, 1980), as it is the most granular emotional state that can be expressed through a text. In particular, core affect has the following properties:

- States of core affect are exceptional due to their relevance for the individual, and require fast reactions and decisions. There is a strong relaxation of these emotions, revealing them as **transient** states.

- Core affect is **not directed**, i.e. there is no clear object attached to the emotional state. We can say that an event induces a state of core affect, but it requires higher levels of cognition to associate emotional states with their possible sources.
- Emotional states of core affect are **atomic** and **disjoint**, i.e. an individual cannot be in a superposition of states of core affect, and these cannot be divided in discrete substates.

In this work, we will use the term *emotion* to refer to core affect, being the basic object of our study. The choice of core affect as theoretical background allows the framing of our research in a wider range of psychological studies, which can be integrated with our work. After defining this level of analysis, a key issue is to choose the model to represent internal emotional states, for which there are competing approaches:

- **Ekman’s basic emotions.** Inspired in Darwin’s work on the facial expression of emotions (Darwin, 1872), Paul Ekman designed the model of 6 basic emotions: anger, disgust, fear, happiness, sadness and surprise (Ekman, 1972). In the context of facial expression of emotions, Ekman’s basic emotions are often referred as the “universal” model, but this universality, even restricted to facial expression of emotions, is currently under question (Jack *et al.*, 2012).
- **Russell’s circumplex.** Focusing on the concept of core affect, James A. Russell developed the circumplex model of affect (Russell, 1980). In this model, emotions are represented in a two dimensional space of *valence*, measuring the pleasure related to the emotion, and *arousal*, i.e. the degree of activity associated with the emotion. Initial empirical works suggested an arrangement of these emotions on a circumference, but they can be extended to a plane defined by both dimensions (Feldman Barrett and Russell, 1999; Scherer, 2005).
- **Watson and Tellegen’s model.** Two independent variables represent emotions in this model, one for positive affect and one for negative affect (Watson and Tellegen, 1985). This way, affective states can be represented as combinations of positive and negative feelings. This model was developed to provide a representation of mood (Watson and Tellegen, 1985; Tellegen *et al.*, 1999), in which complex emotional states can be composed of positive and negative feelings towards different entities.

The most valid representation for core affect is Russell’s circumplex model, which is the representation of emotions that we will adopt for our work. It is important to mention that, after all, these two dimensions are a projection of a the more complex phenomenon

of affect, and there are alternatives to these particular dimensions (Feldman Barrett and Russell, 1999; Yik *et al.*, 1999). Empirical studies show that four dimensions can represent better the variability of emotional experience (Fontaine *et al.*, 2007), but, as we explain below, data analysis techniques are not advanced enough to cope with extra dimensions of emotions (if possible at all). It is important to note that this model is under constant refinement with new experiments improving our understanding of core affect (Yik *et al.*, 2011). Furthermore, it is commonly accepted within the emotion research community, chosen as representation model for emotion lexica (Bradley and Lang, 1999), experiments with emotional pictures (Bradley and Lang, 2007), reports on emotional experiences (Scherer, 2005), monitoring of emotions in everyday life (Kuppens *et al.*, 2010), and psychophysiology of emotions (Kappas *et al.*, 2011).

The social component of emotions

Apart from defining the emotional states of individuals, we need to analyze the social interaction between the users of an online community. In this sense, emotions are internal states that are transmitted to other users through online communication, leading to changes in emotions and subsequent behavior of other users. Social Impact Theory (Latane, 1981) provides a way to measure this kind of social forces, by the quantification of strength, immediacy, and amount of people influenced in a social situation. This theory provides a very general view of social interaction, which does not need to involve emotions. While still valid for the case of online interaction, Social Impact Theory does not provide a specific approach to emotional expression. The relation between emotions and social action can be studied through the concept of Granovetter's threshold model (Granovetter, 1978). In this approach, the individuals involved in a collective social situation would take action according to the strength of social interaction, an internal variable, and a threshold that determines activity. The members of a community would have their internal variable increased by the strength of the interaction with the community as a whole, and when this variable reaches the predefined threshold, the individual is activated.

In a more specific setup, social networks impose certain heterogeneity among the members of an online community, which is taken into account in the cascading process framework of (Lorenz, 2009). While defined for systemic risk, this framework provides a way to analyze how network topologies can influence the spreading of behavior within the threshold model. Different spread rules can be defined, and among them we can analyze non-conservative processes like those present in social interaction. More details on how we use the threshold approach are explained in Chapter 1.

It is acknowledged that the Internet is an important factor in defining present and future societies (DiMaggio *et al.*, 2001), opening relevant questions for sociology. Collective emotions, such as hate, play an important role in the creation of collective identities. Studies in sociology (Adams, 2006) show how emotions are related to group identity through the systematic analysis of Internet-based hate groups. Furthermore, emotions and computer mediated communication have an influence in workgroup efficiency (Flache, 2004), in particular in intensively virtualized work environments. Within sociology, Affect Control Theory (ACT) (Heise, 2006) provides a unifying framework to integrate these empirical observations of emotional interaction. ACT provides a mathematical formulation for the analysis of interpersonal relations, modeling emotional communication through Osgood's semantic differential (Osgood, 1964). Within this theory, any emotional interaction is composed of a set of equations driving the states of the involved users, for which parameters estimations on the empirical data provide plausible explanations (Clay-Warner and Robinson, 2008).

The focus of ACT is to explain the microdynamics of social interaction under the presence of emotions, including a wide range of features like dominance between participants, work roles, family relationships, etc. Most of these characteristics are not measurable, or even present, in online communication, where interaction is often simplified to texts exchanged between users. Our purpose is to explain emotions at the macroscopic level from microscopic interactions (Bar-Tal, 2006), and ACT focuses on interactions between very small and controlled groups of individuals. In general, the usage of ACT is focused towards the explanation of sociological factors rather than human psychology and emotional states. The application of ACT depends on cultural or economical factors between participants, limiting its application for generalized emotion psychology.

Empirical studies in social psychology provide results on how emotions influence social interaction and communication. Numerous works show how emotions elicit social sharing (Rime, 2009), usually looking for functional dependencies between emotional states and social interaction. Some of these results provide an initial ground for the analysis of online emotional interaction. Short stories are more likely to be shared when they carry stronger emotional content (Heath *et al.*, 2001), and this social sharing spreads through social networks in a phenomenon called *secondary social sharing* (Christophe and Rime, 1997). A particularly interesting insight from social psychology is the existence of a *step function* on the relation between emotional intensity and tendency to social interaction (Rime *et al.*, 1998), which is consistent with the threshold models mentioned above, i.e. internal emotional states would lead to the expression of emotions online if their intensity surpasses a given threshold. Our modeling and analysis will be in line with these formulations in social psychology and sociology, aiming at the integration with experimental results.

Quantitative analysis of emotional interaction

To understand collective emotions, we follow a quantitative approach that requires three tools: i) the ability to extract emotions from text, ii) large datasets of emotional interaction in online communities, and iii) a mathematical formalization of the interaction between users in an online community. It is important to emphasize that our work is focused on *emotions* and not on *opinions*, which are also subject of quantitative analysis and modeling. Emotions and opinions are both subjective states of high relevance for the individual, their expression overlaps and is usually associated within the semantic differential (Osgood, 1964; Bradley and Lang, 1999). Apart from these similarities, they differ in very important aspects of their nature:

- Emotions are transient states that require high attention, while opinions last very long and remain latent in the individual under the absence of interaction. This manifests in the relaxation of emotions, which lowers the intensity of emotional states through time (Scherer, 2004). Opinions are not subject of such relaxation, and there is no clear time trend towards more neutral opinions under the absence of stimuli.
- The process of negotiation and consensus formation might lead to a “canceling out” of opinions, in which a final decision is taken by averaging the opinions of individuals. This collective dynamics are commonly absent in emotions, as emotional interaction does not necessarily mean that the emotional states of participants tend to converge. A rational discussion between two individuals with different opinions can be assumed to drive them closer, but this is not the case for the confrontation of emotions.
- While both emotions and opinions can be influenced by herding effects, opinion dynamics is often linked to utilities and preferences, while emotions do not need to follow any particular optimization space.
- Emotions have a physiological manifestation, while opinions are present in the cognitive configuration, or the ideas of the individual.
- In the context of core affect, emotions are very simple states that can be briefly summarized in one or two words, while opinions imply complex thoughts and reasoning, often including conditional statements.
- The opinions of a society can be addressed through the concept of “public opinion”, but there is no clear concept of what is the “public emotion”. On the other hand, we talk about collective emotions as transient states shared by the members of a

community, but these cannot be externalized in a super-entity like the society or the state.

Assuming these fundamental differences, we can explore previous works on the quantization of subjective states, looking for tools and models that can be combined with our work in online collective emotions.

Sentiment analysis

Sentiment analysis provides tools to solve the problem of the mining of subjective states expressed through text (Pang and Lee, 2008). These tools are usually tailored for their application in particular contexts, types of verbal expression, and aspects of sentiment. Many tools and models focus on the case of opinions, in general with supervised techniques that are trained on corpora of annotated texts. Given the differences between opinions and emotions explained above, we focus on the application of tools that aim at the analysis of emotional content from text, rather than opinion mining.

Emotion analysis from text is often implemented through unsupervised techniques, i.e. lexicon-based approaches (Taboada *et al.*, 2011) that use previous knowledge of emotional expression in the form of annotated terms. The reason for not adopting a supervised approach is the absence of a ground truth of emotional expression. A supervised approach requires a set of training tests which are assumed to have perfect tags or classifications, but this does not need to be the case for the perception of emotions from text. There is a non-negligible inconsistency across individuals in the emotions they perceive from a text (Thelwall *et al.*, 2010), difficulting supervised approaches. In particular, the *accuracy* of a tool for emotion analysis is not as relevant as its *validity*, and its performance should be addressed through how well it reproduces the response of human reports and their heterogeneity.

The first and most naïve approach is the application of word frequency techniques, which use previous lexica of emotions to estimate aggregated measures over the text. The ANEW dataset (Bradley and Lang, 1999) has been used to extract valence from texts (Dodds and Danforth, 2009), given the constraint that they have sufficient length. The emotional classes of LIWC (Pennebaker *et al.*, 2001) can be used to estimate the frequency of terms expressing positive or negative affect, which require very large texts to provide an estimation of emotional content. They have been used to analyze a massive dataset of **Twitter** messages (Golder and Macy, 2011), studying the evolution of aggregated mood rather than the dynamics of core affect.

The tools for emotion analysis mentioned so far use general emotion lexica that were not designed for their application in sentiment analysis. More advanced tools use lexica that were designed and tested for a specific tool, allowing the assessment of their accuracy when compared with human annotated texts. For example, OpinionFinder (Wilson *et al.*, 2005) includes a set of indicators associated with emotions, and has been used for the analysis of mood and emotional well-being in **Twitter** messages (O'Connor *et al.*, 2010; Bollen *et al.*, 2011b,a). A competing tool of this kind is SentiStrength (Thelwall *et al.*, 2010), which extracts positive and negative emotions from text. It combines a lexicon of emotional words with rules of negation, amplification, and diminishing of emotions. SentiStrength has been recently used for the analysis of emotional expression in **Yahoo answers** (Kucuktunc *et al.*, 2012), and **Twitter** trends (Thelwall *et al.*, 2011), and its accuracy and validity have been tested across texts from different social media (Thelwall *et al.*, 2012).

Computational social science

The Internet serves as a constant experiment of the largest scale, producing data that can be automatically processed. In a matter of few years, the amount of empirical data on human activities has grown few orders of magnitude in scale and resolution, allowing the quantitative testing of previous hypothesis impossible to validate before (Lazer *et al.*, 2009). Despite of its novelty, this empirical analysis has its limitations, often related to demographic and geographic biases. The potential of these new sources of data grows with each new study (Giles, 2012), in spite of the voices that raised against the quantization of social activity and human sciences (Konnikova, 2012).

In the last years, lots of articles have shown how online data can be used to study human behavior, leading to very relevant results. Computational social science allows the prediction of popularity in online participatory media (Szabo and Huberman, 2010), the study of trend and attention dynamics in social media (Wu and Huberman, 2007; Wang and Huberman, 2012), and testing mechanisms of influence among individuals in online cultural markets (Hedström, 2006; Salganik *et al.*, 2006). Data from **Youtube** shows the different types of collective responses to a video (Crane and Sornette, 2008), which is relevant for viral marketing in the context of information spreading in social networks (Leskovec *et al.*, 2007). Sociological theories have been tested, like Dunbar's number (Goncalves *et al.*, 2011), or the strength of weak links (Szell and Thurner, 2010). Mobile phone data allows the analysis of priority patterns in human communication (Wu *et al.*, 2010), and the emergence of social influence in online interaction can be measured through **Facebook** application installations (Onnela and Reed-Tsochas, 2010), and movie reviews

from `imdb.com` (Lorenz, 2009).

Regarding emotions, most of the studies within computational social science focus on happiness and mood, which are aspects of emotion that do not change as quickly as core affect. Longitudinal analysis of offline social network data shows how happiness is transmitted through the network (Fowler and Christakis, 2008). The network approach to this data allows the study of this infection process (Hill *et al.*, 2010), which would remain unnoticed under a normal survey study. The aggregated happiness of a large amount of Internet users can be measured through data from blogs and **Twitter** (Dodds and Danforth, 2009; Dodds *et al.*, 2011), highlighting the influence of large-scale events in emotional expression. Furthermore, sentiment analysis on **Twitter** data allow the analysis of subjective well-being in this social network (Bollen *et al.*, 2011a), revealing the existence of an assortativity pattern of happiness. With respect to mood, a recent study shows the daily pattern of mood expressions in **Twitter** (Golder and Macy, 2011), showing an increase in positive expression in the morning and in the evening.

The topic of core affect remains less explored in computational social science, as sentiment analysis tools for transient emotional states are a recent development. Initial works could overcome that problem through the manual annotation of blog posts (Sobkowicz and Sobkowicz, 2010), with the important limitation and lack of applicability of such approach. Two independent studies of emotional expression in blogs (Sobkowicz and Sobkowicz, 2010), and (Chmiel *et al.*, 2011b), show that negative emotions drive user interaction in these communities. Further analysis of blogs, digg and BBC fora (Chmiel *et al.*, 2011a) showed the clustering of positive emotions and that discussions that start with emotionally charged messages lead to longer threads. A recent work based on **Twitter** (De Choudhury *et al.*, 2012) extracts the emotions in a text through a machine learning technique trained over texts that contain explicit assessments of the emotions of the writer. Applying it to a large scale dataset from **Twitter**, it was possible to measure the frequency of expression of emotional states in Russell's circumplex model (Russell, 1980).

Agent-based modeling

How do collective phenomena arise from the interaction of many distributed system elements? This question is certainly at the heart of statistical physics. Over the last 150 years it has provided a large set of methodologies applied to physical systems, to infer from the properties of the elements on the micro level on the systems dynamics on the macro level. A very similar question is also asked in different other scientific disciplines. For example, in medicine one wishes to understand the reaction of the immune system

based on the communication and coordinated action of e.g. B or T cells. In economics one is interested in the emergence of systemic risk (Lorenz *et al.*, 2009) in a financial system based on the fault of firms or banks clearing their debts to other firms or banks.

Agent-based modeling is one of the most successful approaches to find the connection between phenomena at the microscopic and macroscopic levels of a system (Schweitzer, 2003). Agent-based modeling is widely applied to the analysis of social phenomena, including cultural dynamics, language adoption, crowd behavior, and opinion dynamics (See (Castellano *et al.*, 2009) for a review). Despite of the differences between opinions and emotions explained above, the case of opinion dynamics provides some background to our work, as both are subjective states influenced by social interaction. The conditions for the emergence of consensus or dissensus in social systems can be understood through agent-based models, in which herding effects and social interactions modify individual opinions. Bounded rationality models, surveyed in (Lorenz, 2007), include certain limitations to opinion dynamics that might be present in emotional interaction, but most of these models do not include any internal relaxation of internal states, which would be necessary for their application to emotions. One notable exception that includes opinion relaxation is the information accumulation model (Shin, 2009), which leads to a nontrivial lower probability of consensus between two communities when they are more connected (Shin and Lorenz, 2010). While this model gives some interesting insight on the effect of internal relaxation of subjective states in the collective result of social interaction, we still require a more precise description of emotion dynamics to represent collective emotional states.

Some models in psychology aim at the dynamics of internal emotions in the absence of social interaction. Appraisal theory (Scherer *et al.*, 2001) provides another promising theoretical perspective: It is based on internal representations of person-environment relations, which can be modeled by so-called BDI (belief-desire-intention) agents. While there is a large body of literature on computational appraisal models (Gratch and Marsella, 2004; Gratch *et al.*, 2009; de Melo *et al.*, 2009) the focus is more on the correct internal representation of emotions and their cognitive consequences, not on the explanation of collective phenomena such as described above. Alternative models include principles of dynamical systems to model emotion dynamics, either within appraisal theory (Sander *et al.*, 2005), or focused on the internal dynamics of emotions without explicit appraisal (Kuppens *et al.*, 2010). This last model of internal dynamics leads to testable formulations of emotion dynamics in terms of valence and arousal, and initial empirical results serve a starting point of our modeling work. These models completely lack of a formulation of the dynamics of social interaction, and thus cannot be directly applied to the question of collective emotions. Nevertheless, they provide an initial understanding on the behavior of individual agents.

Social psychologists are starting to acknowledge the potential of the application of agent-based modeling to social phenomena (Smith and Conrey, 2007). Some initial works on agent-based modeling of emotional interaction focus on the case of modeling the dynamics of negative emotions in forum discussions (Chmiel *et al.*, 2011b), proposing a computational model based on event probabilities. Other models simulate the flow of emotions in artificially generated social networks (Chmiel and Hoyst, 2010), or square lattices (Czaplicka *et al.*, 2010; Czaplicka and Hoyst, 2012). These models can reproduce certain aspects of group emotions, but face certain limitations when applied to collective emotions on the Internet. First, results on these models are based on computer simulations or analytical techniques, but their macroscopic behavior does not have a quantitative correspondence with phenomena observed in online data. Second, agent states are not represented by psychological variables, and the assumptions of agent behavior are not empirically testable in experimental psychology.

Our approach to collective emotions

Given the models and analytical approaches mentioned above, we aim to avoid certain limitations of their application to collective emotions in online communities.

- As some models only define individual dynamics, they lack an explicit formulation of the interaction between individuals. Given that our target is to explain complex collective phenomena, we cannot restrict our models to formulate individual behavior in isolation.
- We cannot directly apply models of opinion dynamics, due to their focus on a phenomenon that is very different from emotions. Certain concepts of emotion dynamics, as part of a wider theoretical approach, are supported by empirical studies and not present in opinion dynamics. In particular, models of emotions and opinions are not equivalent because they do not represent short lived internal states, nonrational behavior, or the stochasticity and nonlinearity of emotions.
- Computational models of emotions can reproduce certain macroscopic behaviors, but often cannot be mathematically analyzed to understand the outcome of simulations (Rank, 2010). We want to have a tractable approach that provides an understanding of collective emotions beyond their reproduction in artificial simulations.
- We want to allow the integration of results from models of emotion dynamics within a wider scientific scope. Most of the current models of emotional behavior are based

on hypotheses that cannot be tested in psychological experiments, and agent states are not based on previous psychological theories.

- Current models of human behavior on the Internet are very dependent on the mechanisms of interaction in the particular community, and cannot be compared across communities. This limits their application and the relevance of any results derived from their analysis.

We aim at overcoming all these limitations through an integrated approach that combines the analysis of emotions in three levels, shown in Figure 3.

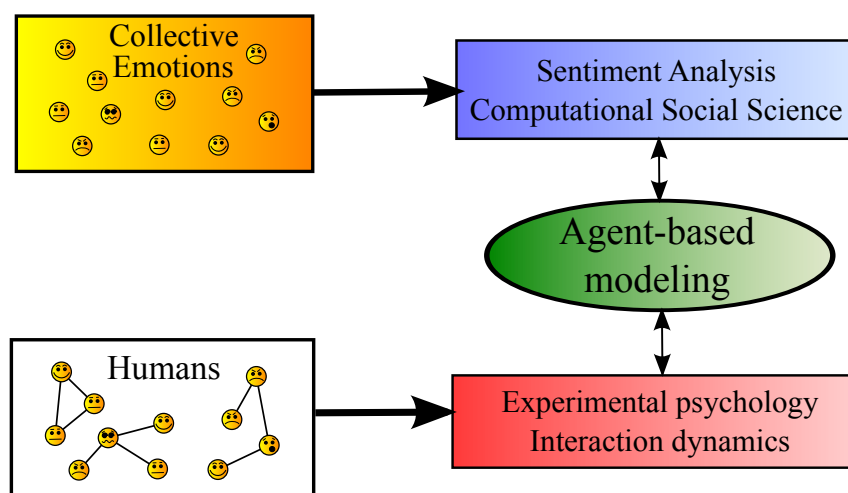


Figure 3: Outline of our approach to collective emotions. We analyze collective behavior in online communities, and analyze it with sentiment analysis and computational social science techniques. We use data from psychological experiments on individual dynamics of emotions, to support the dynamics of agent-based models that reproduce the stylized facts found in data from online sources.

Analysis of online data Large datasets from online communities can provide comments, posts, and messages that can express the emotions of the members of the community. Given the language, length and formality of the texts of an online community, we can select a sentiment analysis tool to extract a quantization of the emotions expressed in each text. These emotional values are expressed by a particular user, have a timestamp, and might be part of a conversation or relate to a particular topic. We apply techniques from computational social science and statistical physics to derive *stylized facts*, i.e. statistical regularities that characterize collective emotions.

Agent-based modeling of emotions Before modeling individual online communities, we define a modeling framework that i) formulates agent states in terms of core affect, ii) has tractable dynamics that allow analytical solutions, and iii) keeps a flexible approach that allows its application to different online communities and is robust to the details of the output of sentiment analysis. This framework is aimed at unifying different models that reproduce collective emotions in a variety of online communities, which should be validated against the stylized facts found in the analysis of online data. If correctly phrased in terms from psychology, this framework can provide testable hypotheses that can be validated with data on individual emotions, serving as a link between the macro and micro levels of emotional behavior.

Empirical analysis of individual emotions Current psychology research provides data on experiments of emotion dynamics. These experimental setups can provide three different kinds of data to test our modeling assumptions: i) data on the dynamics of emotions in everyday life, under the absence of controlled online interaction, ii) subjective reports and physiological signals that measure emotions during the perception of emotionally charged content, and iii) causation between emotional changes and online expression. The analysis of these data can provide the empirical testing of our agent-based models, but also new insights on emotion dynamics not hypothesized during model design. These additional results can be integrated within future models, which then can provide new hypothesis of collective behavior, driving further research on online communities. This closes a research cycle that links the macro and micro levels through agent-based models, with a multidisciplinary feedback between research techniques that can constantly improve each other.

Part I

Modeling collective emotions

Agent-based modeling of collective emotional behavior and empirical analysis of collective emotions in online communities

“Our knowledge has made us cynical, our cleverness hard and unkind. We think too much and feel too little: More than machinery we need humanity; More than cleverness we need kindness and gentleness.”

Charlie Chaplin, *The Great Dictator*, 1940

Chapter 1

A framework for modeling collective emotions in online communities

Summary

In this chapter, we present our agent-based framework to model the emergence of collective emotions, which is applied to online communities. Agents' emotional states are described by their valence and arousal. Using the concept of Brownian agents, these variables change according to a stochastic dynamics, which also considers the feedback from online communication. Agents generate emotional information, which is stored and distributed in a field modeling the online medium. This field affects the emotional states of agents in a non-linear manner. We derive conditions for the emergence of collective emotions, observable in a bimodal valence distribution. Dependent on a saturated or a superlinear feedback between the information field and the agent's arousal, we further identify scenarios where collective emotions only appear once or in a repeated manner. The analytical results are illustrated by agent-based computer simulations. Our framework provides testable hypotheses about the emergence of collective emotions, which can be verified by data from online communities.

1.1 A unified approach to collective emotions

Collective emotions are shared by a large number of individuals as a result of both external events and nonlinear coupling between individuals. Similar to other collective states, also collective emotions can display new, emergent properties which cannot be traced back to individual contributions. Remarkably, the life time of a collective emotion is usually much longer than the one of an individual emotion. On the other hand, individual emotions show a different dynamics in the presence of collective emotions, simply because of the nonlinear feedback of the emergent collective emotion on the individual one.

To study collective emotions, we need an appropriate description of the system elements, which are called agents in the following, and their interactions – but we also need an appropriate framework to predict from these ingredients the possible collective dynamics at the system level. Without such a framework, we are only left with extensive computer simulations of multi-agent systems, in which, for given assumptions of the interactions, we have to probe the entire parameter space, to find out the conditions for certain collective phenomena. Furthermore, collective emotions appear in different online communities, which often have different interaction mechanisms. Models of collective emotions in each of these communities, if designed and analyzed separately, might shed light on the particular properties of collective emotions in each one of them. Such approach, on the other hand, would not allow to draw conclusions on universal properties of collective emotions across communities. If designed within a unifying framework, models of collective emotions in different communities can be compared between different scenarios of online interaction.

In this chapter, we present such a framework to describe collective emotions in online communities through agent-based models. In an agent-based model, we first need to describe the emotional states of individual agents, which should be based on insights obtained in psychology. As explained in the introduction, we follow Russell’s representation of *core affect* (Russell, 1980) , modeling emotions as short-lived psychological states of the individual. This established theoretical perspective is based on two dimensions: *valence*, indicating whether the emotion is pleasant or unpleasant, and *arousal*, indicating the degree of activity or inactivity induced by the emotion. Therefore, the internal states of our agents will be composed of two independent variables of valence and arousal.

Another challenge results from the fact that we need to model the communication between users in online communities. It is not the emotion per se of an user what matters, but its expression in a blog entry, a post etc. This is submitted at a particular time and distributed to the whole online community, where it is perceived by other users with a very different time delay. While modeling a personalized communication would need to know

the underlying social network, in most online communities a particular post is available to everyone (who has subscribed) immediately. For example, product reviews or group chat-room messages can be publicly accessed by all the members of the online community. This justifies the assumption of a mean-field coupling between users, i.e. a medium is updated instantaneously and provides the same information to everyone. Nevertheless, we still have to consider that for example older posts have less impact on users than more recent ones and that positive and negative posts may be submitted with different frequencies. In our modeling framework, we cope with these requirements by introducing an emotional information field generated by the users, which stores and distributes this information accordingly. This idea was already successfully applied in other communication models describing biological or social systems (Schweitzer and Hołyst, 2000; Schweitzer, 2003).

Eventually, we need to model the impact of the emotional information on online users. Unfortunately, the psychological literature does not provide much insight into this problem. Therefore, we are left with providing hypotheses about the feedback between the emotional information and the individual dimensions of arousal and valence. By proposing a very general non-linear feedback, for which different special cases are explored, we are able to derive conditions under which the emergence of collective emotions can be expected. Different scenarios can be obtained: either the repeated occurrence of collective emotions, or the one-time collective emotion. These results can be seen as testable hypothesis, which may be verified either by psychological experiments or by data from online communities. In Chapter 4 and Chapter 5 we provide empirical support from some of the assumptions of emotion dynamics presented here, based on physiological measurement, individual emotion reports, and emotional expression online.

1.2 An agent-based model of emotions

1.2.1 The concept of Brownian agents

Our modeling approach is based on the concept of Brownian agents (Schweitzer, 2003). It allows to formalize the agent dynamics and to derive the resulting collective dynamics in close analogy to methods established in statistical physics. A Brownian agent is described by a set of state variables $u_i^{(k)}$, where the index $i = 1, \dots, N$ refers to the individual agent i , while k indicates the different variables. These could be either *external* variables that can be observed from the outside, or *internal degrees of freedom* that can only be indirectly concluded from observable actions.

Noteworthy, the different (external or internal) state variables can change in the course of

time, either due to influences of the environment, or due to an internal dynamics. Thus, in a most general way, we may express the dynamics of the different state variables as follows:

$$\frac{du_i^{(k)}}{dt} = f_i^{(k)} + \mathcal{F}_i^{\text{stoch}} \quad (1.1)$$

This formulation reflects the *principle of causality*: Any *effect* such as the temporal change of a variable u has some *causes* that are listed on the right-hand side. For the concept of Brownian agents, it is assumed that these causes may be described as a *superposition* of *deterministic* and *stochastic* influences, imposed on agent i . This distinction is based on Langevin’s idea for the description of Brownian motion, which coined the concept. Hence, we sum up influences which may exist on a microscopic level, but are not observable on the time and length scale of the Brownian agent, in a stochastic term $\mathcal{F}_i^{\text{stoch}}$, while all those influences that can be directly specified on these time and length scales are summed up in a *deterministic* term $f_i^{(k)}$. This implies that the “stochastic” part does *not* impose any *directed* influence on the dynamics (which would have counted as deterministic), but on the other hand, it does not necessarily mean a white-noise type of stochasticity. Instead, other types such as colored noise, or multiplicative noise are feasible. Noteworthy, the strength of the stochastic influences may also vary for different agents and may thus depend on local parameters or internal degrees of freedom, as was already used in different applications (Schweitzer, 2003). The *deterministic* part $f_i^{(k)}$ contains all specified influences that cause changes of the state variable $u_i^{(k)}$. This could be nonlinear interactions with other agents $j \in N$ – thus $f_i^{(k)}$ can be in principle a function of all state variables describing any agent (including agent i). But $f_i^{(k)}$ can also describe the response of an agent to available information, as it will be the case for our modeling framework. It should further depend on external conditions – such as forces resulting from external influences (most notably information from mass media). Eventually, $f_i^{(k)}$ may reflect an (external or internal) *eigendynamics* – in the considered case a relaxation of the excited emotional state of an agent (caused by saturation or exhaustion). In order to set up a multi-agent system (MAS) we need to specify the relevant state variables $u_i^{(k)}$ and the dynamics of their change, i.e. $f_i^{(k)}$, which means also to specify the interaction between the agents. We emphasize that the dynamics of the MAS is specified on the level of the individual agent, not on a macroscopic level, so the collective dynamics shall emerge from the interactions of many agents.

1.2.2 Emotional states

To quantify the emotional dynamics of an agent, we consider the following continuous variables, *valence*, $v_i(t)$, and *arousal*, $a_i(t)$. Both define a two-dimensional plane for the

classification of emotions. Valence (x-axis) measures whether an emotion is positive or negative, arousal (y-axis) measures the degree of personal activity induced by that emotion. Hence, an emotional state is defined by $e_i(t) = \{v_i(t), a_i(t)\}$. For example, ‘astonished’ is an emotional state with both positive valence and arousal, ‘satisfied’ has a positive valence, but a negative arousal, ‘depressed’ has both a negative valence and arousal, and ‘annoyed’ has a negative valence and a positive arousal. We note that both valence and arousal describe *internal* variables, i.e. a dynamics inside the agent, which may be only indirectly observable, for example through physiological measurements.

Without any internal or external excitation, there should be no positive or negative emotion, so we assume that in the course of time both valence and arousal relax into an equilibrium state, $e_i(t) \rightarrow 0$, which implies $v_i(t) \rightarrow 0$, $a_i(t) \rightarrow 0$. Hence, in accordance with Equation (1.1) we specify the dynamics of the Brownian agent as follows:

$$\begin{aligned}\dot{v}_i &= -\gamma_{vi} v_i(t) + \mathcal{F}_v + A_{vi} \xi_v(t) \\ \dot{a}_i &= -\gamma_{ai} a_i(t) + \mathcal{F}_a + A_{ai} \xi_a(t)\end{aligned}\tag{1.2}$$

The first term in each equation describes the relaxation into an equilibrium state as an exponential decay of both valence and arousal, if no excitation is given. γ_{vi} , γ_{ai} define the time scales for this relaxation, which are different for valence and arousal and further may vary across individual agents. The second and third term in the equations above describe influences which may induce an emotional state. These can be stochastic influences, expressed by the third term, where $\xi_v(t)$, $\xi_a(t)$ are random numbers drawn from a given distribution of stochastic shocks, with the mean of zero $\langle \xi(t) \rangle = 0$ and no temporal correlations between subsequent events $\langle \xi(t)\xi(t') \rangle = \delta(t-t')$. A_{vi} , A_{ai} denote the strength of these stochastic influences which may again vary across agents. The two functions \mathcal{F}_v , \mathcal{F}_a describe deterministic influences which cause the emotional state. They very much depend on the specific assumptions applicable to collective emotions, in particular the agents’ interaction, access to information, response to the media, but can depend also on internal variables such as *empathy*, i.e. the ability to share the feelings of other agents (Preston and de Waal, 2002), or responsiveness to available information. Most of all, these functions should also reflect a dependence on the emotional state itself, i.e. agents already in a specific mood may be more affected by particular emotions of others. Before we specify these functions in detail, we need to extend the agent description.

1.2.3 Emotional actions

The dynamics of Equations (1.2), already define a stationary state $e_i(t) \rightarrow 0$, given that the deterministic and stochastic influences become negligible. On the other hand, there should

be an excited emotional state of the agent if these influences are large, e.g. if information with a large emotional content becomes available to the agent. Per se, this state is not observable from the outside unless the agent takes any action that *communicates* that emotional state, for example by posting in a blog, etc. Consequently, we assume that the agent expresses its *valence*, i.e. the good or bad feeling, *if* its *arousal*, i.e. the action induced by the emotion, exceeds a certain individual threshold, τ_i :

$$s_i(t + \Delta t) = \text{sign}(v_i(t)) \Theta[a_i(t) - \tau_i] \quad (1.3)$$

Here $\Theta[x]$ is the Heavyside function which is one only if $x \geq 0$ and zero otherwise. If $\Theta[x] = 1$, we make the simplifying assumption that the agent does not communicate all details about his feelings (i.e. the value of v_i) because perfect emotional information cannot be communicated. Instead, the agent communicates only if it is a good or bad feeling, i.e. the sign of v_i , -1 or +1, which is defined as $r_i(t) = \text{sign}(v_i(t))$ in the following. This particularity specified here can be changed to allow $v_i(t)$ to be perfectly communicated, but it is essential to assess if that is in line with the data analysis investigations. 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.

Equation (1.3) further reflects the assumption that the agent does not immediately express its feelings if the arousal hits the threshold at time t , but probably with a certain delay Δt , which may be caused by the fact that the agent has no immediate access to some communication media (computers in the case of cyberemotions) or other things to do. More important feelings should be communicated with a shorter delay. It should vary as well across agents. In accordance with investigations of waiting time distributions in performing human activities (e.g. answering emails), we may assume that Δt can be random drawn from a power-law distribution $P(\Delta t) \propto \Delta t^{-\alpha}$, where α should be empirically determined. Note that the dynamics of the external state variable $s_i(t)$ differ from the form given in Equation (1.1) in that the stochastic influences are not additive, but implicitly present because of the stochastic dynamics for $v_i(t)$ (determining the *sign* of the expression), $a_i(t)$ (determining the *time* of the expression) and Δt (determining the *delay* of the expression).

Based on Equation (1.3), we can define the number of emotional expressions at a given time t as

$$N_s(t) = \sum_i \Theta[a_i(t) - \tau_i] \quad (1.4)$$

Assuming continuous time, the *average* number of expressions per time interval then results

from

$$n_s = \frac{1}{t_{\text{end}}} \int_0^{t_{\text{end}}} N_s(t) dt \quad (1.5)$$

We may calculate this quantity from analytical approximations in Section 1.3.3 and from computer simulations in Section 1.4.

1.2.4 Communicating emotions

By now we have described the (internal) emotional dynamics of an agent that leads to a certain (externally visible) expression of an emotional state. In order to describe cyber-emotions as collective emotions, we now need to specify how this emotional expression is communicated to other agents. In accordance with previous investigations (Schweitzer and Holyst, 2000) we assume that every positive or negative expression is stored in a *communication field* $h_+(t)$ or $h_-(t)$ dependent on its value. $h_{\pm}(t)$ represent the communication media available for the storage and distribution of emotional statements, for example blogs, fora, etc. and simply measure the 'amount' of positive or negative feelings available at a given time. For the dynamics of the field, we propose the following equation:

$$\dot{h}_{\pm} = -\gamma_{\pm} h_{\pm}(t) + s N_{\pm}(t) + I_{\pm}(t) \quad (1.6)$$

Each agent contribution $s_i(t)$ increases the respective field h_+ or h_- by a fixed amount s at the time of expression, which represents the impact of the information created by the agent in the information field, as a time scale parameter. $N_{\pm}(t)$ is the total number of agents contributing positive or negative statements at a given time t , i.e. all the agents with $s_i(t) = 1$ and with $s_i(t) = -1$ respectively.

The relevance of contributions fade out over time as e.g. agents become less affected by old blog entries. This is covered by an exponential decay of the available information with the time scales γ_{\pm} . Eventually, in addition to the agent contributions, positive or negative emotional content from the news may add to the communication field, which is covered by an agent-independent term $I_{\pm}(t)$, which can be modeled for example by a stochastic input.

The main feedback loops of this framework are sketched in Figure 1.1, where we can distinguish between two layers: an internal layer describing the agent (shown horizontally) and an external layer describing the communication process (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 one has its independent dynamics and can, in addition to contributions from other agents,

also consider input from external sources, I . The causality is closed by considering that both valence and arousal of an agent are affected by the communication field.

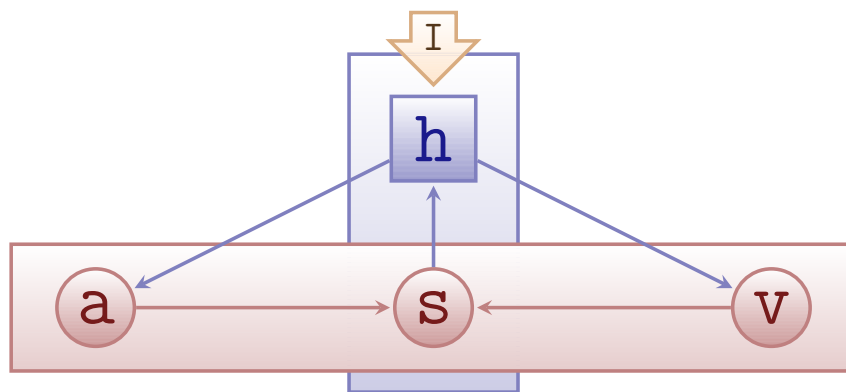


Figure 1.1: Relation between the elements of our agent-based framework. Agents have internal emotional states composed of valence v , and arousal a . These determine the agent expression s , which goes to a communication field h that can influence the emotions of other agents.

In order to complete the model, we need to specify how the available information affects the emotional states of the individual agents, which is covered in the functions \mathcal{F}_v and \mathcal{F}_a .

1.2.5 Emotional feedback

Because we are interested in the outbreak of *collective* emotion, we do not assume the latter as the simple superposition (or addition) of individual emotional states. On the contrary, we assume that an emotional state of one agent, if it is expressed and communicated to other agents, may affect the emotional state of these agents either directly or indirectly. Regarding this effect we are left with hypotheses at the moment. These could be tested in computer simulations to investigate their impact on the possible emergence of a collective emotion – as it is done in the following. But there should be also the possibility to empirically test how individuals are affected by different emotional content, as discussed e.g. in (Gianotti *et al.*, 2008).

With respect to the valence, i.e. the good or bad feeling, we have to take into account that there are two different kind of emotions in the system, positive ones represented by $h_+(t)$ and negative ones represented by $h_-(t)$. Dependent on its own emotional state, an agent may be affected by these information in a different way. If we for example assume that agents with negative (positive) valence mostly respond to negative (positive) emotional

content, we have to specify:

$$\mathcal{F}_v \propto \frac{r_i}{2} \{(1 + r_i)f[h_+(t)] - (1 - r_i)f[h_-(t)]\} \quad (1.7)$$

where $r_i(t) = \text{sign}(v_i(t))$ and $f(h_{\pm}(t))$ are some functions depending either on h_+ or on h_- only. Equation (1.7) then results in $\mathcal{F}_v \propto f[h_-(t)]$ if the agents have a positive valence ($r_i(t) = +1$), and in $\mathcal{F}_v \propto f[h_+(t)]$ in the case of negative valence ($r_i(t) = -1$).

If, on the other hand, there is evidence that agents, independent of their valence, always pay attention to the prevalence of positive or negative emotional content, we may assume:

$$\mathcal{F}_v \propto g[h_+(t) - h_-(t)] \quad (1.8)$$

where g is some function of the difference between the two information available. Other combinations, for example agents with positive (negative) valence pay more attention to negative (positive) emotional content, can be tested as well. Some studies in psychophysiology (Bradley, 2009) provide initial results for heterogeneous emotional attention processes related to valence and arousal.

In the following, we may assume the case of Equation (1.7), i.e. the valence increases with the respective information perceived by the agent. The impact, however, should depend also on the emotional state of the agents in a nonlinear manner. I.e. if an agent is happy (sad), it may become happier (more sad) if receiving information about happy (sad) agents or events, in a nonlinear manner, expressed in the general form:

$$f[h_{\pm}(t), v_i(t)] = h_{\pm}(t) \sum_{k=0}^n b_k v^k(t) \quad (1.9)$$

Here, it is assumed that the coefficients b_k are the same for positive and negative valences, which of course can be extended toward different coefficients.

1.2.6 Arousal and threshold

While the valence expresses the positivity or negativity of the emotion, the arousal measures the degree in which the emotion encourages or discourages activity. Only the latter is important for communicating the emotional content, which happens if a threshold τ_i of the arousal is reached. Certainly, expressing the emotion should have some impact of the arousal, e.g. it is legitimate to assume that the arousal is lowered because of this action, or set back to the initial state in the most simple case. That means we should split the dynamics for the arousal into two parts, one applying before the threshold is reached, the

other one when it is reached. For this, we redefine the arousal dynamics for $a_i(t)$ given in Equation (1.2) as the subthreshold dynamics $\dot{a}_i(t)$ and set:

$$\dot{a}_i = \dot{\bar{a}}_i(t) \Theta[\tau_i - a_i(t)] - a_i(t) \Theta[a_i(t) - \tau_i] \quad (1.10)$$

As long as $x = \tau_i - a_i(t) > 0$, $\Theta[x] = 1$ and the arousal dynamics is given by $\dot{\bar{a}}_i(t)$, explained in Equation (1.2). However, after the threshold is reached, $x \geq 0$, $\Theta[x] = 0$ and $\Theta[-x] = 1$, and the arousal is reset to zero by the $-a_i(t)$ term.

It remains to specify the function \mathcal{F}_a for the subthreshold arousal dynamics. Since arousal measures an activity level, it would be reasonable to assume that agents respond to the sum of both positive and negative emotional content in a way that also depends on their own arousal in a nonlinear manner, regardless of the valence dimension. So, similar as for the valence, we may propose the nonlinear dependence:

$$\mathcal{F}_a \propto [h_+(t) + h_-(t)] \sum_{k=0}^n d_k a^k(t) \quad (1.11)$$

Differently from the above assumption, 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 e.g. that the impact of information increases linearly with the activity level:

$$\mathcal{F}_a \propto [h_+(t) + h_-(t)] a(t) \Theta[a(t)] \quad (1.12)$$

To conclude the above description, we have set out a model where agents emotions are characterized by two variables, valence and arousal. These variables can be psychologically justified and most likely proxied empirically. The combination of these defines what kind of emotional content the agent expresses as an observable output. Again, this output is measurable and can be analyzed. The way the emotional content is stored and distributed to other agents is explicitly modeled as part of a communication dynamics, which can be adjusted to specific practical situations.

1.3 Emergence of collective emotions

1.3.1 Valence dynamics

In our model, a collective emotional state can only emerge if a sufficient number of agents expresses their individual valences, which in turn depends on their arousal. The latter one gets above a critical threshold only if there is sufficient the emotional information $h_+(t)$,

$h_-(t)$ available. However, this information is generated only by the agents. Hence, there is a circular causality between $h(t) = h_+(t) + h_-(t)$ and $a_i(t)$.

In order to get a first insight into the dynamics, let us assume that there exist two different regimes: (i) a 'silent' regime where no sufficient emotional information is available, i.e. $h(t) \rightarrow 0$, and (ii) an 'excited' regime, where $h(t)$ becomes large enough to affect enough agents. To simplify the study, we also assume that each agent is mostly affected by the information that corresponds to its valence state, as given by Equation (1.7).

Then, neglecting any sort of random influences, the dynamics of the valence is expressed by:

$$\dot{v} = -\gamma_v v(t) + h_{\pm}(t) \{b_0 + b_1 v(t) + b_2 v^2(t) + b_3 v^3(t) + \dots\} \quad (1.13)$$

The stationary solutions for the valence then follow from the cubic equation:

$$v^3 + v^2 \{b_2/b_3\} + v \{(b_1 - \gamma_v/h_{\pm})/b_3\} + \{b_0/b_3\} = 0 \quad (1.14)$$

This allows to discuss the following cases:

- In order to allow for a solution $v \rightarrow 0$ as requested, b_0 should tend to zero as well, so we use $b_0 = 0$ here. This leads to

$$v [v^2 + v \{b_2/b_3\} + \{(b_1 - \gamma_v/h_{\pm})/b_3\}] = 0 \quad (1.15)$$

- If positive and negative valences are treated as 'equal', there should be no ab initio bias towards one of them, which implies $b_2 = 0$. This gives, in addition to $v = 0$, the following two solutions:

$$v^2 = \frac{b_1 - \gamma_v/h_{\pm}}{b_3} \quad (1.16)$$

These two solutions become real only if $b_1 > \gamma_v/h_{\pm}$. In this case, we have two equilibrium states for the valences which are symmetrical with respect to zero. Otherwise, $v = 0$ is the only possible real solution.

So, dependent on the value of the information field h_{\pm} we can expect the two regimes: (i) the *silent regime* with $h_{\pm} \rightarrow 0$ and $v = 0$, and (ii) the *excited regime* with the emergence of two different emotions, each of them centered around $\pm b_1/b_3$ (provided the field is large enough). We note that these solutions are symmetrical, which can be changed by considering (a) a bias in the response ($b_2 \neq 0$) or (b) differences in the two informations h_{\pm} (e.g. via different decay rates). It remains to be discussed whether a *coexistence* between the two collective emotional states is possible or the *prevalence* of one of them results. This leads us to the question of path dependence and emotional feedback of section 1.3.2.

If, in addition to the deterministic dynamics specified above, we further consider stochastic influences as specified in the Langevin dynamics, Equation (1.2), we can write up a dynamics for the valence distribution $p(v, t)$. Using Equations (1.2), (1.14), this is given by the following Fokker-Planck equation:

$$\partial_t p(v, t) = -\partial_v [(b_1 h_{\pm} - \gamma_v)v - b_3 h_{\pm} v^3] p(v, t) + \frac{A_v^2}{2} \partial_v^2 p(v, t) \quad (1.17)$$

The stationary solution of the Fokker-Planck equation, $\partial_t p(v, t) = 0$, reads as:

$$p(v) = \frac{1}{\mathcal{N}_v} \exp \left\{ \frac{v^2(b_1 h_{\pm} - \gamma_v) - v^4(b_3 h_{\pm}/2)}{A_v^2} \right\} \quad (1.18)$$

\mathcal{N}_v is the normalization constant resulting from $\int_{-\infty}^{\infty} p(v) dv = 1$. In accordance with the discussion above, the stationary valence distribution $p(v)$ is *unimodal* with the maximum at $v = 0$ if $b_1 < \gamma_v/h_{\pm}$ and *bimodal* with the maxima given by Equation (1.16) if $b_1 > \gamma_v/h_{\pm}$. In both cases, the variance of the distribution is determined by the strength of the stochastic force, A_v^2 . This is shown in Figure 1.2 for two different values of h . The histograms result from computer simulations of the stochastic valence dynamics of 1000 agents for a given h , whereas the solid curves are given by the analytical solution of Equation (1.18).

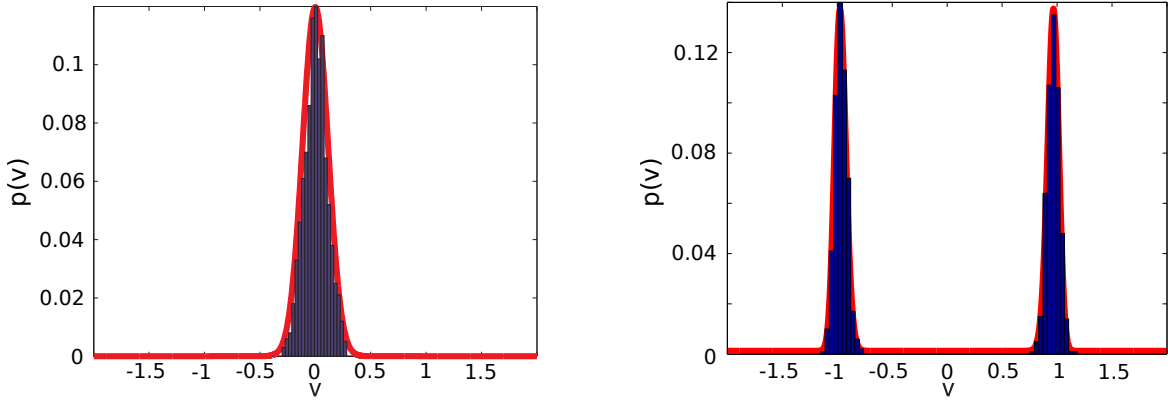


Figure 1.2: Analytical prediction of the valence distribution from Equation (1.18) and histogram of valences of a simulation under $h = 0.25$ (left) and $h = 17.5$ (right) with parameters $N = 1000$, $A_v = 0.15$, $\gamma_v = 0.8$, $b_1 = 1$ and $b_3 = -1$.

To conclude, this analysis has provided us with conditions regarding the influence of emotional information and the response to it, which may lead to the emergence of a collective emotional state. These conditions can be seen as testable hypotheses about the feedback between emotional information and individuals. If they hold true, we are

able to predict the valence of a collective emotional state - which can then be compared to empirical findings. Deviations from these findings, on the other hand, allow us to successively refine the modeling assumptions made. Thus, the framework provided is a useful step toward a thorough understanding of collective emotions.

1.3.2 Arousal dynamics

In the previous section, we have detected an excited regime with a non-trivial valence, based on the assumption that the emotional information $h(t)$ is large enough to affect the agents. The generation of such information, however, depends on the arousal dynamics. Specifically, in our model the arousal needs to reach a certain threshold τ_i at which the agent expresses its emotion. In Section 1.2.6, we have already introduced an arousal dynamics, Equation (1.10), which distinguishes between a subthreshold dynamics, Equation (1.2), and a dynamics at the threshold. By using the nonlinear assumption of Equation (2.7) up to second order and neglecting stochastic influences for the moment, we get for the subthreshold regime (omitting the index i for the moment):

$$\dot{a} = -\gamma_a a(t) + h(t) \{d_0 + d_1 a(t) + d_2 a^2(t) + \dots\} \quad (1.19)$$

where $h(t) = h_+(t) + h_-(t)$. The stationary solutions follow from:

$$a^2 + a\{(d_1 - \gamma_a/h)/d_2\} + \{d_0/d_2\} = 0 \quad (1.20)$$

which allows to discuss the following cases. If we consider only the constant influence of h , i.e. $d_0 \neq 0$, $d_1 = d_2 = 0$, or a linear increase with a , i.e. $d_0 \neq 0$, $d_1 \neq 0$, $d_2 = 0$, we arrive at only *one* stationary solution for the arousal, which depends on h :

$$a(h) = \frac{hd_0}{\gamma_a - d_1 h} \quad (1.21)$$

It means that the agents tend to be always in an 'excited' regime, the level of which is determined by h . If it happens that the arousal reaches a value above the threshold, $a(h) \geq \tau$, then $a(t)$ is arbitrarily set back to zero and then starts to reach $a(h)$, again. The proposed 'silent' regime would then be reached only if $h \rightarrow 0$.

In order to allow for a dynamics where agents can stay at low values of the arousal even if h is large, we have to consider a non-linear influence of the emotional information h , i.e. $d_0 \neq 0$, $d_1 \neq 0$, $d_2 \neq 0$ in the most simple case. Equation (1.20) then has two solutions

$$a_{1,2}(h) = \frac{1}{2} \left(\frac{\gamma_a/h - d_1}{d_2} \right) \pm \sqrt{\frac{1}{4} \left(\frac{d_1 - \gamma_a/h}{d_2} \right)^2 - \frac{d_0}{d_2}} \quad (1.22)$$

which are real only if

$$\left(\frac{d_1 - \gamma_a/h}{2d_2}\right)^2 > \frac{d_0}{d_2} \quad (1.23)$$

From this restriction, we can infer some important conditions on the arousal dynamics. Provided $d_0 > 0$, inequality (1.23) is always fulfilled if $d_2 < 0$ which, for a given h , implies a saturation in the feedback of the arousal on the arousal dynamics. Then we always have two real stationary solutions for the arousal, a positive and a negative one, shown in Figure 1.3. While the positive solution is stable for all values of h , the negative one is always unstable, as verified by the second derivative. This allows to infer the following dynamics for agents expressing their emotions: For agents starting with a small positive or negative arousal, $a > a_2(h)$, $a(t)$ may grow in time up to the stationary value $a_1(h)$, the level of which is determined by the emotional information available at that time. Only if $a_1(h) > \tau_i$, the agent expresses its emotions, which consequently sets back $a_i(t)$ to zero, otherwise it remains at this subcritical arousal level. On the other hand, if the agent, because of some fluctuations, reaches the unstable negative arousal level $a_2(h)$, the feedback of Equation (1.19) will further amplify the negative arousal to $-\infty$. This means that the agent never again expresses its emotions and 'drops out'. If this happens to many agents, a collective emotion cannot be sustained. Which of the two cases is reached, crucially depends on the fluctuation distribution. Looking at the example of Figure 1.3, we can verify that initial fluctuations (for $a = 0$) should not reach the level of 0.1, in order to prevent a 'dropout' of the agents.

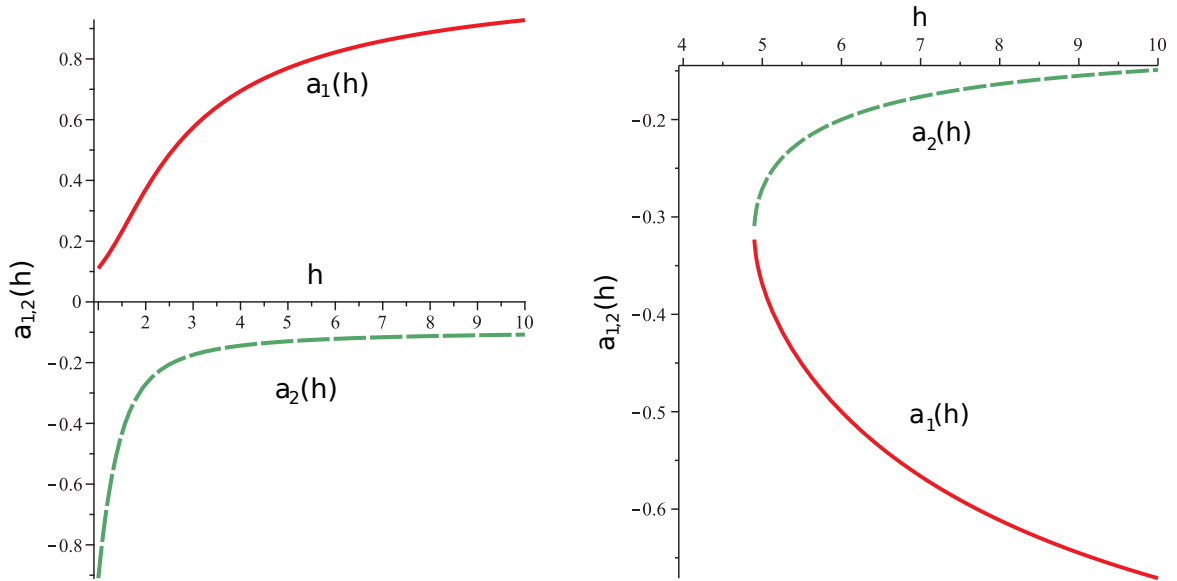


Figure 1.3: $a_{1,2}(h)$ for the parameter set 1.30. There is no bifurcation present when $d_2 = -0.5$ (left) but it appears when $d_2 = 0.5$ (right)

The scenario looks different if, instead of a saturated dynamics with $d_2 < 0$, we assume $d_2 > 0$, i.e. a superlinear growth in the arousal. Inequality (1.23) then defines the range of possible values of d_2 to guarantee two real solutions. As one can verify, there are real solutions already for very small values of h . In the following, we only concentrate on the range of sufficiently large h as shown in Figure 1.3. Then, both real solutions are negative and the one closer to zero is the unstable solution, whereas the most negative solution is stable. For agents starting with a small positive or negative arousal, $a > a_1(h)$, a may further grow independent on the value of h until it reaches the threshold τ_i , at which the agent expresses its emotion, i.e. agents do not remain at a subcritical arousal level forever. If the arousal is set back to zero and because of fluctuations reaches negative values $a < a_1(h)$, it will become more negative, but is always bound by the negative stationary value $a_2(h)$. I.e., the agent never 'drops out' entirely. Instead, even with a negative arousal, it can always get back into an active regime dependent on the fluctuation distribution. Consequently, the 'non-saturated' case defines the scenario where we most likely expect the emergence of collective emotions, where agents regularly express their emotions. However, such a scenario can never be sustained in a purely deterministic dynamics. Instead, spontaneous fluctuations are essential, and our analysis already tells us the critical size of the fluctuations needed (determined by A_a). As we can verify in Figure 1.3, this critical fluctuation level depends on the total information h , which is not unrealistic, because more (diverse) information is also associated with more ambivalence.

1.3.3 Expression of emotions

So far, we have identified critical regimes both in the valence and in the arousal dynamics, provided a given emotional information h . However, as explained above, this information is only generated by the agents above a critical arousal. Consequently, we need to ask what is the minimal time lapse before an agent reaches the threshold τ , contributing to the emotional information. For simplicity, we take the delay time in Equation (1.3) as $\Delta t = 0$ for all expressions. The time lapse to reach the threshold is given by the dynamics of Equation (1.19), which can be solved assuming a given value of h :

$$\int_0^T dt = T = \int_0^\tau \frac{da}{hd_2a^2 + (hd_1 - \gamma_a)a + d_0h} \quad (1.24)$$

This solution assumes that h already exist, either because of an external information, or because it is generated by other agents. Hence, it is an adiabatic approximation of the full dynamics, which assumes $\dot{h} = 0$, this way describing the response of a single agent to the existing (stationary) field. The solution of Equation (1.24) depends on whether the value of $R(h) = 4d_2d_0 - (d_1 - \gamma_a/h)^2$ is positive or negative.

Following the discussion in the previous section, we now have to consider two different regimes for arousal dynamics, the saturated one ($d_2 < 0$) and the superlinear ($d_2 > 0$). In the saturated regime, always $R(h) < 0$ and the solution is given by:

$$T(h, \tau) = \frac{2}{h\sqrt{-R(h)}} \operatorname{arctanh} \left(\frac{\sqrt{-R(h)}}{2d_0/\tau + d_1 - \gamma_a/h} \right) \quad (1.25)$$

For $d_2 > 0$ we can have both $R(h) < 0$ and $R(h) > 0$ dependent on the choice of the other parameters. For $R(h) > 0$, the solution of Equation (1.24) is given by:

$$T(h, \tau) = \frac{2}{h\sqrt{R(h)}} \operatorname{arctan} \left(\frac{\sqrt{R(h)}}{2d_0/\tau + d_1 - \gamma_a/h} \right) \quad (1.26)$$

In the superlinear regime, we expect that the agent is likely to express its emotions more than once (dependent on the fluctuations). In this case, $T(h, \tau)$ gives the (idealized) periodicity of expressing the emotion, i.e. the time after which the agent on average reaches the threshold τ again, after it was set back to zero when expressing the emotion last time. The left panel of Figure 1.4 shows the frequency $f_s(h, \tau) = 1/T(h, \tau)$ at which an agent expresses its emotions, dependent on the (quasistationary) value of h . We note that there is a non-monotonous increase, i.e. below a critical value of $h = h^*$ the frequency is zero, i.e. we do not expect a collective emotional state where agents more than once express their emotions, whereas for $h > h^*$, agents may regularly contribute emotional information, which means a collective emotion is sustained.

Based on the frequency of expression, we are able to calculate the average number of expressions per time interval, n_s , as defined in Equation (1.5). Assuming N agents with a threshold distribution $P(\tau)$, the number of agents with a given threshold τ is $N(\tau) = NP(\tau)$, whereas the frequency $f_s(h, \tau)$ defines how often such agents reach an arousal above the threshold, forcing them to express their emotions. Assuming a uniform threshold distribution for simplicity, we can calculate for a given h :

$$n_s = N \int f_s(h, \tau) P(\tau) d\tau = \frac{N}{\tau_{\max} - \tau_{\min}} \int_{\tau_{\min}}^{\tau_{\max}} \frac{d\tau}{T(h, \tau)} \quad (1.27)$$

where $T(h, \tau)$ is given by Equations (1.25), (1.26). n_s is plotted in the right panel of Figure 1.4. Above a critical value of the field h^* , the number of expressions per time interval increases monotonously, with a noticeable knee at the point where all agents become involved. Obviously, for lower values of h not all agents reach an arousal above the threshold, which prevents them from expressing their emotions. But at a characteristic value \hat{h} , the field is large enough to bring all their arousals above the threshold.

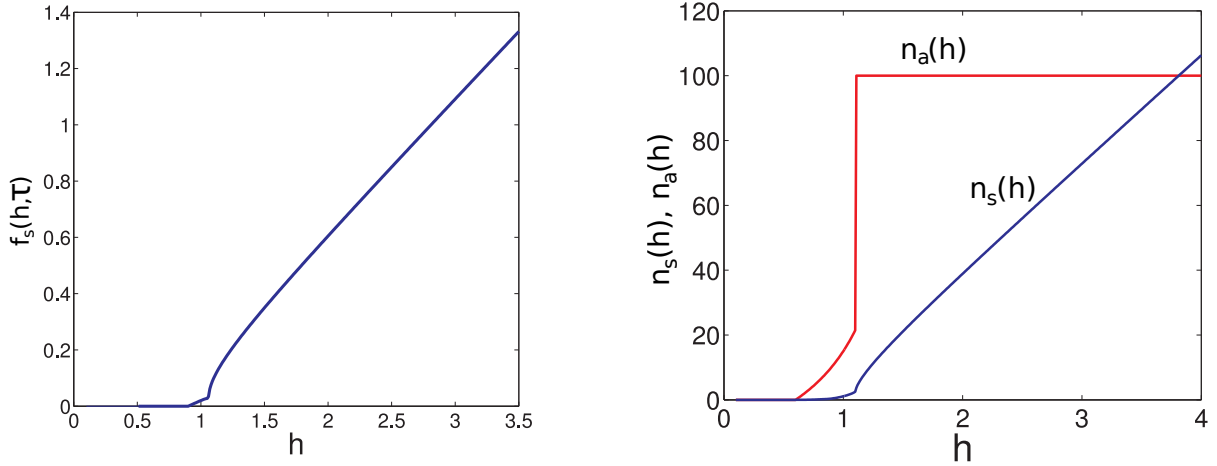


Figure 1.4: Frequency of emotional expression, $f_s(h, \tau) = 1/T(h, \tau)$ for $\tau = 0.5$ (left), $n_a(h)$ and $n_s(h)$ (right) under the parameter set 1.30. Below a critical value of h , an agent with threshold τ would not express its emotions. Above the critical value, the frequency of expression grows with the field h and the amount of agents involved in the conversation grows very fast, until the whole community is involved .

Similar to Equation (1.4), we can also calculate the number of agents expressing their emotions at any given time t as:

$$N_a(t) = \sum_i 1 - \Theta[-f_s(h, \tau)] \quad (1.28)$$

Again $\Theta[x]$ is one only if $x \geq 0$, i.e. for agents with frequency zero the Heavyside function $\Theta[-x]$ returns one. The *average* number of agents expressing their emotions per time interval is then, similar to Equations (1.5),(1.27):

$$n_a = \frac{1}{t_{\text{end}}} \int_0^{t_{\text{end}}} N_a(t) dt = N \int (1 - \Theta[-f_s(h, \tau)]) P(\tau) d\tau \quad (1.29)$$

which can be calculated similar to Equation (1.27). n_a is plotted in the right panel of Figure 1.4 as well, and one clearly identifies the critical \hat{h} , to involve all agents.

1.4 Computer simulations of collective emotions

Based on the analytical insights obtained, we eventually present the results of agent-based computer simulations. This means that we have implemented the individual dynamics given by the stochastic Equations (1.2), of N agents with a heterogeneous threshold distribution $P(\tau)$. The latter one is important as the process of forming a collective emotion

needs generating emotional information, $h(t)$. There could be two possibilities to start this process: (i) an external trigger, expressed by $I_{\pm}(t)$ in Equation (1.6), (ii) initial fluctuations in the arousal which have to be large enough to push some of the agents above the threshold. Very similar to the model of social activation (Granovetter, 1978), it then depends on the distribution of thresholds and the feedback dynamics whether more agents become involved. For our simulations, we have chosen the parameters for the valence and the arousal dynamics, b_k, d_k in such a way that a supercritical feedback between the emotional information generated and the activity of the agents is guaranteed. Specifically, we have chosen:

$$\begin{aligned} \gamma_v = 0.5, A_v = 0.3, b_1 = 1, b_3 = -1, \gamma_h = 0.7, A_a = 0.3, \gamma_a = 0.9, \\ d_0 = 0.05, d_1 = 0.5, \tau_{\min} = 0.1, \tau_{\max} = 1.1, N = 100, s = 0.1, h_0 = 0 \end{aligned} \quad (1.30)$$

That means that, thanks to our analytical efforts, we are likely to expect a collective emotion where most agents express their emotions at least once. Our main focus is therefore on the two different scenarios expressed by the parameter $d_2 \lesseqgtr 0$, which result from the saturated or the superlinear feedback of emotional information on the arousal dynamics.

In the saturated case, $d_2 < 0$, we expect that a collective emotion may appear, but not be sustained because agents have a tendency to 'drop out'. This scenario is illustrated in the left panel of Figure 1.5. Calculating the average number of expressions per time interval, $n_s(t)$, Equation (1.5), we observe an initial burst of activity in the beginning, i.e. many agents contribute their emotional information, which then fades out, only keeping a random level of activity. That means, we observe indeed the emergence of a collective emotion, but this is not sustained because of the assumed saturation. This is also confirmed in the left panel of Figure 1.5, which shows the averaged positive and negative valences of agents. We observe the emergence of a polarized state, where agents with strong positive and mildly negative emotions coexist, i.e. a bimodal valence distribution appears and remains for a while, before it disappears completely because agents 'dropped out'. Consequently, the saturated regime allows the emergence of a collective emotion, but it is restricted to appear once and never again.

In the superlinear case, $d_2 > 0$, we expect the emergence of collective emotions more than once, i.e. they can fade out and be reestablished again. We consider this the more realistic scenario for applications to Internet users, where the up and downs of collective emotions are indeed observed. The right panel of Figure 1.5 illustrates this scenario in a way comparable to case illustrated in the left panel. Here, we see waves of activity indicated by the number of emotional expressions per time interval. The respective averaged positive and negative valences also reflect these waves, i.e. we observe more or less polarized states dependent on the activity.

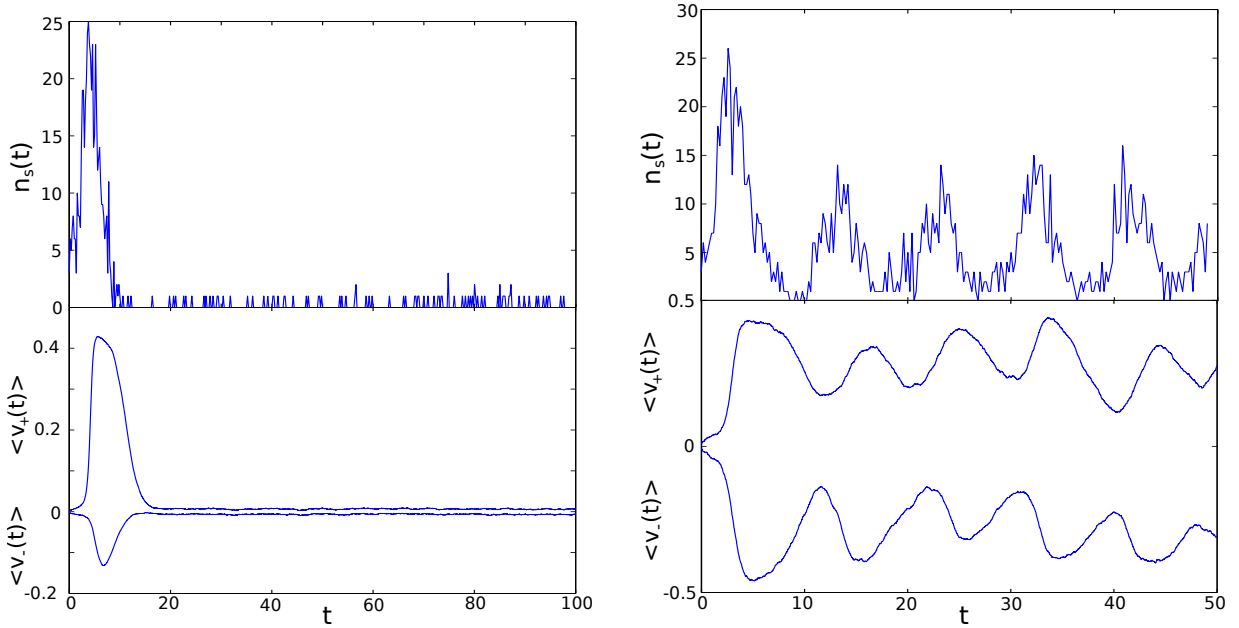


Figure 1.5: Agent expressions (top) binned with $\delta t = 0.2$ and average positive and negative valences (bottom) for a simulation with parameters 1.30 , $d_2 = -0.1$ (left), and $d_2 = 0.5$ (right). In the left panels, a collective emotional state appears and disappears after some time, but never reappears again. In the right panels, collective emotions reappear after the first one.

Consequently, the collective emotions not only emerge once, but are also sustained over a long period. The reason for this was already explained in Section 1.3.2. If agents have expressed their emotions and fall into a 'careless' state characterized by negative arousal, no new emotional information is produced. This in turn lowers the field h , which determines the stationary value of the negative arousal at which agents 'rest'. The lower the field, the larger the stationary arousal, which eventually allows the fluctuations to push agents back into an active regime of $a(t) > a_1(h)$. To illustrate this, we have plotted the arousal of ten randomly chosen agents in Figure 1.6. The typical oscillatory behavior can be clearly seen. If the field is initially low, most likely $a(t) > a_1(h)$, i.e. agents arousal is increased until they express their emotions. This generates a high field. If agents arousal is set back to zero at high h , $a_1(h)$ is almost zero (as can be verified in Figure 1.3), which means that most agents reach the stable stationary level of negative $a_2(h)$, at which they remain until h is lowered again.

The corresponding dynamics of the valence for the randomly chosen agents is also shown in Figure 1.6. One can notice a quite synchronized change of the emotions, which is not surprising as the dynamics mainly depends on the value of h , which is the same for all agents and all other parameters are kept constant. We can, of course, consider

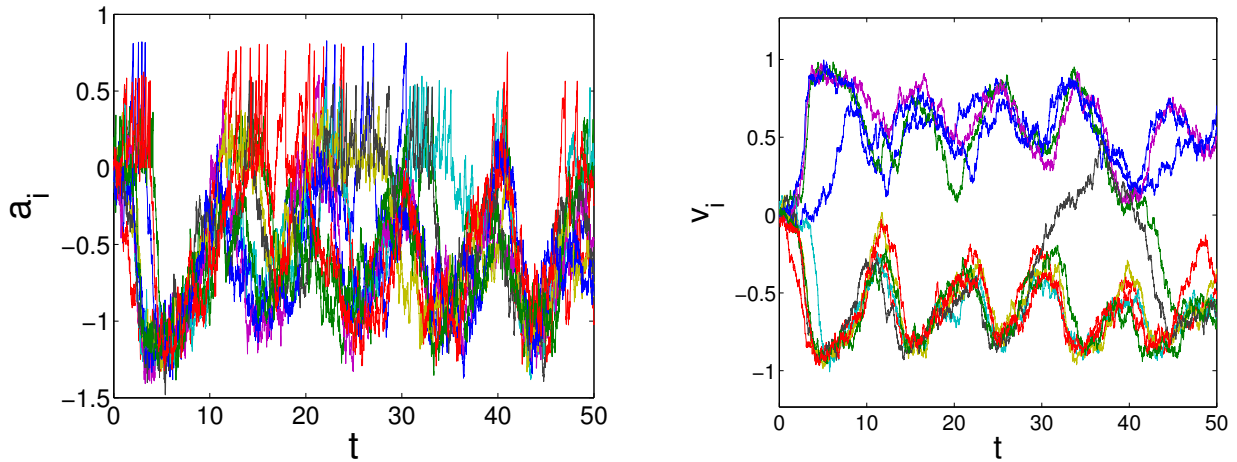


Figure 1.6: Sample trajectories of the arousal (top) and the valence (bottom) of ten agents in a simulation with parameters as in Figure 1.5.

more heterogeneous parameters for the agents, to allow for more diversity. We wish to emphasize that agents do not always have the same emotion over time, some of the sample trajectories clearly show that agents switch from positive to negative emotions and vice versa. To conclude, all simulations are consistent with the analytical results derived in the previous sections. Based on these results, we may be able to derive hypotheses about the behavior of emotional agents, which can be tested e.g. in psychological experiments.

1.5 Discussion and applications of our framework

In this chapter, our aim was to provide a general framework for studying the emergence of collective emotions in online communities. I.e. we are not particularly interested in the most complete description of individual emotions, but rather in an approach that allows to generate testable hypotheses about the conditions under which a particular collective dynamics can be observed. Nevertheless, we refrain from using ad hoc assumptions about the dynamics of individual agents which have been used in 'sociophysics' models of opinion dynamics, etc. Instead, our starting point is indeed a psychological theory of how individual emotional states should be described. Hence, our agent model is based on *psychological variables* such as arousal a and valence v .

As the second important ingredient of our general framework, we explicitly address the *communication* between agents. This is very important to model online communities, where agents do not have a direct, face-to-face communication, but an indirect, time delayed communication, which is mediated by an online medium. The latter stores the

information expressed by the agents (mostly in terms of writings) and allows all agents to get access to this information at the same or a different times. Consequently, this medium provides a mean-field coupling between all agents, which is reflected by the so-called communication field h in our framework. By explicitly modeling the dynamics of the information stored, we consider the decentralized generation of emotional information of different types (h_+ , h_-), at different times, the 'aging' of emotional information (i.e. the decrease in impact of older information) and the distribution of information among agents, for which also other than mean-field assumptions can be used.

As the third ingredient, we eventually model the impact of the emotional information on agents dependent on their emotional states. Here we assume a very general *nonlinear feedback* between the available information h and the individual valence v and arousal a . This allows to derive different hypotheses about the impact, which can be tested e.g. in psychological experiments. On the other hand, if such insight should become available to us, we are able to cope with these findings and to check their consequences on the emergence of collective emotions. Our work in Chapter 4 offers an analysis of datasets of emotional reaction to online content, testing some of the hypotheses formulated in this chapter, and providing new insights on the dynamics of individual emotions for future models within the framework.

It is a strength of our general framework that it allows an analytical treatment, to estimate the range of parameters under which the emergence of a collective emotion can be expected. In particular, we are able to specify the conditions for (a) a polarized collective emotional state (bimodal valence distribution), and (b) scenarios where agents express their emotions either once or consecutively (saturated vs superlinear impact on the arousal dynamics). Such findings are important in order to later calibrate the model parameters against empirical data describing these different regimes. Hence, our modeling approach offers a link to both psychological experiments with individuals, testing the hypotheses about the impact of emotional information, and to data analysis of emotional debates among Internet users, determining model parameters for different scenarios.

With our agent-based model we have provided a general framework to understand and to predict the emergence of collective emotions based on the interaction of agents with individual emotional states. As the framework is very tractable both in terms of mathematical analysis and computer simulations, we are now working on applying it to emotional debates observed in different online communities, such as in blogs, newsgroups, or discussion fora. In this chapter, we have highlighted the features of applying agent-based modeling to the study of collective emotions online. We showed how, under certain conditions, models within our framework are able to reproduce collective emotional states, which in turn

can be understood through analytical solutions of the equations driving agent dynamics. Nevertheless, these results depend on the validation of the assumptions used during the design model. Computer simulations and analytical tools cannot prove the empirical validity of the agent dynamics, which have to be tested versus real world data from online communities and individual emotion dynamics.

Eventually, the general framework provided here is extensible and flexible enough to encompass different situations where collective emotions emerge. This is thanks to the distinction of the two different layers already depicted in Figure 1.1: the internal layer describing the agent and its emotional states, and the external layer describing the communication process of expressing emotional information. The next chapters show our applications of this framework to model emotional interaction in product reviews communities and chatrooms. In addition, our framework has been applied to other types of online interaction, ranging from social networking sites to virtual humans.

Chapter 2

Collective emotions in product reviews communities

Summary

Online communities provide Internet users with means to overcome some information barriers and constraints, such as the difficulty to gather independent information about products and firms. In this chapter, we analyze data of emotions in product reviews, and design a model to reproduce some patterns of the user behavior of these communities. We introduce a new dataset of product reviews from `Amazon.com`, with emotional information extracted by sentiment detection tools, and estimation of mass media attention from `news.google.com`. We find a strong bias towards large values in the expression of positive emotions, while negative ones are more evenly distributed. Furthermore, we find that, depending on the mass media attention towards a product, the emotionality of product reviews increases in time. This supports the hypothesis that the driving force motivating reviewers to contribute new reviews is emotional rather than just informational, and that it can be influenced by marketing campaigns. Analyzing in detail the review dynamics, we found that the distribution of the number of reviews per product is more skewed at early times i.e. the first weeks after the release of the product. In addition, we study the time dynamics of the growth rate of reviews, finding that it decays stronger for products on the top sales ranks. Furthermore, this decay is also stronger for non-emotional reviews, suggesting that the impact of an emotionally charged review stays longer within the reviewer community, triggering future emotional reviews. We describe an agent-based model that can reproduce the scenarios of response to external influences, as well as some properties of the distributions of positive and negative emotions expressed in product reviews.

2.1 Collective emotions of consumers

The behavior of consumer societies is not reducible to the behavior of individual consumers, which is one of the usual limitations of market research (Salganik *et al.*, 2006). Studying the reactions of individual customers to a specific product or trusting expert judgments does not guarantee the acceptance of the product by consumers in the real market. The existence of “unpredictable blockbusters” (Hedström, 2006), i.e. products that reach a great success while their sales expectations are low, still lacks an explanation. In addition, there is empirical evidence that online communication affects the behavior of Internet users with respect to cultural goods (Salganik *et al.*, 2006; Zhu *et al.*, 2012), while social influence between individuals emerges as a collective feature of online interaction (Onnela and Reed-Tsochas, 2010). The impact of emotions and opinions in book and movie reviews has been analyzed in the context of the individual user, (Nahl and Bilal, 2007; Dellarocas *et al.*, 2007) but their collective properties are still to be understood.

Consumer memory about the brand of a product modifies the perceived quality and pleasure one gets from using it. Koenigs and Tranel have shown that the parts of the brain associated with emotional memory change product experience, given that the product’s brand is known (Koenig and Tranel, 2008). Further studies on the neural activity of consumers (Plassmann *et al.*, 2008) show that the experienced pleasantness of a product does not depend only on its intrinsic properties, but it can be influenced by marketing actions, such as price or advertisement. These studies motivate the research on emotions as regulators of consumer decisions, but our research focuses on a very different way to measure consumer emotions: the text they write in the online reviews of the product.

Most e-commerce web sites allow their users to write product reviews, summarizing their experience to inform other potential customers before purchasing a product. It has been shown that users follow universal rules when they give ratings, influencing the expression of other user opinions (Lorenz, 2009; Wu *et al.*, 2010), which can be modeled through agent-based approaches (Li and Hitt, 2008). Thus, it comes as no surprise that in addition to checking the product ratings, most online shoppers state that they read at least four reviews before buying a product (Senecal and Nantel, 2004). This means that the information contained in product reviews is potentially more important than just the rating scores (and it is perceived as such), since through the review consumers express their experience in detail. At the same time, even if the users might not be aware of it, they also express their emotions at the time of writing.

In this chapter, we explore two main factors driving online consumer communication, namely individual emotional expression through product reviews, and mass media reports

close to the release of a product. We provide a description of a new dataset we collected from `Amazon.com`, followed by an analysis of different properties of collective emotional expression towards a large set of products. We continue by providing an agent-based model based on the framework introduced in Chapter 1, which reproduces certain collective features of the reaction of the reviewer community, as well as the distribution of emotions expressed through user reviews.

2.2 The role of emotions in the reviewer paradox

The main interaction mechanism among users of `Amazon.com` is the text they write in their product reviews, which forms most of the user-generated information contained in the site. A common metric to measure the success of a website is its number of visitors, which depends on the content displayed on the site and the user community it addresses. For a product reviews site like `Amazon.com`, encouraging users to write reviews is an essential factor to generate the content it requires to receive visitors, and thus achieve success. Incentivation schemes can be designed to encourage users to write reviews, but it is necessary to learn first what are the reasons that drive customers to write product reviews in the first place.

To explore the reasons for the creation of product reviews, Wu and Huberman (Wu *et al.*, 2010) outlined how product reviews can be seen from a game theoretical perspective. Writing a product review is an cooperation action with a nonzero cost c , which is particularly high for the case of long reviews. For example, an average review from our dataset contains 143.52 words, and its text is usually composed in a consistent and comprehensible manner. Therefore, the time and effort to produce a product review is definitely non negligible, and it can be considered larger than just providing a star rating, or writing a short message in a chatroom. The whole set of n reviews for a product contain certain amount of information $i(n)$. If we assume that there are no misleading or malicious reviews, $i(n)$ is a monotonously increasing function of n , i.e. more reviews can only mean the same or more information. The amount of information contained in the reviews would approach this way a state of perfect information about the product. This can be assumed to have a finite value i_{max} , but not necessarily reachable in a finite amount of reviews, as $i(n)$ might approach i_{max} asymptotically as sketched in Figure 2.1.

A perfectly rational user would take the decision of making a new review if the cost c is less than the increase in payoff perceived by the user, which is the change in i after the review is created. Early reviews can easily have a large impact on i , but if we accept that reviewers provide redundant information, this change in i decreases with the amount

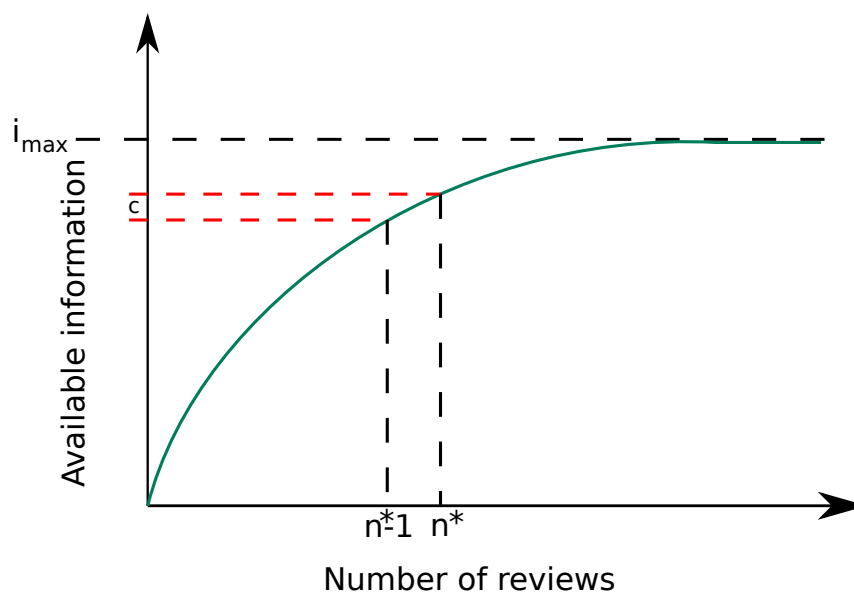


Figure 2.1: Schema of the function of collective available information in the reviews of a product i versus the total amount of reviews n . Given the cost c of making a review, there is a critical value n^* over which a purely rational agent would not create another review.

of previously available reviews. In this situation, there is a critical amount of reviews n^* , after which any new review would increase the total information available in a way that does not pay off for the cost of creating it. As mentioned in (Wu *et al.*, 2010), this kind of paradox is equivalent to the voter's paradox, i.e. after certain amount of reviews, individual users perceive that a new review does not change the collective perception of the product. The large amounts of reviews for certain products, easily having more than 1000 reviews, pose the *reviewer paradox*, as such scenario can easily be beyond the critical value n^* . Our aim is not to measure $i(n)$ or to find n^* , but to analyze the reasons why, under this scenario, members of product reviews communities show this kind of behavior beyond rationality.

One of the possible solutions for such paradox is the heterogeneous perception of the users in terms of their impact on the average rating of a product. Users are encouraged to make reviews that differ from the average, which is empirically validated in different datasets (Wu *et al.*, 2010). This leads to a self-selection behavior in later reviews that can be quantified and modeled (Li and Hitt, 2008), explaining the number of dissenting late reviews. This argument explains part of the late reviews, but fails to provide an answer for all those reviews that *do not differ from the average*. For example, what are the reasons for the existence of the 500th 5-star review for "*Harry Potter and the Sorcerer's Stone*"? One might doubt that the purpose of that review was to influence the already high average rating of the product. The alternative we propose is to study the *nonrational* component

of product reviews.

A valid criticism to the scenario shown in Figure 2.1 is that every review contains a minimum amount of information, in particular regarding the subjective experience of the reviewer. If *I* am writing a product review, even if it happens to be an exact copy of a previous one, it talks about *my* personal experience, and for that reason it contains some relevant information that cannot be included in reviews written by other users. In that situation, the information contained in the review would refer to a subjective state of high relevance for the individual. When this state motivates the action of expressing it through a review, we can refer to it as an *emotion*. This can lead to users writing reviews after the objective information threshold n^* mentioned before, but might also drive the production of early reviews. For example, early product buyers might have emotions driven by marketing campaigns, which is consistent with the previous finding that early reviews are biased towards positive ratings (Li and Hitt, 2008).

This leads us to the hypothesis that emotions compose a feedback mechanism that modifies the individual payoff of a reviewer, leading to the production of reviews even in the presence of a rational paradox. This way, the writing of the emotional experience regarding the product might not only be a cost to pay, but can provide the intrinsic payoff of emotional expression (Rime, 2009). We divide this hypothesis in two subhypotheses along the role of marketing campaigns. First, for most of the products, the emotional content in their reviews should increase for later reviews. Second, another class of products should have decreasing emotional content, and those products should be noticeable through quantization of marketing campaigns and mass media attention.

2.3 A dataset of emotions in product reviews

Amazon.com is not only a top selling platform, it also hosts the largest review community on the Internet, featuring more than 28 million products at the time of analysis. According to Alexa, it is the 16th most visited website, and it reached more than 4.5% of the Internet users, who can purchase and review products, especially books and media. Reviews are always accompanied by a star-rating of the product quality, so shoppers can get a independent, third-person evaluation of the product quality before buying it. By reporting about their experience online, shoppers also have an impact on the market. We have collected a massive dataset of product reviews and sales rankings from Amazon.com, which we combined with a specialized dataset from media attention towards products. Our dataset was generated by three data retrieval programs that were run in sequence:

1. **Product reviews.** During November 2009 we retrieved product reviews of `Amazon.com` through signed requests on its public API. We started by extracting a list of 20000 products from empty searches in the categories of books, music, DVDs, electronics and photography. For each product, we collected the whole set of user reviews available to the data retrieval date, including rating and text for more than 2 million reviews. We followed by performing duplicate deletion, as the Amazon programming interface was delivering repeated products, which contained the same set of reviews for two different product IDs. The way we matched duplicates was by comparing the content of the reviews available for each pair of products, and generating a list that was validated by hand. As a result, our clean dataset on product reviews from `Amazon.com` contains 16670 products with almost 1.8 million reviews. A previous study on `Amazon.com` (Blitzer *et al.*, 2007) only had a maximum of 20 reviews per product, while our work focuses on the whole set of reviews for the listed products. Among the relevant information retrieved for each review, we count with its text, star rating, date, and the amount of helpful and total votes (helpful and unhelpful) provided by other users. Figure 2.2 shows an example of the information displayed on the website along with a product review.

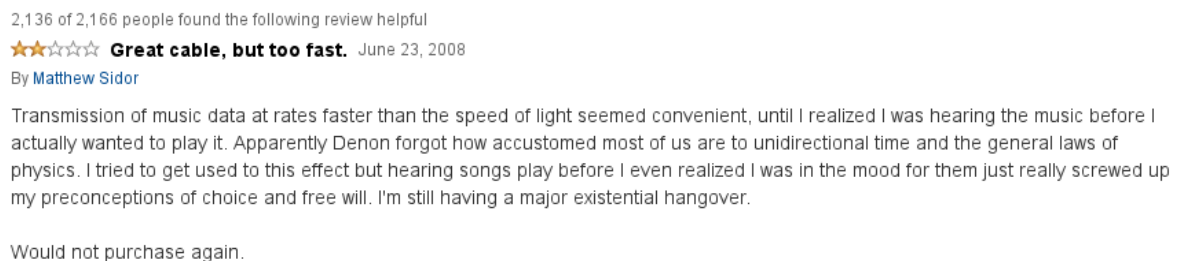


Figure 2.2: Example of product review including text, date, star rating, and counts of helpful and total votes given by other users. The large amount of helpful votes reward this sarcastic review, which points out the impossible properties claimed by the product manufacturer.

2. **Sales rank.** `Amazon.com` does not publicly offer sales numbers for its products, but the weekly product ranks are available for the different categories. For the period between Jan 1st 1999 and Dec 1st 2009, we retrieved the monthly rankings of the 100 top products as sold in `Amazon.com`. This allows us to look up whether products in our dataset are included among the most sold, and how high they are in this ranking.
3. **News volume.** For each product having more than 10 reviews, we obtained the amount of news items appearing in `news.google.com` during the three month in-

terval centered around the date of the first product review. The way to gather this data was by making request to the news archives for the specified months, providing the name of the product as query. Products with a name composed of less than three words were queried by adding the product manufacturer, in order to reduce the amount of false positives in the news search. News items, created mostly by journalists and marketing agencies, are used here to represent the external factors that can influence the response of the reviewers from `Amazon.com`

These three data sources were combined and harmonized in PostgreSQL databases, allowing the matching between product reviews, sales ranks, and news volumes.

2.3.1 Valence in product reviews

We processed the text of each review to produce estimations of the emotional content expressed through it. To do so, we applied the standard procedure introduced in (Dodds and Danforth, 2009), using the ANEW dataset (Bradley and Lang, 1999). This dataset is composed of a lexicon, explained more in detail in Chapter 5, which includes values of valence for more than 1000 English words. This method estimates valence from a text through the naive calculation of the average valence of the words contained in the text, applying

$$v_{\text{text}} = \frac{\sum_{i=1}^n v_i f_i}{\sum_{i=1}^n f_i} \quad (2.1)$$

where v_i is the valence of the i th word of the lexicon, and f_i the frequency of appearance of the word in the analyzed text.

Due to the limited size of the ANEW lexicon, this method should be preferably used on long texts for statistical reasons. To overcome this limitation, we extracted the stem, or root form, of all words in the analyzed text. The stem contains most of the semantic information of a word (and thus its emotional content), and allows us to match similar words rather than exact matches, which eventually improves the statistics. To extract the stem, we used Porter’s Stemming algorithm (Van Rijsbergen *et al.*, 1980), a technique that applies inverse generalized rules of linguistic deflection, mapping deflected words to the same stem. For example, the stem of the words “lovely” and “loving” by that method is “love”, which matches the corresponding word in the ANEW lexicon. This way, the sentiment analysis covers a larger portion of the text and allows to calculate the emotional values for more than 98% of the reviews in our dataset.

The valence estimation for the set of reviews of a product gives us an aggregate of the emotional expression towards the product. As shown in Figure 2.3, the histogram of

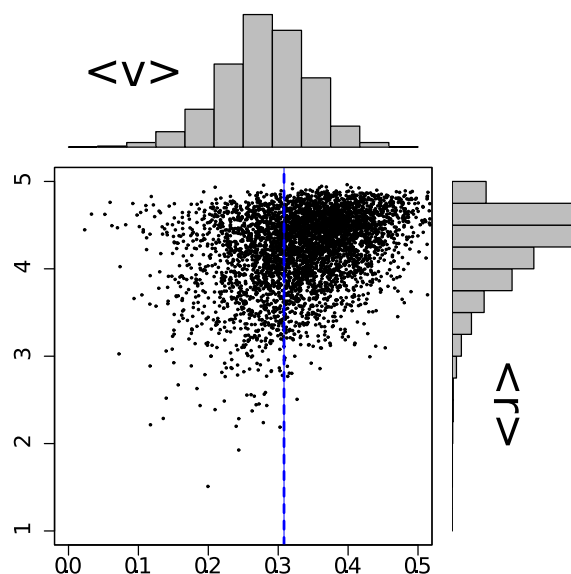


Figure 2.3: Scatter plot of average valence and average rating for each product with more than 100 reviews. The bar-plots on the axes are the average valence distribution (horizontal) and the average rating distribution (vertical). The blue line marks the value $v = 0.31$, the baseline for valence expression in general texts.

average ratings is heavily skewed towards high values, while on the other hand, the average expressed valence is somehow normally distributed around the value $v = 0.31$. This is consistent with our calculations of the baseline of emotional expression in English text from the Internet, as presented in Chapter 5. The average valence of a product review is correlated with its average rating (Pearson's correlation coefficient $\rho = 0.26, p < 10^{-4}$). This correlation is not as strong as we would expect if we assume that both variables measure the same aspect of the subjective experience of the product. In Figure 2.3 we see that products with high average valence are more likely to have a high average rating, but the opposite is not true, i.e. products with a high average rating might have high or low average expressed valence. This means that valence provides additional information that is not solely contained in the star ratings of a product.

In general, the distribution of valences of the reviews for every product seem to follow normal distributions. The inset in the left panel of Figure 2.4 shows a typical valence distribution for a product, with a normal fit calculated using the method of moments. The quality of these fits increases with the amount of available reviews for a product. The left panel of Figure 2.4 shows a barplot of the mean and standard deviation of the R^2 value of the normal fits, binning on the logarithm of the amount of reviews. As expected, the R^2 value of the fit increases with the amount of datapoints used for the estimation,

reaching average values above 0.8 when more than 100 reviews are available.

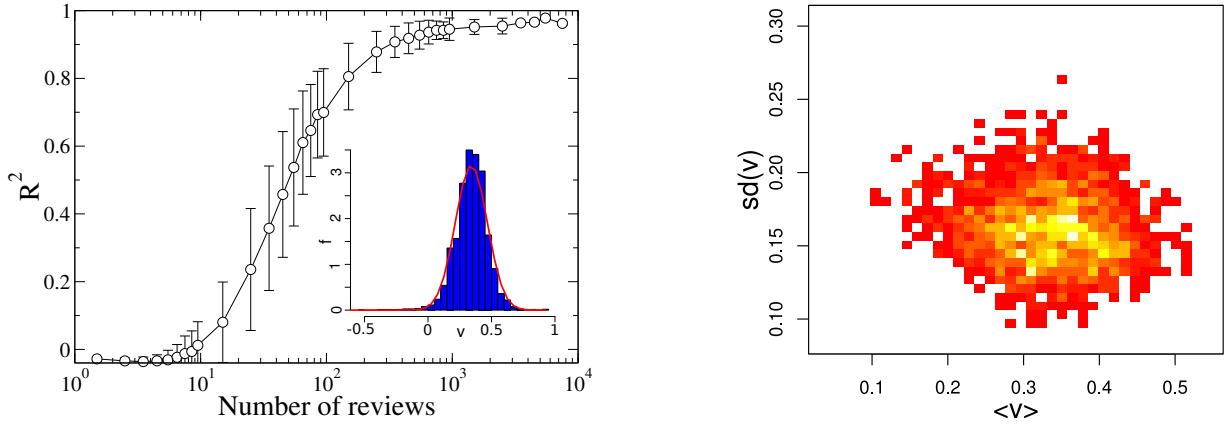


Figure 2.4: Left: Bar-plot of mean R^2 of the normal estimation of the valence distribution versus the amount of reviews of the product. Inset: typical distribution of review valences. The red line is the PDF of the normal estimation. Right: Distribution of the mean and standard deviation of the valence of products with $R^2 > 0.8$.

The variability of the valence distribution gives us an idea of the different possible reactions of the `Amazon.com` community to a product. Assuming a minimum value of $R^2 = 0.8$, we investigate the parameters that define the underlying normal distribution of the valence, i.e. mean and standard deviation. The right panel of Figure 2.4 shows a heatmap of the density of products over the space of mean valence $\langle v \rangle$, and standard deviation of the valence $sd(v)$. This shows a unimodal distribution centered around a mean of 0.3 and standard deviation of 0.15. In addition, all the products have standard deviations of the valence above 0.1, showing the maximum emotional agreement found in the reviews of our dataset.

2.3.2 Collective emotional polarity towards products

Our results based on valence values allow us to verify that emotional expression in product reviews does not show anomalies with respect to generalized emotional expression. However, this analysis does not take into account whether review texts can be considered to contain positive or negative emotions, due to the fact that we did not impose limits given the baseline found in Chapter 5. In our study of emotions in product reviews, we do not need to know a perfect estimation of the valence expressed through a text. Such an estimation might not be feasible with current analysis techniques, or not even possible due to the heterogeneity in human emotions. A recent study (Paltoglou *et al.*, 2012) measures the quality of this method and its relation to the agreement between humans when

rating the emotional content of text. Before aggregating the emotions in the reviews of a product, we process each review by estimating if it contains perceptible emotions, and if these are positive or negative with regard to valence. From each review, we calculate a value of *emotional polarity*, calculated as

$$e_{\text{text}} = \begin{cases} -1 & \text{if } v_{\text{text}} < \mu_{\text{anew}} - \frac{\sigma_{\text{ANEW}}}{\sqrt{n_{\text{text}}}}, \\ 1 & \text{if } v_{\text{text}} > \mu_{\text{anew}} + \frac{\sigma_{\text{ANEW}}}{\sqrt{n_{\text{text}}}}, \\ 0 & \text{otherwise.} \end{cases}$$

where $\mu_{\text{anew}} = 0.31$ and $\sigma_{\text{anew}} = 0.47$ are the mean and standard deviations of the distributions of word valence reweighted by their frequency of appearance on the Internet, as shown in Figure 5.3 of Chapter 5. $n_{\text{text}} = \sum_{i=1}^n f_i$ is the amount of terms from the ANEW dataset detected in the stemmed text. The rationale behind this measure is to make a statistical test of whether the estimated valence is above or below the average of human expression. Consider a text of arbitrary length taken at random from the Internet. Applying the central limit theorem, we know that the valence estimation from such text v_{random} , as calculated in Equation (2.1), approaches a normal distribution of mean μ_{anew} , and standard deviation σ_{anew} divided over the square root of the amount of terms used to make the estimation, n_{text} . Equation (2.3.2) tests whether v_{text} is above or below the mean with a distance of one standard deviation, given n_{text} . This way, a text with a single term from the ANEW dataset would need it to be very high or very low in order to safely infer that it has emotional content.

Given the emotional polarity of the set of reviews for a product R_p , we can aggregate emotions towards the product by calculating the ratios of positive P_p , negative N_p , and neutral U_p reviews, as

$$P_p = \frac{1}{|R_p|} \sum_{j \in R_p} \Theta[e_r^j = 1], \quad N_p = \frac{1}{|R_p|} \sum_{j \in R_p} \Theta[e_r^j = -1], \quad U_p = \frac{1}{|R_p|} \sum_{j \in R_p} \Theta[e_r^j = 0] \quad (2.2)$$

where $|R^p|$ is the number of reviews created about the product p , and $\Theta(x)$ is a Boolean function that returns +1 if the argument is true and zero otherwise. Clearly, $P_p + U_p + N_p = 1$ for all products with at least one review. These three values map each product into a space that can equally represent products that did not elicit collective emotional responses (high U_p), created positive or negative collective emotions (high P_p or high N_p), or created polarized emotional reactions (both high P_p and N_p). The two degrees of freedom of Equation (2.2) can be mapped to a plane where each product is point inside a triangle with a distance to the vertices inversely proportional to U_p for the upper

vertex, N_p for the lower left vertex, and P_p for the lower right vertex. The Cartesian coordinates of this representation are given as $(x, y) = \left(\frac{1}{2}(2P_p + U_p), \frac{\sqrt{3}}{2}U_p\right)$, a system of barycentric coordinates in an equilateral triangle. The left panel of Figure 2.5 shows this transformation of the scatter plot of the product emotions. One can notice the absence of points in the lowest part of the triangle, which corresponds to the cases of simultaneous large ratios of positive and negative reviews.

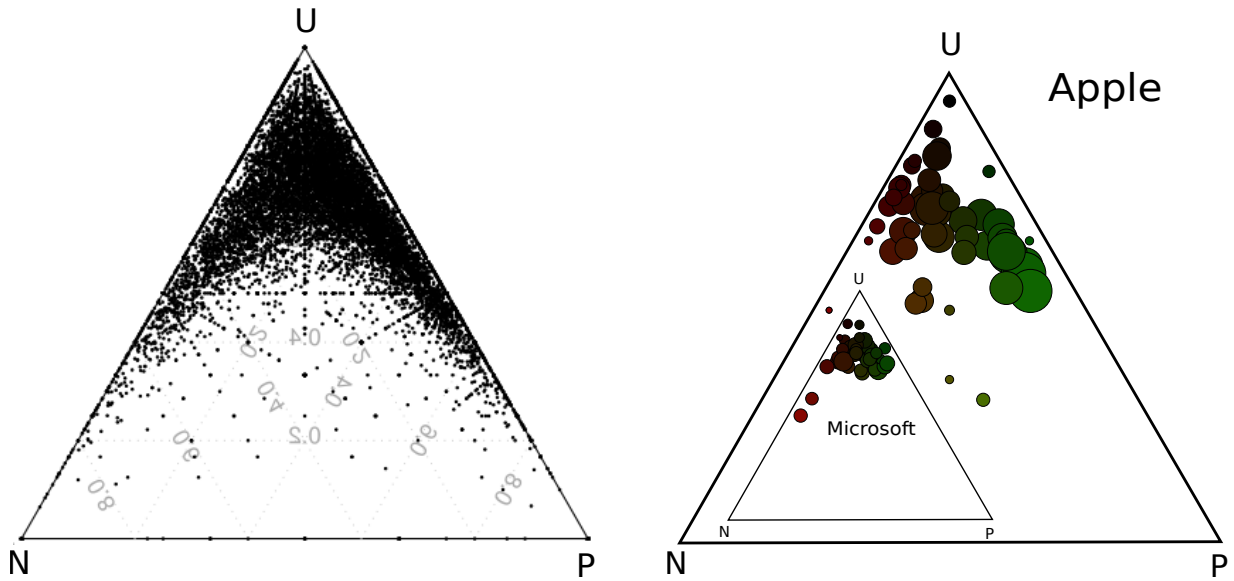


Figure 2.5: Left: Ternary scatter plot of P_p , U_p , and N_p for all the products in our dataset. The distance of the point from each vertex of the triangle is inversely proportional to the corresponding ratio. Right: ternary plot for the products of *Apple*, with points of size proportional to the logarithm of their amount of reviews, and color according to their position in the triangle – a combination of green, black and red for P_p , U_p , and N_p respectively. Inset: same plot for the products from *Microsoft*.

The collective emotions towards the products of **Amazon.com** are arranged along a characteristic curve that avoids simultaneous high P_p and N_p . This shape is in line with the normal distribution of product valences we explained in Figure 2.4. Only bimodal valence distributions would appear in the lower part of the triangle, and the curve in the triangle of Figure 2.5 is essentially a projection of the unimodality of valence in the product space. But this projection is not exempt of useful information, as it lets us compare the collective reaction to different products through their reviews. For example, we can visualize the emotions towards products of different brands, and evaluate brand reputation. As an illustrative example, the right panel of Figure 2.5 shows the cases for *Apple* and *Microsoft*, two direct competitors in the Computer and Electronics category. In this plot, points have a size proportional to the logarithm of the amount of reviews available for the product.

It is simple to note that *Apple* products elicited more positive responses than those of *Microsoft*, with a clear branch towards P_p . These positively accepted products are the Ipod family, sales leaders in portable audio players. For the *Microsoft* case, two products show very high negative responses, which correspond to certain wireless accessories for Xbox that seemed not to work properly.

Market researchers usually find barriers to use data from online origin, as it is not trivial to analyze and summarize them. In general, marketing departments have access to large amounts of data, similarly to our `Amazon.com` dataset, but the lack of a quantitative approach limits their usage. Data visualization of collective emotions, like the one shown in Figure 2.5, can provide means to summarize the analysis of this data as an input for decision support. While these metrics might be interesting for business applications, our focus is to use them to study human behavior as explained in Section 2.2.

2.3.3 Reviewer activity and emotional expression

Once we quantify emotional response through the reviews of a product, we can proceed to test the hypothesis of emotions as motivation for writing product reviews. Our analysis focuses on the time dynamics of the ratios of positive $P_p(t)$, negative $N_p(t)$, and neutral $U_p(t)$ reviews for a product up to time t . These ratios evolve through time, showing us the different trends in emotional expression regarding a product. To compare these trends in emotions, we fixed a time interval of two years after the first review. For this reason, we only included in this analysis products with a review older than two years (6322 products), and we limited our calculations only to the first 2 years of reviews.

For a each product p , we define its ranked time correlation coefficients ρ_N , ρ_U , and ρ_P , as the Pearson's correlation coefficient between the time t and the ranked value of $N_p(t)$, $U_p(t)$, and $P_p(t)$ respectively, for the two years taken into account. For a typical product, the time series of emotion correlations approaches a final saturation value, usually in a nonlinear manner that we simplify through the rank transformation of the ratios. This way, we do not compare absolute values of the emotional ratios, but if later values are higher or lower than previous values. An example of the time series used for this analysis is shown in Figure 2.6, for the case of a DVD movie. Users consistently created more and more negatively emotional reviews, leading to a constant increase in the rank of $N_p(t)$ and a corresponding decrease in $U_p(t)$. The initial ratio of negative reviews was significantly less than the final one, which was not the case for positive reviews that show a $P_p(t)$ without a clear trend. The ranked time correlations for this product, $\rho_p = -0.207$, $\rho_n = 0.981$, and $\rho_u = -0.989$ (all $p < 10^{-7}$), allow us to conclude that later reviews for the product were

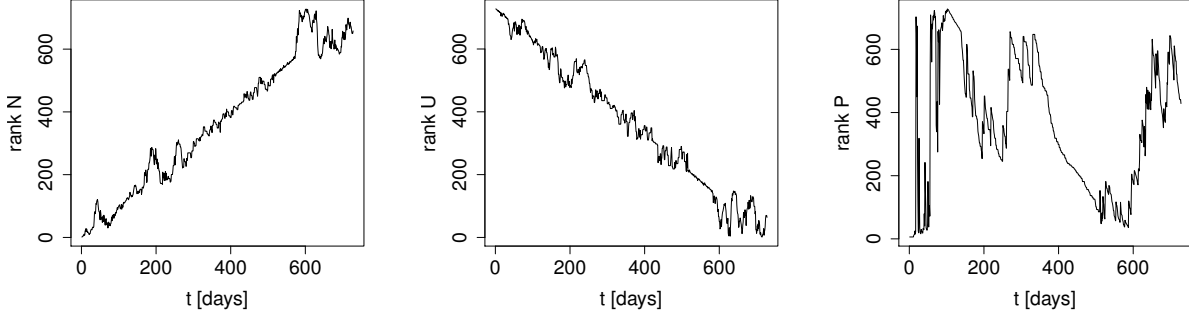


Figure 2.6: Evolution of ratios of emotional expression for the product *Star Wars: Episode 1*. The ratio of negative reviews N_p increased consistently, decreasing U_p along the two years analyzed. There is no clear trend for the case of P_p . The ranked time correlations for this product, $\rho_p = -0.207$, $\rho_n = 0.981$, and $\rho_u = -0.989$, indicate that negative emotions motivated late reviews.

more emotional and more negative than early reviews.

From all the analyzed products, those with non-significant ($p \geq 0.001$) value of ρ_P , ρ_N , or ρ_U , compose 8.64% of our dataset. For all the rest, we find a trend in one or more of the emotional ratios, as shown in the histograms of Figure 2.7. These histograms have the typical gap around 0, as it requires a very large sample to statistically conclude about the significance of very low values of correlation coefficients. Figure 2.7 provides support to our first hypothesis, i.e. most products show increased emotionality in their reviews. The histograms for ρ_N and ρ_P show that more products have increasing values of N_p and P_p over time.

A Student's t-test on the distributions of ρ_N and ρ_P shows that we can reject that their means are 0 with $p < 10^{-15}$, and estimated as $\langle \rho_N \rangle = 0.143$ and $\langle \rho_P \rangle = 0.111$. Furthermore, their medians are 0.38 for ρ_N and 0.31 for ρ_P , leading us to conclude that most of the products have increasing values of ρ_P , ρ_N , or both. For the case of ρ_U , the histogram shows two maxima for largely positive and negative values, but this is higher for negative ρ_U , which corresponds to increasing emotional content (positive and/or negative). A t-test rejects a mean equal to 0 with $p < 10^{-13}$, with a mean of $\langle \rho_U \rangle = -0.062$, and a median of ρ_U at -0.22 . These results lend support for the hypothesis that emotions in general are the motivation for review creation in most of the products.

The sign of ρ_U distinguishes between products for which the emotional content of their reviews increases ($\rho_U < 0$) or decreases ($\rho_U > 0$). Figure 2.8 shows the evolution of the mean value of $U_P(t)$ for each of these product classes, measured on a weekly basis. As

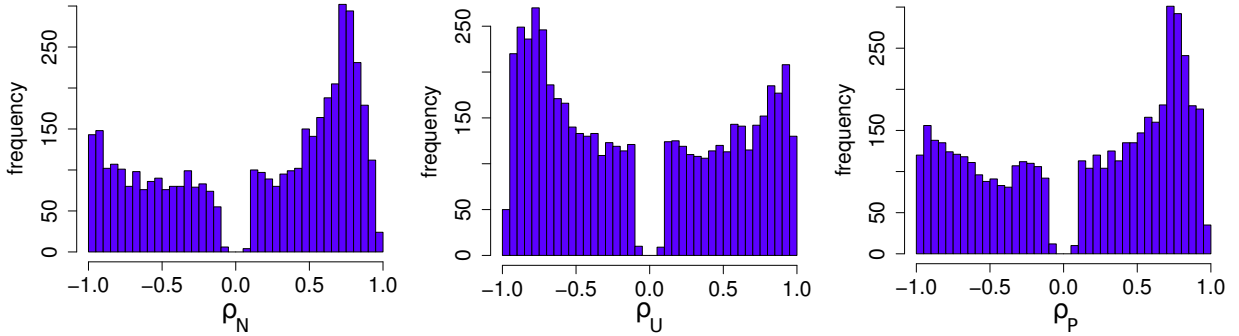


Figure 2.7: Histograms of significant ($p < 0.001$) ranked time correlations ρ_N (left), ρ_U (center), and ρ_P (right).

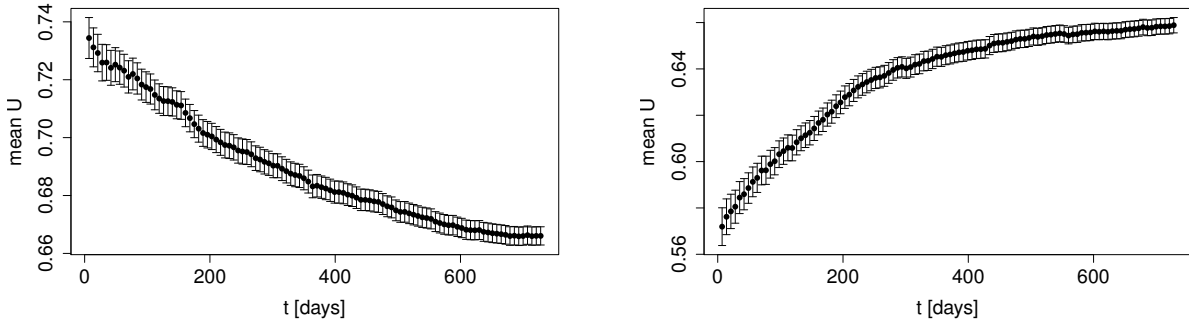


Figure 2.8: Mean value of $U_p(t)$ for products with $\rho_U < 0$ (left) and $\rho_U > 0$ (right) and $p < 0.001$. Error bars show standard error around the mean.

expected, the neutral ratio of products with $\rho_U < 0$ decreases constantly along the two years considered, falling from 0.73 to close to 0.66. This means that the average ratio of emotional reviews increases from 0.27 to 0.34, which shows how emotional expression increases in this majority of products.

The case of $\rho_U > 0$, shown in the right panel of Figure 2.8 shows the opposite trend, with increasing emotionality. Our second hypothesis states that products with $\rho_u > 0$ should receive more emotional reviews in the beginning due to marketing campaigns and traditional media. We validate this hypothesis controlling for mass media attention through our measure of amount of items in `news.google.com`, which we denote as m_p . As shown below, the distribution of m_p is highly skewed, which we need to take into account for statistical tests, as we cannot assume normality or symmetry. We performed a Wilcoxon signed-rank test, comparing the amount of news items for products with $\rho_u > 0$ and $\rho_u < 0$. The test rejects the hypothesis that they are the same ($p = 0.00051$), indicating that mass media attention is larger for the case of decreasing emotionality.

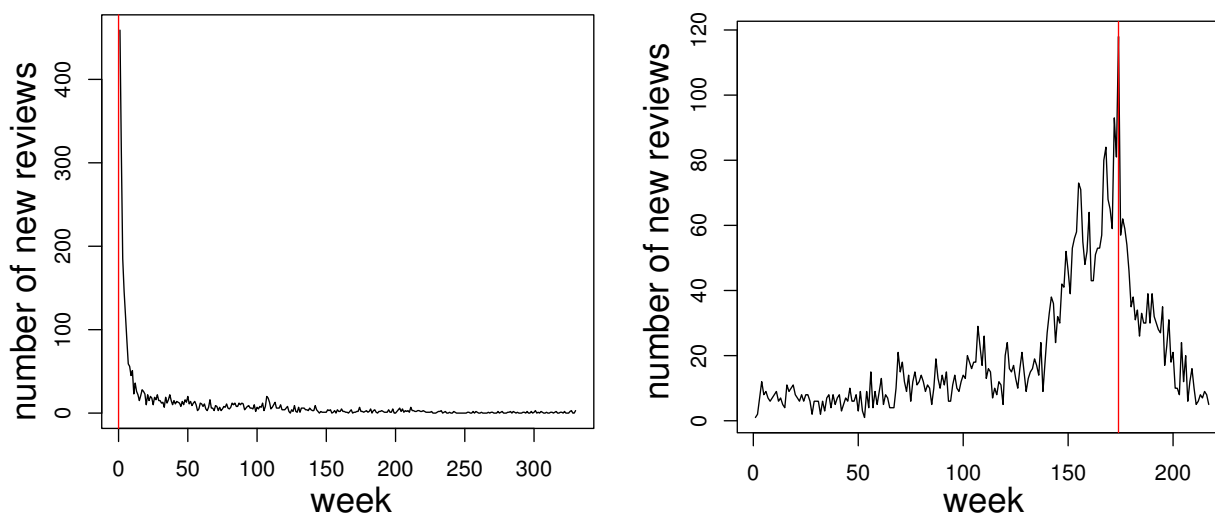


Figure 2.9: Left: weekly amount of new reviews for “Harry Potter and the Deathly Hallows”, as an example of product with a large peak in the first week, followed by a strong decrease. The popularity of the Harry Potter saga and the media attention around its release triggered this behavior. Inset: weekly amounts of new reviews for “Twilight”. The popularity of this book increased through time, creating a peak in the amount of new reviews months after its release.

This analysis is sufficient just to support the hypothesis, but the dependency between the evolution of emotional information and mass media attention requires more precise data sources than our news estimation.

2.4 User activity in a cultural market

Our analysis of product reviews has focused so far on the dynamics of emotional expression in terms of valence. As modeled in our framework, explained in Chapter 1, another important component of emotions is arousal, which drives the activity related to the emotion. In this section, we analyze the collective patterns of activity in product reviews, looking for traces of the influence between users. Our `Amazon.com` dataset contains the dates of all the reviews for the selected set of products, which form a time series of review creation. For example, Figure 2.9 shows the weekly rate of new reviews for the products “Harry Potter and the Deathly Hallows”, and “Twilight”. These examples show the nonstationary nature of the creation of reviews, as users do not simply create reviews at a constant rate. Both examples have strong peaks, marked with a red vertical line at the week with the maximum amount of reviews.

The most striking difference between the time series shown in Figure 2.9 is the time of the activity peak, which is very early for the first but much later for the second. This is a pronounced example of the influence of featured media in the response of the product reviews community, as “Harry Potter and the Deathly Hallows” was subject of a large media campaign, amplified by the success of previous books of its series. On the other hand, “Twilight” took much longer to receive many reviews, as it was not subject of any relevant campaign. For this case, early adopters created reviews that lead others to purchase the product too, cascading into success through word of mouth. Furthermore, the movie version of the novel lead to an even higher increase in the review rates as we see them. This different treatment of products in mass media not only influences the trends of emotional expression, as shown in the previous section, but also influences the time in which users purchase the products and create the reviews.

2.4.1 Product inequality in featured media

The typical time series of review rates for products is nonstationary, and average measures along the lifetime of the product are not indicative of the behavior of the community. On the other hand, the maximum amount of reviews for a day, and the time elapsed for the occurrence of that maximum can provide important information about the relation between user expression and product success. For each product, we have a time series of reviews $n(t)$ created at the week t after its release, which we assume to be the date of the first review. The value of the maximum of this time series, n_{max} , differs a lot across products. The left panel of Figure 2.10 shows the distribution of n_{max} for all the products in our dataset in a log-log scale. The large variance of this distribution is characteristic of a power-law, which we fitted as $P(n_{max}) \sim n_{max}^\alpha$. A maximum likelihood estimation leads us to a value of $\alpha = -2.4 \pm 0.1$. This large variability means that most of the reviews are concentrated in only a small set of products, known as the winner-takes-all effect (Rosen, 1981) of cultural markets (Leskovec *et al.*, 2007). The inter-product dynamics that generates such distribution is outside of the scope of our current analysis, and is left as an open question for future work in marketing and product acceptance.

The distribution of n_{max} is also an interesting indicator of the sales of a product, as it is expected some consumers create reviews in `Amazon.com`. Our dataset contains information on the monthly product rankings, which we use as an estimation of the success of a product in terms of sales within its market. In particular, we find dependencies between the best monthly rank achieved by a product sr_p , and the maximum amount of reviews n_{max} . To estimate the success of products, we defined a cutoff value c_{rank} , which divides the products if $sr_p \leq c_{rank}$, from the rest where $sr_p > c_{rank}$.

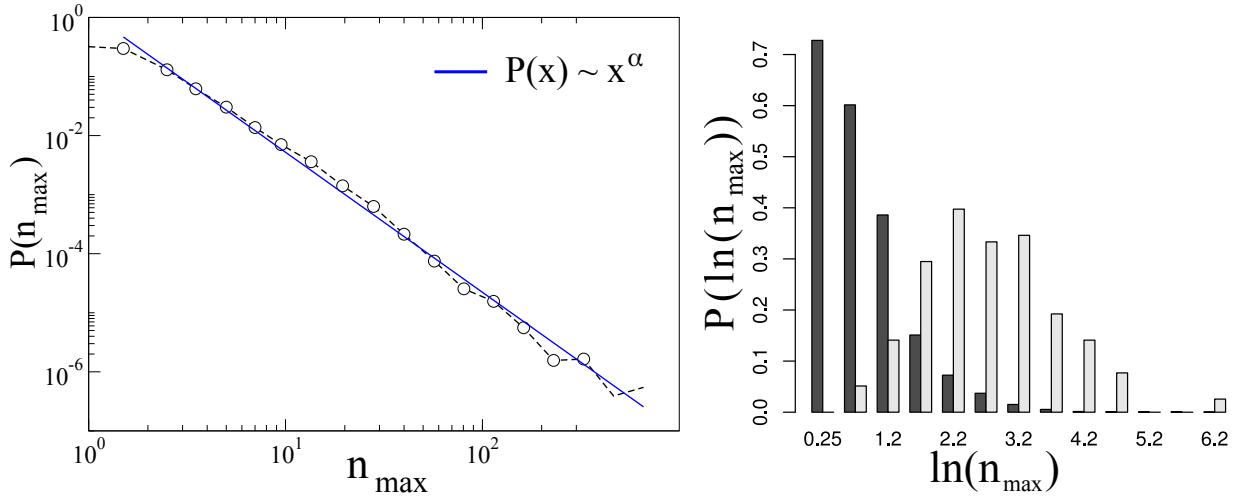


Figure 2.10: Left: peak size distribution $P(n_{max})$, fitted by a power law with exponent $\alpha = -2.4 \pm 0.1$. Right: probability density function of n_{max} for products that reached the top three (gray) and that did not (black). Products that reached a very high sales rank are likely to have much larger reviews peak size than the others.

Each value of the cutoff results in two distributions of n_{max} , one for successful products and another one for the rest. For each value between 1 and 100, we performed Wilcoxon tests on the given distributions of n_{max} , finding that the difference between both is maximum when $c_{rank} = 3$. The right panel of Figure 2.10 shows the rescaled distributions of n_{max} for products that reached the top three ($sr_p \leq 3$), and that did not ($sr_p > 3$). The log-linear scale helps us to notice the large difference between both distributions, showing that products that did not reach the top three have significantly lower n_{max} than those which did.

Apart from n_{max} , a key feature of the reviewing activity of the community is captured by the time when such peak value is reached, which we denote as t_{peak} . The examples shown in Figure 2.9 illustrate the different scenarios that might arise from different t_{peak} , which we assume to be related to mass media attention towards the product. We approximate this attention through the amount of news items in `news.google.com`, m_p , as explained in Section 2.3. The distribution of news items per product follow a highly skewed distribution, as shown in the left panel of Figure 2.11. The news sources of `news.google.com` seem to devote most of their attention to a small set of products, while the majority is largely ignored. Our aim is not to understand the origin of this distribution and its possible power-law properties. We want to understand how the value of m_p influences the reviewing pattern of the community, in particular related with the speed of its response, i.e. through t_{peak} .

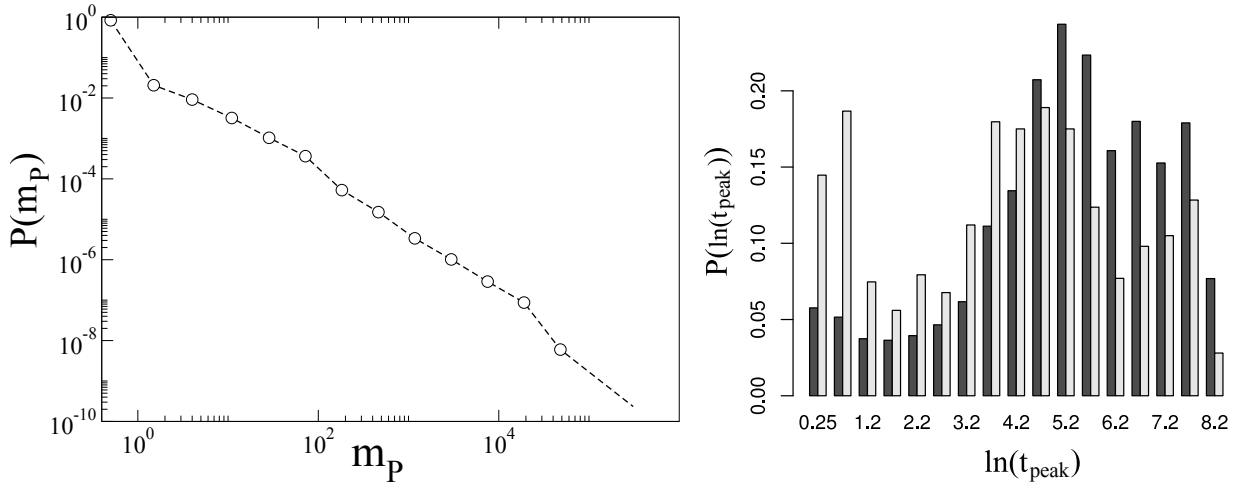


Figure 2.11: Left: Probability density function of the amount of news items m_p for all analyzed products. As we can see, the distribution is very broad, while the jump at the beginning is due to the large number of products without any news about them. Right: probability density function of the time to reach the peak number of review activity, t_{peak} , for products with $m_p \leq 4$ (black), and $m_p > 4$ (gray). For a product, having more news implies a significantly larger probability of having an early peak in the review activity.

We explore the distribution of t_{peak} for products depending on their amount of news items. We divide the set of products in the ones that receive high media attention if their amount of news items is above a cutoff value c_{news} , i.e. $m_p > c_{news}$, and the ones that are not featured in mass media if $m_p \leq c_{news}$. The right panel of Figure 2.11 shows the distributions of t_{peak} for products divided by a cutoff value of $c_{news} = 4$, on a logarithmic scale of t_{peak} . The most striking difference between these distributions is that, for the case $m_p > c_{news}$, it shows a bimodality that does not appear for the case of low media attention. In fact, the probability of having a peak shortly after the release is much lower for products without mass media attention than for the others. Details about the role of the cutoff value in the probability of having an early peak are presented in Appendix A.

As a general observation, we find that the probability of having a peak in activity soon after the release of the product is maximally increased when a product appears in at least one news item, i.e. $c_{news} = 0$. From this analysis we conclude that media attention towards products is very heterogeneous, and that it has a significant impact on the speed of the response of the reviewer community. We quantify this through the peak time t_{peak} , which is more likely to be close to the product release if it has at least one result in news.google.com. This supports our hypothesis of the role of traditional media in reviewing activity, which is consistent with our results of the previous section.

2.4.2 Influence among reviewers

After testing the influence of mass media in the activity of the user community, we ask how users influence each other through their reviews. The statistical techniques to analyze this kind of complex interaction are somewhat different, involving the application of tools from statistical physics. If we conceive the community as a set of users that communicate and couple their activity, we can measure the strength of this coupling through the time when they create reviews. Focusing on a specific product, the time series of amount of new reviews $n_p(t)$ can be seen as a fluctuation on the activity of the community (see Figure 2.9). This kind of fluctuations of human activity can be analyzed through fluctuation scaling (Eisler *et al.*, 2008), which measures the degree of coupling in the activity of the elements of a complex system. This technique is based on the relation between the mean and the standard deviation of the time series during a fixed time range. We calculated the mean activity for a product from $n_p(t)$ as $\langle n_p \rangle = \frac{1}{T} \sum_{t=1}^T n_p(t)$, and its standard deviation as $\sigma_p = \left(\frac{1}{T-1} \sum_{t=1}^T [n_p(t) - \langle n_p \rangle]^2 \right)^{1/2}$. We fixed a time range of two years, i.e. $T = 104$ weeks, focusing only on the products of our dataset that existed in `Amazon.com` for more than that period.

The theory of Fluctuation Scaling lets us expect certain regularities in the relation between $\langle n_p \rangle$ and σ_p . The main one is the relation $\langle n_p \rangle \sim \sigma_p^\beta$, where the exponent β is restricted to the interval $[0.5, 1]$. If the users create reviews in a completely independent manner, σ_p would grow with $(\langle n_p \rangle)^{1/2}$ (Central Limit Theorem). On the other hand, if the users were fully coupled, $\beta = 1$, and the activity of a single user could be deterministically predicted from the activity of the rest of the community. This scaling property holds when the time series of the whole community can be assumed to be stationary, i.e. it has finite mean and standard deviation. For the case of online communities, this approximation is valid when the size of the community is much larger than the maximum amount of users involved in a fluctuation. A previous analysis of Facebook application installations (Onnela and Reed-Tsochas, 2010) shows how Fluctuation Scaling can be applied when this condition is met. This approximation is valid for the case of product reviews when the maximum amount of reviews for a product is much smaller than the total amount of users. `Amazon.com` counts with millions of users, while the maximum amount of reviews of a product does not reach more than few thousand, so we can expect the relation $\langle n_p \rangle \sim \sigma_p^\beta$.

The result of Fluctuation Scaling analysis on the set of studied products reveals a relation $\langle n_p \rangle \sim \sigma_p^\beta$, of $\beta = 0.7638 \pm 0.028$. The binned data and the regression result are presented in Figure 2.12, revealing the influence between the users of `Amazon.com`. The value of the exponent, above 0.5, is typical of complex systems in which there is coupling between the system's elements, which is equivalent to the existence of social influence among the users

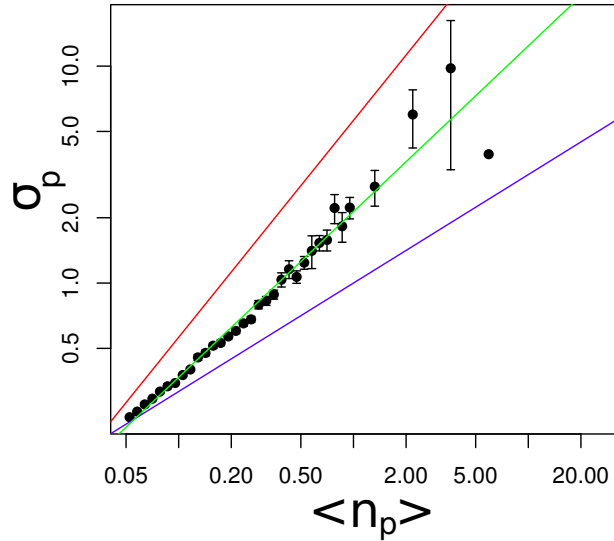


Figure 2.12: Mean amount of weekly reviews per product $\langle n_p \rangle$ versus standard deviation σ_p . Dots show the mean of σ_p on logarithmic binning, and error bars show the standard error in the bin. The green line shows the relation $\langle n_p \rangle = \sigma_p^\beta$, for which we estimated $\beta = 0.7638 \pm 0.028$ ($R^2 = 0.96$). The red and blue lines are visual guides for the case of $\beta = 1$ and $\beta = 0.5$ respectively.

of the product reviews community.

Previous analysis of this kind of data from Facebook applications (Onnela and Reed-Tsochas, 2010) revealed the existence of two regimes, one of exponent 0.55 and another of exponent 0.85. They corresponded to two different kinds of coupling, local versus global respectively. We tested this possibility in our data, applying regression splines with one breakpoint over the logarithmic bins. The result was very similar to the one given by simple linear regression, leading us to conclude that we can safely discard the existence of two different regimes in the coupling of the users. A plausible explanation for this is the fact that product reviews are, in general, written only after the purchase of the product, which is a much more costly action than the free installation of an Facebook application. The coupling between the users of Amazon.com is stronger than the coupling among users of unsuccessful Facebook applications, but weaker than the coupling of users successful ones.

2.4.3 Word of mouth versus mass media

Marketing studies on box sales for movies suggest that marketing campaigns influence movie watchers mostly during the first week (Dellarocas *et al.*, 2007). Due to their high cost, these campaigns are usually focused to create a large attention in short timespan, aiming to trigger *word of mouth effects* across customers. This way, mass media should not only influence the time when reviewing activity is maximum, but also the initial amount of reviews for a product. In this section, we present how mass media attention can be seen as an exogenous influence that drives early reviews, creating large heterogeneity in terms of reviews. We analyze the way the distribution of amount of reviews evolves in time, in order to quantify how user interaction reshapes the variability in the amount of reviews for a product.

We defined the new amount of reviews for a product during the week t after its release as $n(t)$, and the *total* amount of reviews available at that time as $N(t) = \sum_{t'=1}^t n(t')$. By definition, $n(1) = N(1)$, which is the initial amount of reviews created for the product during the first week. If mass media drives this initial attention, m_p and $N(1)$ should be related. Indeed, we find a significant and positive correlation between $N(1)$ and the amount of news items m_p of $\rho = 0.46$ (95% confidence interval: [0.437,0.482]). This means that the amount of items in featured media sources, including marketing campaigns and featured news, boosts the creation of early user reviews. This is also supported by the shape of the distribution of $N(1)$ across products, $P(N(1))$, shown in the left panel of Figure 2.13. The large variance of $P(N(1))$ resembles that of $P(m_p)$, shown in the left panel of Figure 2.11, indicating that mass media creates an initial product inequality typical of the winner-takes-all effect (Rosen, 1981; Leskovec *et al.*, 2007).

Our observations indicate that the large variability of product reviews during the first week is not constant through time. The distribution of the amount of reviews per product two years after the release, $N(104)$, has different properties (note that we restricted our analysis to products that have been in **Amazon** for more than two years). The right panel of Figure 2.13 shows this distribution in a log-linear plot, which does not have the characteristic scaling shape of $P(N(1))$. The amount of reviews after two years seems much closer to a lognormal distribution than to a power-law. This suggests that user interaction through product reviews has different properties than mass media. Following the same methodology for growth rates in social media as in (Asur *et al.*, 2011; Wang and Huberman, 2012), the Q-Q plot of $\ln(N(104))$ shows that the data fits well to a lognormal distribution, with the exception of low quantiles due to its integer nature. This kind of distribution is an indicator that a stochastic multiplicative process could be the driving process behind the creation of new reviews, which we test in the following.

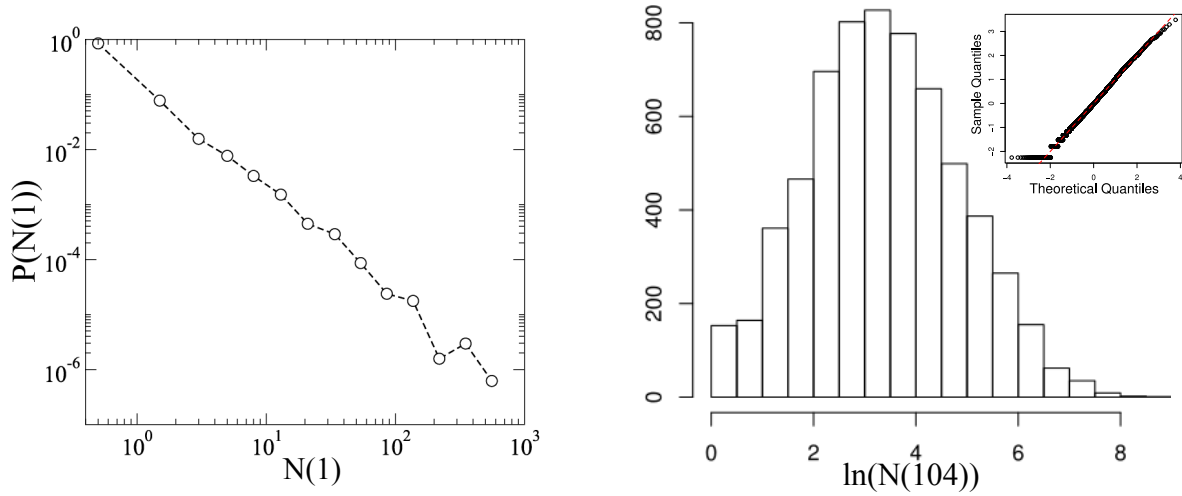


Figure 2.13: Left: distribution of the amount of reviews in the first week $N(1)$ on a log-log scale, with logarithmic binning. Right: distribution of total amount of reviews after two years $N(104)$ on a log-linear scale. Inset: Q-Q plot of $\ln(N(104))$, comparing with a normal distribution.

A multiplicative growth process adds a certain amount of new reviews at every time period, $n(t)$. This amount divided by the accumulated number of previous reviews, $N(t-1)$, gives us the growth rate at time t , $r(t) = \frac{n(t)}{N(t-1)}$. The rationale behind this metric is to approximate the amount of new reviews triggered by a previous one, serving as an estimator of the influence among users. According to (Asur *et al.*, 2011), the null hypothesis would be that the growth rate is normally distributed. We tested this hypothesis against our data to find that the distributions of growth rates are very skewed, and thus, the normality hypothesis clearly fails. After the first weeks, the distribution of $r(t)$ is closer to follow lognormal distributions. Figure 2.14 shows three examples of Q-Q plots for the logarithm of the growth rates after 4 weeks ($r(4)$), 50 weeks ($r(50)$), and 104 weeks ($r(104)$). This result is in line with the recent observations of Wang and Huberman (Wang and Huberman, 2012), analyzing the growth rates of first posting in **Twitter** discussions. The amount of users involved in a **Twitter** discussion also grows with lognormally distributed rates, as a contrast to the *actual* amount of messages, which grows with normally distributed rates. As users generate at most one review per product, our analysis case is equivalent to the one of first posting of (Wang and Huberman, 2012), yielding consistent results.

The quality of the lognormal approximation increases for later t , as the distribution of $N(t)$ is more skewed (Figure 2.13 right) at early times (i.e. the first weeks after the release of the product), but it converges to log-normal at later times due to the stochastic growth process we that drives the evolution of the system. The quality of the log-normal

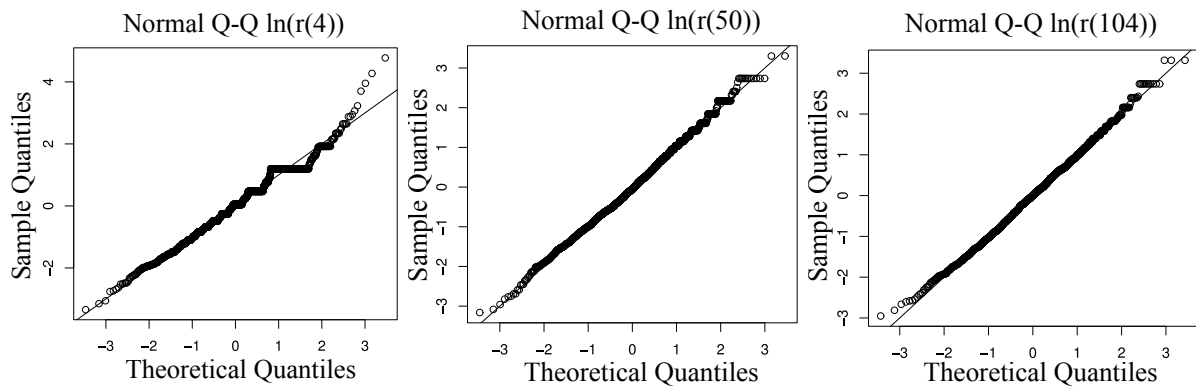


Figure 2.14: Q-Q plots for the growth rates of product reviews during week 4 (left), 50 (middle), and 104 (right). It is clear that for longer time periods the normal fit is more accurate.

approximation is demonstrated in Figure 2.15, where we plot the Kolmogorov-Smirnov distance (Massey, 1951) between the theoretical log-normal distribution with parameters estimated from the data, and the distribution of the logarithm of the weekly growth rate. It is shown that the distance decreases with time and converges to a very low value below 0.05 in less than 12 weeks. This gives us an idea of the speed of user interaction versus the effect of mass media, as the growth rate successfully fits a lognormal in less than 3 months after the release of the product.

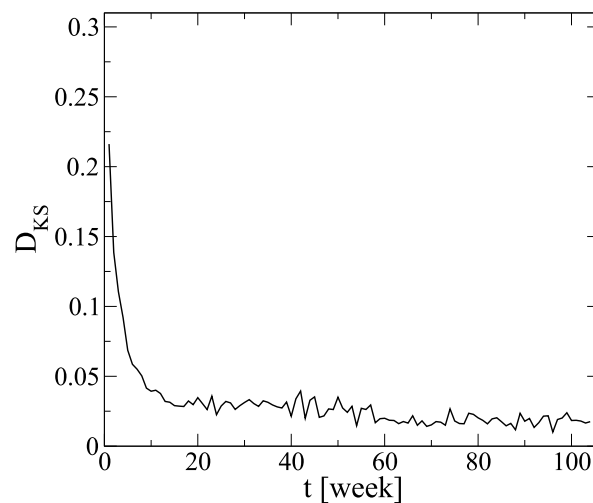


Figure 2.15: Time series of the Kolmogorov-Smirnov distance between the fitted and real logarithm of the growth rate.

Based on the above findings, we propose a statistical model that provides explanation of

how the word of mouth effect within the community changes this distribution of reviews per product. The increase of the amount of reviews for a product, as a growth process is described by the equation: $N(t+1) = [1 + r(t)]N(t)$, where the growth rate at each time corresponds to $r(t) = n(t)/N(t-1)$. Therefore, the amount of reviews at any moment can be expressed as a function of the initial amount of reviews and the time series of growth rates:

$$N(t) = \sum_{i=1}^t n(i) = N(1) \prod_{i=1}^{t-1} [1 + r(i)]. \quad (2.3)$$

In our data analysis, we found that the growth rates at each time step follow a log-normal distribution, i.e. $r(t) = e^{\xi(t)}$, where $\xi(t) \sim \mathcal{N}(\mu(t), \sigma)$. It follows that the amount of reviews at time t is a function of the realization of the values of $\xi(t)$:

$$N(t) = N(1) \prod_{i=1}^t [1 + e^{\xi(i)}]. \quad (2.4)$$

As $r(t)$ follows a log-normal distribution, its mean $\langle r(t) \rangle = e^{\mu(t) + \sigma^2/2}$ and variance $Var[r(t)] = (e^{\sigma^2} - 1)e^{2\mu(t) + \sigma^2}$ are finite. Therefore, a product of a sequence of $1 + r(t)$ converges to a log-normal distribution over time (Lyapunov central limit theorem (Billingsley, 1968)) and $N(t) \approx e^{\xi_a}$, where $\xi_a \sim \mathcal{N}(\mu_a, \sigma_a)$. This explains the lognormal distribution of the amount of reviews after two years $N(104)$, which is shown in the right panel of Figure 2.13

2.4.4 Product success and emotions

Our statistical model includes a time varying mean of the logarithm of the growth rate $\mu(t)$. The left panel of Figure 2.16 shows the time evolution of the empirical estimation of this parameter, which decays as t^α , with $\alpha = -0.91 \pm 0.01$. This type of decay is similar to the one present in the production of **Twitter** messages, where the same t^α shape is present (Asur *et al.*, 2011). Other online communities show different behavior, with decay rates that follow stretched exponentials, like for the case of **Digg** and **Youtube** (Wu and Huberman, 2007; Szabo and Huberman, 2010).

The sales rank allows us to divide the products according to their sales success, to then use this approach in order to measure if there is a quantitative difference in the response of the community. The right panel of Figure 2.16 shows the same mean rates but for two different classes of products, namely, for those that appeared in the top sales ranking of **Amazon** (green) and those that did not (blue). The decay of the growth rate for top products follows a power law with exponent $\alpha = -1.06 \pm 0.01$, while the exponent for the rest products is $\alpha = -0.86 \pm 0.01$. Thus, the growth rate of the top products decays faster than the rest. A plausible explanation for this divergence is that when a product makes

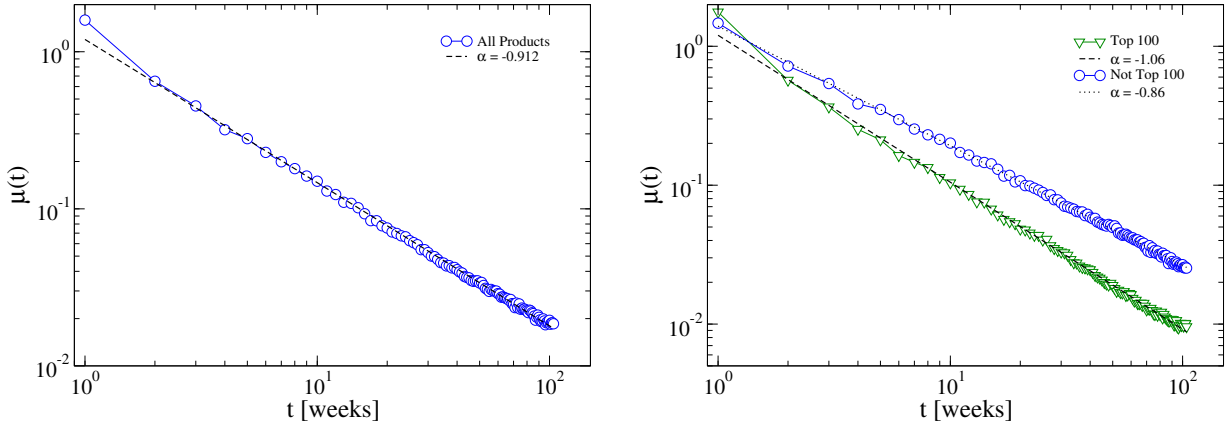


Figure 2.16: Left: time evolution of $\mu(t)$ for all products, decaying as $t^{-0.91}$. Right: different decays for products in the top 100 ranking (green) and outside (blue).

it to the top ranks, its success - which comes with a sudden large amount of purchases - saturates fast the potential consumer community and therefore, accelerates the decay of the growth rate. On the other hand, this burst behavior is absent for products not being able to reach the top ranks. These products have a lower total number of reviews, and this number is decaying with a slower rate, since the potential consumer community is not saturated, i.e. a potential new buyer can emerge at any time.

In our modeling framework, presented in Chapter 1, positive and negative information aggregates in independent information fields. To study the emotional influence among users of **Amazon**, we separate their reviews according to the emotional polarity calculated in Section 2.3.2, and we compute the time evolution of the growth rates by their emotion. For a given product, we define

$$r_-(t) = \frac{N(t) - N(t-1)}{N(t-1)} \quad r_0(t) = \frac{U(t) - U(t-1)}{U(t-1)} \quad r_+(t) = \frac{P(t) - P(t-1)}{P(t-1)} \quad (2.5)$$

with $\mu_-(t)$, $\mu_0(t)$, and $\mu_+(t)$ as the mean of their logarithm as explained in the previous section. These three growth rates follow similar dynamics as the total amount of reviews, of the form of t^α , shown in Figure 2.17. The decay for positive and negative comments have similar power law exponents $\alpha_- = -0.66 \pm 0.01$, and $\alpha_+ = -0.64 \pm 0.01$, but they differ from the decay of the growth rate of neutral comments $\alpha_0 = -0.8 \pm 0.01$. This shows that the impact of an emotionally charged review stays longer in the collective memory of the reviewer community, triggering future emotional reviews. Consistently with our finding on the time evolution of ratios of positive, negative and neutral reviews, this slower decay of growth rate supports the hypothesis that emotions are one driving force of reviewer behavior in online communities.

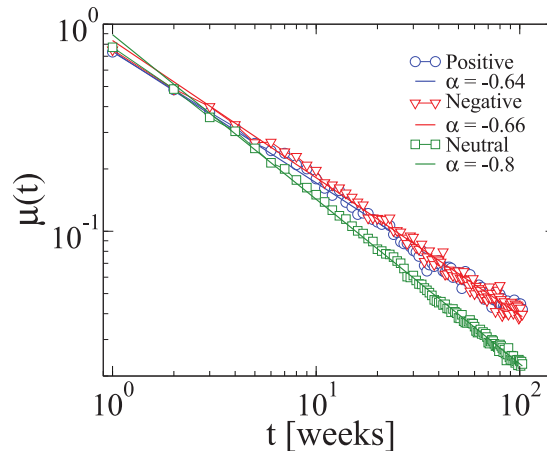


Figure 2.17: Mean of the logarithm of the emotion growth rates versus time: $\mu_+(t)$ in blue, $\mu_0(t)$ in green, $\mu_-(t)$ in red. Error bars showing the standard error are smaller than the size of the symbol, and are not shown.

2.5 Extended analysis of review emotions

2.5.1 Positive and negative emotions in product reviews

The text analysis technique we presented in Section 2.3.1 extracts an estimation of the valence expressed in the text, by means of a naive application of the ANEW lexicon. Such approximation was sufficient for studying collective patterns of emotional expression, but did not suffice for a detailed analysis of the emotions expressed in individual reviews. We combined the ANEW lexicon (Bradley and Lang, 1999) with SentiStrength (Thelwall *et al.*, 2010), a state-of-the-art tool for the extraction of positive and negative emotions from text. SentiStrength has been recently used for the analysis of emotional expression in *Yahoo answers* (Kucuktunc *et al.*, 2012), and *Twitter trends* (Thelwall *et al.*, 2011). This technique uses a human-designed lexicon of emotional terms with a set of amplification, diminishing and negation rules, which are applied if the corresponding terms are detected inside the text. Its output is composed of a positive score ranging from 1 (minimum) to 5 (maximum), and a negative score ranging from -1 if no negative emotions are present, to -5 if negative emotions have maximum strength.

An illustrative example shown on the SentiStrength site¹ is the text “*I really love you but dislike your cold sister*”, which has positive strength 4 and negative strength -3. The term *love* has a value 3 in the lexicon, and the boosting word *really* increases the positive

¹<http://sentistrength.wlv.ac.uk/>

score of the sentence to 4. The words *dislike* and *cold* have lexicon values of -3 and -2 respectively, and SentiStrength provides a negative score corresponding to the maximally negative term.

For our application, we extended the SentiStrength lexicon with the addition of emotional terms from the ANEW dataset (Bradley and Lang, 1999). This dataset is composed of human reports on the valence perceived from a large set of words, on a scale of 1 to 9, being 5 neutral valence. We converted the real values of ANEW to the SentiStrength [-5,5] integer scale by means of a linear map, and we ignored the resulting 0 values. This analysis combines the accuracy of a lexicon-based classifier with an additional psychological support from the survey data of ANEW, as the validity of the classification output relies on the size of the lexicon used. The original lexicon contained 938 terms, and combined with the terms from ANEW, we can use a lexicon with 1430 terms.

For each review, we extracted a positive and a negative score, which refer to the emotional content rather than to the star-rating. The example of Figure 2.2 shows the rest of variables available for each review: (i) emotional variables, i.e. positive and negative score, (ii) the star-rating, and (iii) the helpfulness as amount of helpful votes, and the unhelpfulness as the total amount of votes minus the helpfulness. As a first approach to understand the properties of emotions in individual product reviews, we have calculated the correlation of the mentioned variables. We find that the amount of helpful and unhelpful votes are positively correlated ($\rho = 0.342$), indicating that there are polemic reviews. This means that there is no unique criterion to evaluate the quality of a review, and the usefulness of the information contained in its text might not be the same for two different users. This relation between positive and negative votes was also detected by Van Mieghem (Van Mieghem, 2011) in the votes for stories in the **Reddit** community, calling for a psychological explanation.

We did not find any significant correlation between the emotional scores from the lexicon-based classifier and the helpfulness votes, which means that there is no trivial impact of the emotional content of a review and its usefulness. But the negative emotional score keeps some negative correlation to the rating value ($\rho = -0.186$), i.e. the stronger the negative score, the lower the rating, which is somehow expected. Additionally, positive and negative scores are positively correlated ($\rho = 0.251$), reflecting the coexistence of positive and negative emotions in the same review text.

2.5.2 Positivity and negativity towards products

Given the set of ratings of a product, we calculated the mean and standard deviation of the star-rating values. Their dependence can be seen in the left panel of Figure 2.18. We notice the concentration of a large amount of products close to a mean rating of 4.5 and a standard deviation of 1. The inset of Figure 2.18 shows the distribution of the mean positive score versus its standard deviation, which has a similar shape to the equivalent for rating. Note that individual review positive score and rating are uncorrelated, but the aggregated values show the existence of similar patterns at the product/discussion level. The range of values of both variables are similar, and their deviations are comparable, while this is not the case for the negative score. The right panel of Figure 2.18 shows the same analysis for the negative score, revealing that the range of values for this aggregated measure is much wider, as well as its standard deviation.

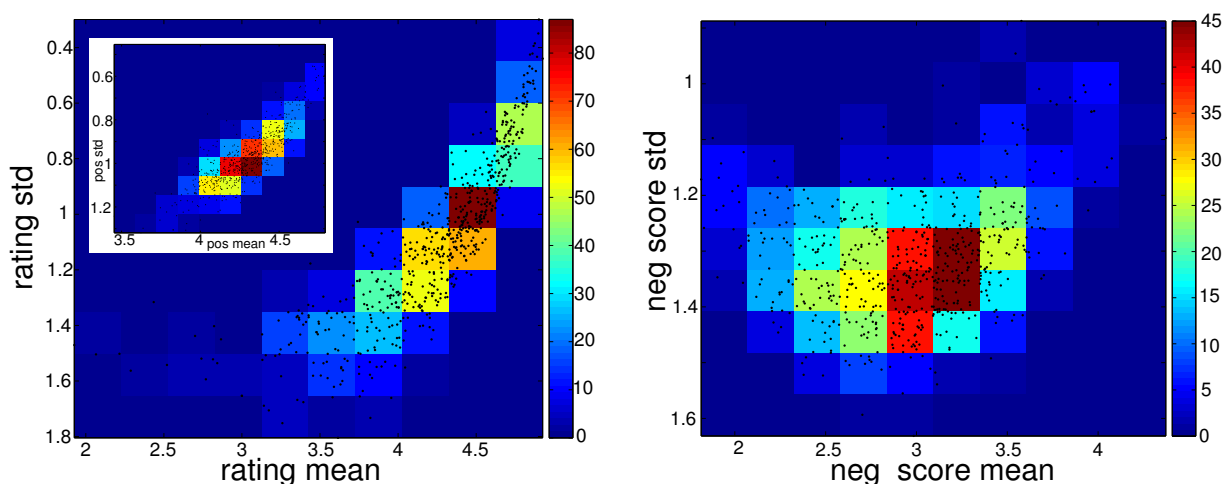


Figure 2.18: Left: mean rating versus standard deviation of the rating for products with more than 50 reviews (collective treatment). Inset: Mean positive score versus standard deviation of the positive score for the same products. Right: mean versus standard deviation of the negative score of products with more than 50 reviews. Color bins over the scatter plot show the amount of points inside.

To gain deeper insight into the emotional content, we studied the distributions of positive and negative scores of all reviews of some products. In general, for larger amount of reviews these distributions follow asymmetric patterns with respect to positive and negative emotions. Positive emotions show some bias to higher values, having a peak in most of the cases at the maximum value and much lower density for the rest. At the same time, negative emotions appear to be much more evenly distributed across possible values. This regularity of emotional expression, present across many products, is the aim of the design

of the agent-based model explained below.

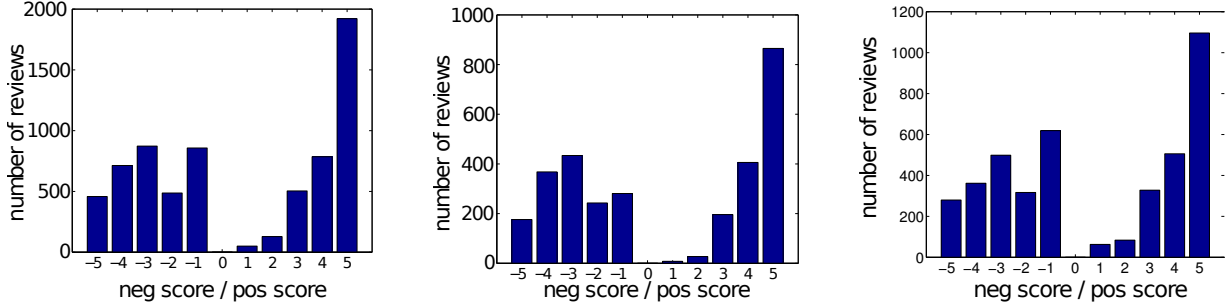


Figure 2.19: Distributions of the emotional scores of reviews for “Harry Potter and The Deathly Hallows”, “Marley and Me”, and “Twilight”.

2.6 An agent-based model of emotions in product reviews

We have designed an agent-based model for collective emotions which builds on the modeling framework described in Chapter 1. The model we present here differs from the one presented in Section 1.3 in the fact that here we try to model collective emotions in product reviews, while the previous model aims at providing a general approach for any online social interaction. As explained later, we introduce particularities of emotional communication in reviews, customer preferences in valence dynamics, and a new activation rule that prevents an agent from creating more than one review per product.

2.6.1 Arousal dynamics

In this model, the arousal, i.e. the degree of activity associated with an emotional state, follows the same dynamics outlined in Equation (1.2), with an exponential decay of parameter γ_a , a deterministic contribution \mathcal{F}_a , which depends on the emotional information h_{\pm} , and a stochastic component $A_a \xi_a(t)$. It is a particularity of product reviews that users review products only once (anomalous behavior excluded). To adapt the general model to this feature, we define the agents’ arousal dynamics and threshold in a way that they express their emotions only once or never. When the arousal reaches the threshold, the following rule is applied:

$$\text{if } a_i(t) > \mathfrak{I}_i \implies \mathfrak{I}_i \leftarrow \infty \quad (2.6)$$

This way, for finite arousals, the agent will never contribute a review again after the first expression. For this case, the thresholds vary among agents in a way that they

follow a normal distribution with mean μ and standard deviation σ , in line with general assumptions present in threshold models for cascading processes (Lorenz, 2009).

The function \mathcal{F}_a depends on the sum of both fields, using the assumption that emotional communication influences activity regardless of its valence sign:

$$\mathcal{F}_a \propto [h_+(t) + h_-(t)] \sum_{k=0}^n (d_0 + d_1 a + d_2 a^2(t)) \quad (2.7)$$

This way, the arousal is affected by an activity baseline d_0 , a linear influence $d_1 a$, and a quadratic saturation if $d_2 < 0$. Our previous analysis in Chapter 1 showed that, for certain values of the parameters d , the expression dynamics had a one-peak behavior similar to the one showed in the time series of Figure 2.9.

2.6.2 Valence dynamics

The dynamics for arousal and valence we propose are supported by empirical studies in which both variables could be approximated with a stochastic process with exponential decay (Kuppens *et al.*, 2010). In our model, the equation for the valence is of the same form as equation 1.2. The influence of the field in the agent's valence \mathcal{F}_v depends on whether its valence is positive or negative. It means that agents with negative experience of the product will likely develop negative feelings and pay less attention to the positive emotions expressed by the reviews, while agents with positive experiences focus more on the positive emotional information than on the negative one.

An exponential function with a cubic decay $b_2 v^3$, b_2 being negative, represents the asymmetry in the perception of the agent dependent on its valence. We assume the function \mathcal{F}_v to be:

$$\mathcal{F}_v(h_+(t), h_-(t), v_i(t)) = \exp\left(\frac{h_+(t) - h_-(t)}{h_+(t) + h_-(t)} \cdot (b_1 v + b_2 v^3)\right) + b_0 \quad (2.8)$$

where the parameters have to satisfy $b_1 > 0$, $b_2 < 0$ and $b_0 < 0$ for the desired behavior. \mathcal{F}_v shown in Figure 2.20 describes the major contribution of these equation for changes in the valence. It is stronger in the positive v , and it diverges for extreme values of v , this way keeping the dynamics of v inside the interval $(-1, +1)$. This holds when the positive field is larger than the negative one, which is the case for marketing campaigns because of their positive impact. Hence, in our model, h_+ is larger than h_- .

User preferences and product properties play a key role in the expression of emotions when writing a review. Manufacturers have to deal with the fact that users are heterogeneous and

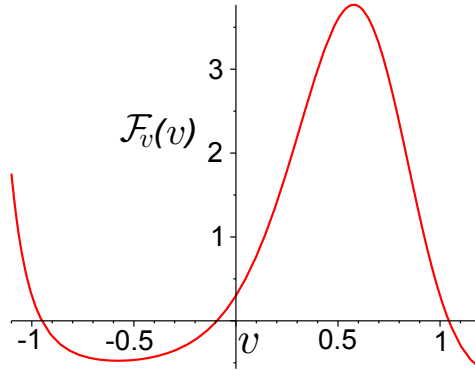


Figure 2.20: \mathcal{F}_v (y axis) versus v (z axis) as in equation (2.8), for $b_0 = -1$, $b_1 = 5$, $b_2 = -5$ under fields $h_+ = 40$ and $h_- = 10$

their preferences vary a lot, so the properties of the product and the marketing campaigns give a starting point for the emotions of the reviewers. In the following, we will model the user preference as a agent internal variable u_i constant in time. To model the heterogeneity of agents, we assume that the preferences are uniformly distributed in the interval $[0, 1]$. Take note that preference is subjective in the sense that it simply determines what is preferred and not what is better or worse. The properties of the product are constant for all agents and are described by a parameter $q \in [0, 1]$ that distinguishes products from others.

Initially, an agent starts experiencing the product, which determines its initial valence v resulting from the difference between the agent's expectation u_i and the product property q :

$$v_i(0) = 1 - \frac{2}{\max(q, 1 - q)} |u_i - q| \quad (2.9)$$

An agent that matches its preferences perfectly with the product has $|u_i - q| = 0$ and its initial valence will be extremely positive ($v_i(0) = 1$). On the other hand, if the product results to be completely opposite to what the agent was expecting, the value of the distance will be maximum (q or $1 - q$) and the initial valence will be -1 .

2.6.3 Intensity of emotional expression

We assume that reviews with higher emotional content have a higher impact on the information field. A review with just factual information is assumed to just have influence on the opinion and information available to the agent, but the influence on the emotions are small. As humans show empathy, the more emotional the review, the larger its influence. When the arousal of an agent reaches its threshold, the agent creates a review with an

emotional content proportional to its valence. For this, we will set the s_i of the agent to a value between 1 and 5. Table 2.1 gives the resulting values dependent on the valence.

Table 2.1: Expression intensity s given valence v .

v interval	s_-	v interval	s_+
$(-\infty, -0.8]$	5	$[0, 0.2)$	1
$(-0.8, -0.6]$	4	$[0.2, 0.4)$	2
$(-0.6, -0.4]$	3	$[0.4, 0.6)$	3
$(-0.4, -0.2]$	2	$[0.6, 0.8)$	4
$(-0.2, 0]$	1	$[0.8, +\infty)$	5

2.6.4 Simulation of emotions in product reviews

Given the initial value of the field and the dynamics of valence and arousal we can reproduce different scenarios of the reaction of the community. As discussed in Section 2.4, reviewer activity might be driven by mass media or word of mouth influence, effects which are illustrated by the examples of Figure 2.9. These two types of collective reaction appear in our simulations, as shown in Figure 2.21. We notice that the presence of a strong initial input could create a high spike in the amount of reviews followed by a fast decay, which can be seen in the left panel of Figure 2.21. In absence of initial information, if the variance of the threshold distribution is large enough, there is a slower increase in the amount of reviews endogenously created within the community. The first agents to write a review are the few ones with a very low threshold, and their activities trigger the purchases and reviews of agents with higher thresholds. The right panel of Figure 2.21 shows how the review frequency increases for this case.

For certain parameter values, our model reproduces the distributions of emotions we measured in real-world review communities. This comparison is shown in Figure 2.22. We find that the distribution of emotions in the simulations have the same bias in the positive part while they are more evenly distributed in the negative part. Our simulations indicate that there is an important herding effect present when perceiving positive emotional content in product reviews, while the expression and perception of negative reviews does not influence the agents' negative valences that much.

To conclude, in Figure 2.22 we show for the distribution of emotions that the outcome of our model has the same macroscopic properties as found in the real world data. This model

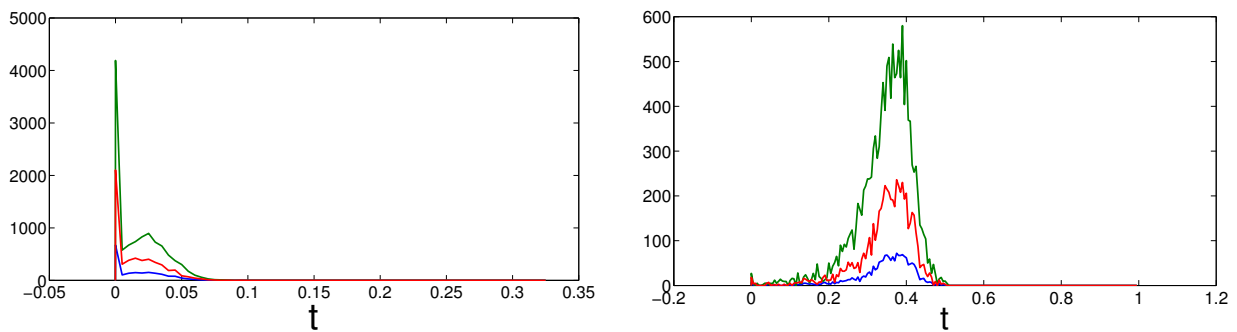


Figure 2.21: Amount of ratings (blue), total positive expression (green) and total negative expression (red) for the simulated time in the case of a strong media impulse (left) and when the emotions spread through the community (right).

provides a phenomenological explanation based on psychological principles that links the dynamics of emotions with the collective behavior observed in product reviews. Here, we verified that the model can have the same qualitative properties as the reviewer community of *Amazon.com*, but its predictive power still needs to be explored. Our model allows for the exploration of a psychometric parameter space, from which the most likely values of the parameters can be found for each product. Following the appropriate approach, future analysis of this model can provide predictive results that could assist community managers and manufacturers in their decisions regarding customer reviews and satisfaction.

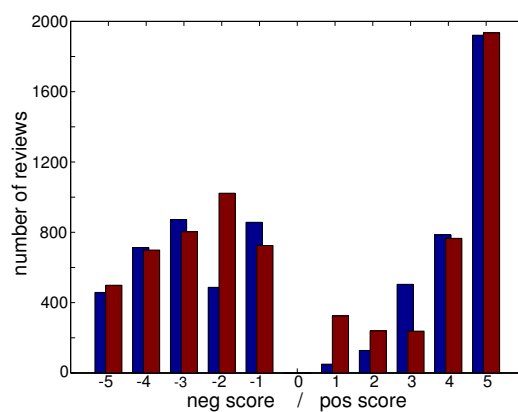


Figure 2.22: Comparison between the emotional distribution of the reviews for "Harry Potter" (blue) and the simulation results (red).

2.7 Conclusions from review emotions

Thanks to a massive retrieval of product reviews available from `Amazon.com`, we were able to mine important information about the patterns of emotional expression through product reviews. First, we tested the hypothesis that emotions are a motivation for the production of late reviews, in particular under the absence of mass media campaigns. By applying sentiment analysis to the text of the reviews and studying their time series, we could find which products have reviews of increasing or decreasing emotional content. Our statistical analysis shows the relation between these trends of emotional content, and the amount of news items for a product in `news.google.com`. These results offer a valuable insight to understand one of the key factors of the success of an online community: the creation of content by its users. Emotional interaction leads to information sharing, which can in turn increase the importance of certain users within the community. Building up on the work presented in this chapter, this relation between social impact and emotional expression has been tested in (Tanase *et al.*, 2012), where a dataset of product reviews in a trust network reveals that central users are more emotional in their reviews.

We deepened our analysis of the time series of creation of reviews, finding a relation between the maximum amount of weekly reviews and the sales success of a product. Furthermore, our analysis revealed that such peak of review creation happens earlier when a product receives attention from mass media sources. After studying these external influences to review creation, we looked into traces of influence among users. The application of Fluctuation Scaling revealed the presence of such influence, in which users create reviews induced by the reviews previously created by other users. In addition, this influence modifies the heterogeneity among products in terms of amount of reviews. When a product is launched, the amount of reviews it receives in the first weeks varies a lot among products, and it depends on the mass media attention it receives. As users purchase and review products within `Amazon.com`, they decrease the inequality in the amount of reviews per product.

We propose a statistical model that explains the emergence of the final distribution of reviews per product, based on a multiplicative process of review creation. This allows us to estimate a parameter of the growth rate that changes through time, finding that it decays rapidly. This decay is different for products that reach the top 100 sales rank, which is faster than for those that do not reach the top list. In addition, we divide the analysis to study the decay in the creation of reviews by their emotional content. We find that the growth of emotional reviews decays slower than that of non-emotional ones, suggesting that emotional reviews keep the attention longer than the ones without emotional content. These patterns call for the application of agent-based models that explain the emotional

interaction among users of the product reviews community.

Our application of SentiStrength shows that, in general, the collective expression of positive emotions in product reviews shows a strong bias towards positive values, while the negative content is more evenly distributed along its possible values. We modeled the behavior of users and product review communities using the modeling framework for collective emotions presented in Chapter 1. This model introduces key features of the interaction between users of a product reviews community, as for example, a user can only review a product once at most. This model can reproduce both possible scenarios in response to external influences, dependent on the presence of marketing campaigns and word of mouth effects. Further, the distributions of expressed emotions have similar properties in the simulations and in the data.

Our model can be fitted to know the internal dynamics of the customers that reviewed a product, as well as the properties of the social interaction they have online. Manufacturers can use this useful information to understand better the way their customers react to their products, and to derive norms and principles to follow in order to maximize customer satisfaction and sales. Our statistical analysis can serve as a comparison for community managers of reviews communities. In particular, understanding the dynamics of user arousal is key for encouraging participation. Our findings can serve as an example for other platforms in order to imitate **Amazon.com**'s success, or to summarize and visualize the collective emotions of the customers that bought a product.

Chapter 3

Modeling emotions in chatrooms

Summary

How do users behave in online chatrooms, where they instantaneously read and write posts? We analyzed about 2.5 million posts covering various topics in Internet relay channels, and found that user activity patterns follow known power-law and stretched exponential distributions, indicating that online chat activity is not different from other forms of communication. Analyzing the emotional expressions (positive, negative, neutral) of users, we revealed a remarkable persistence both for individual users and channels. I.e. despite their anonymity, users tend to follow social norms in repeated interactions in online chats, which results in a specific emotional "tone" of the channels. We provide an agent-based model of emotional interaction, which recovers qualitatively both the activity patterns in chatrooms and the emotional persistence of users and channels. While our assumptions about agent's emotional expressions are rooted in psychology, the model allows to test different hypothesis regarding their emotional impact in online communication.

3.1 The question of online communication

How do human communication patterns change on the Internet? Round the clock activities of Internet users put us into the comfortable situation of having massive data from various sources available at a fine time resolution. But what to look at? Which aggregated measures are most appropriate to capture how new technologies affect our communicative behavior? And then, are we able to match these findings with a dynamic model that is able to generate insights into their origin? In this chapter, we provide both: a new way of analyzing data from online chats, and a model of interacting agents to reproduce the stylized facts of our analysis. In addition to the activity patterns of users, we also analyze and model their emotional expressions that trigger the interactions of users in online chats. Validating our agent-based model against empirical findings allows us to draw conclusions about the role of emotions in this form of communication.

One of the main differences between online and offline communication is that anonymity is much less frequent in the former. Theoretical arguments suggest that, under a high level of anonymity, individual user behavior might differ a lot in comparison with situations in which identities are known (Riva, 2005). A straightforward approach to this difference is the *desinhibiting effect*, (Suler, 2000) which explains this behavioral change as a way to overcome individual limitations in offline life. As an example, a pacific, male white-collar worker would choose to behave as an aggressive, female hunter in a virtual world. As an opposing view, current psychoanalysis goes one step further, reformulating the identity of the user as the actual behavior under anonymity. Thus, anonymity would partial out the effects of punishment or social pressure into the decisions of the individual, allowing a much deeper view of the personality and emotions of a person (Zizek, 1998). Previous studies in psychology indicate that individual preferences are more salient under anonymous, computer-mediated communication than in face-to-face communication (Sassenberg and Boos, 2003). Furthermore, ephemeral and anonymous interaction might foster high levels of activity, often showing behavioral patterns not found in offline interaction (Bernstein *et al.*, 2011).

Online platforms often provide a communication channel that allows the users to define different levels of *onymity* (Vujovic *et al.*, 2011), i.e. how much other users know about their identity. On the highest level of *onymity*, online social networks like **Facebook** or **Google+** aim at the creation of personal profiles. These profiles are designed for users to disclose personal information only to a set of friends or social contacts. Other websites, like fora or product reviews communities, are not focused on user identity but on their interaction, often requiring only a user account to be used. This middle level of *onymity* allows the creation of fake identities and might lower credibility and trust (Rains, 2007).

On the far end of the anonymous spectrum, we have online communities in which there is no guarantee of the identity of any user, often leading to interactions in which not even the name of the user that created a message is relevant or known. Anonymous fora like `4chan.org` or public IRC channels fall in this class. This kind of anonymous online interaction is the subject of our research in this chapter, focusing on what are the similarities and differences with what we know from onymous online and offline interaction.

Another distinguishing feature of online communication is the possible delay of interaction. In some communities, users might take days or weeks to react to another user's post. The users of `Amazon.com`, for example, take days to react to a review, as usually they have to purchase and experience the product in between. But this is not the case for all online communication, some other platforms feature functionalities that lead to fast, real-time communication. This is the case of chatrooms, in which lots of short posts appear and disappear in a short time. We can see this chatrooms as a large-scale social experiment that provides us with data about user activity and interaction. Analysis of the temporal activity patterns of online interaction allow conclusions on how humans organize their time and give different priorities to their communication tasks (Barabasi, 2005; Grinstein and Linsker, 2008; Malmgren *et al.*, 2008; Crane *et al.*, 2010). One particular quantity to describe these patterns is the distribution $P(\tau)$ of the waiting time τ that elapses before a particular user answers e.g. an email. Different studies have confirmed the power-law nature of this distribution, $P(\tau) \sim \tau^{-\alpha}$. Its origin was attributed either to bursts of events (Barabasi, 2005) or to circadian activity patterns (Malmgren *et al.*, 2008). However, the value of the exponent α is still debated. A stochastic priority queue model (Grinstein and Linsker, 2008) allows to derive α by comparing two different rates, the average rate λ of messages arriving and the average rate μ of processing messages. If $\mu \leq \lambda$, i.e. if messages arrive faster than they can be processed, it holds that $\alpha = 3/2$, which is compatible with most empirical findings and simulation models (Barabasi, 2005; Malmgren *et al.*, 2008; Mitrović and Tadić, 2009; Wu *et al.*, 2010). However, in the opposite case, $\mu \geq \lambda$, i.e. if messages can be processed upon arrival, $\alpha = 5/2$ was found together with an exponential correction term. The latter regime, also denoted as the *highly attentive regime*, could be verified empirically so far only by using data about donations (Crane *et al.*, 2010). So, it is an interesting question to analyze other forms of online communication to find more evidence for this second regime.

We analyze data about instant online communication in chatting communities, specifically Internet Relay Chat (IRC) channels, where each channel covers a particular topic. An IRC channel can be created in any server by following a specified protocol, in a similar way as it is done for e-mail communication. There is no monopoly in the creation and usage of IRC channels, and any set of Internet users can create their own one. Prior to the very common

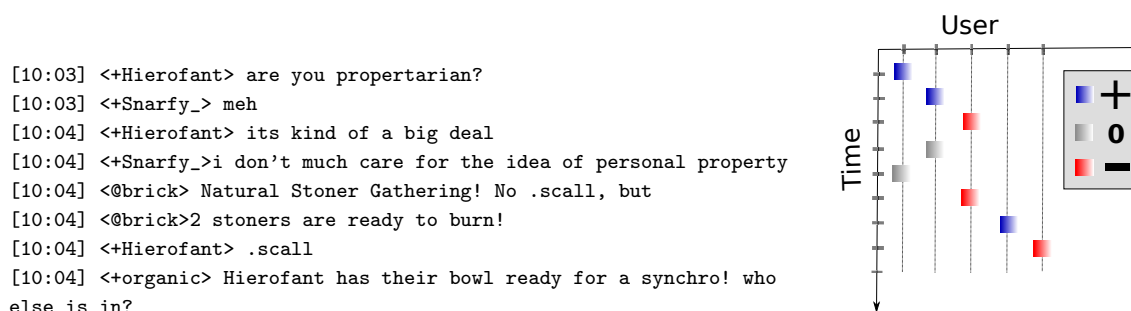


Figure 3.1: Left: example of a discussion from the channel #bonghit. Right: schema of user emotional interaction in a chatroom.

social networking sites of today, IRC channels provided a safe and independent way for users to share and discuss information outside traditional media. Way before the role of **Twitter** in the Arab spring of 2011, IRC channels were used to avoid media blackouts and mass media control. ibiblio.org¹ keeps IRC logs of important events, like the Soviet coup d'etat attempt in 1991, the Gulf War, or the California earthquake of 1994. Nowadays they are still used for a wide variety of reasons, like military purposes, cyberactivism, open source project development, or dating. Being one of the oldest means of communication on the Internet, their influence is still present in the patterns of communication of people in other communities. As an example, the topic hashtags used in **Twitter** were invented earlier in IRC chats, as channel names always start with the character #. Currently, there are more than 3200 IRC servers, and the top 100 IRC networks serve more than 500.000 users (irc.netsplit.de, April 2011).

The left panel of Figure 3.1 shows an example of a public discussion in an IRC chatroom. As expected, our dataset does not include information on private discussions between pairs of users, but includes the public discussion of each chatroom, in which all the users can participate. Different from other types of online communication, such as blogs or fora where entries are posted at a given time (decided by the writer), IRC chats are instantaneous in real time, i.e. users read while the post is written and can react immediately. This type of interaction requires much higher user activity in comparison to persistent communication e.g. in fora. Further, it is more spontaneous, often leading to emotionally-rich communication between involved peers. Consequently, instant communication should require specific tools and models for analysis, that are capable of covering these predominant features. The analysis of IRC channels has generally being restricted to technical issues (Dewes *et al.*, 2003), but quantitative analysis of its social and psychological implications remained unexplored until this work.

¹www.ibiblio.org/pub/academic/communications/logs

We perform our analysis as follows: First, we look into the communication patterns of instant online discussions, to find out about the average response time of users and its possible dependence on the discussed topics. This shall allow us to identify similarities and differences between instantaneous chatting communities and other forms of slower, persistent communication. In a second step, we look more closely into the content of the discussions and how they depend on the emotions expressed by users. Remarkably, we find that most users are very persistent in expressing their positive or negative emotions - which is not expected given the variety of topics and the user anonymity. This leads us to the question in what respect online chats are different from offline discussions, which are influenced by social norms. We argue that even in instantaneous, anonymous online chats users behave very much like "normal" people. Our quantitative insights into user's activity patterns and their emotional expressions are combined to model interacting emotional agents. We demonstrate that stylized facts of the emotional persistence can be reproduced by our model by only calibrating a small set of agent features. This success indicates that our modeling framework can be used to test further hypothesis about emotional interaction in online communities.

3.2 Data from Internet Relay Chats

The data used for this study were retrieved by Marcin Skowron at the Austrian Research Institute for Artificial Intelligence (OFAI). The dataset is based on a large set of public channels from *Efnet* Internet Relay Chats (<http://www.efnet.org>), to which any user can connect and participate in the conversation. Based on the assessment of the initially downloaded set of recordings, 20 IRC channels were selected according to their number of posts. This selection aimed to provide a large number of consecutive daily logs with transcripts of vivid discussions between the channel participants. The finally used data set contained consecutive recordings for 42 days spanning the period from April 4th to May 15th 2006.

The general topics of discussions from the selected channels include: music, sports, casuals chats, business, politics and topics related to computers, operating systems or specific computer programs. This dataset contains 2,688,760 posts. The total number of participants to all this channels is 25,166. However, because some people participate in more than one channel, the total number of unique participants is 20,441. On average, the dataset provides 3055 posts per day. In the recorded period 15 users created more than 10000 posts, and the mean participation per user was 97 posts. The acquired data was anonymized by substituting real user ids to random number references. The text of each

post was cleaned by spam detection and substitution of URL links to avoid them from influencing the emotion classification.

As in Chapter 2, the emotional content was extracted using the SentiStrength classifier (Thelwall *et al.*, 2010), which provides two scores for positive and negative content. IRC posts are very short and informal, which is the type of text SentiStrength was designed for. For each post, we calculate a polarity measure by comparing the two different scores of SentiStrength. The sign of the sum of the positive ($p \in [1, 5]$) and negative ($n \in [-1, -5]$) scores provides an approximation to detect positive ($e = 1$), negative ($e = -1$) and neutral ($e = 0$) posts. The accuracy of this polarity metric was tested against texts tagged by humans and messages including emoticons from MySpace (Paltoglou *et al.*, 2010), Twitter (Thelwall *et al.*, 2011), Youtube and fora (Thelwall *et al.*, 2012), providing accuracies above 88%. This classifier combines an emotion quantization of proved validity with a high accuracy, and is considered the state of the art in sentiment detection (Kucuktunc *et al.*, 2012).

After classification, each post has a value of emotional polarity, a timestamp, and a user id number. User interaction is instantaneous, the post written by user u_1 is immediately visible to all other users logged into this channel, and user u_2 may reply right away. The right panel of Figure 3.1 illustrates the dynamics in such a channel. As time evolves new users may enter, others may leave or stay quiet until they write follow-up posts at a later time.

3.3 User activity patterns

3.3.1 Individual activity

To characterize these activity patterns, we analyzed the waiting time, or inter-activity time distribution $P(\tau)$, where τ refers to the time interval between two consecutive posts of the *same* user in the same channel. We applied the maximum likelihood approach proposed in (Clauset *et al.*, 2009) to test the power-law hypothesis explained in Section 3.1. We found that the power-law nature of the distribution could not be rejected ($p = 0.375$), and that τ is power-law distributed $P(\tau) \sim \tau^{-\alpha}$ with some cut-off, with an exponent $\alpha = 1.53 \pm 0.02$. The left panel of Figure 3.2 shows $P(\tau)$ with logarithmic binning for all users and channels, with a line showing the results of this analysis.

This finding is in line the power-law distribution already found for diverse human activities (Barabasi, 2005; Malmgren *et al.*, 2008; Grinstein and Linsker, 2008; Crane *et al.*, 2010) and classifies the communication process as belonging to the regime where posts arrive

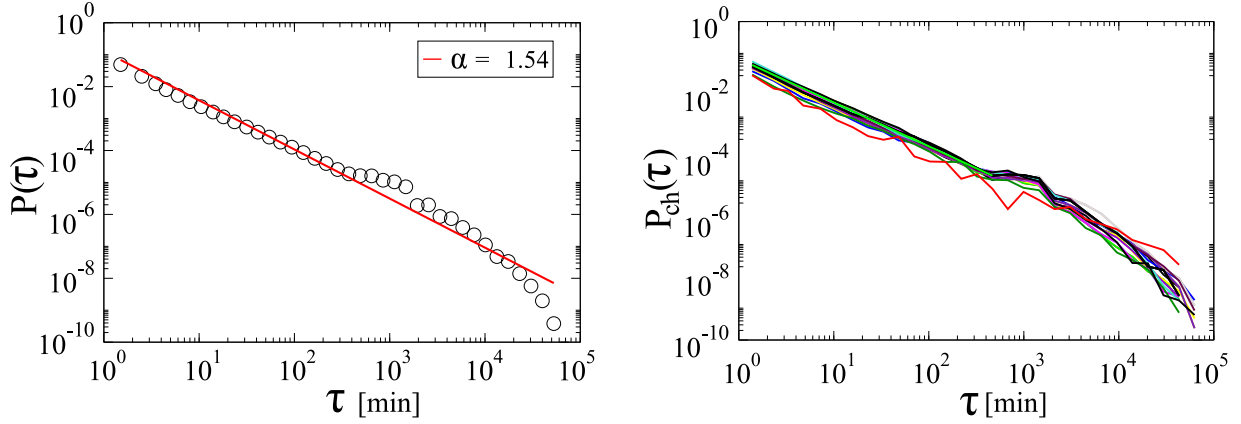


Figure 3.2: Individual user inter-activity time distributions. Left panel: $P(\tau)$ for all users and all channels, fitted as $\tau^{-\alpha}$. The right panel shows the same distribution per channel.

faster than they can be processed. When a sufficiently large amount of users gathers in a chatroom, they can generate posts at very fast rates. This might lead the rest of the users not to reply to all the messages, generally focusing on certain parts of the conversation. In addition, other tasks might have priority over the chat, and their attention can be driven towards other communication mechanisms, e.g. offline communication. Since $\alpha < 2$, the first moment of $P(\tau)$ is not defined, i.e. the average response time is not finite. This is very different from the high attention regime, where $\alpha > 2$, which would imply the existence of a finite value of the expectation of a user to reply in a chat.

It follows to ask whether this general pattern of IRC communication depends on the amount of users connected to a chatroom, as smaller discussions might allow for the appearance of the high attention, unsaturated regime. We explored this possibility by analyzing $P_{\text{ch}}(\tau)$, the individual inter-activity time distribution per channel. The right panel of Figure 3.2 shows these 20 distributions, which are indistinguishable from each other, following the same pattern as $P(\tau)$. This leads us to conclude that all the studied channels are in the saturated regime, but still the other case might exist in channels outside the scope of our analysis.

In the left panel of Figure 3.2 it can be noticed that there is some deviation of the power-law at a time interval of about one day, which shows that some users have an additional regularity in their behavior with respect to the time of the day they enter the online discussion. This “bump” around the one day time interval seems to provide further evidence to the bimodality proposed in (Wu *et al.*, 2010), where the tail of the distribution would correspond to the time delay between bursts of faster communication. This was the case for the analysis of SMS communication, where the tail was well fit by an exponential distribution, which could be explained from a Poisson process of SMS discussion initiation. The

same argument does not hold for users of collective discussions in chatrooms, as in general they only address each other directly through private messages we cannot see. We tested different hypotheses of the shape of the tail through the same maximum likelihood as for the power-law region. Kolmogorov-Smirnov tests on the results show that the tail is closer to a lognormal distribution ($D_{KS} = 0.136$), than to an exponential ($D_{KS} = 0.190$), or a Weibull ($D_{KS} = 0.188$), ruling out the same explanation as for the SMS case. The reasons for this lognormal tail are still an open question, waiting for a hypothetical interaction mechanism that might lead to this result.

3.3.2 Collective discussion activity

We now focus on an important difference between online chats and previously studied forms of communication, such as mail or email exchange, which mostly involve two participants. Due to the collective nature of chatrooms, the system automatically aggregates the posts of a much larger amount of users, which allows us to study their collective temporal behavior. We define the inter-event time of a channel ω_{ch} as the time interval between two consecutive posts in the same channel independent of any user. This should be distinguished from the inter-activity time τ , which characterizes only a single user. We find that the distributions $P_{\text{ch}}(\omega_{\text{ch}})$ of all the channels are still fat-tailed, but do not follow power-laws. The left panel of Figure 3.3 shows these distributions, which could not be confirmed to follow power-laws by the maximum likelihood method. Interestingly, the time interval between posts significantly depends on the topic discussed in the channel. Some "hot" topics receive posts at a shorter rate than others, which can be traced back to the different number of users involved into these discussions. Specifically, we find that the average inter-event time $\langle \omega \rangle_{\text{ch}}$ depends on the amount of users in the conversation and becomes smaller for more popular channels, as one would expect. This size dependence can influence the distribution of $P_{\text{ch}}(\omega_{\text{ch}})$, leading to some stable distribution as a consequence of the combination of a unifying principle, as happens for the case of movie rating distributions (Lorenz, 2009).

If we rescale the channel dependent inter-event time distribution $P_{\text{ch}}(\omega_{\text{ch}})$ using the average inter-event time $\langle \omega \rangle_{\text{ch}}$ per channel, and plot $\langle \omega_{\text{ch}} \rangle P_{\text{ch}}(\omega_{\text{ch}})$ versus $\omega_{\text{ch}} / \langle \omega_{\text{ch}} \rangle$, we find that all the curves collapse into one master curve, as shown in the right panel of Figure 3.3. The general scaling form that we used is $P(\omega_{\text{ch}}) = (1 / \langle \omega_{\text{ch}} \rangle) F(\omega_{\text{ch}} / \langle \omega_{\text{ch}} \rangle)$, where $F(x)$ is independent of the average activity level of the component, and represents a universal characteristic of this particular system. Such scaling behavior was reported previously in the literature describing universal patterns in human activity (Candia *et al.*, 2008). We fit this master curve by a stretched exponential (Altmann and Kantz, 2005; Wang *et al.*,

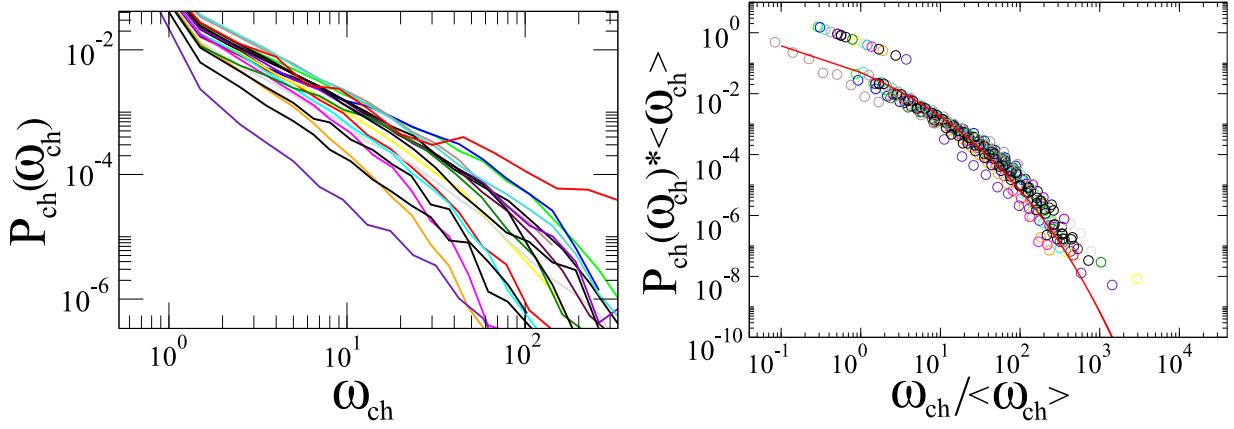


Figure 3.3: Left: distribution of inter-event time $P_{\text{ch}}(\omega_{\text{ch}})$ of each channel. Right: rescaled distribution of the inter-event time by $\langle \omega \rangle_{\text{ch}}$. The red line corresponds to a stretched exponential of parameter $\gamma = 0.21$.

2006)

$$P(\omega_{\text{ch}}) = \frac{a_\gamma}{\langle \omega_{\text{ch}} \rangle} \exp \left[-\beta_\gamma \left(\frac{\omega_{\text{ch}}}{\langle \omega_{\text{ch}} \rangle} \right)^\gamma \right] \quad (3.1)$$

where the stretched exponent γ is the only fit parameter, while the other two factors a_γ and β_γ are dependent on γ through $a_\gamma = \beta_\gamma \frac{\gamma}{\Gamma(1/\gamma)}$, and $\beta_\gamma = \frac{(2^{1/\gamma})^2 \Gamma(\frac{2+\gamma}{2\gamma})}{2\sqrt{\pi}}$ (Altmann and Kantz, 2005). Our regression results give significant values of γ ($p < 0.001$) for every channel, with a mean value of the stretched exponents of $\langle \gamma \rangle = 0.21 \pm 0.05$, not having a very large deviation across channels.

3.3.3 Long range interactions

We note that stretched exponentials have been reported to describe the inter-event time distribution in systems as diverse as earthquakes (Bunde *et al.*, 2005) and stock markets (Wang *et al.*, 2006). These systems commonly exhibit long range correlations which seem to be the origin of the stretched exponential inter-event time distributions (Altmann and Kantz, 2005). Long range correlations have also been reported in human interaction activity (Rybski *et al.*, 2009, 2011), and we tested their presence in the temporal activity over IRC communication. The way to measure this long-range correlations is by calculating the autocorrelation function, defined as $C(\Delta t) = \rho(\omega_{\text{ch}}(t), \omega_{\text{ch}}(t + \Delta t))$, where $\omega_{\text{ch}}(t)$ is the sequence of inter-event times of a particular channel. ρ is the Pearson's correlation coefficient between $\omega_{\text{ch}}(t)$ and its lagged version, $\omega_{\text{ch}}(t + \Delta t)$. This way, for each possible value of Δt , we have an estimation of the correlation in the inter-event times with delay Δt . Figure 3.4 shows an typical example of $C(\Delta t)$ for one of the analyzed channels, verifying

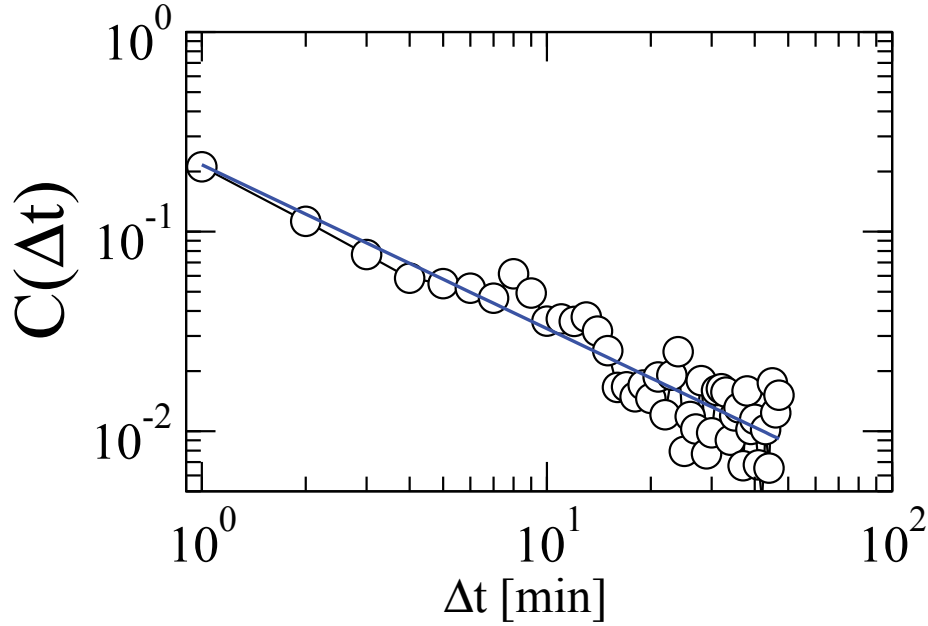


Figure 3.4: Autocorrelation function $C(\Delta t)$ of the sequence of inter-event times of a channel. The scaling relation $C(\Delta t) \sim (\Delta t)^{-\nu_\omega}$ with $\nu_\omega \simeq 0.82$ shows the existence of long range correlations.

the existence of long range correlations in the conversation activity. We found that the decay of the autocorrelation function of the inter-event time interval between consecutive posts within a channel is described by a power-law

$$C(\Delta t) \sim (\Delta t)^{-\nu_\omega} \quad (3.2)$$

with exponent $\nu_\omega \simeq 0.82$. This correlation supports the existence of long range interactions between the users of a chatroom, which is consistent with the stretched exponential function shown in Figure 3.3. In spite of the very fast communication of chatrooms, this analysis shows the existence of user influence at longer time scales, well beyond the quick disappearance of messages in the chat.

In conclusion, our analysis of user activities have revealed a universal dynamics in online chatting communities which is moreover similar to other human activities. This regards (a) the temporal activity of individual users (characterized by a power-law distribution with exponent $3/2$) and (b) the inter-event dynamics across different channels, if rescaled by the average inter-event time (characterized by a stretched exponential distribution with just one fit parameter). We will use these findings as a point of departure for a more in-depth analysis – because obviously the essence of online communication in chatrooms, as compared to other human activities, is not really covered. From the perspective of activity patters, there is not so much new here, which leads us to ask for other dimensions

of human communication that could reveal a difference.

3.4 Patterns of emotional expression

3.4.1 Measuring emotional persistence

We are interested in the role of emotional expression in online chatting communities. Users, by posting text in chatrooms, also reveal their emotions, which in return can influence the emotional response of other users, through conversations like the one illustrated in Figure 3.1. As explained in Section 3.2, we extracted a value of emotional polarity of valence for each post, as a discrete value $e \in \{-1, 0, +1\}$ that characterizes if the post is either negative, neutral, or positive. Instead of using the real time stamp of each post as in the analysis of the user activity, we now use an artificial time scale in which at each (discrete) time step one post enters the discussion, so the number of time steps equals the total number of posts.

To further approach our understanding of individual emotional behavior, we analyze to what extent the users of the chatroom consistently deviate from their average emotional expression. Since our focus is on the user, we reconstruct for every user a time series that consists of all posts communicated in any channel, where the time stamp is given by the consecutive number at which the post enters the user's red. Then, we apply the technique of Detrended Fluctuation Analysis (Peng *et al.*, 1994), to the time series of positive, negative and neutral posts of the user. This method is designed to reveal the long term memory and correlations in a time series (Bunde *et al.*, 2005; Wang *et al.*, 2006; Rybski *et al.*, 2009). The key concept of the method is to map the system to a one-dimensional random walk, comparing the properties of the real time series with the one that would be produced by a random walk.

The DFA of a time series $x(t)$ with length T , which can be divided into N segments is performed as follows: First we integrate the time series, by calculating the profile $Y(t) = \sum_{t'}^t [x(t') - \langle x(t) \rangle]$. Next, we divide the integrated time series into N boxes of equal length Δt . Each box has a local trend, which in a first level approximation, can be fitted by a linear function using least squares. We denote with $y_{\Delta t}(t)$ the y coordinate of the straight line segments that represent the local trend in each box, and we subtract this local trend from the integrated time series $Y(t)$. Next we use the function

$$F(\Delta t) = \sqrt{\frac{1}{N} \sum_{k=1}^N [Y(k) - y_{\Delta t}(k)]^2} \quad (3.3)$$

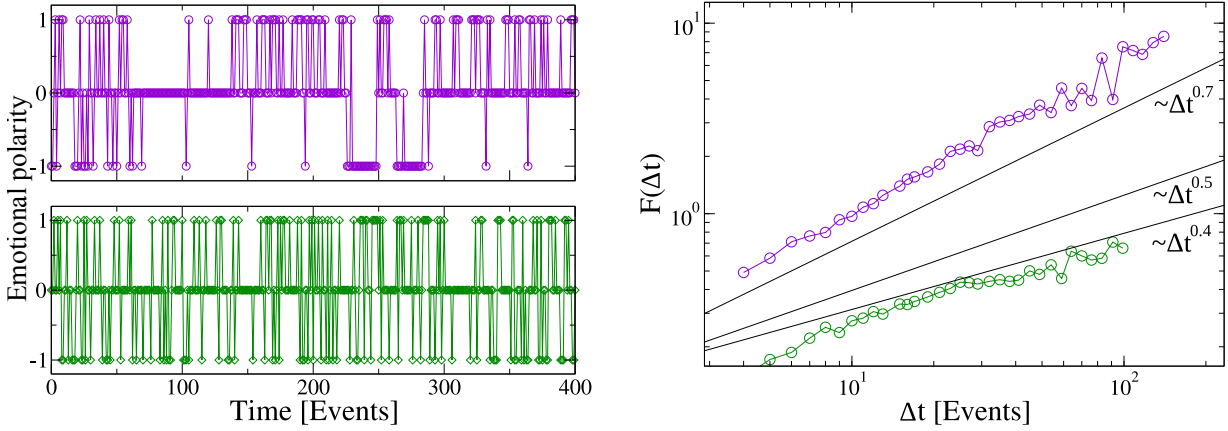


Figure 3.5: Examples of the time series of emotional polarity (left) and scaling of $F(\Delta t)$ (right) of a persistent user (purple) and an anti-persistent user (green).

to calculate the root-mean-square fluctuation of the integrated and detrended time series, and we characterize the relationship between the average fluctuation $F(\Delta t)$, and the box size Δt .

Typically, $F(\Delta t)$ will increase with box size as $F(\Delta t) \sim (\Delta t)^H$, which indicates the presence of power-law (fractal) scaling. Therefore, the fluctuations can be characterized only by the scaling exponent H that is analogous to the Hurst exponent (Hurst, 1951), and it is calculated from the slope of the line relating $\log(F(\Delta t))$ to $\log(\Delta t)$. If only short-range correlations (or no correlations) exist in the time series, then it has the statistical properties of a random walk, i.e. $F(\Delta t) \sim (\Delta t)^{1/2}$. However, in the presence of long-range power-law correlations (i.e. no characteristic length scale) $H \neq 1/2$. A value $H < 1/2$ signals the presence of long range anti-correlations, while a value $H > 1/2$ signals the presence of long range correlations, also known as persistence.

We apply the DFA technique to process the sequence of emotional polarities $e_i(t)$ expressed by user i . Then we quantify the persistence of the user through the Hurst exponent H_i , which estimates how the emotional expression of user i fluctuates around its trend, revealing some properties of the underlying emotional states of the user. $H_i = 0.5$ if the user switches randomly between emotional states, $H_i > 0.5$ if the user shows persistence in its emotional expression, and $H_i < 0.5$ if the user has strong tendency to switch between states.

Figure 3.5 shows two examples of the DFA technique applied to a persistent and an anti-persistent user of the chatroom. The left panel shows the time series of emotional polarity of the two users. For the persistent user (purple), it is easy to notice that it stays longer expressing the same emotional state. On the other hand, the anti-persistent user (green),

switches very quickly between emotional states. The persistent case is analogous to the presence of emotional momentum (Kuppens *et al.*, 2010), which is a the tendency of some users to stay longer in some emotional states. Furthermore, anti-persistence might be a sign of emotional self-regulation (Rime, 2009), leading users to control their expression patterns away from strong emotional content. The right panel of Figure 3.5 shows the computed $F(\Delta t)$ from the expression of the two users of the left panel. Purple dots follow a scaling of parameter close to $H = 0.7$, revealing the persistence of the user expression. Green dots show the scaling of the expression of the anti-persistent user, with $H = 0.4$

3.4.2 Individual emotions

In order to have reliable statistics, we restrict our analysis to only those users with more than 100 posts (which are nearly 3000 users). If we analyze the distribution of the Hurst exponents of all users, shown in the histogram of the left panel of Figure 3.6, we find (a) that the emotional expression of users is far from being random, and (b) that it is clearly skewed towards $H_i > 0.5$, which means that most of the users are quite persistent regarding their positive, negative or neutral emotions. This persistence can be also seen as a kind of memory (or inertia) in changing the emotional expression, i.e. the following post from the same user is more likely to have the same emotional value.

The question whether persistent users express more positive or negative emotions is answered in the right panel of Figure 3.6, where we show a scatter plot of H_i versus the mean value of the emotions expressed by each user, $\langle e_i \rangle$. Again, we verify that the majority of users has $H_i > 0.5$, but we also see that the mean value of emotions expressed by the persistent users is largely positive. This corresponds to the general bias towards positive emotional expression detected in written expression, which is explained in Chapter 5. The lower left quadrant of the scatter plot is almost empty, which means that users expressing on average negative emotions tend to be persistent as well. A possible interpretation for this could be the relation between negative personal experiences and rumination as discussed in psychology (Rime, 2009). Antipersistent users, on the other hand, mostly switch between positive and neutral emotions. The inset of Figure 3.6 shows the ratios of users on each one of the quadrants of H_i and $\langle e_i \rangle$. Among negative users ($\langle e_i \rangle < 0$), there is certain bias towards persistence, i.e. when a user has a tendency towards negative expression, this tends to be persistent too. Popular culture interprets this kind of users as *trolls*, which are commonly accused of creating highly negative messages without really expressing their genuine emotions.

Are the more active users also the emotionally persistent ones? In the left panel of Figure

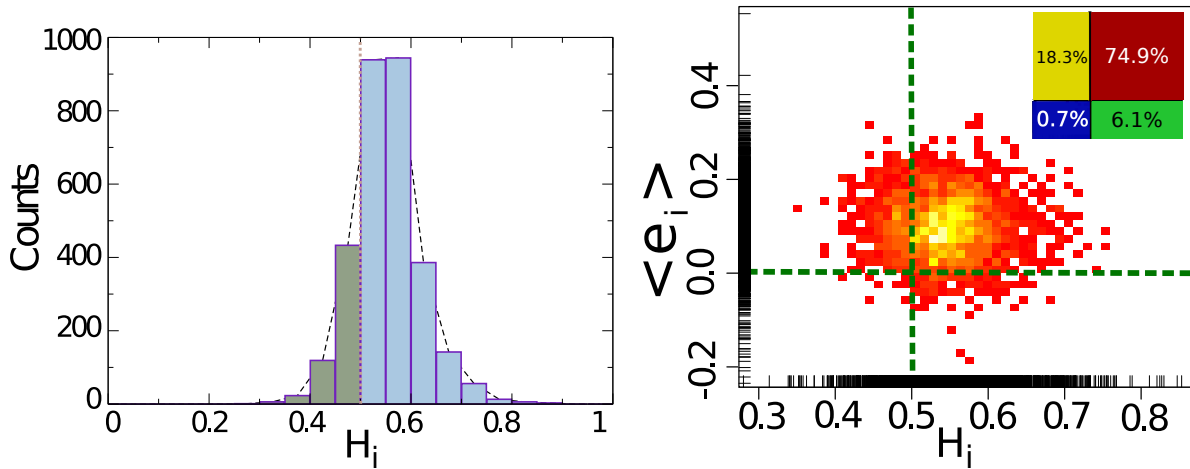


Figure 3.6: Distribution of user persistence $P(H_i)$ (left) for users with more than 100 messages in the recorded dataset. Right: histogram of the density of users according to their persistence H_i and their mean emotional expression $\langle e_i \rangle$. The inset shows the ratios of users on each of the quadrants of the plot.

3.7 we show a scatter plot of the Hurst exponent dependent on the total activity of each user. Even though the mean value of H_i does not show any such dependence, we observe large heterogeneity on the values of H_i for users with low activity. Furthermore, in the right panel of Figure 3.7 we show that the Hurst exponent of a very active user varies only slightly if we divide his time series into various segments and apply the DFA method to these segments. Note that the Hurst exponent is larger up to the usage of approximately 20 segments, which roughly corresponds to segments of two days of length. This could be a sign of the multi-level nature of human emotions, suggesting that at time scales of about 48 hours there are different dynamics of the expression of emotional states. This can be the basis for a differentiation between *mood* (conscious, long emotional states) and *core affect* (short-lived emotional states with a strong physiological component). The combined results of Figure 3.6 and Figure 3.7 allows us to conclude that active users tend to be emotionally persistent and, as most persistent users express positive emotions, they tend to provide some kind of positive bias to the chatroom, whereas users occasionally entering the chat may just try to get rid of some negative emotions.

3.4.3 Emotions at the discussion level

After characterizing the patterns of emotional expression of individual users, we monitor how the total emotion expressed in a given channel evolves over time. We use a moving average approach that calculates the ratio of the amount of posts with a given emotional

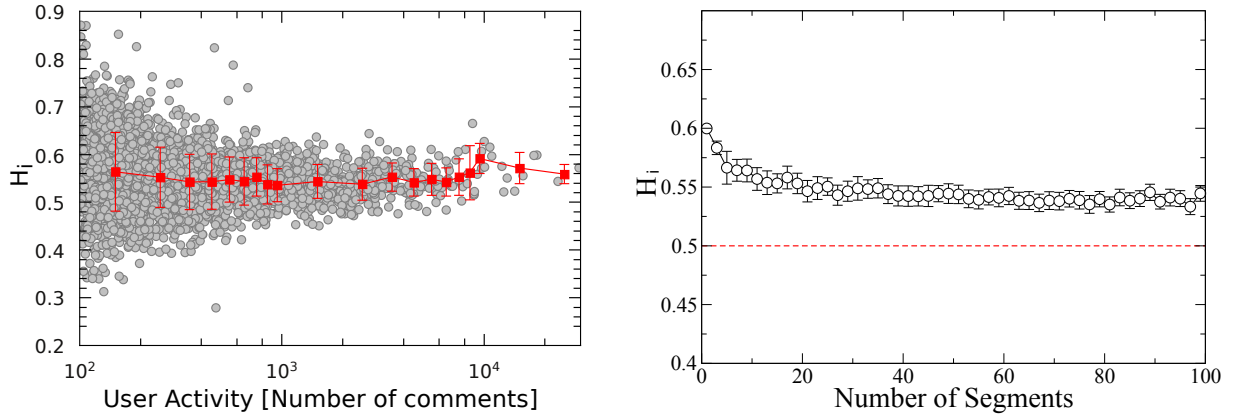


Figure 3.7: Left: scatter plot of H_i versus the level of user activity. Red bars show the mean and standard deviation of H_i for different values of the activity (linear binning, logarithmic scale). Right: mean and standard deviation of H_i calculated over segments of the expression of a user, versus the amount of segments used.

polarity over different time windows, denoted as $N(t)$ for negative expression, $U(t)$ for neutral expression, and $P(t)$ for positive expression (see Section 2.3.2). The three horizontal plots of Figure 3.8 show these fractions of posts as a function of time, for different sizes of the time window. While it is obvious that the emotional content largely fluctuates when using a very small time window, we find that for decreasing time resolution (i.e. increasing time window) the fractions of emotional posts settle down to an almost constant values around which they fluctuate. From this, we can make two interesting observations: (i) the emotional content in the online chats does not really change in the long run (one should notice that times of the order 10^3 are still large compared to the time window $DT = 50$ used), i.e. we observe fluctuations that depend on the time resolution, but no "evolution" towards more positive or negative sentiments. (ii) For the low resolution, the fraction of neutral posts dominates the positive and negative posts at all times. In fact there is a clear ranking where the fraction of negative posts is always the smallest. Both observations become even more pronounced when averaging over the whole time scale for each of the 20 IRC channels, as shown in the right column of Figure 3.8

Our findings differ from previous observations of emotional communication in blog posts and forum comments which identified a clear tendency toward negative contributions over time, in particular for periods of intensive user activity (Mitrović and Tadić, 2009; Chmiel and Hoyst, 2010). Such findings suggest that an increased number of negative emotional posts could boost the activity, and extend the lifetime of a forum discussion. However, blog communication in general evolves slower than online chats. Hence, we need to better understand the role of emotions in real-time Internet communication, which obviously

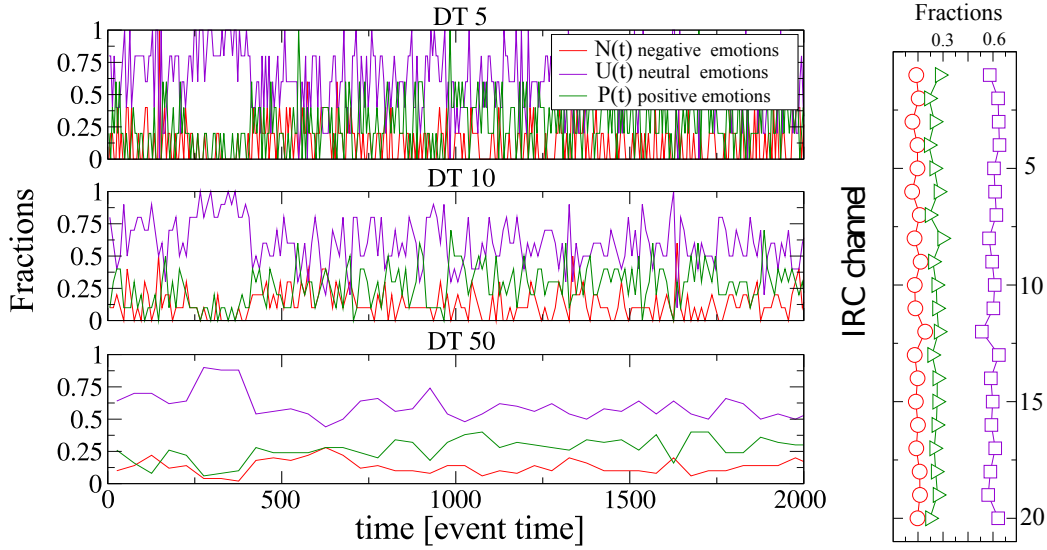


Figure 3.8: Fraction of negative, neutral, and positive emotional expressions measured under different time scales for a typical IRC channel. The right column shows the overall ratios of emotional expression for all the analyzed channels.

differs from the persistent and delayed interaction in blogs and fora.

This leads us to the question how persistent the emotional bias of a whole discussion is. For each channel, we calculated its emotional persistence H_{ch} by applying the DFA method over the whole set of emotional expressions in the channel, regardless of their author. While Figure 3.6 has shown the persistence with respect to the different users, Figure 3.9 plots this collective discussion persistence. For each channel, we divided its set of emotional expressions in 10 different segments, applying DFA to each one. The barplots of Figure 3.9 represent the mean H_{ch} over the 10 segments of the channel, and the bar around it shows the standard deviation. To control for possible methodological biases, we also calculated the Hurst exponent of shuffled versions of the channels, destroying any temporal correlations. These calculations show that the conversation persistence is significantly larger than the one in the shuffled data, which is close to 0.5, as expected.

The analyzed channels deal with very different topics, but most of them show similar values of H_{ch} , always above 0.5. We conclude that the persistence of the discussion per se (which is different from the persistence of the users which can leave or enter a arbitrary times) reflects a certain narrative memory. Precisely, for each chat, we observe the emergence of a certain (emotional) "tone" in the narration which can be positive, negative or neutral, dependent the emotional expressions of the (majority of) persistent users. We note that we could not find evidence of correlations using the autocorrelation function $C(\Delta t)$ of the emotion time series of the channel, while the observed persistence in the fluctuations of

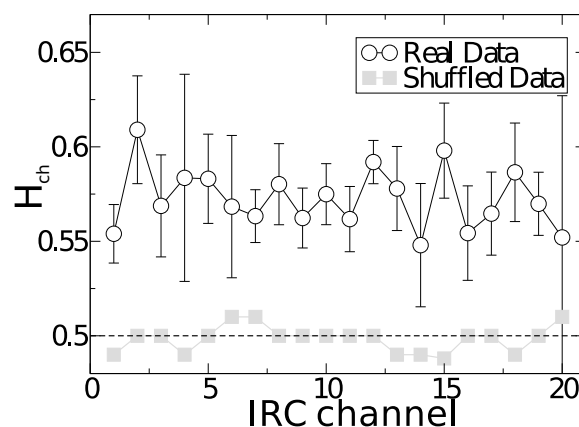


Figure 3.9: Hurst exponents of the discussions in each channel. Error bars show the mean H calculated over 10 different segments of the channel, with bars that show the standard deviation. Grey squares show the Hurst exponent for the shuffled discussion.

user emotional expression, as captured by H_{ch} is very robust. This indicates that the chat community assumes an emotional memory locally encoded in the current messages (from the user perspective), while the size of the conversation is too large to detect it through averaging techniques, or long range correlation analysis.

3.5 Modeling persistence in chatrooms

3.5.1 An agent-based model for chat communities

After identifying both the activity patterns, and the emotional expression patterns of users in online chats, we setup an agent-based model that is able to reproduce these stylized facts. We applied the general framework introduced in Chapter 1, designed to model and explain the emergence of collective emotions in online communities through the evolution of psychological variables. The case of IRC channel communication is of particular interest because of its fast and ephemeral nature. Thus, we have designed a model for IRC chatrooms, outlined in Figure 3.10, which shows the main causation among the elements of this model.

Agents are characterized by two variables, their emotionality, or valence, v which is either positive or negative and their activity, or arousal, which is represented by the time interval τ between two posts s in the chatroom. The valence of an agent i , represented by the internal variable v_i , changes in time due to a superposition of stochastic and deterministic

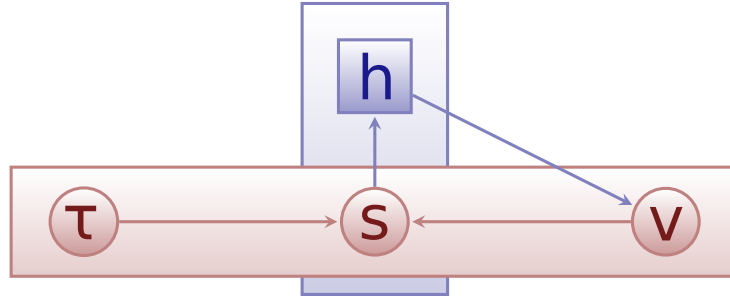


Figure 3.10: Schema of the agent-based model for emotions in chatrooms. Agents have an expression s which emotionality is determined by the internal valence v , and the arousal driven by the inter-activity time τ . The information field h can be perceived by all agents, and influences their valence.

influences (Schweitzer, 2003):

$$\dot{v}_i = -\gamma_v v_i + b * (h_+ - h_-) * v + A_v \xi_i \quad (3.4)$$

The stochastic influences are modeled as a random factor $A_v \xi_i$ normally distributed with zero mean and amplitude A_v , and represent all changes of the individual emotional state apart from chat communication. The deterministic influences are composed of an internal decay of parameter γ_v , and an external influence of the conversation. The change in the valence caused by the emotionality of the field ($h_+ - h_-$) is measured in valence change per time unit through the parameter b . The previous models under this framework had an additional saturation term in the equation of the valence dynamics. This way the positive feedback between v and h was limited when the field was very large. But, as we show in Figure 3.8, chatrooms do not show the extreme cases of emotional polarization observed in other communities. Thus, we simplify the dynamics of the valence without using any saturation terms, since a large imbalance between h_+ and h_- is unrealistic given our analysis of real IRC data.

In general, the level of activity associated with the emotion, known as arousal, can be explicitly modeled by stochastic dynamics as well, as we did in Chapter 1 and Chapter 2. Here, the activity of an agent is estimated by the inter-activity time distribution that triggers the expression of the agent, i.e. by the power-law distribution $P(\tau) \sim \tau^{-1.53}$ shown in Figure 3.2. Assuming that an agent becomes active and expresses its emotion at time t , it will become active again after a period τ . The agent then writes a post in the collective discussion with an emotional content determined by its valence (see below). This information is stored in an external field common for all agents, which is composed of two components, h_- and h_+ , for negative and positive information, and their difference measures the emotional charge of the communication activity. Since we are interested in

emotional communication, we assume that all neutral posts entered, or already present, in a chatroom do not influence the emotions of the agents participating to the conversation. Thus, the dynamics of the field is influenced only by the amount of agents expressing a particular emotion at a given time: $N_+(t) = \sum_i (1 - \Theta(-1 * s_i))$ and $N_-(t) = \sum_i (1 - \Theta(s_i))$, where Θ is the Heaviside step function. Therefore, the time dynamics of the fields can be described as:

$$\dot{h}_{\pm} = -\gamma_h h_{\pm} + c * N_{\pm}(t) \quad (3.5)$$

These two field components, h_+ and h_- , decay exponentially with a constant factor γ_h , i.e. their importance decays very fast as they move further down the screen (posts never disappear, but become less influential very quickly). Each field increases by a fixed amount c from every post stored in it. The values of the valence of the agents are changed by the field components, as described by Equation (3.4). In contrast with traditional means of communication, online social media can aggregate much larger volumes of user-generated information. This is why h is defined without explicit bounds. Chatrooms pose a special case to this kind of communication, as they can contain large amount of posts but limited amount of users. Most IRC channels have technical limitations for the amount of users that can be connected at once, which in turn is reflected in the total amount of posts present in the general discussion. In our model, h might take any value, but the empirical activity pattern combined with the fixed size of the community dynamically constraints it to limited values.

Whenever an agent creates a new post in an ongoing conversation, its external emotional expression, s_i , changes its value in the following way:

$$s_i = \begin{cases} -1 & \text{if } v_i < V_- \\ +1 & \text{if } v_i > V_+ \\ 0 & \text{otherwise.} \end{cases} \quad (3.6)$$

, where $s_i = 0$ for all the moments in which the agent is not expressing any emotion. The thresholds V_- and V_+ represent a limit value of the valence that determines the emotional content of each post, and in general can be asymmetric, as humans tend to have different thresholds for the triggering of positive and negative emotional expression (Christophe and Rime, 1997; Rime, 2009). Each action contributes to the amount of information stored in the information field of the conversation, increasing h_- if $s = -1$ or h_+ if $s = +1$. When a post is created, its emotional polarity is determined by the valence, and then aggregated in the collective discussion which can be perceived by all agents.

The effect of chatroom communication on an agent's emotionality is modeled as an empathy-driven process (Preston and de Waal, 2002) that influences the valence. In the valence dynamics we propose in Equation (3.4), agents perceive a positive influence when their

emotional state matches the one of the community, and a negative one in the opposite case. All the assumptions of our model are supported by psychological theories. Parameter values and dynamical equations can be tested against experiments in psychology, as we do in Chapter 4 with data from the appropriate experiments (Kappas *et al.*, 2011). Furthermore, our model provides a consistent view of the emotional behavior in chatrooms leading to testable hypotheses that can drive future psychology research.

3.5.2 Simulation of emotions in a chatroom

Given the full description of the agent-based model, we performed a set of simulations to explore its behavior at the individual and collective levels. First, we analyzed individual users in isolation, i.e. $b = 0$. The internal parameters A_v and γ_v were set to values uniformly distributed in the ranges $A_v \in [0.2, 0.5]$ and $\gamma_v \in [0.2, 0.5]$. The activity rule of the agents, given by $P(\tau)$ was unchanged, and the expression of each agent was recorded during 45000 time units (31.25 days if $t=1$ min). For each agent, we applied the DFA method to its sequence of emotional expressions $s_i(t)$, and then calculated the Hurst exponent H_i . We discarded the results with $R^2 \leq 0.9$, as the realization of $P(\tau)$ might induce some agents to make very few expressions. For this simulation setup, the distribution of H_i is shown in the left panel of Figure 3.11. This distribution is shifted towards positive values similar to the one observed in real data, this way reproducing the emotional persistence of the conversation without assuming any time dependence between user expressions.

We performed extensive computer simulations using different parameter sets on a computer cluster composed of 64 cores. This time we aimed at understanding the patterns emerging from the interaction of agents through the chatroom, i.e. $b > 0$. The role of the values of these parameters in the collective behavior of the model is discussed in the next section. Thanks to this exploration we focused on a particular set of values to simulate conversations similar to actual IRC chats. We used 10000 agents in a conversation lasting 45000 time units, and performed 10 realizations of the model using the following set of parameters:

$$V_- = -0.15, \quad V_+ = 0.05, \quad \gamma_v = 0.2, \quad A_v = 0.2, \quad b = 0.01, \quad c = 0.05, \quad \gamma_h = 0.9 \quad (3.7)$$

We analyzed the activity of the simulated discussion and its persistence across simulations. The main results of these simulations are the following:

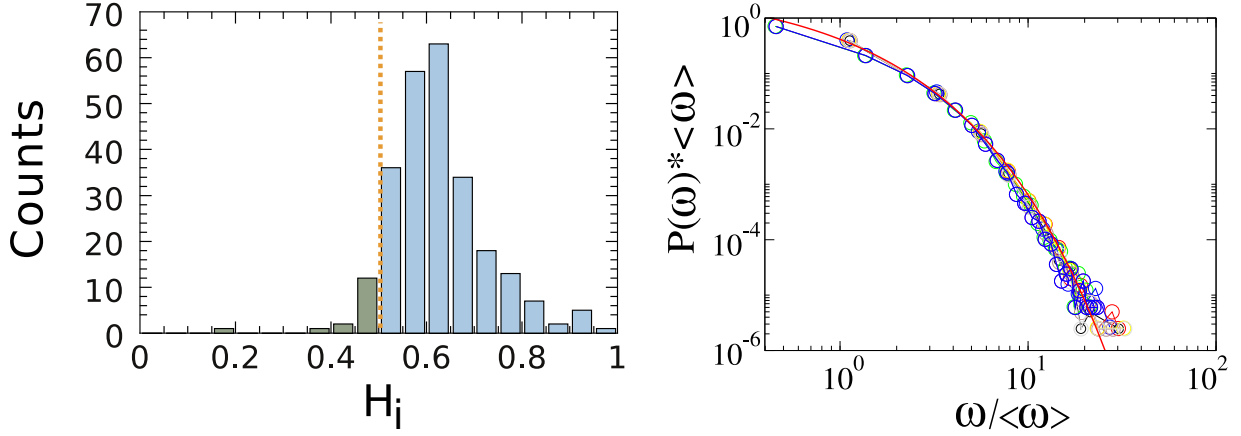


Figure 3.11: Left: Distribution of the Hurst exponent of simulation of agents in isolation with $A_v \in [0.2, 0.5]$ and $\gamma_v \in [0.2, 0.5]$. Only Hurst exponents with $R^2 > 0.9$ are reported. Right: distribution of rescaled inter-event times ω for 10 simulations of the model with parameter set 3.7. The red line corresponds to a stretched exponential fit of $\gamma = 0.59$ ($p < 0.001$).

- Conversation inter-event time distribution.** For each simulation, we calculate the mean inter-event time $\langle \omega \rangle$, in order to rescale it as we did with the empirical data. The rescaled inter-event time distribution of the simulations is well fit by a stretched exponential distribution, as in the real data. The right panel of Figure 3.11 shows this distribution, with a stretched exponential fit of $\gamma = 0.59$ ($p < 0.001$).
- Conversation persistence** The mean Hurst exponent for the 10 simulated channels is $\langle H_{sim} \rangle = 0.567 \pm 0.007$, whereas for the real IRC channels $\langle H_{ch} \rangle = 0.572 \pm 0.021$. This result suggests that our agent-based model reproduces qualitatively the emergence of emotional persistence in the IRC conversation and thus, based on all findings, is able to capture the essence of emotional influence between users in chatrooms. This persistence at the collective level reflects the coherence in emotional states emerging from the interaction of the agents of our model.

3.5.3 Computational analysis of the model

In order to understand how each parameter affects the ratio of emotion polarities in the posts, and the user and conversation persistence as well, we performed a large set of simulations using different combinations of parameter values. For each combination of values we performed 10 simulations, recording agent activity to calculate the ratios of emotional expression (N, P, U), the individual persistence H_i , and the conversation persistence H_{ch} .

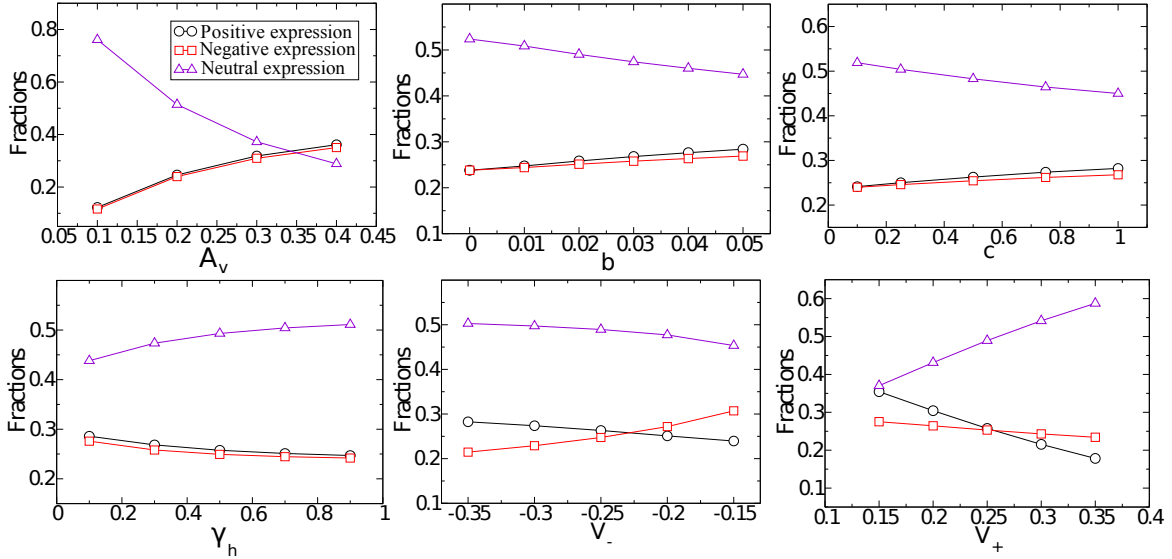


Figure 3.12: Fractions of positive, negative and neutral posts for different values of the parameters in our simulations.

We measure the above metrics for each value of a parameter, including all the simulations in which that parameter was set to that particular value. This way we explore the dependence of an individual parameter in the context of the whole set of simulations, which aimed to be as general as possible with our computational resources.

Figure 3.12 shows the changes in N , P , and U according to the value of each parameter. The ratio of emotionally charged expressions (positive and negative) increases with the amplitude of the valence stochastic component A_v . The explanation for this relation is that higher A_v implies larger deviation in the valence of the agents, which leads to more emotional messages if the mean of the distribution of agent valences lies between the V_- and V_+ thresholds. This ratios of emotional messages are also slightly increased by the coupling parameters b and c , which make stronger the influence of communication on the valence. The inverse is true for the parameter γ_h , i.e. the ratio of neutral posts increases the larger the decay of the field. Clearly a faster decay in the influence between agents should also decrease their overall emotional expression. An increase in the threshold V_{\pm} leads to lower frequency of expression of the corresponding sign, in line with the fact that the agents would require stronger emotional states to overcome higher thresholds.

We also calculated the mean conversation and user persistences across all the simulations that had certain parameter values, which are summarized in Figure 3.13. Under the influence of an emotional field, the user persistence increases with A_v , meaning that a stronger stochastic component can lead to conversations more similar to the observations in real data. The coupling parameters c and b increase both mean persistence, as expected

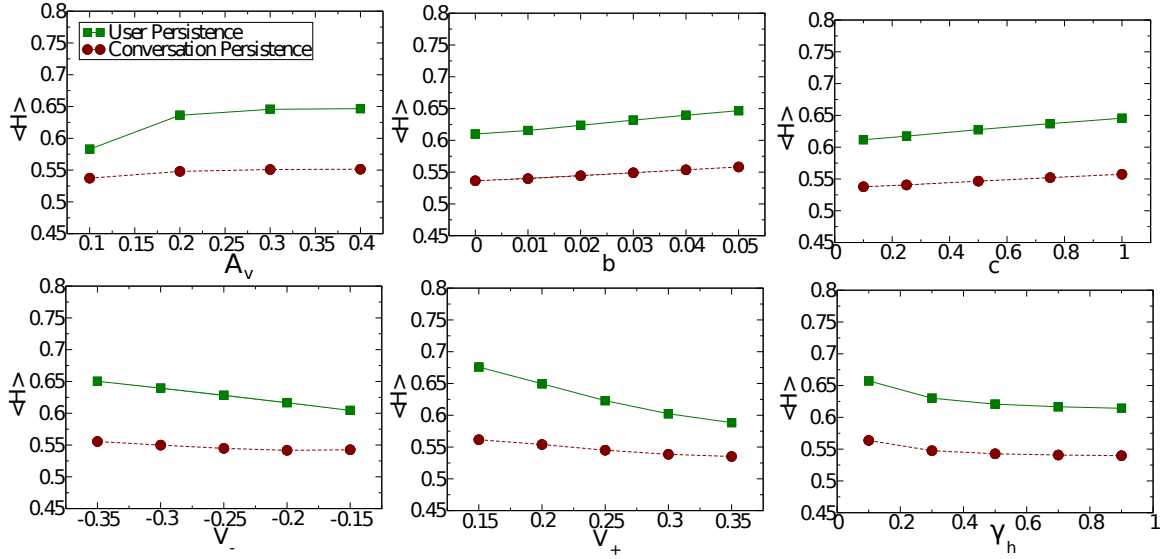


Figure 3.13: Mean value of the Hurst exponents of the emotional expression of agents and conversations for different values of the simulation parameters. The error bars showing the standard error of the mean value are smaller than the symbol size and are not visible.

from more coherent similarity in the emotional states of the agents. The effect of larger γ_h is the inverse, the stronger the decay of information, the weaker the persistence. Larger positive thresholds V_+ lead to lower user persistence, while the inverse is true for the negative threshold V_- . This might be related to the dominance of positive expression over negative one, which in turn shifts the difference between h_+ and h_- , leading to more positive valences.

Finally we look at the role of the internal valence decay, γ_v from a different perspective. As mentioned in Section 3.4.2 with relation to Figure 3.7, the empirical analysis of individual persistence suggests the existence of two different timescales of emotions, a slower one related to mood, and a faster one related to core affect. We studied two possible values of this decay, $\gamma_v = 0.1$ corresponding for the case of mood (slower decay), and $\gamma_v = 0.5$ for the case of core affect (faster decay). For both cases, the distributions of individual and discussion persistence are shown in Figure 3.14.

The Kolmogorov-Smirnov distance between the distribution of individual persistences in a simulation with $\gamma_v = 0.1$ and the real data is $D_{KS} = 0.845$, while between the distribution for $\gamma_v = 0.5$ and the real data is $D_{KS} = 0.519$. Furthermore, the mean Hurst exponent of the simulated conversations is also closer to the values from the 20 IRC channels for the case of $\gamma_v = 0.5$, implying that the relaxation speed of the valence of chatroom users is fast. This means that, within our framework, the empirical behavior of chat users is

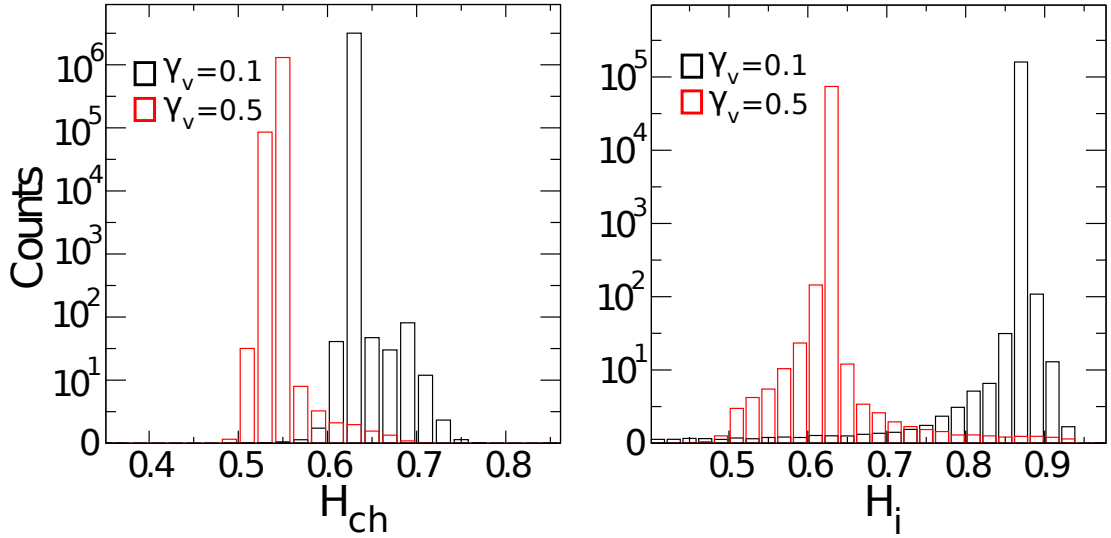


Figure 3.14: Distribution of the Hurst exponents for the simulated conversations (left) and agents (right) for the cases of $\gamma_v = 0.1$ (black) and $\gamma_v = 0.5$ (red).

better explained by the theoretically fast dynamics of core affect than by the slower, more complex mood. This result serves as a robustness check for our framework, which was originally phrased within the context of core affect, as explained in Chapter 1.

3.5.4 Model limitations

Our results have shown that our model is able to reproduce the emergence of emotional persistence in a chatroom discussion, building on dynamical principles that also lead to individual persistence. However, there are features of the activity patterns of users that are not fully grasped by our model. For example, the stretched exponent of the inter-event time distribution of our simulations, $\gamma = 0.59$ ($p < 0.001$), is different from real IRC channels where the average value was $\langle \gamma \rangle = 0.21$, i.e. there is a faster decay in the simulations. This could be explained by the fact that in the real chat users usually write after they have read the previous post, i.e. there are additional correlations in the times users enter a chat. These, however, are not considered in the simulations, because agents post in the chat at random after a given time interval τ , i.e. there is no additional coupling in posting times.

To explore the possible source of this differences we applied DFA to the sequence of inter-event times $\omega(t)$ of the IRC channels and of our simulations. The Hurst exponent obtained for the IRC channels is $H_\omega \simeq 0.6$, which is the slope of the left panel of Figure 3.15. The origin of this persistence could be due to synchronized bursts of activity leading

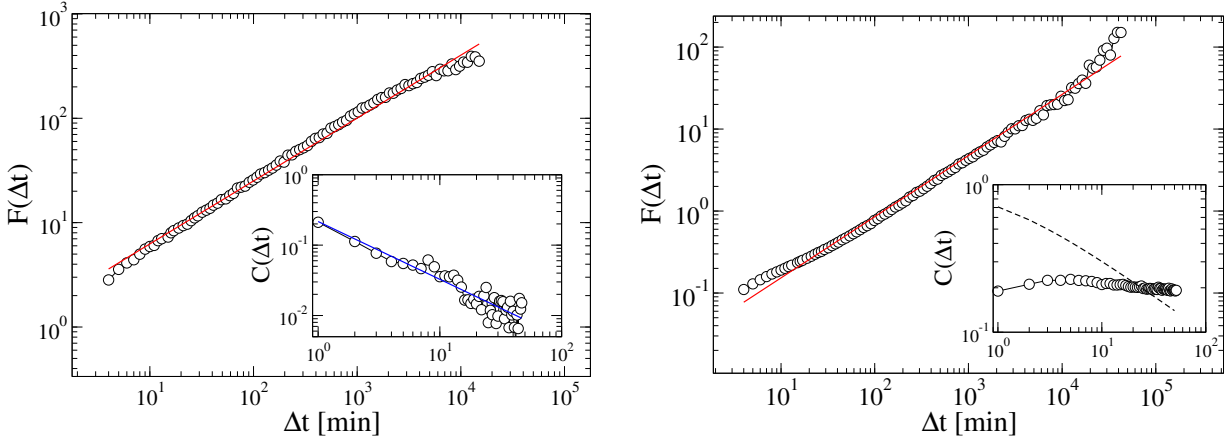


Figure 3.15: DFA fluctuation function calculated using the inter-event times of a real IRC channel (left) and a simulation with the parameter set 3.7 (right). The insets show the autocorrelation for both cases, with a dotted line showing the expected decay of the autocorrelation function according to the scaling relation $\nu_{\omega'} = 2 - 2H_{\omega}$, present when there are long range dependencies in expression times (Kantelhardt, 2009).

to persistent dependencies over different time scales, or due to the broad distribution of inter-event times $P(\tau)$, or to a combination of both. The existence of dependencies in the activity is highlighted by a power-law decaying autocorrelation function $C(\Delta t)$, shown in the left inset of Figure 3.15 and already commented in Figure 3.4. The empirical estimation of the decay exponent of $C(\Delta t)$ is $\nu_{\omega} \simeq 0.82$, following the scaling relation with the Hurst exponent $\nu_{\omega} = 2 - 2H_{\omega}$ explained in (Kantelhardt, 2009).

This metrics are different for the case of our simulations, shown in the right panel of Figure 3.15. We found that, in the simulation inter-event time series, $H_{\omega} = 0.75$, however, we did not observe a power-law decay of the autocorrelation function. This suggests that the simulated correlations in event times are due to the power-law distributed inter-activity times used as input to our model ($P(\tau)$), and it is in line with the above discussion about the absence of coupling, which also explains the difference in the stretched exponents. The question of how to find this correlations in an agent-based model within our framework remains open. We should note that these kind of scaling relations can be obtained with agent-based models (Wu *et al.*, 2010), at least for the case of dyadic interactions. The arousal dynamics of the agents can be defined in a way that these activity correlations are present, allowing the testability of how emotions lead to long range correlations in user activity.

3.6 Discussion

We started with the question to what extent human communication patterns change on the Internet. To answer this, we used a unique dataset of online chatting communities with about 2.5 million posts on 20 different topics. Our analysis considered two different dimensions of the communication process: (a) activity, expressed by the time intervals τ at which users contribute to the communication, and ω at which consecutive posts appear in a chat, and (b) the emotional expressions of users. With respect to activity patterns we did not find considerable differences between online chatrooms and other previously studied forms on online and offline communication. Specifically, both the inter-activity distribution of users and the inter-event distribution of posts followed known distributions. Thus, we may conclude that humans do not really change their activity patterns when they go online. Instead, these patterns seem to be quite robust across online and offline communication.

The picture differs, however, when looking at the emotional expressions of users. While we cannot directly compare our findings on emotional persistence to results about offline communication, we find differences between online chatrooms and other forms online communication, such as blogs, or fora. While the latter could be heated up by negative emotional patterns, we observe that online chats, which are instantaneous in time, very much follow a balanced emotional pattern across all topics (shown in the emotional persistence of the channels), but also with respect to individual users, which are in their majority quite persistent in their emotional expressions (mostly positive ones).

This observation is indeed surprising as online chats are mostly anonymous, i.e. users do not reveal their personal identity. However, they still seem to behave according to certain social norms, i.e. there is a clear tendency to express an opinion in a neutral to positive emotional way, avoiding direct confrontations or emotional debates. One of the reasons for such behavior comes from the "repeated interaction" underlying online chats. As the daily "bump" the activity patterns also suggest, most users return to the online chats regularly, to meet other users they may already know. This puts a kind of social pressure on their behavior (even in an unconscious manner) to behave similar to offline conversations. In conclusion, we find that the online communication patterns do not differ much from common offline behavior if a repeated interaction could be assumed.

Eventually, we argue that the emotional persistence found is indeed related to the nature of human conversations. After all, the correlations shown in the emotional expressions of different users indicate that there is some form of emotional sharing between participants. This suggests the presence of social bonds among users in the chatroom (Rime, 2009) and

confirms similarities between online and offline communication.

The fact that we could reveal patterns of emotional persistence both in users and in discussed topics, does not mean that we also understand their origin. One important step towards this "microscopic" understanding is provided by our agent-based model of emotional interactions in chatrooms. By using assumptions about the agent's behavior which are rooted in research in psychology, we are able to reproduce the stylized facts of the chatroom conversation, both for the activity in channels and for the emotional persistence. Specifically, our model allows us to test hypotheses about the emotional interaction of agents against their outcome on the systemic level, i.e. for the chatroom simulation. This helps to reveal what kind of rules are underlying the online behavior of users which are hard to access otherwise. But our agent-based model, like all the models, is a simplified view of the reality in order to explain the statistical regularities found in our observations. We performed additional analysis on the correlation between inter-event times, finding that our model is not able to reproduce long range influence in activity times. This calls for an extension of the model that introduces arousal dynamics, which can lead to coupling in the times when agents express their emotional states.

Part II

Describing the emotions of the individual

Quantitative analysis of the emotions of individuals through changes in emotional states and patterns of emotional expression.

“All the knowledge I possess every one else can acquire, but my heart is exclusively my own. ”

Johann von Goethe, *The Sorrows of Young Werther*, 1774

“Happiness is beneficial for the body, but it is grief that develops the powers of the mind.”

Marcel Proust, *The Past Recaptured*, 1927.

Chapter 4

Dynamics of individual emotions

Summary

We study the dynamic changes of emotional states induced by reading and participating in online discussions, which define the individual microdynamics of our modeling framework. Using principles of nonlinear dynamics, we quantify changes in valence and arousal through subjective reports and physiological signals, as recorded in four independent studies including 268 participants (150 female), 20 threads, 18 IAPS images, and 12000 reports from daily experience. In absence of controlled online stimuli, valence and arousal decay exponentially toward fixed values. While reading emotionally charged online discussions, valence dynamics show the existence of strong negative and positive attractors. Arousal is increased with emotional content regardless of its polarity, and the user intention to participate in a discussion increases with experienced arousal. After writing a post, participants' arousal significantly decreases as an additional relaxation mechanism. We present a method for the processing of EMG signals that combines stationary and non-stationary properties of its time series. This method allows the processing of Zygomaticus Major responses for conventional extraction of peak activation, as well as improved analysis of local maxima during weak negative stimuli. We analyze activation at Corrugator Supercilii for stationary properties allowing detection of dynamic changes of valence in participants reading online discussions. These results allow the design of agent based models to reproduce and analyze the emergence of collective emotions in online communities.

4.1 Empirical study of emotion dynamics

The purpose of the agent-based models presented in Chapter 2 and Chapter 3 was to reproduce the stylized facts observed in data from online communities. Such agent-based models can be very useful to predict future user behavior, or to design mechanisms that optimize certain aspects of the community. But this cannot be achieved if the microdynamics that drive agent actions are not validated beyond computer simulations or analytical results. It is relatively simple to design a model that, based on *ad hoc* assumptions, reproduces any observed macroscopic behavior. Our modeling framework is explicitly designed to avoid this pitfall, with agent dynamics built on testable assumptions. As we explain in Chapter 1, this framework links the collective behavior of the community with individual emotion dynamics and interactions, which can be empirically tested in experimental studies. The dynamics of these emotions are phrased within the psychological theory of *core affect* (Russell, 1980), which provides a unified representation of the kind of emotions we refer to.

Thanks to our framework, the agent dynamics of our models can be phrased as testable hypothesis, formulated in terms that allow their framing within a broader psychological context. Valence and arousal are common variables taken into account in psychological studies, which can provide data to test the dynamics proposed by our models. In general, previous studies on emotions focus on the role of emotions in other aspects of human activity. These works study emotions in a static manner, without analyzing the way in which they change. Agent-based approaches are rather scarce in empirical psychology (Kuppens *et al.*, 2010), a field in which most of the research is based on observations and data analysis, but not on computational models to reproduce those observations. However, there is a changing trend in this respect, and new works on the agent-based approach are emerging within the psychology community. (Smith and Conrey, 2007) suggests that that agent-based models can provide new insights into social psychology, unifying the current models of social interaction. This view is not precisely focused on emotion, but provides some grounding to the application of our techniques in social psychology. Regarding emotional dynamics, some works within mathematical psychology justify the usage of agent-based modeling (Scherer, 2009). Furthermore, some models already include principles of the stochastic dynamics of emotional states, including emotion dynamics as CUSP catastrophes (Sander *et al.*, 2005), and emotional state dynamics from a Brownian agent approach (Kuppens *et al.*, 2010; Oravecz *et al.*, 2011). This last case provides special support to our modeling framework, which we discuss below.

We proceed to study the dynamics of emotions in our framework through the empirical analysis of various datasets of different origin. We organize the dynamics of our modeling

framework in groups according to the type of data and experiments they require to be tested:

1. **Eigendynamics:** the internal dynamics of psychological states, which are not induced by any external influence. This requires data on long periods in which the participants of the experiment do not need to be using a computer (but might be). In our modeling framework, the main component of the internal dynamics is the exponential relaxation towards a neutral state ($-\gamma_v v(t)$, $-\gamma_a a(t)$ in Equation (1.2)), which means that states of high emotionality do not last for very long. This kind of data would allow us to measure the stochastic component of valence and arousal dynamics ($A_v \xi_v(t)$, $A_a \xi_a(t)$ in Equation (1.2)), if there is enough experimental data available.
2. **Perception.** In our framework, external influences on the agent states are generated by online interaction, which in this case consists mostly on posts written by other users. Monitoring the emotions of individuals while exposed to online content would allow us to measure valence and arousal change (\mathcal{F}_v , \mathcal{F}_a in Equation (1.2)). Furthermore, under particular conditions, we can measure how emotional states influence the attention levels and the choices of individual users towards different types of content.
3. **Production.** In our framework, agents create posts according to a rule that depends on their emotional state. We divide the dynamics of production in three subsets:
 - *Antecedents.* Which is the function of the internal state of agents that determines when they create posts in the online discussion? Our initial conjecture is that, when the arousal of an agent reaches a threshold (τ_i in Equation (1.10)), then the agent expresses its emotions through a post.
 - *Consequences.* After expression, we hypothesized the existence of a feedback mechanism that regulated the arousal, decreasing it to 0 (Equation (1.10)). This can be tested if the emotions of the participants of an experiment are monitored before and after creating posts.
 - *Post composition.* The emotional content of a post created by an agent is a function of its emotional state. In our model, the external variable s_i is set according to valence (Equation (1.3)), which is a way to transform internal emotions into the measurable data we can extract from the Internet. Testing the way emotions are encoded in a text heavily depends on the sentiment analysis tool used to process the posts, for which particular experimental setups are

necessary. Validating these tools is outside the scope of this research, but it is important to address their limitations every time we extract emotions from user-generated content.

4. **Online communication.** In the first applications of our model, agents communicate through an information field that can be equally accessed and modified by all of them. This design corresponds to online communities such as fora, product reviews, or chatrooms. This might, however, be different in other platforms like online social networks. We defined a set of assumptions about the dynamics of this communication, including an exponential decay on the impact of information ($-\gamma_h h$ in Equation (1.6)), and different field components for emotional polarity. Testing these assumptions depends on the particular online community under study, and are outside the scope of individual emotion dynamics.

In this chapter, we present an analysis of empirical data that explores emotion eigendynamics, influence of perception, and antecedents and consequences of emotional expression through posts. Our results are based on a set of experimental studies that monitor the emotional states of their participants in different ways. One of these datasets we use was produced by an *in vivo* experiment, in which participant emotions were recorded through reports during their daily life, without any experimental setup. We also count with data from other three experiments in which the participants were exposed to online emotional content, and asked to write posts. These experiments were designed to test the dynamics of online emotional interaction, and provide data in the form of subjective assessment of emotions and physiological data measurements. We use this set of datasets to investigate emotion dynamics within our agent-based framework, testing hypotheses and estimating the values of the parameters that drive individual emotions.

4.2 Experimental setups

Two groups of psychologists provided us with empirical datasets of emotional dynamics. In this section, we outline their experimental setups, the kind of stimuli that were used, and technical details about size and quality of the data. Figure 4.1 summarizes the different dynamics and stimuli used in these experiments, which are detailed in the following:

- **Dynaffect2**, produced the Research Group of Quantitative Psychology and Individual Differences, in KU Leuven ¹ contains valuable data on the eigendynamics of

¹<http://ppw.kuleuven.be/okp>

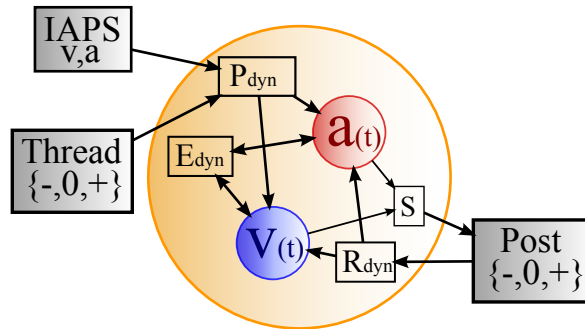


Figure 4.1: Schema of the types of dynamics tested in this Chapter. Emotion eigendynamics E_{dyn} represent the internal changes in emotional states in the absence of online interaction. Perception dynamics P_{dyn} change valence $v(t)$ and arousal $a(t)$ by testing two kinds of inputs: threads of known emotional polarity and IAPS pictures with values of valence and arousal. Production rules S determine when a participant wants to create a message, and the feedback after emotional expression is given by R_{dyn} .

valence and arousal (E_{dyn}). In this study, 60 students (40 female) used a portable device that was asking their emotional state (valence and arousal) 50 times a day during the waking hours of 4 days. The high frequency of this dataset makes it specially useful to test the dynamics of *core affect*. The response scale used for this experiment was $[1, 99]$, which we rescale to $[-1, 1]$ for the integration with our other analysis. The Dynaffect2 dataset was already used to test various versions of a Brownian agent model without social interaction (Kuppens *et al.*, 2010; Oravecz *et al.*, 2011). We reproduce these results within our modeling framework, allowing the combination of previous empirical findings with our further analysis of online interaction.

- **CS2.** The next three datasets, produced by the group of Cognition, Emotion and Social context of Jacobs University Bremen ², include valuable data on the dynamics of emotions under the perception and production of online content. The first dataset, CS2, was produced in a study in which 91 students (54 female) read 20 threads predetermined to have certain emotional content (9 negative, 9 positive, 2 neutral). This experiment was carried online, with the participants reading the threads outside any laboratory. Figure 4.2 shows the example of a post as seen by the participants, belonging to a thread preselected as negative from a real BBC forum discussion. After reading the posts of a thread, the participants provided subjective reports of their emotions on a Likert scale of 1 to 7, which we transformed to our scale $[-1, 1]$. Participants also answered a questionnaire which, among other questions, asked what

²http://www.jacobs-university.de/shss/akappas/workgroup_kappas/home

was their intention to continue reading and to participate in the conversation. We interpreted these answers as continuing and participating probability estimations. The subjective assessments of emotions of this study are useful to test the influence of emotionally charged threads (P_{dyn}), and the answers to the questionnaires provide data on how users decide when to create posts (S).

- **CS3.** This experiment is an improved version of CS2, adding the measurement of physiological signals. The 7 threads used for this experiment are a subset of those for CS2 (3 positive, 3 negative, 1 neutral), and 18 pictures from the International Affective Picture System (IAPS) (Bradley and Lang, 2007) were also used to measure the physiological reaction to emotional images, which can also be part of online communication. These pictures contain values of valence and arousal generated in previous survey studies, and allow us to test short term physiological reactions to emotional images. CS3 adds data on 53 participants (26 female), which in combination with CS2, compose two independent experiments which study perception (P_{dyn}) and production rules (S).
- **CS4.** The same pictures as in CS4 were also shown in CS4, providing an independent experiment for the emotional influence of IAPS pictures (P_{dyn}). This experiment, based on 65 participants (30 female), did not include the stimulus of emotionally charged threads, but focused on the production of emotional posts. Participants were asked to write posts expressing certain emotions as initiators of discussions, or as replies to previous posts in a discussion. The participants provided subjective assessments of their emotions before and after writing the posts, which we will use to study the feedback in emotions due to online expression (R_{dyn}).

4.2.1 Measurements of emotions

The experimental data included in this study is composed of two types of measurements: i) participant reports that included the subjective assessment of emotions, and ii) physiological signals in controlled environments. These two kinds of measurements of emotions have different advantages and disadvantages, and the usage of both can lead to more robust results. Figure 4.3 outlines the relationship between internal variables and measurement signals. Q represents subjective assessments of emotions, which are typically retrieved through individual reports on emotional states, quantified as valence and arousal. The four experimental datasets analyzed here include this type of data, usually through Likert scales of different granularity. Subjective assessments of emotions provide the emotional

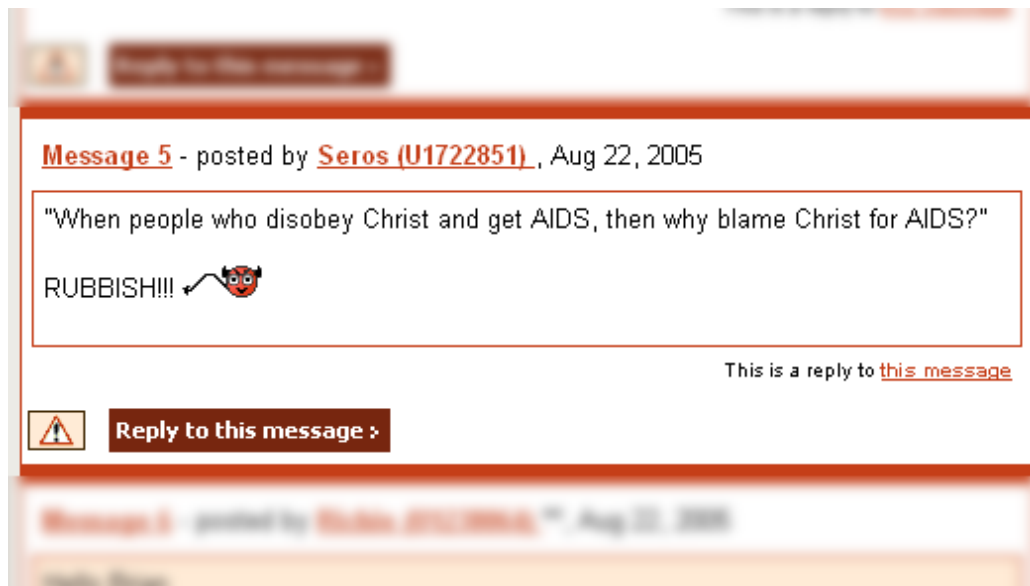


Figure 4.2: Example of a post from a thread used as negative stimulus in CS2 and CS3. This thread was extracted from the BBC forum, on a topic related to religion.

state of an individual at an instant in time, and they are a rather inexpensive way to estimate emotions. On the other hand, these subjective reports do not let us monitor emotions in a continuous manner, as they require the participant to shift its attention to provide the values. Additionally, this kind of subjective reports can be subject of biases that need to be carefully taken into account when designing the experimental setup. These biases can include tendencies to select responses that are more socially desirable (Robinson *et al.*, 1991), drive the participants towards positive values (Yik *et al.*, 2011), or depend on the framing of the question (Bryan *et al.*, 2011). The response formats used in the experiments analyzed here were designed by professional psychologists, following methodologies that avoid as much as possible the influences of these kind of biases (Robinson *et al.*, 1991).

The second source of data about emotional states are physiological signals controlled by the participant's nervous system. The autonomic nervous system is divided in the parasympathetic and sympathetic nervous systems, each one controlling different kind of reactions (Dorland, William and Alexander, Newman, 1980). The first controls actions that do not require an immediate reaction (i.e. sweating), and the second drives fast physiological responses (i.e. smiling), in general related to urgent reactions of the organism. Our dataset combines physiological signals from both systems, which do not need to be stationary, in particular for the sympathetic system. During experiments CS3 and CS4, the physiological responses of each participant were recorded while viewing pictures and reading threads. The recorded physiological signals are the following:

1. *Zygomatikus Major* (Zyg): the main muscle involved in the action of smiling. The activity of this muscle is measured through Electromyography (EMG), which measures the electrical potential in muscle cells.
2. *Corrugator Supercilii* (Cor), the frowning muscle, which should be an indicator of negative valence. The same EMG technique is applied to measure the activity of this muscle.
3. Heart rate (HR), related to arousal, is measured through a sensor in the participants' fingers, measuring the amount of heartbeats per minute (BPM).
4. Skin conductance (SC), is the sweating response to the stimulus. It is assumed to be related to arousal, and it is measured in experimental setups through the Electrodermal response (EDR), measured in μV .

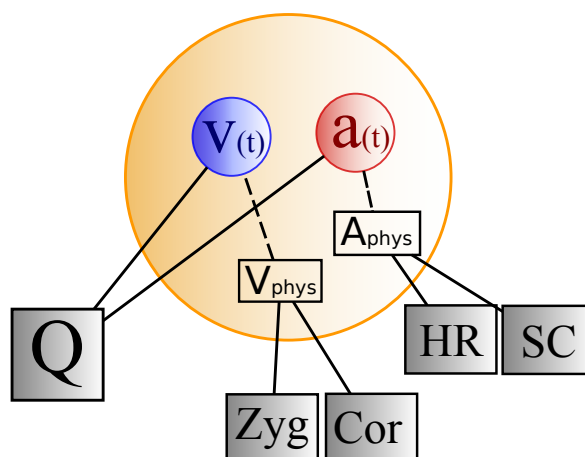


Figure 4.3: Schema of the measurements present in the available datasets. These are i) subjective assessments of emotional states (Q), and ii) a set of physiological measurements, composed of the EMG of *Zygomatikus Major* and *Corrugator Supercilii*, Skin Conductance, and Heart Rate.

The relation between physiological signals and internal emotional states is a current research subject with many open questions (Bradley *et al.*, 1996). Empirical data shows how Zyg and Cor activities are related to valence, in particular Zyg to positive valence and Cor to negative valence (Bradley, 2009). HR and SC are associated with arousal, working at different time scales (Jones and Troen, 2007). The relation between the internal emotional state of an individual ($v(t)$, $a(t)$) and these measurements is represented in Figure 4.3 through the indirect relations V_{phys} and A_{phys} . These physiological signals offer a great chance to study emotional reactions that do not need to be conscious, while

subjective assessment through reports always require the participant to be aware of their emotional state. On the other hand, the nature of their relations to the real underlying emotion (V_{phys} and A_{phys}) are not fully understood, and they are highly dependent on the individual and the experimental conditions. Our work on these physiological signals aims at advancing the understanding of these relations, improving techniques for monitoring internal emotions during online interaction.

4.3 Physiology of online interaction

We focus on the analysis of the physiological signals of CS3 and CS4, to present our analysis of the data from subjective assessments of emotions for the next section. The set of stimuli presented to the participants included IAPS pictures, and preselected threads with different emotional contents. The IAPS picture database contains values of valence, arousal and dominance for each picture, coming from survey studies in controlled setups. The pictures selected for the study were chosen to evenly span the range of valence, concentrating on neutral arousal. Each picture was shown for a fixed interval of 6 seconds, and baselines of physiological activity were calculated before each stimulus. The threads were preselected according to their positive, negative, or neutral valence content. We focus on exploring the connection between the physiological manifestation of valence states and their evolution while visualizing images and reading threads.

4.3.1 The Zygomaticus peak

Some of the physiological signals recorded during the experiments belong to the sympathetic nervous system, which usually behaves in a non-stationary manner as a fast response to external stimuli. This is the case for the smiling reaction, quantified by the EMG activity of *Zygomaticus Major* (Zyg). Figure 4.4 shows the typical reaction of $Zyg(t)$ to the visualization of a picture with high positive valence (left), compared with the case in which the participant did not show any smiling reaction (right). The time series $Zyg(t)$ on the left has a much larger scale than the one on the right (40 vs 0.6), but this is not present as a stable increase. $Zyg(t)$ shows a clear peak, suggesting that this signal represents a nonlinear response.

Traditional metrics used in psychophysiology are time averages and standard deviations, which assume that the signal is stationary. Inspired by reactions like the one shown in the left panel of Figure 4.4, we propose an exponential regression model to better estimate the

nonstationary properties of $Zyg(t)$:

$$Zyg(t) = \begin{cases} z_0 e^{(\beta_1(t-\Theta))} & t \leq \Theta, \\ z_0 e^{(\beta_2(t-\Theta))} & t > \Theta. \end{cases} \quad (4.1)$$

This model has an exponential growth (if $\beta_1 > 0$) until it reaches the maximum value of the signal, located at $t = \Theta$ with $z(\Theta) = z_0$. After the peak, there is an exponential decay if $\beta_2 < 0$. The values of z and Θ are approximated from the data, and the growth and decay values are calculated through nonlinear regression. The red lines of Figure 4.4 show the regression results. In the left panel, we appreciate that the model regression successfully fits the peaked signal, explaining a large fraction of its variance. The right panel illustrates the case of the absence of a peaked response, which leads to a regression result that tries estimate a peak from the noise.

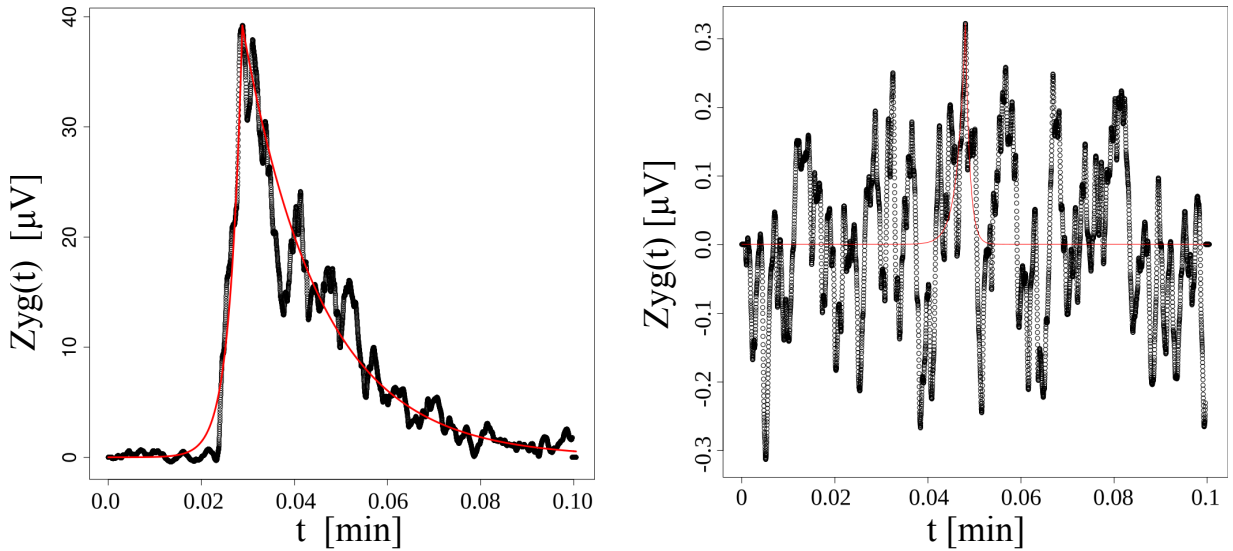


Figure 4.4: Examples of time series of the *Zygomatikus Major* activity of one participant while viewing two different IAPS pictures. In red, regression line as estimated from Equation (4.1). The example on the left shows a clear peaked activity (high R^2) while the example of the right does not (negative R^2).

Not all the physiological responses to IAPS pictures are like the one showed on the left panel of Figure 4.4, and our model helps us to find when they have distinguishable peaks. We can filter which responses had a significant peak through the regression quality of the model, which we use in the following to find improved results on the relation between Zyg and the valence of shown pictures.

4.3.2 Response to pictures

Using our peak model, we classified the $Zyg(t)$ responses according to the R^2 value given by the fit of Equation (4.1), classifying them as *peaked* if $R^2 > 0$ and *not peaked* otherwise. 20% Zyg responses to IAPS pictures were classified as *peaked* in CS3, and 12.3% in CS4. We also applied the same model to the rest of the signals (Cor, HR, SC), finding that not even 5% could be considered as peaked. Visual inspection of these other signals confirmed that stationary approaches are more appropriate, leaving this peak analysis for the Zyg response to IAPS pictures.

In these experiments, each picture was presented for a fixed interval of 6 seconds, and it has an assigned value of valence v_{pic} in the scale $[1, 9]$, being 5 the neutral valence. We analyzed the relation between the model parameters and the valence of the shown picture. We found that the most important parameter is z_0 , which corresponds to the maximum value in the peak regression. Figure 4.5 shows the relation between z_0 and v_{pic} for the cases that were classified as *peaked* (red) compared to the maximum value of $Zyg(t)$ when it was classified as *not peaked* (black).

For both experimental setups, the value of z_0 for *peaked* responses is larger than the maximum value when the time series is *not peaked*, serving as a check that no peak detection is fitting white noise like the one shown in the right panel of Figure 4.4. The model highlights the existence of smiling responses to slightly negative pictures ($v_{pic} \in [2, 3]$), compared to the responses to neutral pictures ($v_{pic} \in [5, 6]$). Smiling during the presentation of positive pictures increased in both studies, but this was more evident in CS3 than in CS4, where it decreased for the highest valence. These results support the hypothesis of a nonlinear dependence between valence and Zygomaticus Major activity.

In addition, we systematically analyzed the stationary properties of HR, SC and Cor, finding that the mean frowning activity $\langle Cor_{pic} \rangle$ has a large increase for very negative images in both studies, as shown in Figure 4.6. Both studies show a nonlinear relation between valence and $\langle Cor_{pic} \rangle$ for large valence that might depend on the selected pictures, opening the discussion and asking for more empirical results.

4.3.3 Response to threads

In addition to IAPS pictures, experiment participants of CS3 read threads with a preclassified emotional content. The usual reading time of the threads was too long to apply the peak model of Equation (4.1), as more than one peak could appear while reading a single thread. Threads were presented in a random order, and participants provided subjective

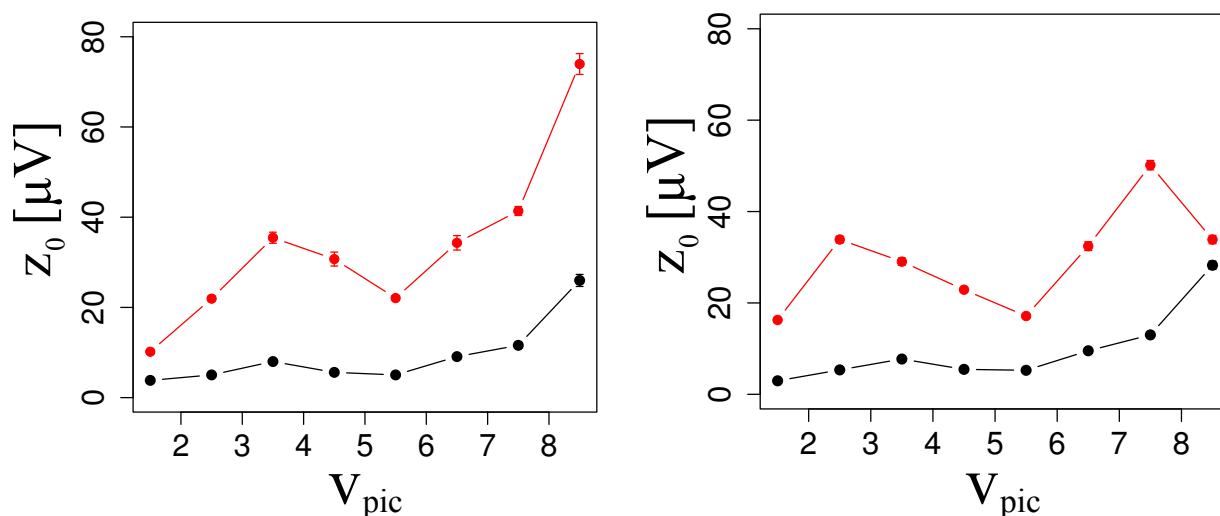


Figure 4.5: In red: mean value of the parametric estimation of z_0 in CS3 (left) and CS4 (right) versus v_{pic} when the Zyg response is classified as peaked. Error bars show standard error around the mean value, and are often difficult to see compared to symbol size. For comparison, black lines show the mean maximum value of Zyg versus v_{pic} when the Zyg response was classified as not peaked.

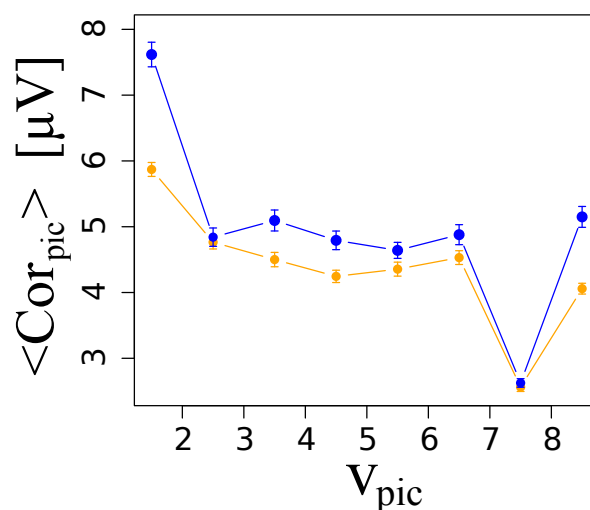


Figure 4.6: Dependence between mean Corrugator activity $\langle Cor_{pic} \rangle$ and valence of stimulus v_{pic} for CS3 (blue) and CS4 (yellow). Points show the average over valence values and error bars the standard error.

reports of their valence and arousal between threads. This data offers us the chance to explore the physiological manifestation of valence changes, which are triggered by reading threads with a given valence polarity.

We focused on the relation with physiological signals and the valence values before and

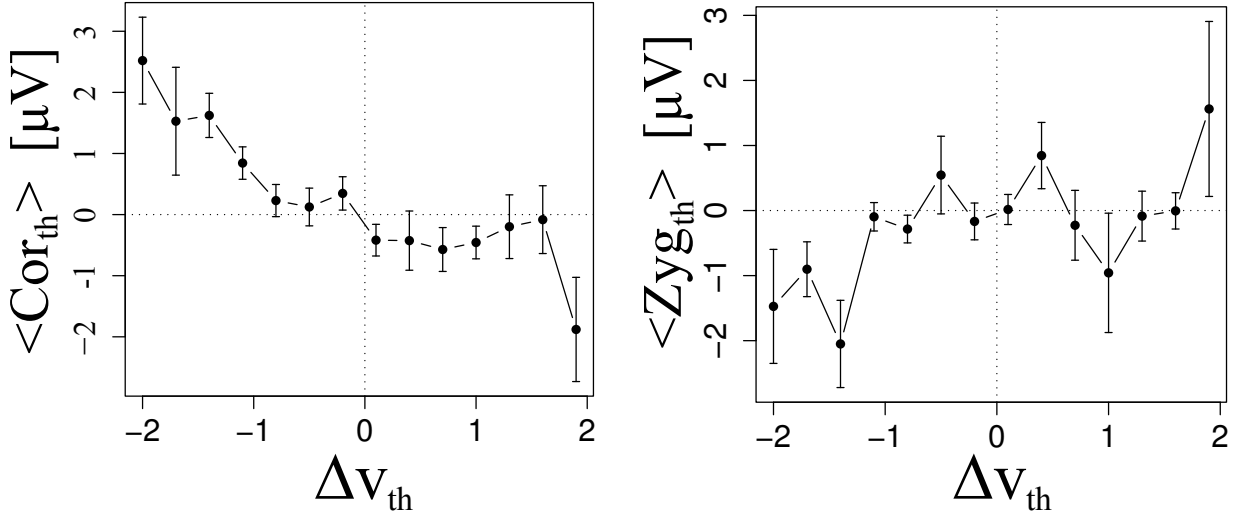


Figure 4.7: Mean activity of Corrugator Supercilii $\langle Cor_{th} \rangle$ (left) and Zygomaticus Major $\langle Zyg_{th} \rangle$ (right) during thread reading in CS3 versus valence change during the reading Δv_{th} . Error bars represent standard error over the mean of the binned values of $\langle Cor_{th} \rangle$ and $\langle Zyg_{th} \rangle$.

after reading a thread, as well as with the difference between both, through Spearman's correlation coefficients. The detailed results are reported in Appendix C, reporting here the most significant findings. The mean activity of the Corrugator Supercilii, $\langle Cor_{th} \rangle$, is negatively correlated with the valence change while reading the thread Δv_{th} ($\rho = -0.32, p < 10^{-9}$). In the left panel of Figure 4.7 we can see the relation between $\langle Cor_{th} \rangle$ and Δv_{th} , where the mean Corrugator activity is indistinguishable for changes in valence in the range $[-1, 1]$, but for stronger changes the frowning activity increases significantly if negative, and decreases if positive. These correspond to changes of valence polarity, when the valence of the individual passed from positive to negative and vice versa.

A linear regression with the $\langle Cor_{th} \rangle$ as dependent variable and the valence reports before v_{pre} and after reading the thread v_{post} shows the following results:

	Estimate	sd	t value	$Pr(> t)$
Intercept	0.1985	0.1099	1.807	0.0716
v_{pre}	-0.7313	0.1604	-4.559	7.02e-06
v_{post}	0.7396	0.1665	4.441	1.18e-05

Table 4.1: Regression results for the mean Cor activity versus the valence reports before and after reading the threads. $R^2 = 0.1033, F(2, 368) = 21.2, p < 10^{-8}$

The regression results show a significant dependence of $\langle Cor_{th} \rangle$ on the valence value before and after reading the thread, with similar magnitude but different sign. This, combined with the results of Figure 4.7 leads us to conclude that the mean Cor activity is a valid indicator of strong valence changes during the reading of a thread.

The second most relevant physiological signal was the mean response in the *Zygomatikus Major*, with a Spearman's correlation coefficient of $\rho = 0.121 (p < 0.004)$. In the right panel of Figure 4.7, we can see that the mean value of the *Zygomatikus Major* $\langle Zyg_{th} \rangle$ allows the detection of large decreases in the valence, but the pattern of relation with valence is not as clear as for the peak values detected while showing IAPS pictures. This result could be improved with an extension of our peak model, including the possibility of the detection of multiple peaks within the same time series. We open this question for further research, as additional assumptions are necessary for such statistical model.

4.3.4 Physiological memory of thread experience

The regression results of Table 4.1 suggest the existence of certain physiological memory of the previous emotional state, which can be noticed when it changes. A very recent study of brain activity, measured through event-related potentials (ERP), studied possible memory effects during the visualization of emotional images (Schupp *et al.*, 2012), finding no evidence of emotional memory in picture visualization. On the other hand, previous works on EMG suggest the opposite (Bradley *et al.*, 1996), supporting the hypothesis of the existence of emotional memory in the activity of Corrugator Supercilii. We extend these works by analyzing the emotional memory while reading threads. We calculate the autocorrelation function $C(\Delta t)$ of the four physiological signals while reading threads. This autocorrelation measure is the same one as used in Section 3.3.3, which we applied in the context of the activity of an online community. However, we do not expect to find scaling properties of $C(\Delta t)$ in physiological signals, as they do not come from the interaction of many agents on large time scales. We use this tool to measure the timespan of coherent behavior in physiological activity, which would correspond to the length of physiological memory when reading a thread.

For each signal $sig \in [Cor, Zyg, HR, SC]$, and each thread polarity $p \in [-, 0, +]$, we compute the autocorrelation function of all the activity of sig during the reading of threads of polarity p , which we denote as $C_{sig}^p(\Delta t)$. Fig 4.8 shows the autocorrelation functions of a participant for *Cor* during the reading of negative threads (left) and during the reading of positive threads (right). For each Δt , the value of the corresponding Pearson's correlation coefficient has a confidence interval at the 95% level. The dashed horizontal lines of

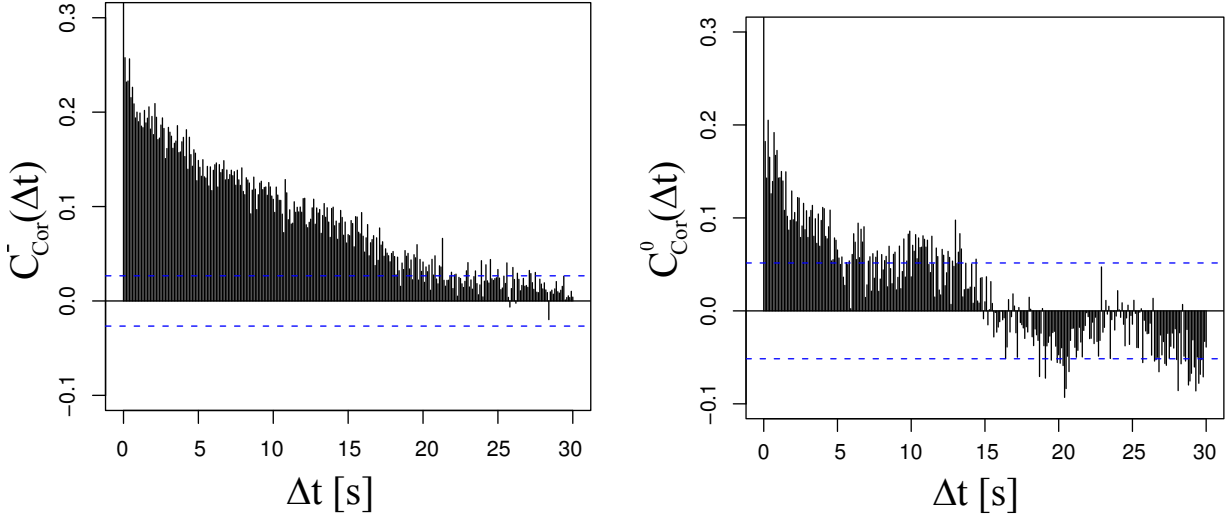


Figure 4.8: Examples of the autocorrelation function $C_{Cor}(\Delta t)$ of the Cor signal of a participant during the visualization of negative threads (left) and neutral threads (right). Dashed bars show the limit of significance, being reached in the neutral case much earlier than in the negative case.

Figure 4.8 shows the limit value below which any value of $C_{Cor}(\Delta t)$ is not statistically significant. These examples show a clear difference in the time when the autocorrelation function reaches a value that cannot be considered larger than 0. For the case of a neutral thread, $C_{Cor}^0(\Delta t)$, the autocorrelation disappears in less than 5 seconds, leaving only spurious correlations afterwards. The opposite case is present during the reading of negative threads, $C_{Cor}^-(\Delta t)$, which takes more than 15 seconds to reach this threshold. This moment when the autocorrelation vanishes is an indicator of the memory of the physiological signal, measuring the time length of the coherent behavior of the signal under different emotional stimuli. We denote the first time when $C_{sig}(\Delta t)$ does not have a significant value as m_{sig} , i.e. the memory length of sig measured in seconds.

If there is a coherent memory on the activity of a physiological signal, we expect it to be longer while the participant is reading threads containing emotional expression. To test this hypothesis, for each participant and type of stimulus, we compute the individual memory length of each physiological signal m_{sig} while reading positive, negative, and neutral threads. Figure 4.9 shows the distribution of these values for the four different signals and the three thread polarities. A visual inspection reveals that, at least for the case of Cor and SC, the probability of having a longer memory is larger under the presence of emotionally charged threads (positive and negative) than in neutral ones.

We verify these observations with nonparametric tests on the difference between the distributions shown in Figure 4.9. As these do not show any signs of normality, we choose to

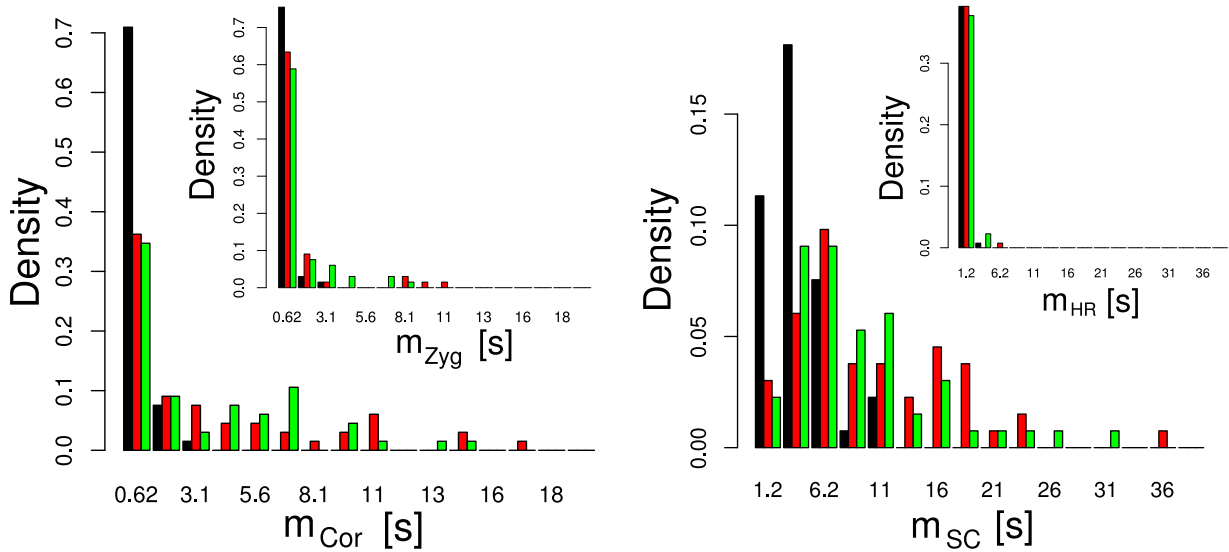


Figure 4.9: Rescaled histograms of the memory length of the four signals while reading neutral threads (black), positive threads (green) and negative threads (red).

run Mann-Whitney U tests and to compare their medians in order to test whether one is above the other. These tests revealed that, for Cor, Zyg and SC signals, we can consider that the memory while reading emotionally charged threads is longer than when reading neutral threads ($p < 10^{-8}$, $p < 10^{-6}$, and $p < 10^{-8}$ respectively), but this is not the case for HR ($p > 0.05$). The hypothesis that the memory for positive threads is the same than for negative threads could not be rejected for any of the physiological signals ($p = 0.61$, $p = 0.43$, $p = 0.54$, and $p = 0.64$), indicating that this method is not able to distinguish differences in physiological memory while reading threads of different valence polarity.

Comparing the medians, we can get an idea of the size of the distance between the distributions that passed the previous test. The median of m_{Cor} was 1.2 seconds for neutral threads, while it was 18.2 and 22.2 seconds for negative and positive threads respectively. The median of m_{Zyg} was 2.3 seconds for neutral threads, 4.4 seconds for negative threads, and 4.7 seconds for positive threads. The median of m_{SC} was 37.5 seconds for neutral threads, while it was 82.4 and 70 seconds for negative and positive threads respectively. This longer memory of the Skin Conductance was expected, as this response is regulated by the parasympathetic nervous system, implying slower adaptation to the environment. These results show the time range of the physiological manifestation of the emotions elicited by reading an online discussion, which is significantly larger when reading an emotional discussion than when reading a neutral one.

These results could be used in industrial applications that aim at estimating user emotions from physiological signals, for example measuring skin conductance through the mouse,

or smiling through facial recognition. In addition, future improvements can provide useful insights in the unconscious nature of emotion dynamics, avoiding the errors introduced by participant reports.

4.4 Emotion eigendynamics

Our work in the previous section provided new insights on the physiological manifestation of the changes of emotional states, focusing on those triggered by the visualization of pictures and the reading of online discussions. During the experiments outlined in Section 4.2.1, participants also reported subjective assessments of their emotional states (Q), which provide further data on the way emotions evolve during online communication. In particular, we can combine this reports across different experimental setups, using our findings to support a unified agent-based model of emotion dynamics.

4.4.1 The emotion relaxation equation

In their original study, Kuppens et. al. (Kuppens *et al.*, 2010) studied the dynamics of individual emotions in everyday life. The method for the gathering of experimental data was based in participant reports at random moments during waking hours, asking for valence and arousal. In their exploration of the dynamics of emotions, they estimated the existence of a *home base* and an *attractor strength* as the key components of emotion dynamics. The purpose of this section is to incorporate their analytical results into our modeling framework, to define the way emotions evolve when users are not interacting through the Internet. The dataset provided by Peter Kuppens contains data from two independent studies, in which the emotional states of the participants were sampled with different frequencies. We focus in their second study, DynAffect2, which contained data sampled at the highest frequency. This dataset is closer to capture the dynamics of core affect (Russell, 1980), rather than the dynamics of mood (Yik *et al.*, 1999, 2011), which work at a slower speed.

Dynaffect2 contains no information about user interaction, but covers most of the waking hours of the participant. We study the eigendynamics and stochastic component of the changes in emotional states, as part of the framework presented in Chapter 1. In this context of individual dynamics, the equations for the valence and arousal changes over

time can be expressed as:

$$\begin{aligned}\dot{v}_i(t) &= \gamma_v \cdot [b - v_i(t)] + A_v \xi \\ \dot{a}_i(t) &= \gamma_a \cdot [d - a_i(t)] + A_a \xi\end{aligned}\quad (4.2)$$

where γ_v and γ_a are the relaxation parameters towards the emotional ground state $(v_i, a_i) = (b, d)$ under a stochastic influence of amplitudes A_v and A_a .

Each participant of DynAffect2 generated a time series of values for $(v_i(t), a_i(t))$, with different time increments between timestamps. To estimate the parameter values of the emotion dynamics, we use the solution of the stochastic differential equations 4.2 in time as explained in (Oravecz *et al.*, 2011). This internal dynamics follow a diffusion process belonging to the family of Ornstein-Uhlenbeck processes (Doob, 1942), evolving as a normal distribution with mean that approaches the ground state, and standard deviation according to the amplitude of the stochastic component (Gardiner, 2004). Thus, after a time Δt , the valence would follow:

$$v(t + \Delta t) \sim \mathcal{N}\left(b + [v(t) - b] \cdot [e^{-\gamma_v \Delta t}], \frac{A_v}{2\gamma_v} [1 - e^{-2\gamma_v \Delta t}]\right) \quad (4.3)$$

The empirical estimations of mean and standard deviation of the valence distribution allow us to test the validity of this equation, and to estimate the parameter values. From the mean part of Equation (4.3), we can provide estimators of the parameters b , and γ_v through the expected valence changes:

$$\frac{v(t + \Delta t) - v(t)}{\Delta t} = [v(t) - b] \cdot [e^{-\gamma_v \Delta t} - 1] \quad (4.4)$$

For the case of the stochastic component, its amplitude A_v can be jointly estimated with γ_v through the relation:

$$A_v = \frac{2\gamma_v \sqrt{\epsilon_v^2(\Delta t)}}{1 - e^{-2\gamma_v \Delta t}} \quad (4.5)$$

where $\epsilon_v(\Delta t)$ is the standard deviation of $v(t + \Delta t)$ in the data. The exact estimation of A_v requires a large amount of high resolution data, because it requires a valid estimation of the standard deviation per time increment $\epsilon_v(\Delta t)$. While DynAffect2 can be considered as providing enough data for this kind of approximation, the datasets of online perception, CS2 and CS3, do not allow a valid estimation of A_v at this level. As we want to unify the results of these different datasets, we choose to have a common approach to all of them. Therefore, we approximate A_v under the assumption of a large enough Δt , which leads to $e^{-2\gamma_v \Delta t} \rightarrow 0$. We can estimate the amplitude of the stochastic component through

$$A_v = 2\gamma_v \sqrt{\epsilon_v^2} \quad (4.6)$$

where ϵ_v^2 is the standard deviation of v across all observations. Note that this same formulation can be applied to the arousal dynamics, exchanging v for a in Equations (4.3 - 4.6).

4.4.2 Eigendynamics parameter estimation

The estimated parameter values and the regression quality over the intraday emotional changes of DynAffect2 are summarized in Table 4.2. R_v^2 and R_a^2 measure the amount of variance explained by the deterministic dynamics, and their values around 0.2 show us that relaxation towards the baseline is measurable from the data, but other components of emotion dynamics are present. This was expected from this longitudinal data, including all the external influences in the stochastic component of this analysis. In addition, $R_v^2(\xi)$ and $R_a^2(\xi)$ measure the quality of the assumption that the stochastic component is composed of white noise, estimated through a fit of the regression error to a normal distribution. Both $R^2(\xi)$ are above 0.9 suggesting that this model is providing a good approximation for the internal agent eigendynamics.

b	γ_v	A_v	R_v^2	$R_v^2(\xi)$	d	γ_a	A_a	R_a^2	$R_a^2(\xi)$
0.195	0.03	0.0324	0.207	0.909	-0.098	0.027	0.0322	0.186	0.96

Table 4.2: Parameter estimations and R^2 values for valence and arousal in DynAffect2. All parameter estimations of b , γ_v , d , γ_a have $p < 10^{-10}$, and their time scale is measured in min.

For each one of the emotion dimensions, the regression leading to these estimations has two dependent variables: the previous value, and the time increment Δt . The properties of the relaxation dynamics explained here can be seen in Figure 4.10, where the average change over time is presented versus the previous value. The red lines show the analytical solutions given by the parameter estimations from table 4.2, for $\Delta t = \langle \Delta t \rangle$ in Equation (4.4).

For both valence and arousal, we can notice the existence of nonzero baselines close to $v = 0.2$ and $a = -0.1$, which correspond to the ground state $(v_i, a_i) = (b, d)$, present in the parameter estimation of Table 4.2. The fit of the linear function of the attraction to the baselines is better for moderated values, in $[-0.5, 0.5]$. The reason lies on the fact that the original dataset provides more points in this range than in extreme values, and our regression method gave the same weight to each experimental observation. With more data available on very high and low values, it would be possible to reweight the

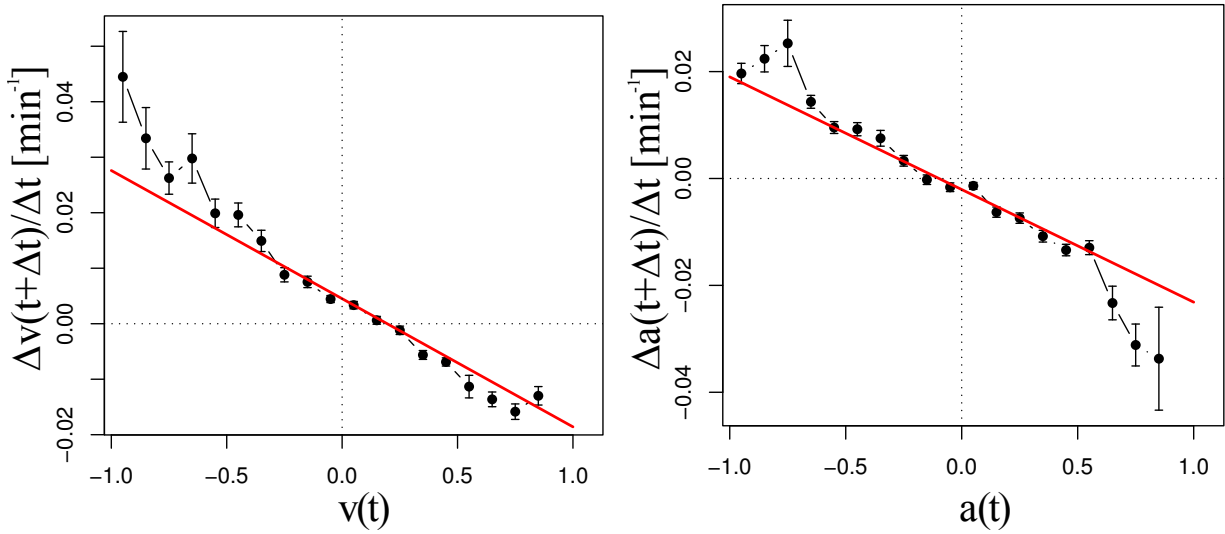


Figure 4.10: Valence and arousal change diagrams for DynAffect2. The mean change per time unit in valence and arousal is inversely proportional to its previous value. Error bars show standard error of the mean change in the empirical data. The red line shows the solution of Equation (4.4) for $dt = \langle \Delta t \rangle$.

regression in order to increase the leverage of extreme values, which is left open for future experiments.

It is still left to ask if we can observe any other deterministic influences in this data. The joint distribution of valence and arousal, shown in the kernel density plot of Figure 4.11, has two local maxima, one with a positive arousal and one with a negative one. Previous studies on this same dataset showed the dependence of the arousal on daily patterns (Oravecz *et al.*, 2011). This suggests the existence of nonlinear deterministic dynamics on the arousal, and an extended model accounting for them could increase the $R_a^2(\xi)$ of the model. This kind of measures have already been studied in the context of expressions of mood (Golder and Macy, 2011), but our modeling framework does not include periodic influences. In most of the cases, the geographic location of a member of an online community cannot be known, making it impossible to introduce this daily time pattern in our analysis of online data.

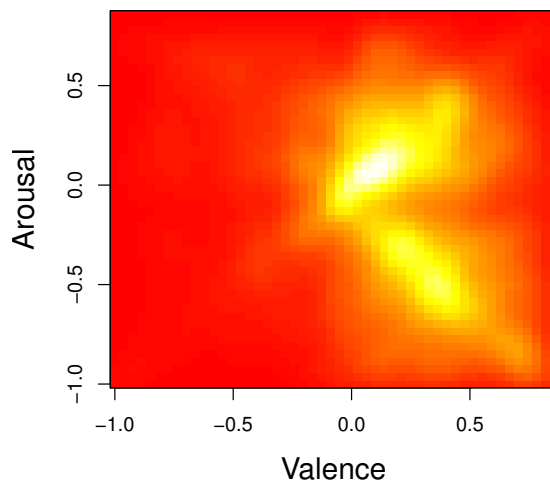


Figure 4.11: Kernel density plot of the reported values of valence and arousal in DynAffect2. Density color ranges from red (minimum) to white (maximum) in a grid of 50x50. Two attractors can be noticed in the positive and negative arousal parts of positive valence.

4.5 Emotion dynamics under perception of online content

4.5.1 Emotional states after reading forum discussions

In the experiments CS2 and CS3, the participants were reading a randomly ordered series of threads with preassigned emotional content. After reading each thread, each user reported their experienced valence and arousal. This succession of stimuli allows us to reconstruct the changes induced in valence and arousal after reading each kind of stimulus (positive, negative or neutral), in order to know how online discussions influence the internal state of a user. These influences, represented as P_{dyn} in Figure 4.1, compose the perception dynamics of online emotional interaction.

To have an idea of the influence on emotions that a thread can have, we divided the final reports by the polarity of the thread, and we produced kernel density plots. The result is shown in Figure 4.12, for both experimental setups and the three kinds of threads. An initial visual inspection shows that negative threads shift the mean value towards negative valences, while positive threads do the exact opposite. Neutral threads keep values concentrated around 0 valence, as expected from the preclassification of the presented threads.

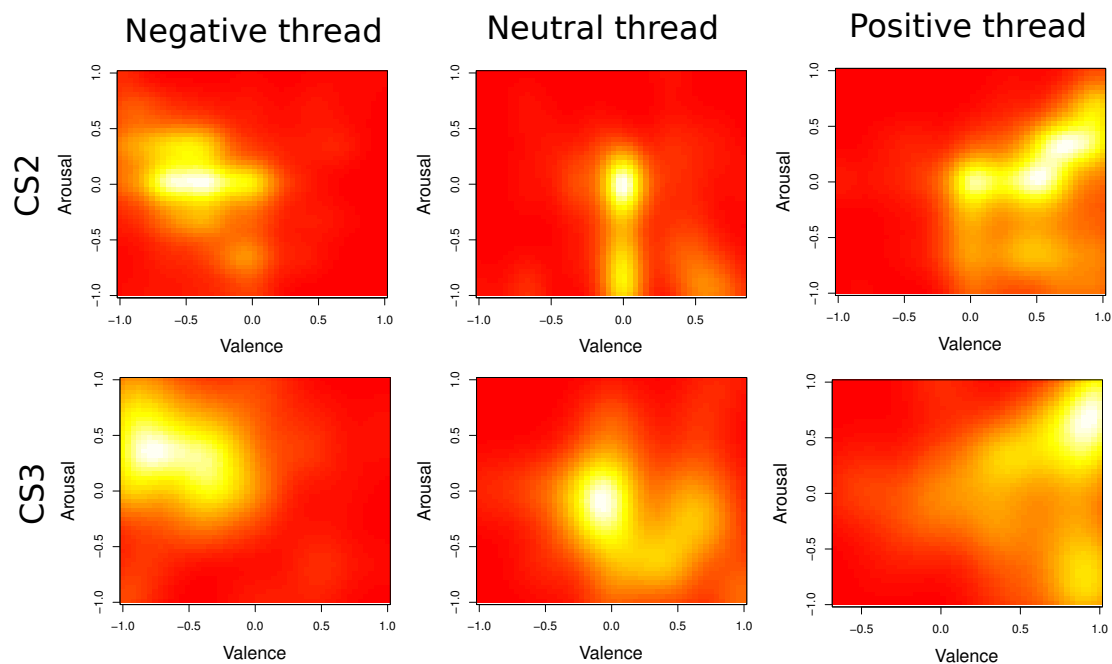


Figure 4.12: Kernel density plots of valence and arousal values for participant reports in CS2 (top) and CS3 (bottom) after reading negative, neutral, and positive threads. Density color ranges from red (minimum) to white (maximum) in a grid of 50x50.

The spread of arousal while reading positive and negative threads follows the patterns predicted by the theory of appetitive attention, which proposes that under the presence of external stimuli, arousal is correlated with the absolute value of the valence (Bradley, 2009). The reason for this pattern in emotions is grounded in biology: Organisms need to have proactive reactions to negative influences, such as threats, while they also need to keep their attention directed towards positive stimuli, such as food. This is in line with our results in language analysis and the relation between the emotions expressed by words and their information content, presented in Chapter 5.

Note that the histogram of valence and arousal values for neutral threads in CS2, shown in the second plot of the first row in Figure 4.12, shows a second peak with negative arousal. In CS2, participants were reading the threads from their computers at home, while in CS3 they were in a controlled setup in Jacobs University Bremen. The appearance of this negative arousal attractor might be related to the fact that participants were in a more natural environment in CS2, and they were able to get bored by reading a non-emotional thread. On the other hand, the experimental setup of CS3 could make them excited and mask the existence of this negative arousal pattern. In our empirical analysis of the dynamics of valence and arousal, we will include nonlinear factors to explore if

we can provide an explanation for this bimodality in the arousal dimension, testing the assumption made in the first exploration of the modeling framework in Chapter 1.

4.5.2 Changes in emotional states due to reading forum discussions

In CS2 and CS3, the time intervals between the participant reports were quite short in comparison to the ones in DynAffect2. The amount of time needed by the users to read a thread was heterogeneous, but not deviating as much as in DynAffect2. Data in CS2 and CS3 is rather scarce compared with DynAffect2, but it allows us to assume a constant Δt in our analysis. Instead of fitting the solution of the equations of emotion dynamics to the data, we used a Linear Mixed Model (Oravecz and Tuerlinckx, 2011), estimating the changes in the variables as a combination of a linear regression with a stochastic component. This has the advantage of not needing to have an explicit solution of the Ordinary Differential Equation driving the emotion dynamics, allowing us to test nonlinear equations of emotional influence. The main drawback is the inability to approximate the dynamics of individual participants, which would allow the estimation of the heterogeneity in the reaction to emotional content. As we do not have a large amount of data per participant, this would not be possible anyway, leaving the Linear Mixed Model as the best choice.

In our framework, we assumed that emotional content changed the internal state of the agents through an information field h , which aggregates the current state of such emotional information. We test this microdynamics through the equations:

$$\begin{aligned} dv(t)/dt &= \gamma_v \cdot [b - v(t)] + \mathcal{F}_v(h, v(t)) + A_v \xi \\ da(t)/dt &= \gamma_a \cdot [d - a(t)] + \mathcal{F}_a(h, a(t)) + A_a \xi \end{aligned} \quad (4.7)$$

In our experimental setup we only have three cases of h , corresponding to the fields generated by positive, negative and neutral threads. These threads are all of similar length, so we can assume that the levels of negative and positive information in the emotionally-charged cases are similar. Following our framework, we will test the functions \mathcal{F}_v and \mathcal{F}_a as polynomials of the form:

$$\begin{aligned} \mathcal{F}_v(h, v(t)) &= b_0(h) + b_1(h)v(t) + b_2(h)v(t)^2 + b_3(h)v(t)^3 \\ \mathcal{F}_a(h, a(t)) &= d_0(h) + d_1(h)a(t) + d_2(h)a(t)^2 + d_3(h)a(t)^3 \end{aligned} \quad (4.8)$$

We fit the model for polynomials of degree 3 for the different cases of h by regression on the datasets from CS2 and CS3. The superposition of the effects of Equation (4.8) and the

first term of Equation (4.7) produces the changes observed in the empirical data. Thus, we can generate estimator from polynomial regression of third order as:

$$\frac{\Delta v(t)}{\Delta t} = \sum_{i=0}^3 \hat{b}_i v(t)^i + A_v \epsilon \quad \frac{\Delta a(t)}{\Delta t} = \sum_{i=0}^3 \hat{d}_i a(t)^i + A_a \epsilon \quad (4.9)$$

where the different estimators for the polynomial coefficients provide values to estimate the parameters of Equation (4.8):

$$\begin{aligned} \hat{b}_0 &\simeq b_0(h) + \frac{b}{\gamma_v}, & \hat{b}_1 &\simeq b_1(h) - \gamma_v, & \hat{b}_2 &\simeq b_2(h), & \hat{b}_3 &\simeq b_3(h), \\ \hat{d}_0 &\simeq d_0(h) + \frac{d}{\gamma_a}, & \hat{d}_1 &\simeq d_1(h) - \gamma_a, & \hat{d}_2 &\simeq d_2(h), & \hat{d}_3 &\simeq d_3(h). \end{aligned} \quad (4.10)$$

Our results of Section 4.4 provide estimators for the parameters of the internal dynamics of emotions, shown in Table 4.2. We took $b = 0.195$, $\gamma_v = 0.03$, $d = -0.098$, and $\gamma_a = 0.027$, and estimated the values of the rest of the parameters of \mathcal{F}_v and \mathcal{F}_a through the relations of Equation (4.10). We calculated the values of the coefficients \hat{b}_i and \hat{d}_i through least squares regression of the polynomials of Equation (4.9), estimating the values of $b_i(h)$ and $d_i(h)$. The detailed regression results are given in Appendix C, comparing the results in both experimental setups.

For the valence, the regression results provided significant values in both CS2 and CS3 for $b_0(h)$ and $b_1(h)$, but these were not significant for orders above one. The estimation of A_v , given by Equation (4.6), kept between values of 0.097 and 0.144. Additionally, the R_v^2 of this regression indicated that our model explains more than 60% of the variance of emotion dynamics while reading threads. The case of arousal was similar, parameter estimations were significant and consistent across experiments for $d_0(h)$ and $d_1(h)$, but higher orders of the dynamics kept low significance. A_a kept values between 0.12 and 0.182, indicating a slightly larger variance in arousal than in valence dynamics. The R_a^2 has values between 0.34 and 0.5, suggesting that there are additional factors that influence arousal dynamics. This was expected, as the threads were preselected by their polarity in terms of valence, and not by the arousal expressed in the discussion.

The comparison between our regression results and the data can be observed in Figure 4.13 and Figure 4.14. We illustrate the dynamics of an emotion variable x by studying the function of its change per time unit $(\Delta x(t + \Delta t)/\Delta t)$ versus its current value $(x(t))$, under the different polarities of h studied here. This approach is the discrete time version of the one we used for the simulation results shown in Chapter 1, and the dynamics of emotional interaction in product reviews of Chapter 2 (e.g. Figure 2.20).

In Figure 4.13, error bars show binned data on participant emotion reports, and dashed lines shown the prediction of valence changes with parameters up to order 1, choosing $dt = \langle \Delta t \rangle$ to produce the plot. Both experiments show similar valence dynamics, with negative fixed points of the valence for the case of negative threads, and positive solutions for the case of positive threads. This means that the ground state of emotions is shifted towards more positive values while reading positive discussions, while this ground state switches to a point of negative valence while reading threads preclassified as negative. We can also observe that the slopes of the changes are similar for the cases of reading positive and negative threads, suggesting that the strength of this attraction is similar in both. For clarity reasons, the dynamics of valence for neutral threads are not shown in Figure 4.13, but they are shown in an aggregated case explained below.

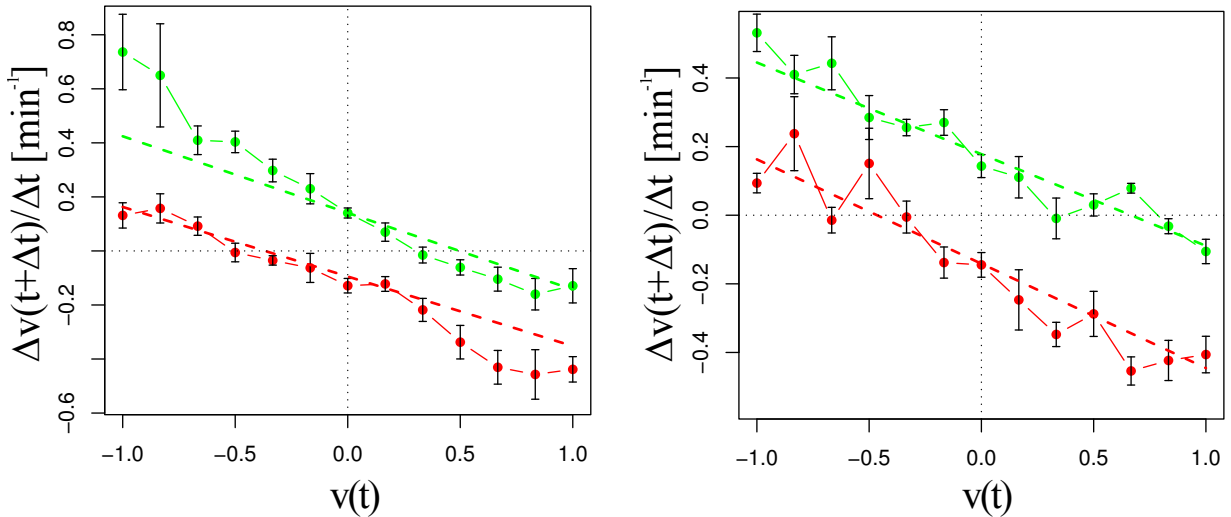


Figure 4.13: Mean change in valence per time unit while reading threads versus previous reported valence in CS2 (left) and CS3 (right). Error bars represent standard error. Red bars indicate values after reading negative threads, green bars after reading positive threads. Lines show the changes predicted by Equation (4.7) with $dt = \langle \Delta t \rangle$.

The changes induced in arousal due to reading threads are clearly different to the case of valence. Figure 4.14 shows the changes in arousal versus its previous value while reading the three types of threads. For both experiments, the arousal dynamics have solutions close to 0 while reading positive and negative threads, while this solution seemed to be negative while reading neutral threads. The dynamics of arousal is very similar when perceiving emotional content, regardless of its valence polarity, which was one of the assumptions of our modeling framework. While reading neutral content, the influence in arousal is much lower or it even has a negative solution that drives agents away from online interaction,

which we analyze in Section 4.6.

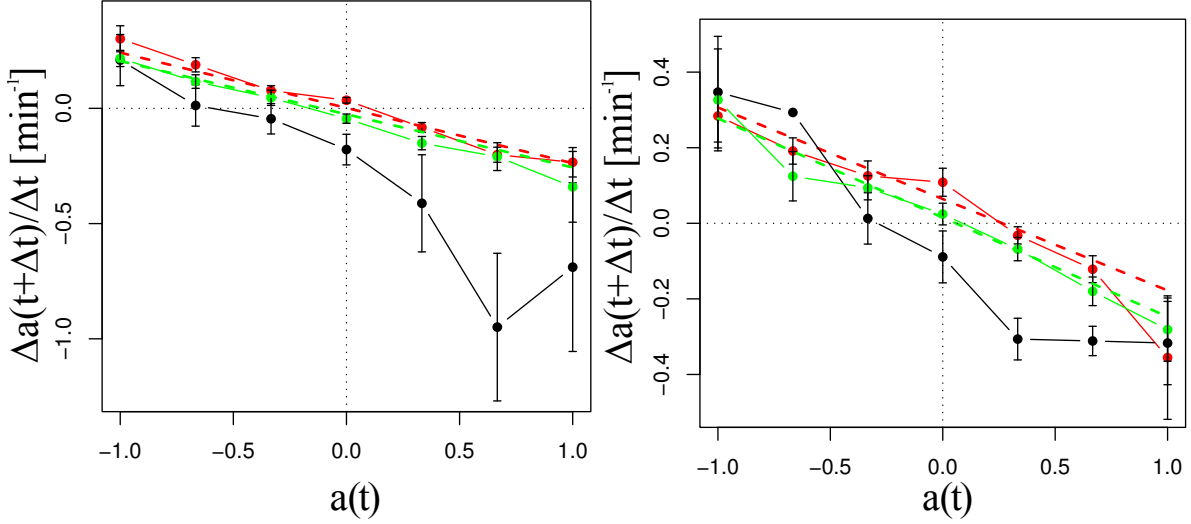


Figure 4.14: Mean change in arousal per time unit while reading threads versus previous reported arousal in CS2 (left) and CS3 (right). Error bars represent standard error. Red bars indicate values after reading negative threads, green bars after reading positive threads, and black bars after reading neutral threads. Lines show the changes predicted by Equation (4.7) with $dt = \langle \Delta t \rangle$.

4.5.3 Parameter estimation for agent-based models

One of our purposes for studying individual emotion dynamics is providing empirical support to agent-based models of emotions. We studied the properties of these dynamics in the data from two independent studies, one in which the participants were in a controlled setup (CS3), and another in which they were reading the threads at home (CS2). Separated statistical analysis on both datasets revealed the existence of certain dynamics, in particular supporting linear dependencies of previous to future states. To make future agent-based models of emotion dynamics as realistic as possible, we combined both datasets in one analysis. This way we produced parameter estimations for a minimal version of emotional influence (Equation (4.8)), up to order 1.

Our estimations of the parameters of valence and arousal dynamics are shown in Table 4.3, providing significant values for all the coefficients. In this combined analysis, our previous observations of the solutions of valence and arousal dynamics hold. Reading positive threads drive the valence dynamics towards positive values, with a factor 0.148. The opposite is present for negative threads, while the case of neutral threads shifts the

baseline of valence by a value close to 0. Across all the three polarity cases, the strength of this attraction was very similar, with $b_1(h)$ around -0.265 . The estimated arousal parameters show biases over the baseline for a value close to 0 in both positive and negative threads, while this is driven towards slight negative values $d_0(h) \sim -0.1$ while reading neutral threads. This shows how non-emotional content can lead arousal to points below its natural baseline, as Internet users have some tendency to get bored or to lose attention.

h	$b_0(h)$	$b_1(h)$	A_v	R_v^2	$R_v^2(\xi)$
Pos	0.148***	-0.273***	0.122	0.637	0.663
Neu	0.041**	-0.265***	0.116	0.644	0.457
Neg	-0.118***	-0.264***	0.125	0.608	0.706
h	$d_0(h)$	$d_1(h)$	A_a	R_a^2	$R_a^2(\xi)$
Pos	-0.014*	-0.190***	0.151	0.351	0.774
Neu	-0.095***	-0.208***	0.14	0.447	0.517
Neg	0.015**	-0.193***	0.132	0.444	0.853

Table 4.3: Parameter estimations and R^2 values for valence and arousal on the combined dataset of CS2 and CS3. * $p < 0.1$, ** $p < 0.001$, *** $p < 10^{-10}$

These observations can be appreciated in Figure 4.15, which summarizes the combination of datasets into the same dynamics of valence and arousal. Valence dynamics show the solutions mentioned above, with the neutral one close to 0. Arousal dynamics show the negative solution present in non-emotional threads, which can be seen as a version of the negative solution we found through the analysis of the equations driving arousal dynamics in Chapter 1, shown in Figure 1.3. Nevertheless the conditions for the existence of a stable negative solution of the arousal are different in this analysis than in our first approximation. Our theoretical approach proposed the presence of this state under high h , while in this empirical analysis we find it for the reading of neutral threads, i.e. low h . This finding will allow future versions of the model to have more realistic arousal dynamics, inspired in the empirical findings presented here.

The R^2 coefficients of Table 4.3 give us an idea of the quality of our model when compared to empirical data. One of the advantages of an agent-based approach is that it allows us to implement computational equivalents, comparing empirical and simulation results. Not only we can provide a quantitative explanation for the observed data, but we can reproduce its behavior in simulations, or even apply it in the field of affective computing (Rank, 2010). We verify the validity of our model by setting up simulations of the behavior of individual agents under the influence of the three kinds of threads, with valence and arousal driven by Equation (4.7) with the parameter estimations of Table 4.3. We simulated 100000 agents

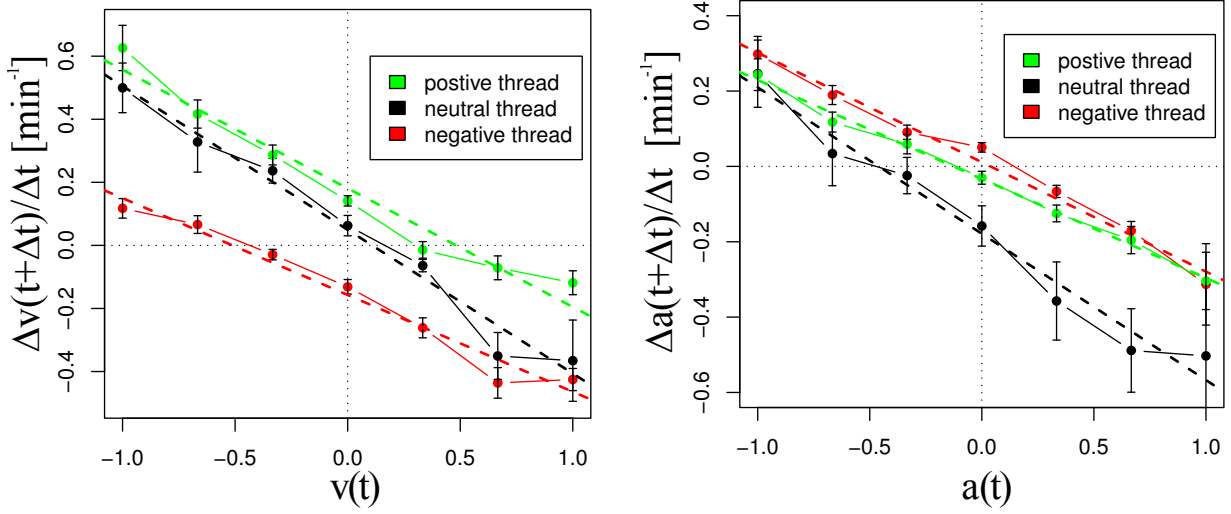


Figure 4.15: Mean change per time unit in valence (left) and arousal (right) while reading threads versus previous reported valence and arousal in the joint data of studies CS2 and CS3. Error bars represent standard error. Red bars indicate values after reading negative threads, green bars after reading positive threads, and black bars after neutral threads. Lines show the changes using the parameter estimations of Table 4.3.

per type of thread, recording their final valence and arousal values. The comparison between the subjective reports of emotions and the final emotional states of simulated agents are shown in Figure 4.16, binned with the same granularity as the empirical data.

These simulation results show that, in general, there is a good level of agreement between the emotions reported by experiment participants, and the artificial agents we implemented. The positive and negative biases in emotions due to reading threads are present, but the apparent bimodality of arousal in neutral threads is due to the binning used in the histograms. Some disagreements are present in this analysis: Empirical data shows higher bins for $a = 0$ in neutral and negative threads, and for $v = 0$ in neutral threads. This tendency, if explicitly introduced in valence and arousal dynamics, would improve the quality measures of our model fit. However, this bias towards answering in the middle might have their origin in the design of the experimental setup, and future experiments should combine different response scales to control for possible experimental biases. On the other hand, there can be an additional explicit tendency towards 0 valence and arousal, but so far we lack any plausible psychological explanation beyond the exponential relaxation already included in our model.

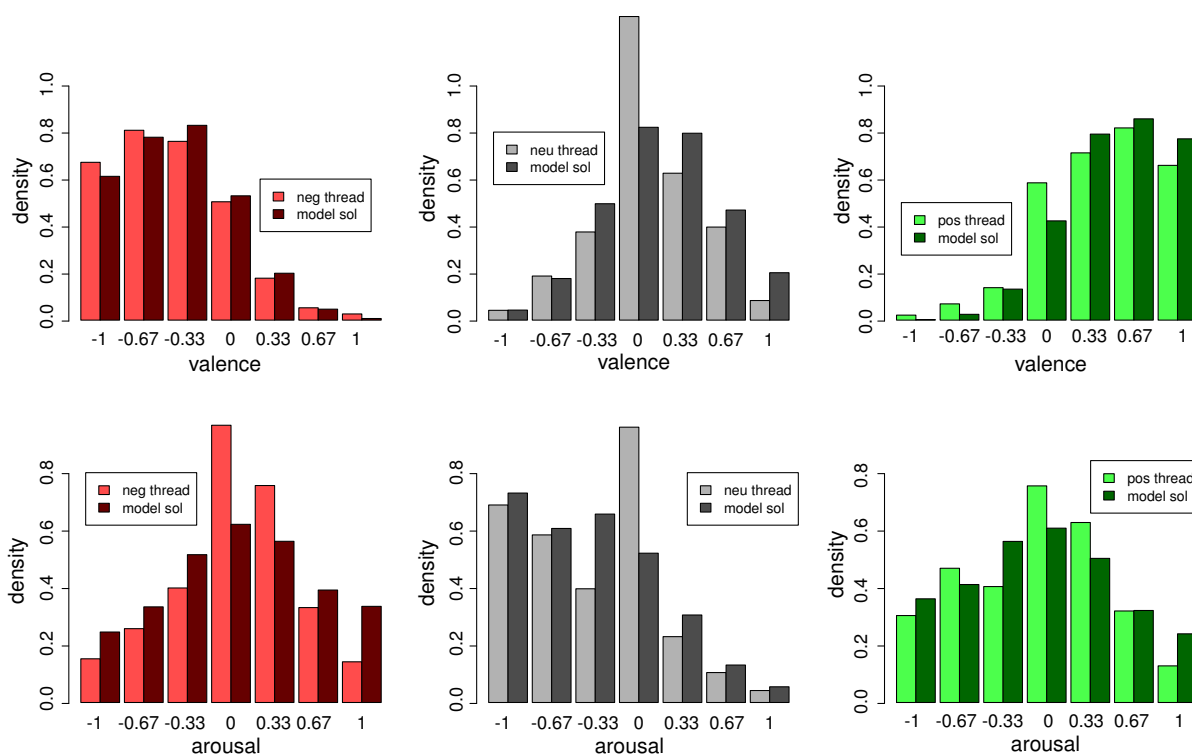


Figure 4.16: Distributions of valence (top row) and arousal (bottom) for participant reports while reading threads (light bars) versus agent simulations with our parameter estimations (dark bars).

4.6 Antecedents and consequences of expression

In the reports of CS2 and CS3, participants were asked about their intention to participate in a given discussion. One of the main assumptions of our general framework is that arousal is the driving force behind user participation in online discussions. By looking at the relation between the participant reports about arousal and willingness to participate, we find that indeed user participation is heavily influenced by his or her arousal, as shown in Figure 4.17. Up to a threshold value of the arousal, users share the same tendency not to participate at all. Above such value, participation probability increases proportionally to arousal. A Wilcoxon test shows that a participant with an arousal above 0 has a probability of participation that is 0.3 higher than one with negative arousal in CS3 and 0.16 higher in CS2.

There is a discrepancy between CS2 and CS3 in the maximum value of the arousal, leading to a lower probability of posting in CS2. This can be related to the lower amount of data for high arousal states in CS2, and the differences between experimental setups. Nevertheless,

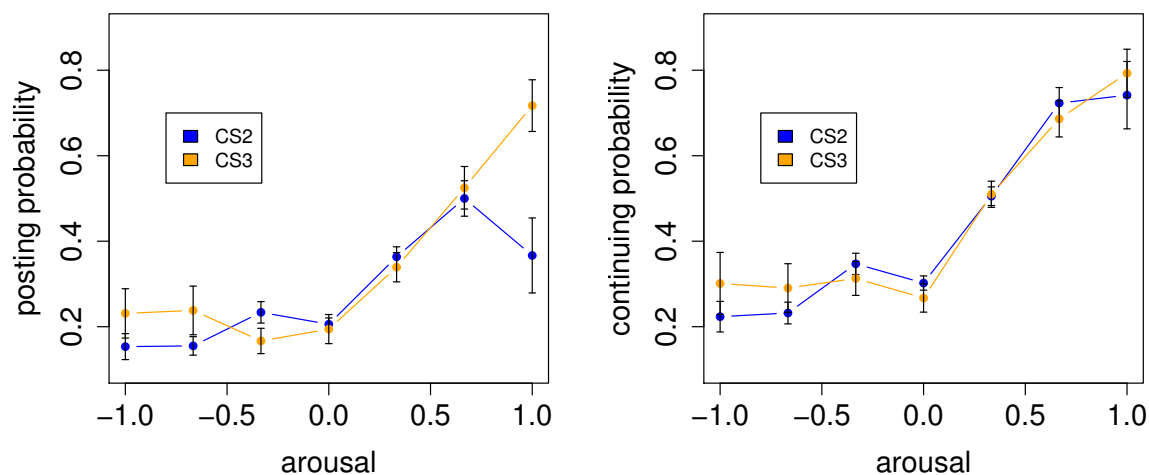


Figure 4.17: Mean reported participation and continuation intention given arousal for CS2 (yellow) and CS3 (blue). Error bars show standard error.

this nonlinearity in the dependence between arousal and participation could be explained by principles of emotion regulation, which would lead users to inhibit emotional expression under states of high excitation. This would be in line with the behavior of antipersistent users of chat communities, discussed in Chapter 3, but more empirical data is necessary to verify such theory.

In addition to the intention to participate and continue reading a thread, the participants of CS2 and CS3 were asked questions about how interesting and relevant was the thread. Figure 4.18 shows an overview of the relationships between valence, arousal, and the replies to these questions. The first column shows how the intention of participating and continuing reading a discussion depend on the valence. For the case of probability of posting on the thread, there is no clear influence of the valence, separating posting dynamics from the particular value of valence. This was one of the assumptions of our modeling framework, in which the rule that determined when an agent expressed its emotions only depended on its internal arousal, and not on its valence. The situation is different for the case of the intention to continue reading. The data from CS3 shows a larger intention to continue for the participants that reported a high valence after reading the thread, while this influence is similar but softer for the case of CS2. This might depend on the selection of threads, which was larger for CS2 than for CS3, but nevertheless this poses some initial evidence that Internet users drive their attention towards positive content.

The reports for relevance and interest in the thread were positively correlated with re-

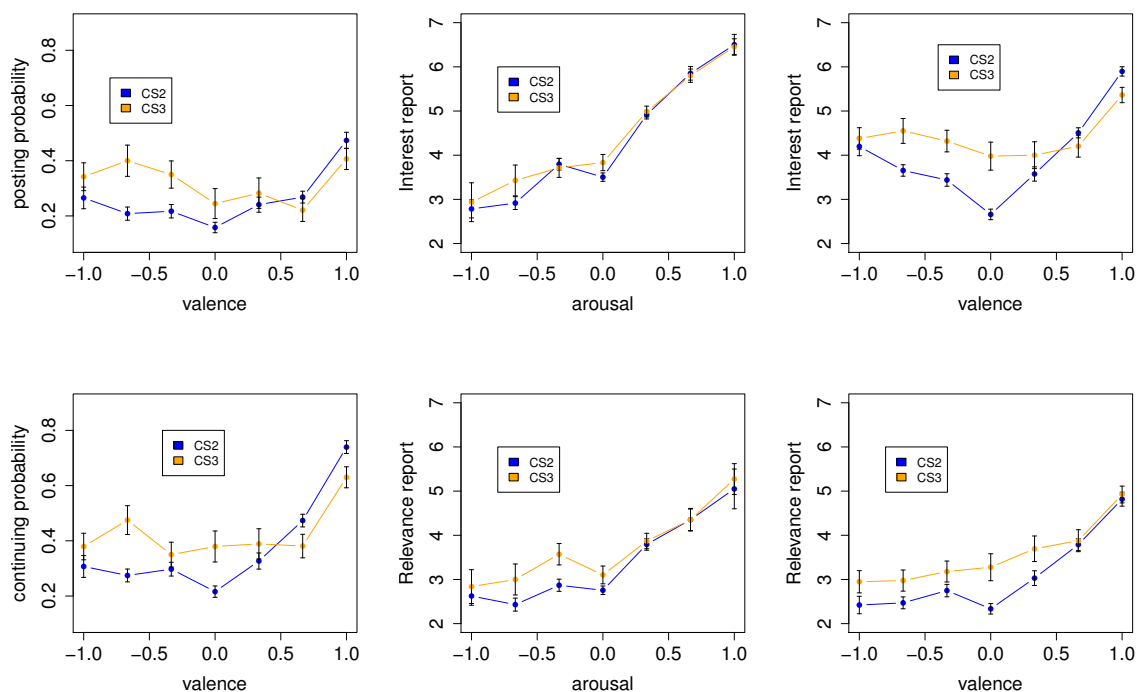


Figure 4.18: Mean reported participation, continuation intention, interest, and relevance reports given valence and arousal for CS2 (yellow) and CS3 (blue). Error bars represent standard error.

ported arousal, in line with the principle that higher arousal states are more relevant and important for the individual, as they require immediate action. On the other hand, relevance was not so much influenced by valence, and the relationship with interest reports differed between experiments. In CS3, the interest report of a participant did not seem to depend on its valence, while in CS2, the relation between them has V-shape. This reported interest is thus correlated with the absolute value of the valence, making threads that elicit stronger valences more interesting. Again, this relationship seemed highly dependent on the particular content of the thread, disappearing for the control case of CS2.

The experimental setup of CS4 focused on the production of posts, with subjective assessment of emotions before and after writing the post. In a similar way as we did for the changes after reading a thread, we calculate the effect of expression in the emotional state of an individual, represented as R_{dyn} in Figure 4.1. We divided the valence and arousal reports in negative if they belonged to the lower third of their range, neutral if in the middle, and positive if they were in the higher third of their possible values. Our findings are shown in Figure 4.19. After replying to a post, users with initially negative

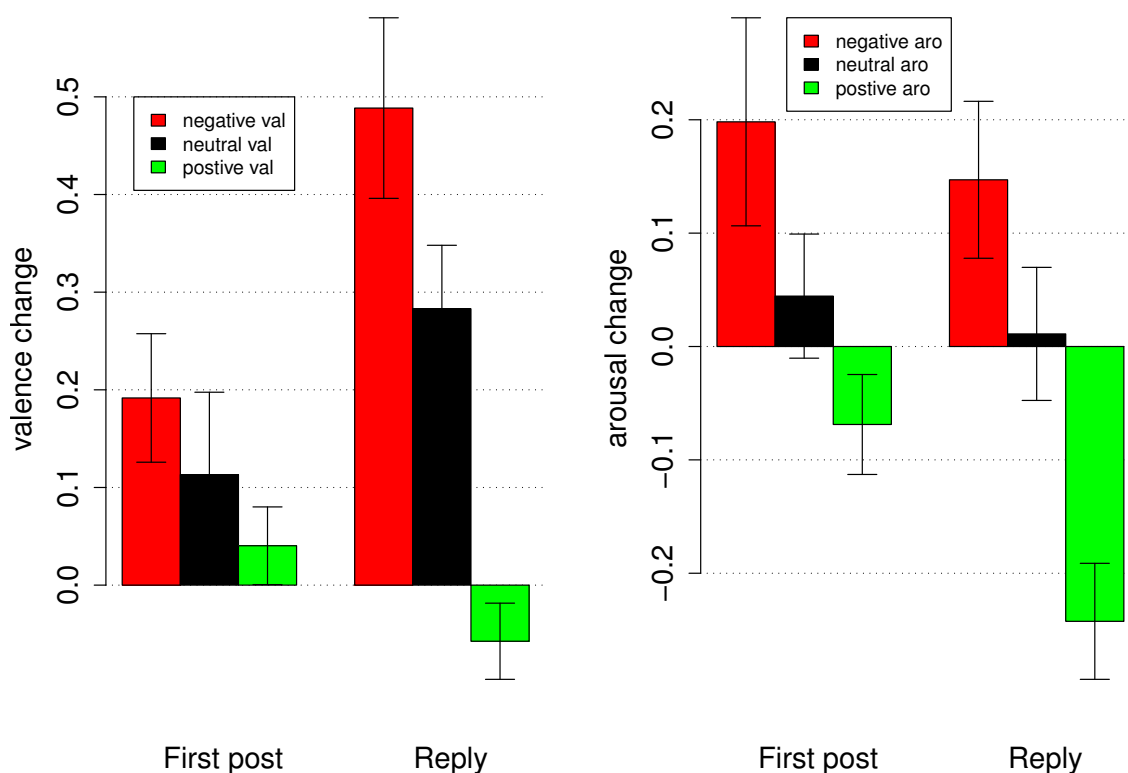


Figure 4.19: Change in valence and arousal after starting a thread and replying to a post, given the polarity of their previous value. Red bar shows the change in valence and arousal if its previous value was in the lower third of its possible range, black bars if the previous value was in the middle third, and green bars if their previous value was in the highest third of their possible value. Error bars show the standard error around the mean changes over the intervals.

or neutral valence experience an increase on it, while users with positive valence experience no significant change. The observed pattern is not so strong after posting the first message, which suggests that participating in the conversation increases negative valences more than starting a conversation. The change in arousal shows a slight increase after writing a post, when the participant is in negative arousal states, for both replies and thread initiations. We did not find any significant changes in this patterns when dividing by polarity of the message replied to, or the first message posted.

The most interesting insight from this analysis is that the participants, when in a positive arousal state, experience a large decrease in arousal after replying to a post. This change is also present when initiating a conversation, but of much smaller magnitude. This

supports our hypothesis of a decrease in arousal after posting, when this one has been triggered by a high arousal state, as shown in the right part of Figure 4.19. Nevertheless, this change in arousal was not a sharp rule of a reset to 0 after a post, which was our initial assumption. On the contrary, humans seem to experience a smaller decrease on their arousal after replying to a post, which still leaves room for further postings in the absence of interaction. This difference in posting dynamics should be taken into account in future models within our framework, but the causality between participation and arousal is supported by these results. When an Internet user perceives emotional content (P_{dyn}), its arousal increases, leading to higher chances of participating in the discussion (S). This participation induces an instant decrease in arousal (R_{dyn}), which, in combination with the internal relaxation of the arousal (E_{dyn}), would decrease the probability of further participation. If other users create more emotional content, the arousal of the agent would increase again, leading to the coupling of user behavior that explains collective behavior in online communities.

4.7 Conclusions

The work presented here showed how online communication influences the emotions an Internet user, and how these emotions change in time. We applied techniques from statistical physics to empirical psychology, finding new results in the measurement of physiological data. Our application of nonlinear regression to the activity of Zygomaticus Major reveals strong relations to valence, allowing us to detect the existence of peaks in smiling activity. In addition, we applied the autocorrelation function to reveal the timespan of physiological memory while reading emotionally charged threads. These techniques provide new metrics of the physiological manifestation of emotional states, which can be applied to future analysis of emotion dynamics in psychophysiology (Lodewyckx *et al.*, 2011).

Our work aimed at testing the assumptions used to design our agent-based framework of collective emotions. Analyzing subjective assessments of emotions, we found strong support for the presence of an exponential relaxation of emotions towards a ground state. In terms of online communication, we found that valence changes according to the polarity of threads, and that threads with emotional content lead the arousal to higher values than threads with neutral content. The agents in our framework express their emotions when their arousal reaches a particular threshold. This is verified by inspecting the dependence between reported arousal and intention of participating in a conversation, which reveals that such intention increases linearly when the arousal is beyond a threshold value. Another assumption of our modeling framework was the effect of writing a post in arousal,

which we hypothesized as an instant decrease of arousal. Our empirical results indicate the existence of this decrease after replying to posts in an online discussion.

The analysis of these datasets led to new findings about emotion dynamics that can improve future models within our framework. We replicated the findings of (Kuppens *et al.*, 2010), showing the existence of a nonzero baseline of the valence. Thanks to the fact that our framework was designed based on psychological variables, now we can include this kind of internal asymmetries in our models. New models within the framework, focused on emotions in blogs (Mitrović and Tadić, 2012), in social networking sites (Šuvakov *et al.*, 2012), and in trust networks (Tanase *et al.*, 2012), include the existence of this kind of ground state, proving the flexibility of our framework in incorporating empirical findings. In addition, we provide empirical support to the microdynamics of emotions in our models, using empirical parameter estimations. We simulated a minimal model of the dynamics of individual emotions during the perception of online content, where valence and arousal of simulated agents are similar to the participant reports after reading threads.

The dynamics of emotional states during online interaction shows that there is a negative solution of the arousal for non-emotional threads, which we did not take into account in the initial models of our framework. In addition, we found that replying to a post creates a valence increase for users that have a negative valence before interaction, as a beneficial result of the participation in an online discussion. Furthermore, our assumption of the decrease in arousal after expression is hereby extended, as we find that such a decrease does not reset the arousal to 0, but lowers it by some fixed amount. Further experiments shall focus on how simultaneous positive and negative emotional content influences the emotions of the readers of a thread. While still some of these hypotheses remain untested, the results shown here allow the design of better agent-based models of emotions.

Chapter 5

The positive bias of emotional expression

Summary

We show that the frequency of word use is not only determined by the word length (Zipf, 1935) and the average information content (Piantadosi *et al.*, 2011b), but also by its emotional content. We have analyzed three established lexica of affective word usage in English, German, and Spanish, to verify that these lexica have a neutral, unbiased, emotional content. Taking into account the frequency of word usage, we find that words with a positive emotional content are more frequently used. This lends support to Pollyanna hypothesis (Boucher and Osgood, 1969) that there should be a positive bias in human expression. We also find that negative words contain more information than positive words, as the informativeness of a word increases uniformly with its valence decrease. Our findings support earlier conjectures about (i) the relation between word frequency and information content, and (ii) the impact of positive emotions on communication and social links.

5.1 The role of emotional expression in social interaction

One would argue that human languages, in order to facilitate social relations, should be biased towards positive emotions. This question becomes particularly relevant for sentiment classification, as many tools assume as null hypothesis that human expression has neutral emotional content (Pang and Lee, 2008; Thelwall *et al.*, 2011), or reweight positive and negative emotions (Taboada *et al.*, 2011) without a quantification of any possible bias of emotional expression. An emotional bias in written expressions, however, would have a strong impact, as it shifts the balance between positive and negative expressions. Thus, for all researchers dealing with emotions in written text it would be of particular importance to know about such bias, how it can be quantified, and how it affects the baseline, or reference point, for expressed emotions.

In our empirical analysis of emotional expression in online communities, we noticed the existence of such a bias towards positive expression. To the datasets already explained in Chapter 2 and Chapter 3, we analyzed additional datasets of emotional expression from other communities. Two of these datasets were provided by the Statistical Cybernetics Research Group of the University of Wolverhampton¹, with messages from **MySpace** wall posts and **Twitter** microblogging *tweets*. These two datasets already included emotional scores calculated with SentiStrength (Thelwall *et al.*, 2010). During the course of this research, we retrieved three more datasets with comments for **Youtube** videos, short posts in the anonymous forum **4chan.org**, and bug discussions on open source projects with **Bugzilla**². We also processed all these datasets with SentiStrength. For each of the messages, SentiStrength provided two scores of positive emotions $p \in [+1, +5]$ and of negative emotions $n \in [-1, -5]$. We calculated the valence value v of a message as $v = p + n$, which constrains it to the interval $[-4, 4]$. For the cases of IRC chat messages, **4chan.org** posts, and **Bugzilla** bug comments, we excluded the messages that had both $p = 1$ and $n = -1$, as most of them correspond to code or image exchanges.

The general bias towards positive emotional expression can be observed in Figure 5.1. With the exception of **4chan.org**, all the datasets show a larger amount of messages with positive valence than with negative valence. This bias is specially strong for the product reviews from **Amazon.com**, and softer for other communities like **Bugzilla**. As mentioned above, this positive bias might be related to social relations. This hypothesis is consistent with the negativity of **4chan.org**, a completely anonymous community that does not create well-defined social bonds, and characterized by its levels of negative expression (Bernstein *et al.*, 2011). The sentiment analysis of SentiStrength, as a lexicon-based tool,

¹<http://cybermetrics.wlv.ac.uk/>

²We would like to thank M. Zanetti, I. Scholtes, and E. Sarigol for providing the Bugzilla dataset

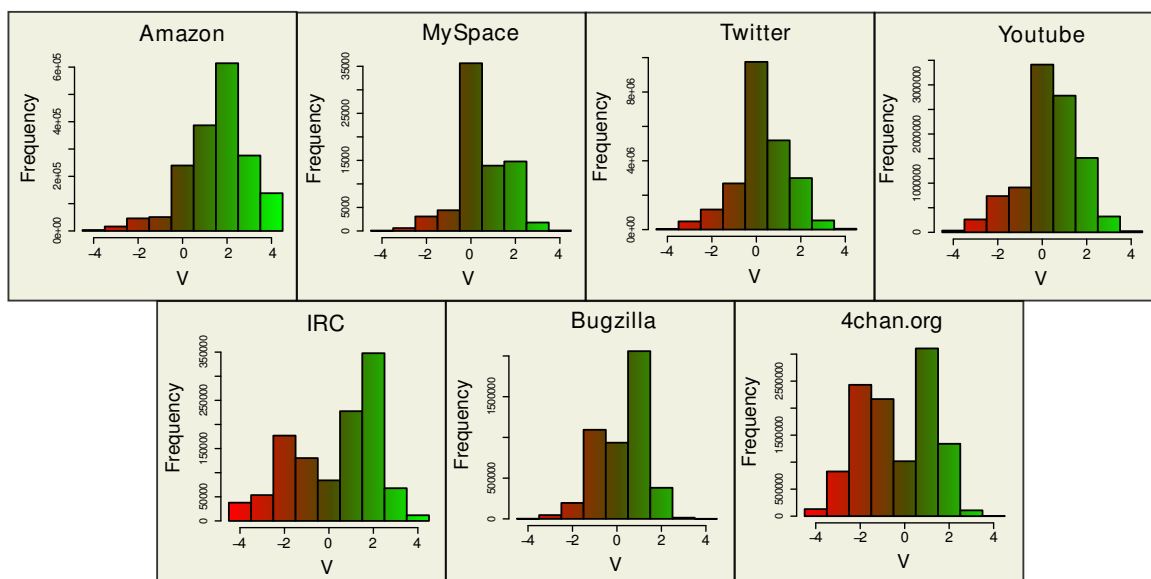


Figure 5.1: Distributions of valence across different datasets of online expression. Top row: valence distributions from the **Amazon.com** reviews of Chapter 2, **MySpace** wall posts (Thelwall *et al.*, 2010), **Twitter** messages (Thelwall *et al.*, 2011), and **Youtube** video comments. Bottom row: filtered distributions of valence in emotional messages in the IRC channels studied in Chapter 3, **Bugzilla** discussions of open source projects, and short posts in **4chan.org** threads.

is based on the presence of positive and negative emotion-bearing terms, which lead us to ask the following questions: Are positive words more frequent on the Internet than negative words? If so, how universal is this bias towards positive expression? How does it influence the communication process between Internet users?

Historically, the *frequency* of words was first analyzed by Zipf (Zipf, 1935, 1949) showing that frequency predicts the *length* of a word as result of a principle of least effort. Zipf's law highlighted fundamental principles of organization in human language (Ferrer i Cancho and Sole, 2003), and called for an interdisciplinary approach to understand its origin (Hauser *et al.*, 2002; Kosmidis *et al.*, 2006; Havlin, 1995) and its relation to word meaning (Piantadosi *et al.*, 2011a). Recently Piantadosi *et al.* (Piantadosi *et al.*, 2011b) extended Zipf's approach by showing that, in order to have efficient communication, word length increases with information content. Further discussions (Griffiths, 2011; Reilly and Kean, 2011; Piantadosi *et al.*, 2011a) highlighted the relevance of *meaning* as part of the communication process as, for example, more abstract ideas are expressed through longer words (Reilly and Kean, 2007). Our work focuses on one particular aspect of meaning, namely the *emotion* expressed in a word, and how this is related to word frequency and information content. This approach requires additional data beyond word length and frequency,

which became available thanks to large datasets of human behavior on the Internet, like the ones mentioned here.

In order to link the emotionality of each word with the information it carries, we build on the recent work of Piantadosi et al. (Piantadosi *et al.*, 2011b). This way, we reveal the importance of emotional content in human communication which influences the information carried by words. While the rational process that optimizes communication determines word lengths by the information they carry, we find that the emotional content affects the word frequency such that positive words appear more frequently. This points towards an emotional bias in used language and supports Pollyanna hypothesis (Boucher and Osgood, 1969), which asserts that there is a bias towards the usage of positive words. Furthermore, we extend the analysis of information content by taking into account word context rather than just word frequency. This leads to the conclusion that positive words carry less information than negative ones. In other words, the informativeness of words highly depends on their emotional polarity.

5.2 Frequency of emotional words

Our investigation is devoted to this problem by combining two analysis, (i) quantifying the emotional content of words in terms of valence, and (ii) quantifying the frequency of word usage in the whole indexable web (Brants and Franz, 2009). We provide a study of the baseline of written emotional expression on the Internet in three languages that span more than 67.7% of the websites³: English (56.6%), German (6.5%), and Spanish (4.6%). These languages are used everyday by more than 805 million users, who create the majority of the content available on the Internet.

Before explaining the details of our analysis, we wish to emphasize that our work distinguishes itself both regarding its methodology and its findings from a recent article (Kloumann *et al.*, 2012). There, the authors claim a bias in the amount of positive versus negative words in English, while no relation between emotionality and frequency of use was found. A critical examination of the conditions of that study shows that the quantification of emotions was done in an uncontrolled setup through the Amazon Mechanical Turk. Participants were shown a scale similar to the ones used in previous works (Bradley and Lang, 1999; Vo *et al.*, 2009; Redondo *et al.*, 2007), as explained in (Dodds *et al.*, 2011). Thanks to the popular usage of the Mechanical Turk, the authors evaluated more than 10.000 terms from the higher frequency range in four different corpora of English expression. However, the authors did not report any selection criterion for the participant

³en.wikipedia.org/wiki/Global_Internet_usage

reports, opposed to the methodology presented in (Bohannon, 2011) where up to 50% of the participants had to be discarded in some experiments.

Because of this lack of control in their experimental setup, the positive bias found in (Kloumann *et al.*, 2012) could be easily explained as an acquiescent bias (Knowles and Nathan, 1997; Bentler, 1969), a result of the human tendency to agree in absence of further knowledge or relevance. In particular, this bias has been repeatedly shown to exist in self assessments of emotions (Russell, 1979; Yik *et al.*, 2011), requiring careful response formats, scales, and analysis to control for it. Additionally, the wording used to quantify word emotions in (Kloumann *et al.*, 2012) (*happiness*), could imply two further methodological biases: The first one is a possible social desirability bias (Robinson *et al.*, 1991), as participants tend to modify their answers towards more socially acceptable answers. The positive social perception of displaying happiness can influence the answers given by the participants of the study. Second, the choice of the word *happiness* implies a difference compared with the standard psychological term *valence* (Russell, 1980). Valence is interpreted as a static property of the word while happiness is understood as a dynamic property of the surveyed person when exposed to the word. This kind of framing effects have been shown to have a very large influence in survey results. For example, a recent study (Bryan *et al.*, 2011) showed a large change in the answers by simply changing *voting* for *being a voter* in a voter turnout survey.

Hence, there is a strong sensitivity to such influences which are not controlled for in (Kloumann *et al.*, 2012). There are alternative lexica contain information related to positive and negative emotional content (Pennebaker *et al.*, 2001; Stone *et al.*, 1966; Bradley and Lang, 1999). For the case of valence, as an element of *core affect*, the current standard lexicon for English emotional words is ANEW (Bradley and Lang, 1999), which has equivalents in German (Vo *et al.*, 2009) and Spanish (Redondo *et al.*, 2007) These lexica were produced in three controlled, independent setups, and provide the most reliable estimation of word emotionality for our analysis. In detail, they contain 1034 English words, 2902 German words and 1034 Spanish words, together with their emotional scores obtained from extensive human ratings. These lexica have effectively established the standard for emotion analysis of human texts (Dodds and Danforth, 2009). Each word in these lexica is assigned a set of values measuring different aspects of word emotionality. The three independent studies that generated the lexica for English (Bradley and Lang, 1999), German (Vo *et al.*, 2009), and Spanish (Redondo *et al.*, 2007) used the Self-Assessment Mannequin (SAM) method to ask participants about the different emotion values associated to each word in the lexicon. One of these values, a scalar variable v called valence, represents the degree of pleasure induced by the emotion associated with the word, and it is known to explain most of the variance in emotional meaning (Russell, 1980). For our analysis, we

use v to quantify word emotionality.

In each lexicon, words were chosen such that they evenly span the full range of valence. In order to compare the emotional content of the three different languages, we have rescaled all values of v to the interval $[-1,1]$. As shown in the upper panel of Figure 5.3, indeed, the average valence, as well as the median, of all three lexica is very close to zero, i.e. they do not provide an emotional bias. This analysis, however, neglects the actual frequency of word usage, which is highly skew distributed (Zipf, 1935, 1949). For our frequency estimations we have used Google’s N -gram dataset (Brants and Franz, 2009) which, with 10^{12} tokens, is one of the largest datasets available about textual expression. The source of this dataset is the whole Google crawl, which aimed at spanning a large subset of the web, providing a wide point of view on how humans write on the Internet. This offers us the chance to test if the biases shown in Figure 5.1 are a universal feature of online emotional expression.

For our analysis, we have studied the frequency of the words which have an affective classification in the respective lexicon in either English, German, or Spanish. Figure 5.2 shows emotion word clouds for the three languages, where each word appears with a size proportional to its frequency. The color of a word is chosen according to its valence, ranging from red for $v = -1$ to green for $v = +1$. It is clear that green dominates over red in the three cases, as positive emotions predominate on the Internet. Some outliers, like “home”, have a special higher frequency of appearance in websites, but as we show later, our results are consistent with frequencies measured from traditional written texts like books.

In a general setup, the different usage of words with the same valence is quite obvious. For example, both words “party” and “sunrise” have the same positive valence of 0.715, however the frequency of “party” is 144.7 per one million words compared to 6.8 for “sunrise”. Similarly, both “dead” and “distressed” have a negative valence of -0.765, but the former appears 48.4 times per one million words, the latter only 1.6 times. Taking into account all frequencies of word usage, we find for all three languages that the median shifts considerably towards positive values. This is shown in the right panel of Figure 5.3. Wilcoxon tests show that the means of these distributions are indeed different, with an estimated difference in a 95% confidence interval of 0.257 ± 0.032 for English, 0.167 ± 0.017 for German, and 0.287 ± 0.035 for Spanish. Hence, with respect to usage we find evidence that the language used on the Internet is emotionally charged, i.e. significantly different from being neutral. This affects quantitative analysis of the emotions in written text, because the “emotional reference point” is not at zero, but at considerably higher valence values (about 0.3).

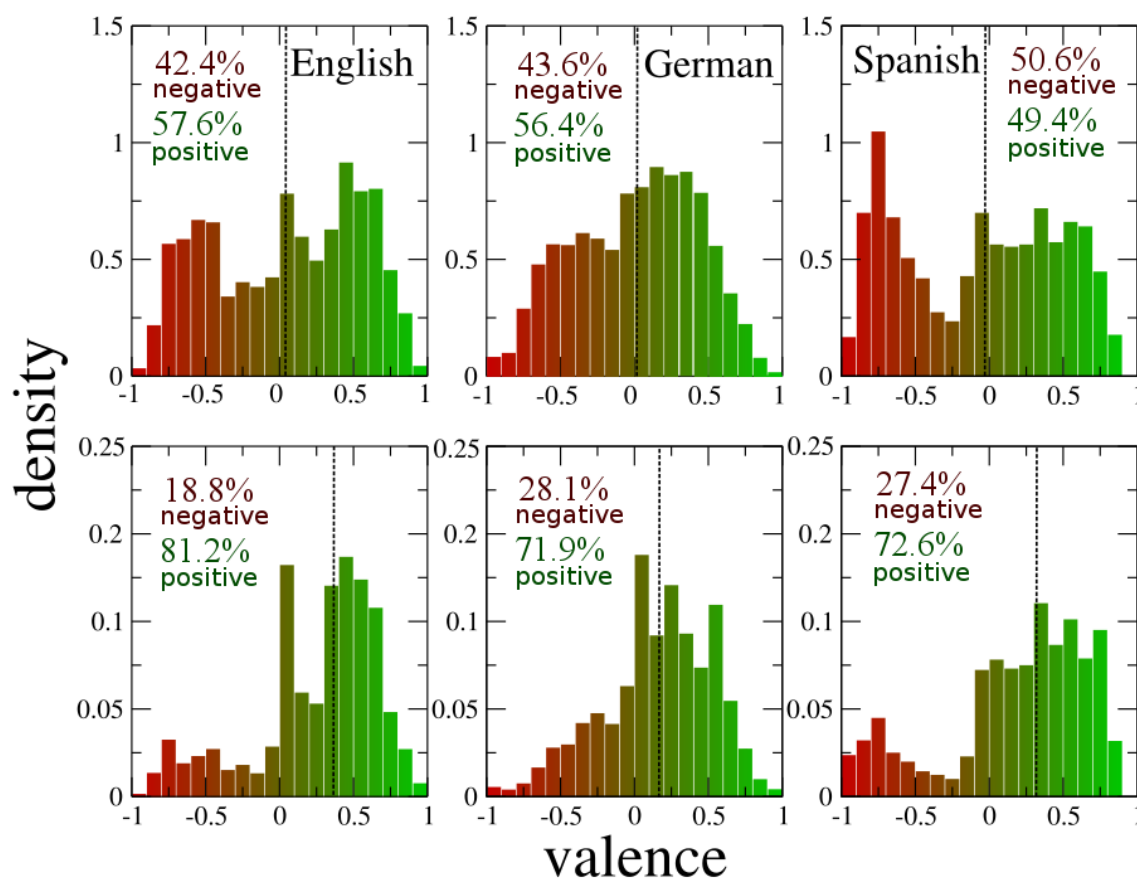


Figure 5.3: (Upper panel) distributions of reported valence values for words in English (left panel, lexicon: (Bradley and Lang, 1999), 1034 entries), German (middle panel, lexicon: (Vo *et al.*, 2009), 2902 entries), and Spanish (right panel, lexicon: (Redondo *et al.*, 2007), 1034 entries), normalized by the size of the lexica. Average valence (median) 0.048 (0.095) for English, 0.021 (0.067) for German, and -0.065 (-0.006) for Spanish. (bottom 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.

sequences. First we discuss the relation between word valence and information content estimated from the simple word occurrences, namely self-information. Then we explain how this extends when the information is measured taking into account the different contexts in which a word can appear. The self-information of a word, $I(w)$ (Cover and Thomas, 1991) is an estimation of the information content from its probability of appearance, $P(w)$, as

$$I(w) = -\log P(w) \quad (5.1)$$

Frequency-based information content metrics like self-information are commonly used in computational linguistics to systematically analyze communication processes. Information content is a better predictor for word length than word frequency (Piantadosi *et al.*, 2011b; Florian Jaeger, 2010), and the relation between information content and meaning, including emotional content, is claimed to be crucial for the way humans communicate (Griffiths, 2011; Reilly and Kean, 2011; Piantadosi *et al.*, 2011a). We use the self-information of a word as an estimation of information content for a context size of 1, to build up later on larger context sizes. This way, we frame our analysis inside the larger framework of N-gram information measures, aiming at an extensible approach that can be incorporated in the fields of computational linguistics and sentiment analysis.

For the three lexica, we calculated $I(w)$ of each word and linked it to its valence, $v(w)$. As defined in Equation (5.1), very common words provide less information than very unusual ones, but this nonlinear mapping between frequency and self-information makes the latter more closely related to word valence than the former. The first two lines of Table 5.1 show the Pearson's correlation coefficient of word valence and frequency $\rho(v, f)$, followed by the correlation coefficient between word valence and self-information, $\rho(v, I)$. For all three languages, the absolute value of the correlation coefficient with I is larger than with f , showing that self-information provides more knowledge about word valence than plain frequency of use.

Figure 5.4 shows in detail the relation between v and I . From the clear negative correlation found for all three languages (between -0.3 and -0.4), we deduce that words with less information content carry more positive emotions, as the average valence decreases along the self-information range. As mentioned before, the Pearson's correlation coefficient between word valence and self-information, $\rho(v, I)$, is significant and negative for the three languages (Table 5.1). Our results outperform a recent finding (Augustine *et al.*, 2011) that, while focusing on individual text production, reported a weaker correlation (below 0.3) between the logarithm of word usage frequency and valence. This previous analysis was based on a much smaller data set from Internet discussions (in the order of 10^8 tokens) and the same English lexicon of affective word usage (Bradley and Lang, 1999) we used. Using a much higher accuracy in estimating word frequencies and extending the analysis to three different languages, we were able to verify that there is a significant relation between the emotional content of a word and its self-information, impacting the frequency of usage.

Eventually, we also performed a control analysis using alternative frequency datasets, to

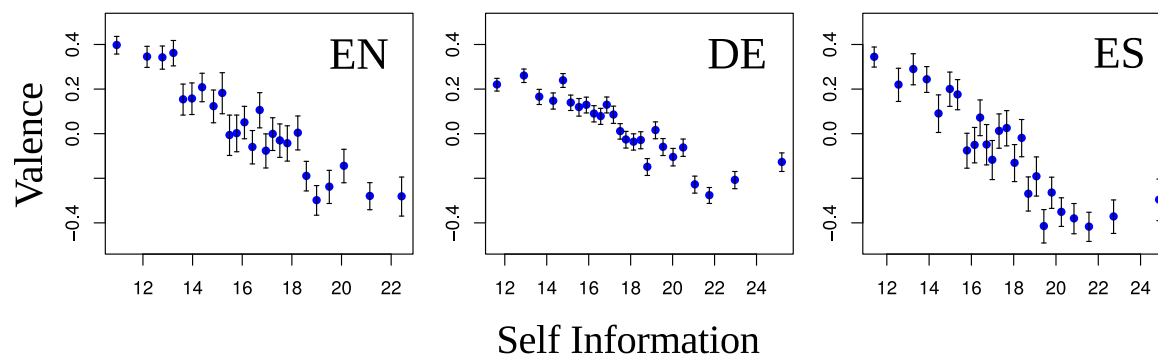


Figure 5.4: Relation between word self-information and valence for English (left), German (middle), and Spanish (right). Average valence is shown for bins that contain 5% of the data, with error bars showing the standard error. For all the three languages, valence clearly decreases with the self-information of the word, i.e. positive words carry less information than negative words.

	English	German	Spanish
$\rho(v, f)$	0.222 **	0.144 **	0.236 **
$\rho(v, I)$	-0.368 **	-0.325 **	-0.402 **
$\rho(v, I')$	-0.294 **	-0.222 **	-0.311 **

Significance levels: * $p < 0.01$, ** $p < 0.001$.

Table 5.1: Correlation coefficients of the valence (v), frequency f , self-information I from the N -gram dataset (Brants and Franz, 2009), and self-information I' measured from the frequencies reported in (Kucera and Francis, 1967; Baayen *et al.*, 1993; Sebastián *et al.*, 2000).

account for possible anomalies in the Google dataset due to its online origin. We used the word frequencies estimated from traditional written corpora, i.e. books, as reported in the original datasets for English (Kucera and Francis, 1967), German (Baayen *et al.*, 1993), and Spanish (Sebastián *et al.*, 2000). Calculating the self-information from these and relating them to the valences given, we obtained similar, but slightly lower Pearson's correlation coefficients $\rho(v, I')$ (see Table 5.1). So, we conclude that our results are robust across different types of written communication, for the three languages analyzed.

5.4.1 Context-based information measures

It is not surprising to find a larger self-information for negative words, as their probability of appearance is generally lower. The amount of information carried by a word is also

highly dependent on its context. Among other factors, the context is defined by the word neighborhood in the sentence. For example, the word “violent” contains less information in the sentence “dangerous murderers are violent” than in “fluffy bunnies are violent”, as the probability of finding this particular word is larger when talking about murderers than about bunnies. For this reason we evaluate how the context of a word impacts its informativeness and valence. The intuition behind measuring information depending on the context is that the information content of a word depends primarily on i) the amount of contexts it can appear and ii) the probability of appearance in each one of these contexts. Not only the most infrequent, but the most specific and unexpected words are the ones that carry the most information.

Taking into account the contexts in which a word w can appear, the information content of w is defined as:

$$-\sum_c P(C = c|W = w)\log(P(W = w|C = c)) \quad (5.2)$$

where C and W are the random variables for contexts and words respectively. As explained in (Piantadosi *et al.*, 2011b), for each context c_i where a word w appears, its information content can be approximated from a corpus as:

$$-\frac{1}{N}\sum_{i=1}^N \log(P(W = w|C = c_i)) \quad (5.3)$$

where N is the total frequency of the word in the corpus used for the estimation. These values were calculated as approximations of the information content given the words surrounding w up to size 4.

We analyzed how word valence is related to the information content up to context size 4 using the original calculations provided by Piantadosi *et al.* (Piantadosi *et al.*, 2011b). This estimation is based on the frequency of sequences of N words, called N -grams, from the **Google** dataset (Brants and Franz, 2009) for $N \in \{2, 3, 4\}$. This dataset contains frequencies for single words and N -grams, calculated from an online corpus of more than a trillion tokens. For each size of the context N , we have a different estimation of the information carried by the studied words, with self-information representing the estimation from a context of size 1.

Figure 5.5 shows how valence decreases with the estimation of the information content for each context size. Each bar represents the same amount of words within a language and has an area proportional to the rescaled average information content carried by these words. The color of each bar represents the average valence of the binned words. The decrease of average valence with information content is similar for estimations using 2-grams and

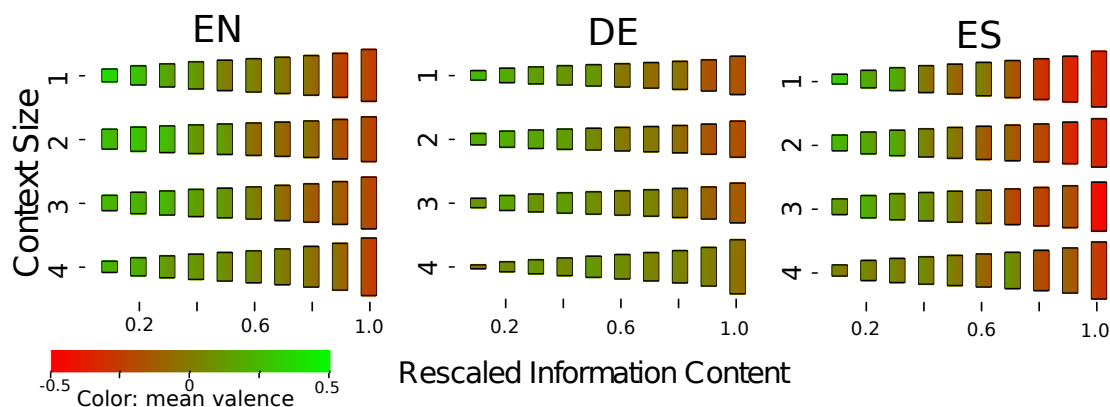


Figure 5.5: Relation between valence and information content measured up to a context of size four. Each bar represents a bin containing 10% of the words in the lexicon, with a size proportional to the average information content of the words in the bar. The color of each bar ranges from red to green, representing the average valence of the words in the corresponding bin. Each bar has a color gradient according to the standard error of the valence mean. Information content has been rescaled so it can be compared among context sizes. For all three languages and context sizes, negativity increases with information content.

3-grams. For the case of 4-grams it also decreases for English and Spanish, but this trend is not so clear for German. These trends are properly quantified by Pearson’s correlation coefficients between valence and information content for each context size (Table 5.2). Each correlation coefficient becomes smaller for larger sizes of the context, as the information content estimation includes a larger context but becomes less accurate.

	English	German	Spanish
$\rho(v, I_2)$	-0.332 **	-0.301 **	-0.359 **
$\rho(v, I_3)$	-0.313 **	-0.201 **	-0.340 **
$\rho(v, I_4)$	-0.254 **	-0.049 *	-0.162 **

Significance levels: * $p < 0.01$, ** $p < 0.001$.

Table 5.2: Correlation coefficients of the valence v and information content measured for 2-grams I_2 , 3-grams I_3 , and 4-grams I_4

5.4.2 Additional analysis of valence, length and self-information

In order to provide additional support for our results, we tested different hypotheses impacting the relation between word usage and valence. First, we calculated Pearson’s

and Spearman's correlation coefficients between the absolute value of the valence and the self-information of a word, $\rho(abs(v), I)$ (see Table 5.3). We found both correlation coefficients to be around 0.1 for German and Spanish, while they are not significant for English. The dependence between valence and self-information disappears if we ignore the sign of the valence, which means, indeed, that the usage frequency of a word is not just related to the overall emotional intensity, but to the positive or negative emotion expressed by the word.

	English	German	Spanish
$\rho(abs(v), I)$	0.032 [°]	0.109 ***	0.135 ***
$\rho(l, I)$	0.378 ***	0.143 ***	0.361 ***
$\rho(v, l)$	-0.044 [°]	-0.071 ***	-0.112 ***
$\rho(v, I l)$	-0.379 ***	-0.319 ***	-0.399 ***
$\rho(l, I v)$	0.389 ***	0.126 ***	0.357 ***

Significance levels: [°] $p < 0.3$, * $p < 0.1$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.3: Correlation coefficients of the valence (v), absolute value of the valence ($abs(v)$), and word length (l) versus self-information (I). Partial correlations are calculated for both variables ($\rho(v, I|l), \rho(l, I|v)$), and correlation between valence and length ($\rho(v, l)$).

Subsequently, we found that the correlation coefficient between word length and self-information ($\rho(l, I)$) is positive, showing that word length increases with self-information. These values of $\rho(l, I)$ are consistent with previous results (Zipf, 1935; Piantadosi *et al.*, 2011b). Pearson's and Spearman's correlation coefficients between valence and length $\rho(v, l)$ are very low or not significant. In order to test the combined influence of valence and length to self-information, we calculated the partial correlation coefficients $\rho(v, I|l)$ and $\rho(l, I|v)$. The results are shown in Table 5.3, and are within the 95% confidence intervals of the original correlation coefficients $\rho(v, I)$ and $\rho(l, I)$. This provides support for the existence of an additional dimension in the communication process closely related to emotional content rather than communication efficiency. This is consistent with the known result that word lengths adapt to information content (Piantadosi *et al.*, 2011b), and we discover the independent semantic feature of valence. Valence is also related to information content but not to the symbolic representation of the word through its length.

Finally, we explore the sole influence of context by controlling for word frequency. In Table 5.4 we show the partial correlation coefficients of valence with information content for context sizes between 2 and 4, controlling for self-information. We find that most of the correlations keep significant and of negative sign, with the exception of I_2 for English. The weaker correlation for context sizes of 2 is probably related to two word constructions

such as negations, articles before nouns, or epithets.

	English	German	Spanish
$\rho(v, I_2 I)$	-0.034 [°]	-0.100 ***	-0.058 *
$\rho(v, I_3 I)$	-0.101 **	-0.070 ***	-0.149 ***
$\rho(v, I_4 I)$	-0.134 ***	-0.020 *	-0.084 **

Significance levels: [°] $p < 0.3$, * $p < 0.1$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.4: Partial correlation coefficients of the valence (v) and information content measured on different context sizes (I_2, I_3, I_4) controlling for self-information (I).

These high-frequent, low-information constructions lead to the conclusion that I_2 does not explain more about the valence than self-information in English, as short range word interactions change the valence of the whole particle. This finding supports the assumption of many lexicon-based unsupervised sentiment analysis tools, which consider valence modifiers for two-word constructions (Thelwall *et al.*, 2011; Taboada *et al.*, 2011). On the other hand, the significant partial correlation coefficients with I_3 and I_4 suggest that word information content combines at distances longer than 2, as longer word constructions convey more contextual information than 2-grams. Knowing the possible contexts of a word up to distance 4 provides further information about word valence than sole self-information.

5.5 Discussion

Our analysis provides strong evidence that words with a positive emotional content are used more often. This lends support to Pollyanna hypothesis (Boucher and Osgood, 1969), i.e. positive words are more often used, for all the three languages studied. Our conclusions are consistent for, and independent of, different corpora used to obtain the word frequencies, i.e. they are shown to hold for traditional corpora of formal written text, as well as for the Google dataset and cannot be attributed as artifacts of Internet communication.

Furthermore, we have pointed out the relation between the emotional and the informational content of words. Words with negative emotions are less often used, but because of their rareness they carry more information, measured in terms of self-information, compared to positive words. This relation remains valid even when considering the context composed of sequences of up to four words (N -grams). Controlling for word length, we

find that the correlation between information and valence does not depend on the length, i.e. it is indeed the usage frequency that matters.

In our analysis, we did not explore the role of syntactic rules and grammatical classes such as verbs, adjectives, etc. However, previous studies have shown the existence of a similar bias when studying adjectives and their negations (Rozin *et al.*, 2010). The question of how syntax influences emotional expression is beyond the scope of the present work. Note that the lexica we use are composed mainly of nouns, verbs and adjectives, due to their emotional relevance. Function words such as “a” or “the” are not considered to have any emotional content and therefore were excluded from the original studies. In isolation, these function words do not contain explicit valence content, but their presence in text can modify the meaning of neighboring words and thus modify the emotional content of a sentence as a whole. Our analysis on partial correlations show that there is a correlation between the structure of a sentence and emotional content beyond the simple appearance of individual words. This result suggests the important role of syntax in the process of emotional communication. Future studies can extend our analysis by incorporating valence scores for word sequences, exploring how syntactical rules represent the link between context and emotional content.

The findings reported in this chapter suggest that the process of communication between humans, which is known to optimize information transfer (Piantadosi *et al.*, 2011b), also creates a bias towards positive emotional content. A possible explanation is the basic impact of positive emotions on the formation of social links between humans. Human communication should reinforce such links, which it both shapes and depends on. Thus, it makes much sense that human languages on average have strong bias towards positive emotions, as we have shown (see Figure 5.3). Negative expressions, on the other hand, mostly serve a different purpose, namely that of transmitting highly informative and relevant events. They are less used, but carry more information.

Our findings are consistent with emotion research in social psychology. According to (Rime, 2009), the expression of positive emotions increases the level of communication and strengthens social links. This would lead to stronger pro-social behavior and cooperation, giving evolutionary advantage to societies whose communication shows a positive bias. As a consequence, positive sentences would become more frequent and even advance to a social norm (cf. “Have a nice day”), but they would provide less information when expressed. Our analysis provides insights on the asymmetry of evaluative processes, as frequent positive expression is consistent with the concept of *positivity offset* introduced in (Miller, 1961) and recently reviewed in (Norman *et al.*, 2011). In addition, Miller’s *negativity bias* (stronger influence of proximal negative stimuli) found in experiments provides

an explanation for the higher information content of negative expression. When writing, people could have a tendency to avoid certain negative topics and bring up positive ones just because it feels better to talk about nice things. That would lower the frequency of negative words and lower the amount of information carried by positive expression, as negative expression would be necessary to transmit information about urgent threats and dangerous events.

Eventually, we emphasize that the positive emotional “charge” of human communication has a further impact on the quantitative analysis of communication on the Internet, for example in chatrooms, fora, blogs, and other online communities. Our analysis provides an estimation of the emotional baseline of human written expression, and automatic tools and further analysis will need to take this into account. In addition, this relation between information content and word valence might be useful to detect anomalies in human emotional expression. Fake texts supposed to be written by humans could be detected, as they might not be able to reproduce this spontaneous balance between information content and positive expression.

Chapter 6

Discussion and conclusions

Science fiction often conceives future societies as dually composed by two separated worlds, a *real world* in which humans physically interact with each other, and a *cyberspace* in which certain technology allows humans to communicate in a different manner. In general, these two worlds are separated by a sharp barrier that distinguishes what is *reality* from what is *fiction*; what happens in classical human interaction, and what happens in the *virtual* world. A paradigmatic example of this kind of society is shown in the movie “The Matrix”, in which these two worlds are clearly separated. Regardless of whether human communication technologies have reached or not such power, this hypothetical dual society has other possible readings. Slavoj Žižek, commenting about this precise example, states:

“The choice between the blue and the red pill is not really a choice between illusion and reality. Of course Matrix is a machine for fictions, but these are fictions which already structure our reality. If you take away from our reality the symbolic fictions that regulate it, you lose reality itself.”

Slavoj Žižek, The Pervert’s Guide to Cinema, 2007

In our current society, online interaction is present in a way such that it is embedded within our everyday interaction. The pervasiveness of information technologies do not draw sharp boundaries between what is *cyberspace* and what is not. The Internet might be full of lies and fake personalities, but those are very real and they have an important impact in our life. There is no difference between the offline and online societies, as both are heavily interrelated in a manner that does not divide them. This does not leave human societies unchanged, the Internet serves as a tool to overcome cognitive biases in the way humans communicate with each other. In online communities, the accessibility

to large amounts of people is not restricted to influential individuals with access to the mass media. Often, we can trace back our words and the words of others thanks to stored discussions in online platforms, or we can cope with information overloads through personalized recommendations, or search for what we *guess* we want.

With this work, we wanted to push Science beyond fiction, taking a small step into the understanding of human emotions, and how this *Internet society* is shaping the way we interact. Our objects of research are the collective emotions that emerge from the interaction of many users, analyzing the traces that millions of users leave online. But we did not stop in the mere observation of human behavior. We created agent-based models that simulate emotions as far as we understand them, reproducing certain aspects of their collective behavior. The question of whether we have provided machines with emotions is rather irrelevant, as explained in words from Noam Chomsky:

“The question of whether a computer is playing chess, or doing long division, or translating Chinese, is like the question of whether robots can murder or airplanes can fly [...]. These are questions of decision, not fact; decision as to whether to adopt a certain metaphoric extension of common usage.”

Noam Chomsky, Powers and Prospects, 1996.

Nobody would say that the agents of our models possess emotions, they are a computational equivalent of how emotions are represented within our knowledge. The applications of this approach are limited to our previous understanding about emotions, but it provides a platform from which we can derive new knowledge. Furthermore, these models allow us to formulate new hypotheses that can lead further research. This closes the cycle between theory, empirical analysis, and modeling that has driven our research towards a more complete understanding of human emotions.

Contributions in a nutshell

Thanks to the combination of computational social science with sentiment analysis, we have analyzed an amount of empirical data that allows us to quantify various patterns of emotional interaction. We combined principles from psychology with tools from statistical physics, to connect collective emotional states with individual emotions and interaction among humans. We designed a framework to create agent-based models that, when simulated, reproduce some facets of human emotions, in particular online collective emotions. In combination with experimental studies on the dynamics of individual emotions, our modeling framework provides a unified view to collective emotions in online communities.

Statistical analysis of online emotions

Emotions in product reviews. We started our empirical study of collective emotions in online communities by focusing on product reviews communities. We retrieved a new and unique dataset from `Amazon.com`, which we processed with different sentiment analysis techniques. We combined this dataset with an estimation of mass media attention from `news.google.com`, leading to the following results:

- We tested the hypothesis that emotions are a driving force behind the creation of product reviews. Using correlations on the time series of ratios of emotional reviews, we found that most of the analyzed products have increasing levels of emotionality in their reviews (Figure 2.7).
- Using statistical analysis inspired in complex systems, we tested the existence of influence among reviewers, i.e. how the presence of previous reviews affects the future behavior of the users (Section 2.4.2).
- Our analysis indicates that mass media creates an initial distribution of reviews per product of a very large variance, but the collective dynamics among reviewers reshape this distribution into another one, shown in Figure 2.13.
- We analyzed user interaction through reviews as a growth process, measuring the time evolution of the influence among reviewers of a product, as formalized in Equation (2.3). We noticed that, for the case of emotional reviews, their influence decays slower than for non-emotional ones, as shown in Figure 2.17.
- We could distinguish products on the top sales ranking, which show a faster decay of the growth rates of reviews (Figure 2.16).

Emotional persistence in chatroom communities. The second type of community we analyzed is a set of publicly available chatrooms. The communication mechanism of chatrooms is very different from product reviews in terms of anonymity, purpose, formality, and speed, providing us a complementary view to our previous findings.

- Some features of human communication remain unchanged in chatrooms. The distribution of the time between the messages of the same user follows the power-law distribution shown in Figure 3.2, typical of a saturated regime.
- An interesting feature of these communities is the presence of a collective, real-time discussion in which many member are involved. Looking into the time between

user expression in such collective discussion, we found certain universal features that reveal the long range influence among users, as discussed in Section 3.3.2.

- Detrended Fluctuation Analysis of the emotional expression of chatroom users shows that users are mostly persistent (Figure 3.6), which suggests that users share emotional experiences in a group chat.
- We measured persistence at the level of the whole general discussion, revealing the existence of collective emotional states of the community, persistently shared among the involved users (Section 3.4.3).
- The overall positive tone of the user interaction in chatrooms suggests that the users tend to follow social norms, in spite of being almost anonymous for the rest.

The set of statistical tools we have used for our analysis is not exclusive of these two communities, and opens the possibility to apply it to many more. We compared data from different communities to find general rules of emotional expression, which showed us the presence of a bias towards positive expression.

Modeling emotional interaction

A framework for modeling collective emotions in online communities. We designed an agent-based framework that provides a common language for the definition of models of emotions during online interaction. The main features of this model are:

- Our framework is defined in a flexible manner, such that it can be applied to a large variety of online communities. Different models can be designed according to the communication mechanisms present in the online community, or the sentiment analysis tools available.
- In a first exploration of this framework, we showed how, under particular dynamics of emotions, collective emotional states can emerge from the interaction of many agents as shown in Figure 1.6.
- The modeling framework is analytically tractable. We illustrate this feature through analysis of the emergence of collective emotional polarization, which is matched with the outcomes of model simulations (Figure 1.5).
- The dynamics driving user interaction and emotions in our framework are predicated within theories of emotions in psychology. This allows further testing of such microdynamics in experimental setups, and open questions for further research.

Modeling emotions product reviews and chatrooms The first applications of our modeling framework aimed at reproducing collective emotions in product reviews communities and chatrooms, which were quantified through the stylized facts we found in our analysis of empirical data.

- Simulations of our product reviews model reproduced the typical distributions of positive and negative emotional expression in product reviews. This comparison can be seen in Figure 2.22.
- The patterns of collective reactions in our simulations are similar to our observations from empirical data, representing cases of mass media influence versus word-of-mouth effects.
- Our model of chatroom users used an empirical estimation of the time intervals between the messages created by users, focusing on their emotional expression. Individual agent simulations show that this model can reproduce the persistence we measure in individual chatroom users, as shown in Figure 3.11
- Simulation of chatroom discussion show how our model reproduces emotional persistence at the collective level, as trace of collective emotional states, as explained in Section 3.5.2.
- A careful analysis of the inter-event times of real and simulated chatroom discussions reveal that our model does not reproduce certain properties of the empirical data, opening the field for improved models (Section 3.5.4).

Understanding individual emotions

Emotion dynamics in experimental psychology We combined datasets from 4 independent experimental setups that provided data on different aspects of emotional dynamics, which we used to test the dynamics proposed in our modeling framework:

- We were able to measure the relaxation of emotions towards a ground state, independent of online interaction. This is formalized in Equation (4.2) and shown in Figure 4.10.
- We verified that the valence of an Internet user is influenced by the emotional content of online discussions. We could quantify this influence, which shifts the stable point of valence dynamics.

- Our analysis of arousal dynamics shows that reading emotional threads increases arousal, regardless of the valence expressed in those threads (Figure 4.14).
- We verified that arousal determined the intention of participating in an online discussion, increasing linearly after a threshold value, as shown in Figure 4.17.
- After replying in an online conversation, the participant arousal decreases by certain amount.
- We formalized all these results and we made parameter estimations for the dynamics of emotions while reading very positive, very negative, and neutral threads. Individual agent simulations under these dynamics are in agreement with emotional reports after reading threads, shown in Figure 4.16

Furthermore, the statistical analysis of this data revealed new principles that drive emotion dynamics, which can be used in new models within our framework:

- The ground state of emotional relaxation has a positive valence, unlike the 0 value we previously hypothesized.
- Valence has a strong influence on the decision whether to continue or not reading an online discussion, shown in Figure 4.18.
- After replying to a post in an online discussion, users with negative valences experience a positive influence (Figure 4.19).
- The decrease in arousal after writing a post does not reset it to 0, but it decreases by a significant value.

Our empirical analysis of emotion dynamics included data on physiological responses to online interaction, which we used to measure the physiological manifestation of changes in internal emotions. We developed a nonlinear model to estimate the presence of smiling responses, which showed that slightly negative and highly positive images elicit smiles (Figure 4.5). The combined analysis of physiological activity showed how frowning is related to changes in valence while reading threads. In addition, we measured the memory of physiological signals, to find that reading emotional threads leads to longer frowning responses and changes in skin conductance, as explained in Section 4.3.4.

Measuring patterns of generalized emotional expression In addition to experimental data from psychology, we studied some properties of individual emotions through large datasets of emotional expression. We focused on the empirical testing of the Pollyanna hypothesis, which suggests the existence of a positive bias in the usage of emotional words. Our results are the following:

- We analyzed emotional expression in seven different communities, finding that six of them have larger amounts of positive messages, as shown in Figure 5.1.
- In the most general setup, we studied the relation between word frequencies in the whole indexable web, and their valence content according to psychological data. We find that the distributions of word valences in English, German, and Spanish are similarly biased towards words with positive emotional content, as explained in Section 5.2.
- In Section 5.3 we explain how we use information estimations that take into account word context, to find that the information content of a word is inversely proportional to its valence, i.e. positive words carry less information than negative words.

Applications of the framework for collective emotions

The main target of the framework presented in Chapter 1 is to unify future models of emotional interaction in a wide range of online communities. We presented two models designed within this framework, one focused on emotions in product reviews communities (Chapter 2), and another to reproduce emotional persistence in chatroom discussions (Chapter 3). Our aim with this framework is not only to use it for our models, but to reach a larger research community, which can study collective emotions in other communities rather than product reviews and online chats. Here we summarized the explicit applications of our framework, which was used to design agent-based models of emotional dynamics in communities not covered by this thesis.

Emotions in social networks

The current success of social networking sites makes them a very interesting case of online communication. Our first models of emotional interaction included an information field that could be accessed and modified by all the users, representing a collective discussion like those present in fora or chats. On the other hand, social networking sites display

information in an asymmetric manner, allowing users to control which other users have access to the content they produce. Friendship or follower links serve as building blocks for an information filtering mechanism, which delivers content to users based on their social contacts. How this communication is included in our framework depends on the particular design of the social networking site.

One of the first applications of our modeling framework provided a model for emotional interaction in **MySpace**, one of the largest social networking sites. A histogram of the empirical values of valence and arousal for each message, estimated with text analysis techniques, is shown on the left panel of Figure 6.1. The valence distribution has a strong positive bias, consistent with the results discussed in Chapter 5. In addition, the application of similar techniques as the ones used for chatrooms in Chapter 3 reveal the existence of long-range interactions among **MySpace** users, creating cascades of messages through the social network.

This agent-based model, presented in (Šuvakov *et al.*, 2012), builds on empirical findings, simulating emotions on the real social network retrieved from **MySpace**. The activity in this model is driven by the entry process of users into the social network, which was empirically estimated from the available data. To model how agents are affected by the messages they perceive, there are three levels of aggregated information: (i) messages on the agent's wall (h_i), (ii) messages perceived on the friends' walls (\bar{h}_i), (iii) and an aggregation of messages on all walls, i.e. a mean-field information h_{mf} that captures a kind of "atmosphere" of the community.

The network topology is given empirically and does not change through the simulations. What changes instead is the activity of the agents and their emotional states. The right panel of Figure 6.1 shows the aggregated emotional content of the messages exchanged during a simulation. The links are directed, their width is given by the amount of messages sent through that link, and their color represents the average valences of the messages sent along the link. It is obvious that there is a strong bias towards generating messages with positive emotions (indicated by green). Deeper analysis of the simulated networks of discussions reveals that they have similar properties as the original data, in particular related to the cascades of messages in the social network. This model includes a control parameter, which represents an emotional influence external to the community, which can be a news event or a particular consequence of the design of the social networking site. Different simulations of the model show the effect of this parameter in the collective behavior of the community.

Inspired in similar communication mechanisms, another agent-based model was designed to reproduce emotional interaction in blog communities (Mitrović and Tadić, 2012). Sta-

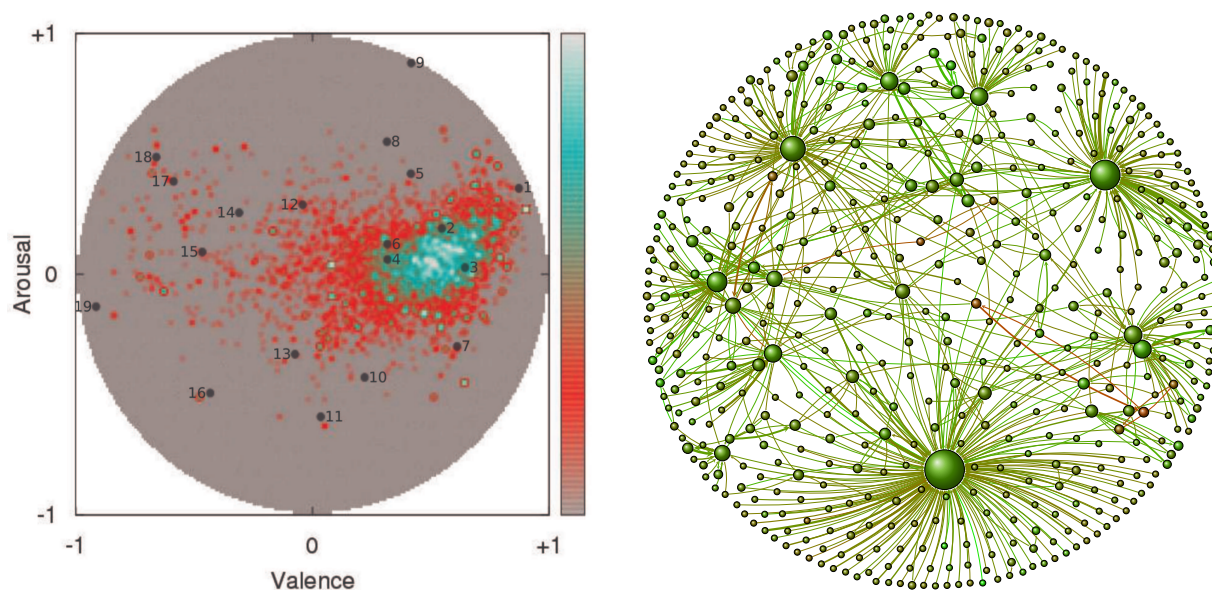


Figure 6.1: Left: Circumplex map of the emotions extracted from the MySpace dataset, as shown in (Šuvakov *et al.*, 2012). The frequency is indicated by the color scale from 0 (gray) to 50 (white). Points illustrate examples of emotions known in psychology (Scherer, 2005). Right: An example of simulated dialogs between the agents as nodes on the part of MySpace network. Two nodes are connected by a directed edge that symbolizes the communication from one agent to another during a simulation. Edge width is proportional to the number of messages sent along the the link, and the edge colors are chosen according to the average valence expressed in these messages. Similarly, node colors represent the average valence expressed by the agent during the simulation time, from red (-1) to green (+1).

tistical analysis of datasets from BBC blogs and digg.com showed how a bipartite network of users and blog posts is divided in communities, which have cascades of messages fueled by negative emotions. An agent-based model within our framework defines emotion dynamics that create the links between users and posts, leading to long-range interactions and an emergent network with a community structure comparable to the empirical system. The model shows how post popularity is related with the prevalence of negative emotions, and how the emergent community structures can be controlled by certain parameter of the model.

Ongoing works apply our modeling framework to emotional activity in product reviews communities that incorporate a trust network. In particular, these directed networks have a structure in which users that produce more helpful and emotional reviews are more central, and therefore more important (Tanase *et al.*, 2012). Empirical analysis shows

that valence influence among users happens at the level of the whole discussion about the product, while arousal influence heavily depends on the trust network. This can be formalized as an agent-based model of model of emotional interaction that shows how emotions flow through the unidirectional links of the trust network.

Emotion modeling in virtual human platforms

Computational social science is not the only field where our modeling framework can be applied. The testable formulation of emotion dynamics is also valuable for the implementation of platforms for online interaction, i.e. a 3D chatting forum. Our framework has been applied to design a model for the data-driven simulation of the emotions of virtual humans (Gobron *et al.*, 2012; Ahn *et al.*, 2012). Virtual human platforms are 3D environments in which users interact through avatars, which are designed to behave in the most realistic manner. One of the challenges of the design of such platforms is nonverbal communication, including facial expression and body language (e.g head, arms gesture). As emotions are one of the driving components behind facial expression, virtual human platforms need to take them into account in order to provide a pseudo-realistic, consistent environment.

The particular model designed for the emotions of virtual humans follows a data-driven approach, in which the verbal expression of the agents is given by user interaction. These expressions are processed via three linguistic analysis techniques, extracting values of emotional polarity, valence, arousal, and dominance, a third dimension of emotions that differentiates anger from fear. A real-time simulation of the emotional state of the avatar is run during the user interaction, and this state can be modified in three different ways: i) by explicit input from the user, ii) by estimations of the user emotional state from her messages, and iii) from emotional dynamics based on the communication of the user with others. This last influence in emotional states is formalized through functions which can be empirically tested within our framework, as we did in Chapter 4.

This model provides the most granular level of description of the microdynamics of emotions, defining emotion dynamics based on dyadic interactions. The exchange of emotional information is model-led through three independent components of the information field, related to information directed to an individual agent, information contained in a discussion, and information related to the feedback of an agent's expression on its own emotional state. The left diagram of Figure 6.2 shows the design of this model, as implemented by the former Virtual Reality Lab (now Immersive Interaction Group) in EPFL (Gobron *et al.*, 2012; Ahn *et al.*, 2012). Each avatar is associated with a simulated agent, for which

its emotional state is constantly updated. This state serves as the input to determine the facial expression and body gestures of the avatar, given previous models of the relation between emotions and nonverbal communication.

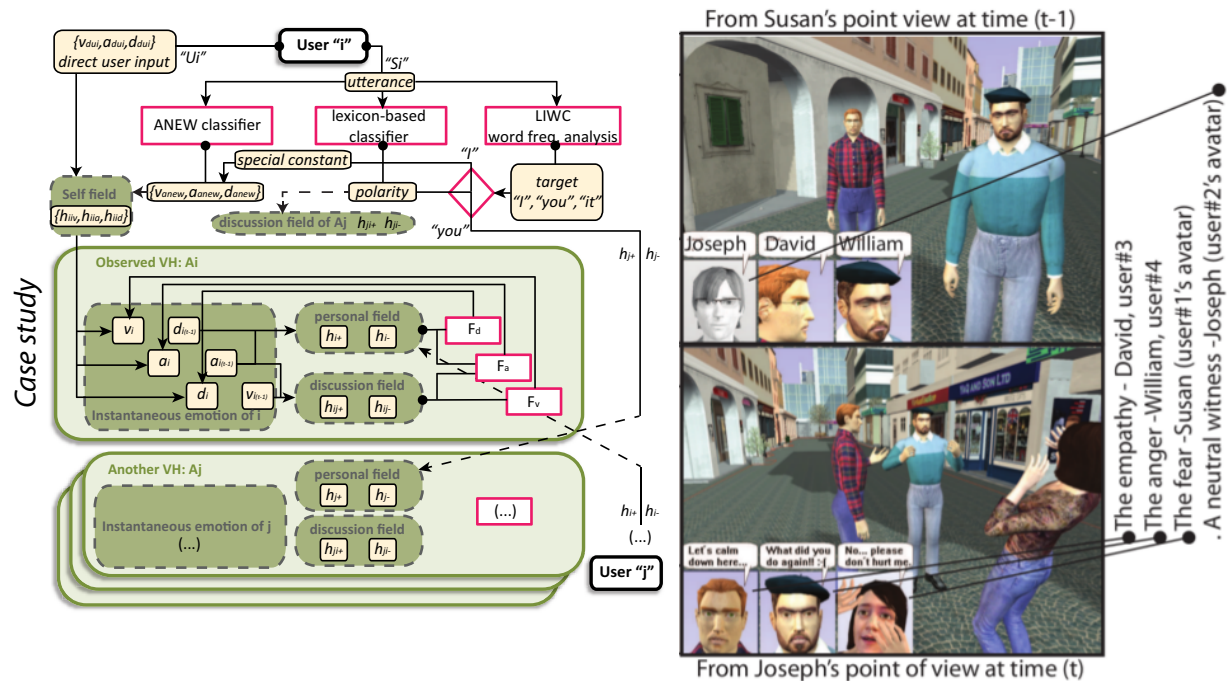


Figure 6.2: Left: schema of the agent internal states and interaction mechanisms in the virtual humans model, as explained in (Gobron *et al.*, 2012; Ahn *et al.*, 2012). Right: example of a user test in which four avatars interact in an emotion eliciting conversation -top image, Susan's view; bottom image, Joseph's view.

This model of virtual human emotions has been fully implemented and tested with participants that interacted in a virtual 3D world. The right picture of Figure 6.2 shows an example of the view of two participants (Susan and Joseph) during a discussion with four avatars at two consecutive moments, with various cases of facial expression and body gestures. Beyond the ICT and industrial applications of this model, experimental setups with such virtual human platform can provide useful empirical data. The detailed interaction between users can be recorded, and their explicit inputs serve as subjective assessments of emotions that can be integrated within our framework. Furthermore, these experiments might serve as an "emotional Turing test", in which we test the validity of these emotion dynamics. Imaging techniques can map the real facial expression of a user to its avatar, and experiment participants should evaluate if they can distinguish between simulated emotional reactions and genuine emotions from experiment participants.

Emotions in online political campaigns

The impact of our contribution is not limited to modeling applications. Our statistical analysis of collective emotions has produced a set of tools that can be applied to other online communities. An example of the application of our analytical approach is the study of emotional expression in online political campaigns. The online interaction between voters reshapes the way politicians conceive their own campaigns, and also provide data for the quantification of the reach and success of a campaign. We recently applied sentiment analysis to the comments in the Youtube videos of presidential candidates in the United States, for which we count with data on the 2008 and 2012 elections (Garcia *et al.*, 2012).

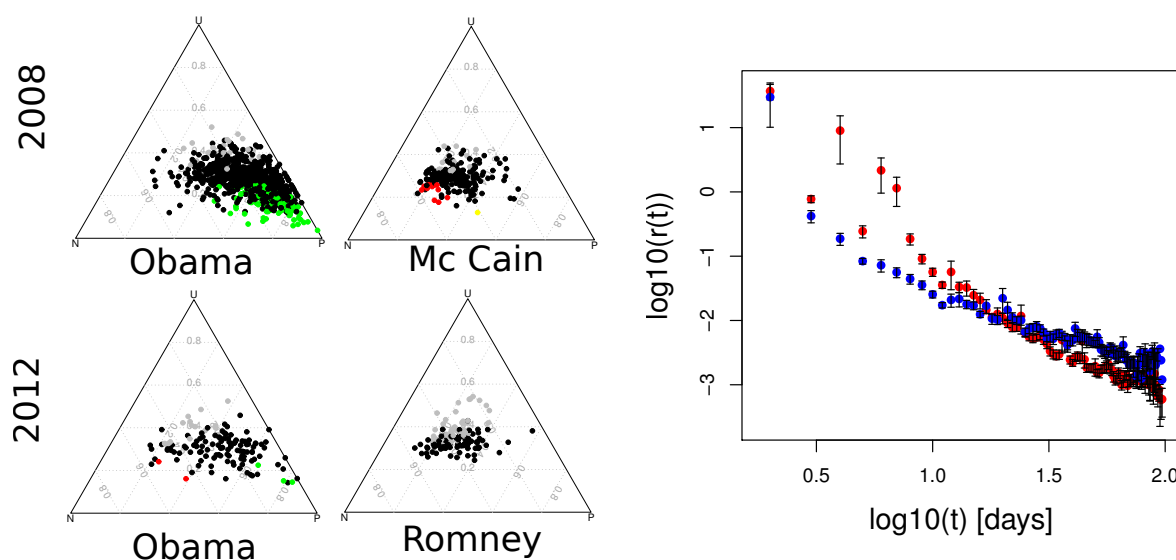


Figure 6.3: Left: triangular projection of collective emotions in the Youtube videos of the democrat and republican candidates of the 2008 and 2012 US elections. Right: time evolution of the mean growth rates of the video views of the 2012 democrat (red) and republican (blue) campaigns.

For each video, we calculated the ratios of emotional expression, described in Section 2.3.2. We focused on the republican and democrat campaigns of 2008 and 2012, summarizing the emotions expressed in the comments for the videos in the triangles shown in the left panel of Figure 6.3. Each point represents a video, and its color is determined according to statistical tests on its emotional ratios. Black points represent videos with comments not differing from normal Youtube comments, green points are videos with noticeable large ratios of positive emotions, and red videos correspond to the negative emotional case. The winning campaign of Obama in 2008 shows a clear bias towards positivity, which seemed

to be also present in the partial data from the 2012 campaign.

We also calculated growth ratios of the video views of the 2012 campaigns, in the same way as we did for the creation of product reviews in Section 2.4. The right panel of Figure 6.3 shows the time evolution of these growth rates for the democrat campaign in red, and for the republican campaign in blue. Our analysis shows that the Youtube users viewing the videos of both campaigns behave in a very different way. Barack Obama's videos have growth rates above 1 for the first week, decaying faster than a power-law afterwards. The case is very different for Mitt Romney's views, which grow slower than Obama's during the first weeks, but keep a more steady growth. Our complex systems approach to the online behavior of potential voters reveals that the critically of their reactions is very different. Some of Obama's videos have explosive growths of views that have a fast decay in their impact on the Youtube user community, while Romney's videos are slowly flowing within a coupled social network. These results quantify the different kinds of *virality* of political videos, measuring the evolution of the attention towards these campaigns.

Relevance in other fields

In every field dealing with human behavior, emotions are a relevant factor due to their importance in almost every human activity, including decision making, cooperation, verbal expression, etc. We provide an overview of some of the future applications that can use our modeling framework or our results on collective emotions.

Emotion aware ICT

In addition to the applications to virtual human platforms we explained above, our modeling framework can improve other types of Information and Communications Technologies, in particular when large amounts of users are involved. One example of this application is the analysis of emotions in Open Source Software communities. These communities interact in two different ways: i) through their collaboration in software development, and ii) through their social interaction in emails, fora, etc. In general, the act of collaboration in software does not need to communicate emotions between developers, but their interaction through online development platforms can be very rich in emotional expression. This emotional interaction can influence the development of the software project, improving or jeopardizing it. For example the pidgin¹ project, an open source instant messaging tool,

¹www.pidgin.im

suffered a fork into two subprojects due to a big fight on a forum discussion ². This poses negative collective emotions as risk factors of the project, and their understanding can help the managers to measure the risk of different design decisions.

Most of the collaborators of Open Source Projects are unpaid, and they usually invest their free time in reporting or fixing bugs. As there is no clear economical motivation, these projects require a different kind of reward mechanism that motivates the participants. Incentive schemes usually include reputation metrics, but emotional rewards among the members might be necessary to maintain an Open Source developer community. In this context, we can ask: Is the productivity of a developer dependent on the positive expression he or she receives from other developers and users of the software? Are bug reports with very negative emotional expression less likely to be solved by developers? These questions can be addressed through the analysis of emotional expression bug reports, on datasets like the ones used in (Zanetti and Schweitzer, 2012).

Other ICT tools that are starting to apply emotion dynamics are dialog systems, computer programs that chat with users of an online community. The agent-based models of our framework can be adapted to work as an input that drives the expression of a dialog system (Rank, 2010), opening the field of emotionally intelligent chatbots. Our findings from the analysis of chatroom communities, explained in Chapter 3, can enrich the behavior of these kind of systems (Skowron and Rank, 2012). For example, the emotional content of an utterance can be selected according to the persistence distribution we found in Section 3.4.2, and the time between utterances of the dialog system can be sampled from the inter-activity time distribution of Figure 3.2. Furthermore, dialog systems might be useful to maintain chat discussions within normal limits of emotional expression, for example by maintaining the emotional persistence we found at the discussion level. These dialog systems can be used to improve informational chatbots, which automatically provide information and answers to user questions.

Adding emotion awareness to computer systems, known as affective computing, can make use of our empirical findings and model results. Data-driven simulations can make use of real-time monitoring of the emotions of users, providing a precise estimation of the emotions elicited by human-computer interactions. Not only sentiment analysis, but other tools like Facial expression of emotions can be tracked by video recognition, as is done by Nviso³, a start-up company that sets up systems that recognize user emotions through a webcam. But this is not limited to gesture based computing, emotions can be measured by physiological signals like the ones studied in Chapter 4, as planned in IBM's Blueeyes

²developer.pidgin.im/ticket/4986

³www.nviso.ch

project⁴. Different devices could serve this purpose, such as a sensitive mouse⁵.

Applications in finance, management, and security

The impact of collective emotions is of great relevance for many other social and economical systems, for example it has been shown that the collective mood in **Twitter** has an influence in general indicators of the stock exchange (Bollen *et al.*, 2011b). In finance, collective emotions might serve as a prediction tool for bubbles, which could be created by generalized wishful thinking; as well as crises and flights to quality, which are influenced by mass hysteria. The ongoing EU Forecasting Financial Crises project⁶(FOC) combines web mining with sentiment analysis to aggregate opinions and emotions in blogs and social media, with the purpose of making predictions about possible economic crises. Agent-based models that explain how the interaction between users leads to collective emotions can help to predict this kind of effects.

The problem of finding the best solution for viral marketing in a network has been raised by researchers from Google and Microsoft together (Even-dar and Shapira, 2007), leading to the hypothesis that emotional interactions are the building blocks of inter-consumer influence. Similarly, online community management is emerging as a new need within the Internet society. This issue, of which many aspects are still unknown, tackles the purpose of maintaining a growing and active community of users, for which encouragement schemes are of great importance (Chen *et al.*, 2010). Websites provide means of influence between users that have nontrivial consequences. In addition, collective emotions can be a very important factor for product acceptance and satisfaction, as shorter product lifecycles have an important impact in consumer feelings about their purchases.

Management sciences can also profit from models of emotional interaction, as the importance of emotions in work relations is key for the economic success of an organization. Efficiency measures can be subject to interaction between the members of a workgroup, coupled with the emotion dynamics of its members. The target of such models is the simultaneous optimization of work performance and emotional atmosphere, formulated under assumptions that depend on the performed task. The emotions of the members of collaboration teams can be monitored in noninvasive ways, supporting decision rules that keep individual emotions in desirable ranges. These applications would require ethical supervision to ensure the subjective well-being of the persons involved. Agent-based models have been proposed in sociology (Flache, 2004) to understand how distributed workgroup

⁴www.almaden.ibm.com/cs/BlueEyes

⁵www.ecse.rpi.edu/homepages/qji/HCI/mouse.html

⁶<http://www.focproject.net/>

emotions might affect productivity, as well as create conflicts.

Beyond the already mentioned applications to financial crises, collective emotions can be seen as one of the driving forces behind certain cyber risks. For example, unforeseeable bursts of activity can drive many users to visit the same page at the same time. This phenomenon, known as the *Digg effect* (Szabo and Huberman, 2010), is originated by the emergence of very popular content in social media. Emotion analysis tools can help to predict when such behavior emerges from user interaction, predicting when this large amount of traffic is going to happen and adapting server bandwidth to cope with it. This same approach might help the monitoring of rumor spreading in social networks, which can have important consequences. Collective emotions are important for civil security, for example the rumor spreading during the riots in Kyrgyzstan in 2010. Fake information about ethnic spread through fora and mobile phone messages, leading to an escalation of conflict in various zones of the country.

Future work

Our approach to collective emotions in online communities can provide further insights on other aspects of online communication, and their impact on other aspects of society. To the current applications to political sciences commented here, models within our framework can add explanations on how emotional interaction in political discussions influences the whole political climate of a community. When election results are known, the analysis of more detailed datasets from websites like **Youtube** can be used to test the success of political campaigns. Such dataset can also provide unique data on individual user motivation, in particular if centered around *videoblogs*, i.e. channels where a single user posts videos of herself on a constant basis. It is remarkable that **Youtube** allows the expression of *dislikes* to these personal videos, and user comments tend to be more negative than normal expression. Advanced versions of models within our framework can integrate how these discussions affect the owner of the channel, and what is their influence on the creation of future videos.

The set of tools and models we presented in this work can be applied to other interesting questions. For example, we plan to study the conditionants for the success of crowdsourcing projects in **kickstarter**, or for non-profit campaigns like the ones in **change.org**. Users discuss about these topics in a wide variety of social networking sites, and a unified approach that studies their emotional expression can shed light on the reasons why some projects are successful, and why some others are not. The dynamics of the popularity of these projects can be analyzed through a complex system approach, aiming at prediction

results that can serve as advice for project and website developers.

In Chapter 3, we studied the question of how online anonymity influences the behavior of users of IRC chatrooms. But there are still many open questions about anonymity and online communication. For example, we noticed that users of the chats tend to have positive emotional expression, but other anonymous communities like `4chan.org` show the opposite behavior, as shown in Chapter 5. The reasons for these different patterns of emotional expression under anonymity can be also explored through experiments, setting participants in scenarios in which there are controlled levels of anonymity. Studying the emotions of the participants of these experiments can be used to further understand how online anonymity influences the way we behave.

Another relevant application for the type of analysis we showed here is the study of how emotions influence collective information processing. Online social interaction leads to users sharing content across social links, a process that might be influenced by the emotional content of the shared information. Initial results show the existence of these dependencies in the spreading of information in `Twitter` (Pfitzner and Garas, 2012). Furthermore, the choice of the links to use to transmit information might depend on the emotions involved in past interactions. This question is specially relevant for the case of Open Source Software communities, which rely on these processes of information exchange to organize software projects. The emotions involved in the interaction between software developers can be used to explore the role of emotions in developer motivation, as well as in the whole software project involving users.

Emotions at a large scale can be also useful to understand how human cultures change through time. We are currently analyzing emotional expression in the comments of `Youtube` videos from the Eurovision Song Contest, relating this emotions to the final results of the contest. Such analysis serves as an indicator of the cultural similarities between European countries, and how international events reshape the collective emotions of one society towards another. We can apply our insights on word emotion frequencies to large scale datasets like the `Google books` dataset. Ranging over decades of written expression in books, we can evaluate if the bias of emotional expression we found is a universal property of the analyzed languages, or if it has changed through time.

Finally, our usage of agent-based models has not finished with the results shown here. Our empirical findings from Chapter 4 should be used to define new models of emotional dynamics, which draw a more realistic picture of collective emotions on the Internet. Thanks to the analytical tractability of our framework, we will be able to derive new collective results from these individual dynamics. Further analysis of data from online communities can serve as an empirical validation for such results, improving our future

understanding of collective emotions in online communities.

Appendix

Appendix A

Supplementary Material to Chapter 2

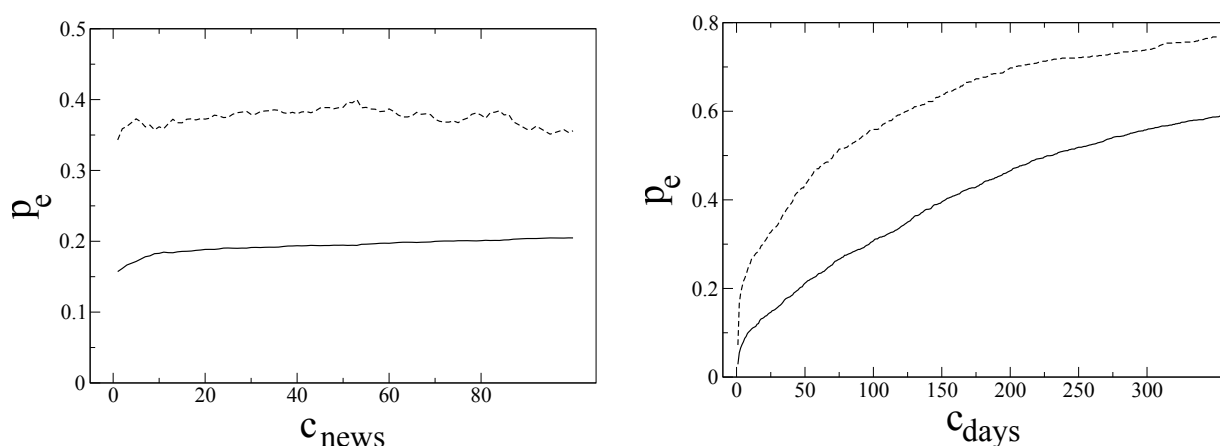


Figure A.1: Left: Probability p_e of finding the peak activity within less than a month after the first review, for products with $m_p > c_{news}$ (dashed line) and $m_p \leq c_{news}$ (solid line), plotted versus various values c_{news} . It is clear that products with more news have always higher probability for an early peak in the review activity. Right: p_e for products without any news item (solid line) and for products with at least one news item (dashed line), plotted versus different values of c_{days}

A product has an early peak of the reviewing activity if t_{peak} is below another cutoff c_{days} , and late otherwise. We explore the role of the news cutoff value c_{news} in the speed of reviewing activity, by measuring the probability p_e of having a peak at time t_{peak} before c_{days} days after the release of the product. This probability corresponds to the area below the probability density function of t_p for the distributions given c_{news} . We ranged through the possible values for c_{news} and c_{days} , calculating p_e on the given t_{peak} distributions. Fig. A.1 shows the values of p_e for products with m_p above and below c_{news} depending on this

news cutoff on the left panel, and depending on c_{days} on the right panel. Fixing $c_{days} = 30$, we find that the value of the cutoff in news c_{news} does have a significant impact on p_e . What we find is a significant difference when $c_{news} = 0$, i.e. the behavior of the reviewer community changes as soon as the product appears in at least one news item. This is consistent across different values of c_{days} , which takes us to fix $c_{news} = 0$ in our analysis. Given this news cutoff value, we explore the influence of c_{days} by ranging through its possible values, as shown in the right panel of A.1. The result is the p_e for each set of products as a pair of monotonously increasing functions, having products with $m_p > 0$ always a larger chance to have $t_{peak} < c_{days}$.

Appendix B

Supplementary Material to Chapter 3

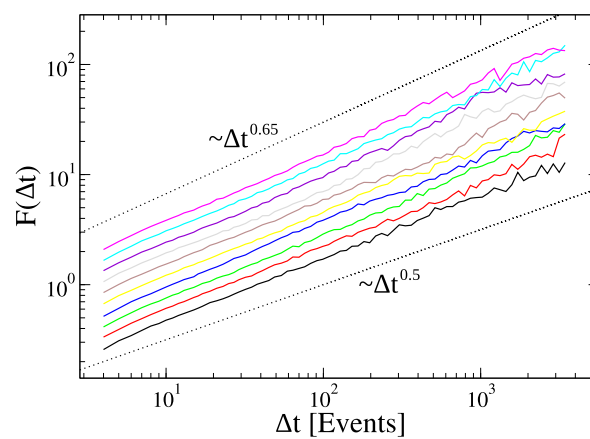


Figure B.1: DFA functions for 10 segments of the collective discussion in an IRC channel, shifted to compare the slopes to the dashed lines.

Appendix C

Supplementary Material to Chapter 4

<i>var</i>	<i>f(var)</i>	$\rho(f(var), v_{prev})$	$\rho(f(var), v_{post})$	$\rho(f(var), v_{post} - v_{prev})$
Zyg	max	0.005°	0.09*	0.06°
	mean	-0.088*	0.088*	0.121*
	min	-0.171***	0.049°	0.151**
Cor	max	0.205***	-0.08°	-0.195***
	mean	0.229***	-0.235***	-0.321***
	min	0.13*	-0.169**	-0.207***
HR	max	-0.01°	0.083°	0.065°
	mean	0.012°	0.057°	0.032°
	min	-0.012°	0.0348°	0.033°
SC	max	-0.064°	0.084°	0.104*
	mean	-0.071°	0.076°	0.102*
	min	-0.084°	0.07°	0.107*

Table C.1: Pearson's correlation coefficients between aggregates f of physiological signals var and values of the valence before reading the thread v_{pre} , after reading the thread v_{post} , and their difference. ° $p < 0.95$, * $p < 0.1$, ** $p < 0.01$, *** $p < 0.001$.

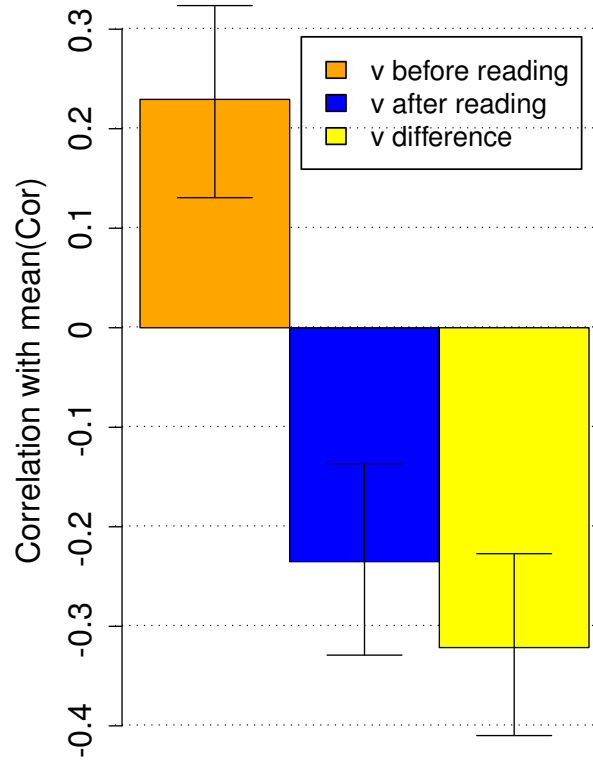


Figure C.1: Barplot of the Pearson's correlation coefficient between the mean activity of the Corrugator Supercilii and the valence values before reading the thread, after reading the thread, and their difference. Errorbars represent 95% confidence intervals. The most determining variable is the difference, but it has a strong overlap with the valence after the thread.

Study	h	$b_0(h)$	$b_1(h)$	$b_2(h)$	$b_3(h)$	A_v	R_v^2	$R_v^2(\epsilon)$
CS2	Pos	0.11***	-0.278***	0.079***	0.043°	0.115	0.641	0.727
CS3	Pos	0.171***	-0.186***	0.039°	-0.081°	0.125	0.693	0.593
CS2	Neu	0.021°	-0.318***	0.014°	0.153*	0.097	0.687	0.592
CS3	Neu	0.081**	-0.406***	-0.027°	0.154°	0.135	0.668	0.908
CS2	Neg	-0.081***	-0.216***	-0.06**	-0.009°	0.114	0.619	0.92
CS3	Neg	-0.135***	-0.369***	-0.013°	0.132*	0.144	0.615	0.855

Table C.2: Parameter estimations and R^2 values for the valence dynamics under the perception of online content. ° $p < 0.9$, * $p < 0.1$, ** $p < 0.01$, *** $p < 0.001$

Study	h	$d_0(h)$	$d_1(h)$	$d_2(h)$	$d_3(h)$	A_a	R_a^2	$R_a^2(\epsilon)$
CS2	Pos	-0.029**	-0.194***	0.016°	-0.001°	0.137	0.34	0.925
CS3	Pos	0.018°	-0.239***	-0.024°	-0.01°	0.182	0.4	0.893
CS2	Neu	-0.081***	-0.188**	-0.133**	-0.096°	0.133	0.491	0.317
CS3	Neu	-0.099***	-0.338***	0.082°	0.174°	0.135	0.503	0.889
CS2	Neg	0.01°	-0.172***	-0.053*	-0.068*	0.120	0.481	0.55
CS3	Neg	0.071***	-0.2***	-0.044°	-0.018°	0.151	0.447	0.851

Table C.3: Parameter estimations and R^2 values for the arousal dynamics under the preception of online content. ° $p < 0.9$, * $p < 0.1$, ** $p < 0.01$, *** $p < 0.001$

Appendix D

Response in the media

Media coverage to “Emotional persistence in online chatting communities”

Nature.com

Language: Online chat behaviours tend to follow social norms



Figure D.1: The article was featured in the frontpage of Nature.com

Medical daily

Social Pressure Makes People Behave in Chat-Rooms

“Contrary to popular belief that people who chat online tend to behave badly, recent research has shown that the pressure to act civil holds good even in online chat-rooms.”

ETH life

Soziales Verhalten trotz Anonymität

“Die Kommunikation von Menschen in Internet-Chatrooms folgt allgemeinen gültigen Mustern, die vorhersagbar sind. Dies zeigen ETH-Forscher unter der Leitung von Frank Schweitzer in einer neuen Studie. ” (Fabio Bergamin)

Radio the voice of Russia

“Are we losing our humanity in the vast spaces of the digital realm? Maybe not. A team of scientists from Austria and Switzerland recently published their work entitled ” Emotional persistence in online chatting communities”. Basically they tackled the myth that depersonalized communication online is drastically different from the real world. [...]

Turns out, online chat activity is not different from other forms of communication. How did they actually gauge that? Well, there’s a lot of science at play here and if you’re interested in specifics, hit up the website for ”Scientific Report” found at nature.com. I’m familiar with both statistics, calculus and sociology, but the full text reminded me of the worst days of my university education - so if you’re not a numbers geek, just bear with me and I’ll give you the basics.

One of the simplest forms of translating communicational dynamics into quantifiable numbers is the time between consecutive messages from one user. We’re talking statistical regularity here, so what we have is a window between two points of time. Well, users responses measured in this manner present a downward-directed exponential distribution. In plain English - people prefer to either respond right away or not give any answer at all - the same distribution applies for other forms of commutation, including email and even snail-mail - but of course, in the latter scenarios, the values are different, giving a larger window in which people are likely to respond. Even simpler - chat dynamics online are the same than when people are talking in real life.” (Peter Lekarev)

20 minuten

Guter Benimm trotz Pseudonym

“Eine ETH-Studie fördert bemerkenswertes zu Tage: In Internet-Chats verhalten sich die meisten Teilnehmer anständig. Auch dann, wenn sie ein Pseudonym benutzen.” (Fee Riebeling)

20min.ch

ETH nimmt Trolle unter die Lupe

“In der Anonymität der Internet-Chatrooms fallen doch nicht alle Hemmungen: Zu diesem Schluss kommt eine Studie der ETH Zürich. Und: Auch im virtuellen Leben sind die Stänkerer in der Minderzahl.”

ORF.at

Sozial trotz Anonymität

“Ein anonymes Auftreten in Internet-Chats geht nach einer Studie nicht mit hemmungslosen Pöbeleien einher. Ganz im Gegenteil: Die meisten Nutzer geben sich positiv, manchmal neutral und selten negativ, haben Forscher bei einer Untersuchung von virtuellen Diskussionsräumen herausgefunden.

Auch in Chatrooms halten sich Menschen weitgehend an soziale Normen, schlossen daraus die Wissenschaftler um David Garcia Becerra von der Eidgenössischen Technischen Hochschule (ETH) Zürich.”

Focus online

Internet: Studie: Auch in anonymen Chats wird kaum gepöbelt

20 minutes

Peu de rôleurs dans les chatrooms

Media coverage to “Positive words carry less information than negative words”

Science Daily

Positive Words: The Glue to Social Interaction

“Words charged with a positive emotional content are used more frequently, thus enhancing human communication.

Scientists at ETH Zurich have studied the use of language, finding that words with a positive emotional content are more frequently used in written communication. This result supports the theory that social relations are enhanced by a positive bias in human communication. The study by David Garcia and his colleagues from the Chair of Systems Design is published in the first issue of the new SpringerOpen journal EPJ Data Science.”

Psych central

Written Communication Typically Takes a Positive Spin

“A new study on the use of language discovers that words with a positive emotional content are more frequently used in written communication. The encouraging tenor is believed to enhance human communication. Researchers believe the finding supports the theory that social relations are enhanced by a positive bias in human communication. The study by David Garcia and his colleagues from the Swiss Federal Institute of Technology is published in the journal EPJ Data Science.” (Rick Nauert)

Psyweb.com

“Positive words, positive feelings, little meaning”

Fachzeitungen.de

“Positive Wörter: der Klebstoff für soziale Interaktion”

Eleconomista.es

“Hablar con palabras positivas: el ‘pegamento’ de la sociedad”

List of Figures

1	User-generated images illustrating the collective emotions around Boxyy . . .	4
2	Schema of social tie strength versus amount of ties	6
3	Diagram of our approach to collective emotions in online communities . . .	17
1.1	Schema of the relation between elements of our framework	28
1.2	Simulated valence distributions and analytical solution	32
1.3	Bifurcation diagrams of arousal	34
1.4	Analytical solution of expression frequency under field	37
1.5	Simulated time series of collective emotional expression	39
1.6	Sample trajectories of arousal and valence	40
2.1	Schema of the paradox of product reviews	46
2.2	Example of product review	48
2.3	Scatter plot of average valence and rating	50
2.4	Analysis of valence distributions per product	51
2.5	Ternary scatter plots of collective emotions towards products	53
2.6	Time series of emotional expression for a product	55
2.7	Histograms of emotion correlations	56
2.8	Mean emotions in correlated product reviews	56
2.9	Example of products with different review dynamics	57
2.10	Review peak size distributions	59
2.11	Distributions of media attention and peak time	60
2.12	Fluctuation scaling of review activity	62

2.13	Distributions of reviews after one week and two years	64
2.14	Q-Q plots of logarithmic growth rates	65
2.15	KS distance versus time	65
2.16	Decay of growth rates given product success	67
2.17	Decays of growth rates of reviews per emotion	68
2.18	Histograms of mean ratings and emotional scores	70
2.19	Distributions of emotions for three products	71
2.20	Influence function of information field on the valence	73
2.21	Time series of simulated products	75
2.22	Comparison between empirical and simulated emotion distributions	75
3.1	Example of chatroom discussion	82
3.2	Inter-activity time distributions	85
3.3	Inter-event time distributions	87
3.4	Autocorrelation function of collective activity	88
3.5	Examples of emotional persistence	90
3.6	Individual persistence distributions	92
3.7	Individual persistence analysis	93
3.8	Emotions averages in chatroom discussions	94
3.9	Collective emotional persistence	95
3.10	Schema of our model for emotions in chatrooms	96
3.11	Simulated persistence and inter-event time distributions	99
3.12	Relation between parameters and emotion fractions	100
3.13	Relation between parameters and persistences	101
3.14	Distributions of persistences given decay parameter	102
3.15	Empirical and simulated long-range dependencies	103
4.1	Schema of the analyzed emotion dynamics	113
4.2	Thread post example from an experiment	115
4.3	Schema of experimental measurements	116

4.4	Examples of smiling responses and model fit	118
4.5	Smiling peak versus picture valence	120
4.6	Mean frowning versus picture valence	120
4.7	Mean EMG signals versus change in valence	121
4.8	Examples of autocorrelation functions for Corrugator Supercilii	123
4.9	Distributions of memory lengths per thread type	124
4.10	Dynamic functions of valence and arousal	128
4.11	Kernel density plot of valence and arousal	129
4.12	Kernel density plots of emotions after reading threads	130
4.13	Mean valence changes while reading threads	133
4.14	Mean arousal changes while reading threads	134
4.15	Combined valence and arousal dynamics	136
4.16	Comparison of simulated and empirical valence and arousal distributions	137
4.17	Activity as a function of arousal	138
4.18	Activity and interest as a function of emotions	139
4.19	Change in emotions after writing posts	140
5.1	Emotional expression biases in online communities	145
5.2	Emotion word clouds	149
5.3	Lexicon valence distributions and bias in usage	150
5.4	Relation between word information content and valence	152
5.5	Diagram of information contents and mean valence	154
6.1	Emotions in MySpace and simulated discussion network	167
6.2	Application of the framework to virtual humans	169
6.3	Emotions in Youtube political campaigns	170
A.1	Analysis of news and peak time cutoffs	179
B.1	Examples of DFA functions	181

C.1	Correlations between frowning activity and valences	184
D.1	Nature frontpage	187

List of Tables

2.1	Expression intensity given valence	74
4.1	Regression results for mean frowning versus valence	121
4.2	Parameter estimations for emotion eigendynamics	127
4.3	Parameter estimations of emotion dynamics in experimental setups	135
5.1	Correlation coefficients of word emotion and information content	152
5.2	Correlations between valence and other estimations of information content	154
5.3	Partial correlations between length, information content, and valence	155
5.4	Partial correlation coefficients of valence and information contents	156
C.1	Correlations between aggregates of physiological signals and valences	183
C.2	Detailed regression results of valence	184
C.3	Detailed regression results of valence	185

Bibliography

Adams, J. (2006). White supremacists, oppositional culture and the World Wide Web. *Social Forces* **84**(2), 759.

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 VII Conference on Articulated Motion and Deformable Objects AMDO 2012*. Mallorca, Spain: Springer, pp. 47–57.

Alexander, J. C.; Eyerman, R.; Giesen, B.; Smelser, N. J.; Sztompka, P. (2004). *Cultural Trauma and Collective Identity*. University of California Press, 304 pp.

Altmann, E.; Kantz, H. (2005). Recurrence time analysis, long-term correlations, and extreme events. *Physical Review E* **71**(5), 56106.

Asur, S.; Huberman, B. A.; Szabo, G.; Wang, C. (2011). Trends in Social Media: Persistence and Decay. In: *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. pp. 434–437.

Augustine, A. A.; Mehl, M. R.; Larsen, R. J. (2011). A Positivity Bias in Written and Spoken English and Its Moderation by Personality and Gender. *Social Psychological and Personality Science* **2**(5), 508–515.

Baayen, R. H.; Piepenbrock, R.; van Rijn, H. (1993). *The CELEX lexical database [CD-ROM]*. *Tech. rep.*, University of Pennsylvania, Linguistic Data Consortium, Philadelphia.

Bar-Tal, D. (2006). Bridging Between Micro and Macro Perspectives in Social Psychology. In: *Bridging social psychology: The benefits of transdisciplinary approaches*, Lawrence Erlbaum Associates. pp. 341–346.

Barabasi, A.-L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature* **435**(7039), 207–11.

- Bentler, P. M. (1969). Semantic space is (approximately) bipolar. *Journal of Psychology: Interdisciplinary and Applied* **71(1)**, 33–40.
- Bernstein, M.; Monroy-Hernández, A.; Harry, D.; André, P.; Panovich, K.; Vargas, G. (2011). 4chan and /b/: An Analysis of Anonymity and Ephemerality in a Large Online Community. In: *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. Barcelona, pp. 50–57.
- Billingsley, P. (1968). *Convergence of probability measures*. No. 1 in Wiley Series in Probability and Statistics, John Wiley & Sons, 277 pp.
- Blitzer, J.; Dredze, M.; Pereira, F. (2007). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In: *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. Prague, Czech Republic, vol. 45, pp. 440–447.
- Bohannon, J. (2011). Social Science for Pennies. *Science* **334**, 2011.
- Bollen, J.; Gonçalves, B.; Ruan, G.; Mao, H. (2011a). Happiness is assortative in online social networks. *Artificial life* **17(3)**, 237–51.
- Bollen, J.; Mao, H.; Zeng, X.-j. (2011b). Twitter mood predicts the stock market. *Journal of Computational Science* **2**, 1–8.
- Boucher, J.; Osgood, C. E. (1969). The pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior* **8(1)**, 1–8.
- Bradley, M.; Cuthbert, B.; Lang, P. J. (1996). Picture media and emotion: Effects of a sustained affective context. *Psychophysiology* **33(1)**, 662–670.
- Bradley, M. M. (2009). Natural selective attention: orienting and emotion. *Psychophysiology* **46(1)**, 1–11.
- Bradley, M. M.; Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings*. Tech. rep., The Center for Research in Psychophysiology, University of Florida.
- Bradley, M. M.; Lang, P. J. (2007). The International Affective Picture System (IAPS) in the Study of Emotion and Attention. In: M. M. Bradley; P. J. Lang (eds.), *Handbook of emotion elicitation and assessment*, Series in Affective Science, Oxford University Press, chap. 2. pp. 29–46.

- Brants, T.; Franz, A. (2009). *Web 1T 5-gram, 10 European languages version 1*. Linguistic Data Consortium, Philadelphia.
- Bryan, C. J.; Walton, G. M.; Rogers, T.; Dweck, C. S. (2011). Motivating voter turnout by invoking the self. *Proceedings of the National Academy of Sciences* **108**(31), 12653–12656.
- Bunde, A.; Eichner, J.; Kantelhardt, J.; Havlin, S. (2005). Long-Term Memory: A Natural Mechanism for the Clustering of Extreme Events and Anomalous Residual Times in Climate Records. *Physical Review Letters* **94**(4), 48701.
- Candia, J.; González, M. C.; Wang, P.; Schoenharl, T.; Madey, G.; Barabási, A.-L. (2008). Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical* **41**(22), 224015.
- Castellano, C.; Fortunato, S.; Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics* **81**(2).
- Chen, Y.; Harper, F. M.; Konstan, J. A.; Li, S. X. (2010). Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens. *American Economic Review* **100**(4), 1358–1398.
- Chmiel, A.; Hoyst, J. A. (2010). Flow of emotional messages in artificial social networks. *International Journal of Modern Physics C* **21**(5), 593–602.
- Chmiel, A.; Sienkiewicz, J.; Thelwall, M.; Paltoglou, G.; Buckley, K.; Kappas, A.; Hoyst, J. A. (2011a). Collective Emotions Online and Their Influence on Community Life. *PLoS ONE* **6**(7), e22207.
- Chmiel, A.; Sobkowicz, P.; Sienkiewicz, J.; Paltoglou, G.; Buckley, K.; Thelwall, M.; Hoyst, J. A. (2011b). Negative emotions boost user activity at BBC forum. *Physica A: Statistical Mechanics and its Applications* **390**(16), 2936–2944.
- Christophe, V.; Rime, B. (1997). Exposure to the social sharing of emotion: Emotional impact, listener responses and secondary social sharing. *European Journal of Social Psychology* **27**(1), 37–54.
- Clauset, A.; Shalizi, C. R.; Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. *SIAM Review* **51**(4), 661.
- Clay-Warner, J.; Robinson, D. T. (2008). *Social Structure and Emotion*. Elsevier, 365 pp.

- Cover, T. M.; Thomas, J. A. (1991). *Elements of Information Theory*, vol. 6 of *Wiley Series in Telecommunications*. New York, USA: John Wiley & Sons, Inc.
- 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.
- Crane, R.; Sornette, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences of the United States of America* **105(41)**, 15649–15653.
- Czaplicka, A.; Chmiel, A.; Hołyst, J. A. (2010). Emotional agents at the square lattice. *Acta Physica Polonica A* **117(4)**, 688–694.
- Czaplicka, A.; Hołyst, J. A. (2012). Modeling of Internet influence on group emotion. *International Journal of Modern Physics C* **23(3)**, 16.
- Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*, vol. 232. John Murray, 477 pp.
- De Choudhury, M.; Counts, S.; Gamon, M. (2012). Not All Moods Are Created Equal! Exploring Human Emotional States in Social Media. In: *Sixth International AAAI Conference on Weblogs and Social Media*. pp. 66–73.
- Dellarocas, C.; Zhang, X. M.; Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing* **21(4)**, 23–45.
- Dewes, C.; Wichmann, A.; Feldmann, A. (2003). An analysis of Internet chat systems. In: *Proceedings of the 2003 ACM SIGCOMM conference on Internet measurement - IMC '03*. New York, New York, USA: ACM Press, p. 51.
- DiMaggio, P.; Hargittai, E.; Neuman, W. R.; Robinson, J. P. (2001). Social implications of the Internet. *Annual review of sociology* **27**, 307–336.
- Dodds, P. S.; Danforth, C. M. (2009). Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents. *Journal of Happiness Studies* **11(4)**, 441–456.
- Dodds, P. S.; Harris, K. D.; Kloumann, I. M.; Bliss, C. A.; Danforth, C. M. (2011). Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter. *PLoS ONE* **6(12)**, e26752.

- Doob, J. L. (1942). The Brownian Movement and Stochastic Equations. *Annals of Mathematics* **43(2)**, 351–369.
- Dorland, William and Alexander, Newman (1980). *Dorland's Medical dictionary*. Saunders Press ; distributed by Holt, Rinehart & Winston, Philadelphia, Pa. : New York .:
- Eisler, Z.; Bartos, I.; Kertész, J. (2008). Fluctuation scaling in complex systems: Taylor's law and beyond1. *Advances in Physics* **57(1)**, 89–142.
- Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In: *Nebraska Symposium on Motivation*. pp. 207–282.
- Even-dar, E.; Shapira, A. (2007). A Note on Maximizing the Spread of Influence in Social Networks. In: *WINE 2007, LNCS 4858*. pp. 281–286.
- Feldman Barrett, L.; Russell, J. A. (1999). The Structure of Current Affect: Controversies and Emerging Consensus. *Current Directions in Psychological Science* **8(1)**, 10–14.
- Ferrer i Cancho, R.; Sole, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences of the United States of America* **100(3)**, 788–91.
- Flache, A. (2004). How May Virtual Communication Shape Cooperation in a Work Team? *Analyse & Kritik* **26**, 258.
- Florian Jaeger, T. (2010). Redundancy and reduction: speakers manage syntactic information density. *Cognitive psychology* **61(1)**, 23–62.
- Fontaine, J. R. J.; Scherer, K. R.; Roesch, E. B.; Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological science* **18(12)**, 1050–1057.
- Fowler, J. H.; Christakis, N. A. (2008). Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. *British Medical Journal* **337(dec04 2)**, 2338–2338.
- Garcia, D.; Mendez, F.; Serdült, U.; Schweitzer, F. (2012). Political Polarization and Popularity in Online Participatory Media: An Integrated Approach. In: *to appear in: Proc PLEAD 12, Workshop in Politics, Elections adn Data*.
- Gardiner, C. W. (2004). *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences*. Springer, Springer Series in Synergetics.

- Gianotti, L. R. R.; Faber, P. L.; Schuler, M.; Pascual-Marqui, R. D.; Kochi, K.; Lehmann, D. (2008). First valence, then arousal: the temporal dynamics of brain electric activity evoked by emotional stimuli. *Brain topography* **20(3)**, 143–156.
- Giles, J. (2012). Computational social science: Making the links. *Nature* **488(7412)**, 448–450.
- Gobron, S.; Ahn, J.; Garcia, D.; Silvestre, Q.; Thalmann, D.; Boulic, R. (2012). An Event-Based Architecture to Manage Virtual Human Non-Verbal Communication in 3D Chatting Environment. In: *Proceedings of the VII Conference on Articulated Motion and Deformable Objects AMDO 2012*. Mallorca, Spain: Springer, pp. 58–68.
- Golder, S. A.; Macy, M. W. (2011). Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures. *Science* **333(6051)**, 1878–1881.
- Goncalves, B.; Perra, N.; Vespignani, A. (2011). Validation of Dunbar’s number in Twitter conversations. *Arxiv preprint arXiv:1105.5170* , 1–8.
- Granovetter, M. S. (1978). Threshold Models of Collective Behavior. *The American Journal of Sociology* **83(6)**, 1420–1443.
- Granovetter, M. S. (1983). The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory* **1**, 201–233.
- Gratch, J.; Marsella, S. (2004). A domain-independent framework for modeling emotion. *Cognitive Systems Research* **5(4)**, 269–306.
- Gratch, J.; Marsella, S.; Wang, N.; Stankovic, B. (2009). Assessing the validity of appraisal-based models of emotion. *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops* , 1–8.
- Griffiths, T. L. (2011). Rethinking language: how probabilities shape the words we use. *Proceedings of the National Academy of Sciences of the United States of America* **108(10)**, 3825–6.
- Grinstein, G.; Linsker, R. (2008). Power-law and exponential tails in a stochastic priority-based model queue. *Physical Review E* **77(1)**, 2007–2009.
- Hauser, M. D.; Chomsky, N.; Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* **298(5598)**, 1569–79.
- Havlin, S. (1995). The distance between Zipf plots. *Physica A: Statistical and Theoretical Physics* **216(1-2)**, 148–150.

- Heath, C.; Bell, C.; Steinberg, E. (2001). Emotional Selection in Memes : The Case of Urban Legends. *Journal of Personality* **81(6)**, 1028–1041.
- Hedström, P. (2006). Experimental macro sociology: predicting the next best seller. *Science* **311(5762)**, 786–787.
- Heise, D. R. (2006). *Expressive Order: Confirming Sentiments in Social Actions*. Springer, 178 pp.
- Hill, A. L.; Rand, D. G.; Nowak, M. A.; Christakis, N. A. (2010). Emotions as infectious diseases in a large social network: the SISa model. *Proceedings of The Royal Society B* **277(1701)**, 3827–35.
- Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *American Society of Civil Engineering* **116**, 770–808.
- Jack, R. E.; Garrod, O. G. B.; Yu, H.; Caldara, R.; Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences of the United States of America* **109(19)**, 7241–4.
- Jones, C. M.; Troen, T. (2007). *Biometric valence and arousal recognition*. OZCHI '07, New York, New York, USA: ACM Press, 191–194 pp.
- Kantelhardt, J. W. (2009). Fractal and Multifractal Time Series. In: *Encyclopedia of Complexity and Systems Science*, Springer. p. 59.
- Kappas, A.; Tsankova, E.; Theunis, M.; Kuester, D. (2011). CyberEmotions: Subjective and physiological responses elicited by contributing to online discussion forums. In: *Poster presented at the 51st Annual Meeting of the Society for Psychophysiological Research, Boston, Massachusetts*.
- Kloumann, I. M.; Danforth, C. M.; Harris, K. D.; Bliss, C. A.; Dodds, P. S. (2012). Positivity of the english language. *PloS one* **7(1)**, e29484.
- Knowles, E. S.; Nathan, K. T. (1997). Acquiescent Responding in Self-Reports: Cognitive Style or Social Concern? Eric S. Knowles and Kobi T. Nathan. *Journal of Research in Personality* **301(31)**, 293–301.
- Koenig, M. D.; Tranel, D. (2008). Prefrontal cortex damage abolishes brand-cued changes in cola preference. *Social cognitive and affective neuroscience* **3(1)**, 1–6.

- Konnikova, M. (2012). Humanities aren't a science. Stop treating them like one. URL <http://blogs.scientificamerican.com/literally-psyched/2012/08/10/humanities-arent-a-science-stop-treating-them-like-one/>. Scientific American blogs.
- Kosmidis, K.; Kalampokis, A.; Argyrakis, P. (2006). Statistical mechanical approach to human language. *Physica A: Statistical Mechanics and its Applications* **366**, 495–502.
- Kucera, H.; Francis, W. N. (1967). *Computational analysis of presentday American English*. Providence: Brown University Press.
- Kucuktunc, O.; Cambazoglu, B. B.; Weber, I.; Ferhatosmanoglu, H. (2012). A large-scale sentiment analysis for Yahoo! answers. In: *Proceedings of the fifth ACM international conference on Web search and data mining - WSDM '12*. New York, New York, USA: ACM Press, p. 633.
- Kuppens, P.; Oravecz, Z.; Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of personality and social psychology* **99(6)**, 1042–60.
- Latane, B. (1981). The Psychology of Social Impact. *American Psychologist* **36(4)**, 343–356.
- Lazer, D.; Pentland, A.; Adamic, L. A.; Aral, S.; Barabasi, A.-L.; Brewer, D.; Christakis, N. A.; Contractor, N.; Fowler, J. H.; Gutmann, M.; Jebara, T.; King, G.; Macy, M. W.; Roy, D.; Van Alstyne, M. (2009). Computational social science. *Science* **323(5915)**, 721–723.
- Leskovec, J.; Adamic, L. A.; Huberman, B. A. (2007). The dynamics of viral marketing. *ACM Transactions on the Web* **1(1)**, 39.
- Leskovec, J.; Backstrom, L.; Kleinberg, J. (2009). Meme-tracking and the dynamics of the news cycle. In: *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09*. New York, New York, USA: ACM Press, p. 497.
- Li, X.; Hitt, L. M. (2008). Self-Selection and Information Role of Online Product Reviews. *Information Systems Research* **19(4)**, 456–474.

- Lodewyckx, T.; Tuerlinckx, F.; Kuppens, P.; Allen, N.; Sheeber, L. (2011). A hierarchical state space approach to affective dynamics. *Journal of mathematical psychology* **55(1)**, 68–83.
- Lorenz, J. (2007). Continuous opinion dynamics under bounded confidence: a survey. *International Journal of Modern Physics C* **18(12)**, 1819–1838.
- Lorenz, J. (2009). Universality in movie rating distributions. *The European Physical Journal B* **71(2)**, 251–258.
- Lorenz, J.; Battiston, S.; Schweitzer, F. (2009). Systemic risk in a unifying framework for cascading processes on networks. *The European Physical Journal B* **71(4)**, 441–460.
- Malmgren, R. D.; Stouffer, D. B.; Motter, A. E.; Amaral, L. A. N. (2008). A Poissonian explanation for heavy tails in e-mail communication. *Proceedings of the National Academy of Sciences of the USA* **105(47)**, 18153–18158.
- Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association* **46(253)**, 68–78.
- de Melo, C.; Zheng, L.; Gratch, J. (2009). Expression of Moral Emotions in Cooperating Agents. *Intelligent Virtual Agents* , 301–307.
- Miller, N. E. (1961). Some recent studies of conflict behavior and drugs. *American Psychologist* **16(1)**, 12–24.
- Mitrović, M.; Tadić, B. (2009). Bloggers behavior and emergent communities in Blog space. *The European Physical Journal B* **73(2)**, 293–301.
- Mitrović, M.; Tadić, B. (2012). Dynamics of bloggers communities: Bipartite networks from empirical data and agent-based modeling. *Physica A: Statistical Mechanics and its Applications* **391(21)**, 5264–5278.
- Nahl, D.; Bilal, D. (2007). *Information and Emotion: The Emergent Affective Paradigm in Information Behavior Research and Theory*. Information Today, Inc., 359 pp.
- Norman, G. J.; Norris, C. J.; Gollan, J.; Ito, T. a.; Hawkley, L. C.; Larsen, J. T.; Cacioppo, J. T.; Berntson, G. G. (2011). Current Emotion Research in Psychophysiology: The Neurobiology of Evaluative Bivalence. *Emotion Review* **3(3)**, 349–359.

- O'Connor, B.; Balasubramanyan, R.; Routledge, B. R.; Smith, N. A. (2010). From Tweets to Polls : Linking Text Sentiment to Public Opinion Time Series. In: *Fourth International AAAI Conference on Weblogs and Social Media*. pp. 122–129.
- 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.
- Oravecz, Z.; Tuerlinckx, F. (2011). The linear mixed model and the hierarchical Ornstein-Uhlenbeck model: some equivalences and differences. *The British journal of mathematical and statistical psychology* **64(Pt 1)**, 134–60.
- Oravecz, Z.; Tuerlinckx, F.; Vandekerckhove, J. (2011). A hierarchical latent stochastic differential equation model for affective dynamics. *Psychological Methods* **16(4)**, 468–490.
- Osgood, C. E. (1964). Semantic Differential Technique in the Comparative Study of Cultures1. *American Anthropologist* **66(3)**, 171–200.
- Paltoglou, G.; Gobron, S.; Skowron, M.; Thelwall, M.; Thalmann, D. (2010). Sentiment analysis of informal textual communication in cyberspace. In: H. Springer (ed.), *In Proc. Engage 2010, Springer LNCS State-of-the-Art Survey*. Citeseer, pp. 13–25.
- Paltoglou, G.; Theunis, M.; Kappas, A.; Thelwall, M. (2012). Prediction of valence and arousal in forum discussions. Forthcoming.
- Pang, B.; Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval* **2(12)**, 1–135.
- Peng, C.-K.; Buldyrev, S. V.; Havlin, S.; Simons, M.; Stanley, H. E.; Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E* **49(2)**, 1685–1689.
- Pennebaker, J. W.; Francis, M. E.; Booth, R. J. (2001). *Linguistic Inquiry and Word Count: LIWC 2001*. Erlbaum Publishers.
- Pfifner, R.; Garas, A. (2012). Emotional divergence influences information spreading in Twitter. In: *AAAI ICWSM 2012*. pp. 2–5.
- Piantadosi, S. T.; Tily, H.; Gibson, E. (2011a). Reply to Reilly and Kean: Clarifications on word length and information content. *Proceedings of the National Academy of Sciences* **108(20)**, 109–109.

- Piantadosi, S. T.; Tily, H.; Gibson, E. (2011b). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences* **108**(9), 3526.
- Plassmann, H.; O'Doherty, J.; Shiv, B.; Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences of the United States of America* **105**(3), 1050–1054.
- Preston, S. D.; de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *The Behavioral and brain sciences* **25**(1), 1–20; discussion 20–71.
- Rains, S. A. (2007). The Impact of Anonymity on Perceptions of Source Credibility and Influence in Computer-Mediated Group Communication: A Test of Two Competing Hypotheses. *Communication Research* **34**(1), 100–125.
- Rank, S. (2010). *Docking agent-based simulation of collective emotion to equation-based models and interactive agents*. New York, New York, USA: ACM Press, 8 pp.
- Redondo, J.; Fraga, I.; Padron, I.; Comesana, M. (2007). The Spanish adaptation of ANEW (affective norms for English words). *Behavior research methods* **39**(3), 600–5.
- Reilly, J.; Kean, J. (2007). Formal distinctiveness of high- and low-imageability nouns: analyses and theoretical implications. *Cognitive science* **31**(1), 157–68.
- Reilly, J.; Kean, J. (2011). Information content and word frequency in natural language: word length matters. *Proceedings of the National Academy of Sciences of the United States of America* **108**(20), E108; author reply E109.
- Rime, B. (2009). Emotion Elicits the Social Sharing of Emotion: Theory and Empirical Review. *Emotion Review* **1**(1), 60–85.
- Rime, B.; Finkenauer, C.; Luminet, O.; Zech, E.; Philippot, P. (1998). Social Sharing of Emotion: New Evidence and New Questions. *European Review of Social Psychology* **9**, 145–189.
- Riva, G. (2005). Virtual reality in psychotherapy: review. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society* **8**(3), 220–30; discussion 231–40.
- Robinson, J. P.; Shaver, P. R.; Wrightsman, L. S. (1991). *Measures of Personality and Social Psychological Attitudes*. Elsevier, 750 pp.

- Rosen, S. (1981). The Economics of Superstars. *The American Economic Review* **71**(5), 845–858.
- Rozin, P.; Berman, L.; Royzman, E. (2010). Biases in use of positive and negative words across twenty natural languages. *Cognition & Emotion* **24**(3), 536–548.
- Russell, J. A. (1979). Affective Space is Bipolar. *Journal of Personality and Social Psychology* **37**(3), 345–356.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology* **39**(6), 1161–1178.
- Rybski, D.; Buldyrev, S. V.; Havlin, S.; Liljeros, F.; Makse, H. A. (2009). Scaling laws of human interaction activity. *Proceedings of the National Academy of Sciences* **106**(31), 12640.
- Rybski, D.; Buldyrev, S. V.; Havlin, S.; Liljeros, F.; Makse, H. a. (2011). Communication activity in social networks: growth and correlations. *The European Physical Journal B* **84**(1), 147–159.
- Salganik, M. J.; Dodds, P. S.; Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science* **311**(5762), 854–856.
- Sander, D.; Grandjean, D.; Scherer, K. R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural networks : the official journal of the International Neural Network Society* **18**(4), 317–52.
- Sassenberg, K.; Boos, M. (2003). Attitude Change in Computer-Mediated Communication: Effects of Anonymity and Category Norms. *Group Processes & Intergroup Relations* **6**(4), 405–422.
- Scherer, K. R. (2004). Emotions in everyday life: probability of occurrence, risk factors, appraisal and reaction patterns. *Social Science Information* **43**(4), 499–570.
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information* **44**(4), 695–729.
- Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition & Emotion* **23**(7), 1307–1351.

- Scherer, K. R.; Schorr, A.; Johnstone, T. (2001). *Appraisal processes in emotion: Theory, methods, research. Series in affective science*. New York: Oxford University Press.
- Schupp, H. T.; Schmälzle, R.; Flaisch, T.; Weike, A. I.; Hamm, A. O. (2012). Affective picture processing as a function of preceding picture valence: An ERP analysis. *Biological psychology* , 1–7.
- Schweitzer, F. (2003). *Brownian Agents and Active Particles. Collective Dynamics in the Natural and Social Sciences*. Springer Series in Synergetics, Berlin: Springer, 1st editio edn., 420 pp.
- 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.
- Sebastián, N., M.; Carreiras, M., A.; Cuetos, M., F. (2000). *LEXESP: Léxico informatizado del español*. Barcelona: Ediciones de la Universitat de Barcelona.
- Senecal, S.; Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing* **80(2)**, 159–169.
- Shin, J. (2009). Information accumulation system by inheritance and diffusion. *Physica A: Statistical Mechanics and its Applications* **388(17)**, 3593–3599.
- Shin, J. K.; Lorenz, J. (2010). Tipping diffusivity in information accumulation systems: more links, less consensus. *Journal of Statistical Mechanics: Theory and Experiment* **2010(06)**, 6005–6020.
- Simon, H. A. (1997). *Models of Bounded Rationality, Vol. 3: Emperically Grounded Economic Reason*. MIT Press, 479 pp.
- Skowron, M.; Rank, S. (2012). From dyads to group interactions with affective dialog systems. In: *AISB/IACAP World Congress 2012 - Alan Turing 2012, Symposium on Linguistic and Cognitive Approaches To Dialog Agents - LaCATODA 2012*.
- Smith, E. R.; Conrey, F. R. (2007). Agent-based modeling: a new approach for theory building in social psychology. *Personality and social psychology review* **11(1)**, 87–104.
- Sobkowicz, P.; Sobkowicz, A. (2010). Dynamics of hate based Internet user networks. *The European Physical Journal B* **73(4)**, 633–643.

- Stone, P. J.; Dunphy, D. C.; Smith, M. S. (1966). *The General Inquirer: a Computer Approach to Content Analysis*. Oxford, England: M.I.T Press, 651 pp.
- Suler, J. R. (2000). Psychotherapy in Cyberspace: A 5-Dimensional Model of Online and Computer-Mediated Psychotherapy. *CyberPsychology & Behavior* **3(2)**, 151–159.
- Szabo, G.; Huberman, B. A. (2010). Predicting the popularity of online content. *Communications of the ACM* **53(8)**, 80.
- Szell, M.; Thurner, S. (2010). Measuring social dynamics in a massive multiplayer online game. *Social Networks* **32(4)**, 313–329.
- Taboada, M.; Brooke, J.; Voll, K. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics* **37(2)**, 267–307.
- Tanase, D.; Garcia, D.; Garas, A.; Schweitzer, F. (2012). Emotional reviews boost user visibility in Social Recommender Systems. *Forthcoming* .
- Tellegen, A.; Watson, D.; Clark, L. A. (1999). On the Dimensional and Hierarchical Structure of Affect. *Psychological Science* **10(4)**, 297–303.
- Thelwall, M.; Buckley, K.; Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology* **62(2)**, 406–418.
- Thelwall, M.; Buckley, K.; Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology* **63(1)**, 163–173.
- Thelwall, M.; Buckley, K.; Paltoglou, G.; Cai, D.; Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology* **61(12)**, 2544–2558.
- Van Mieghem, P. (2011). Human Psychology of Common Appraisal: The Reddit Score. *IEEE Transactions on Multimedia* **13(6)**, 1404–1406.
- Van Rijsbergen, C. J.; Robertson, S. E.; Porter, M. F. (1980). *New models in probabilistic information retrieval*. Cambridge, England: Computer Laboratory, University of Cambridge, 123 pp.
- Vo, M. L.-H.; Conrad, M.; Kuchinke, L.; Urton, K.; Hofmann, M. J.; Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior research methods* **41(2)**, 534–8.

- Šuvakov, M.; Garcia, D.; Schweitzer, F.; Tadić, B. (2012). Agent-based simulations of emotion spreading in online social networks. *Forthcoming* .
- Vujovic, L.; Tsankova, E.; Kappas, A.; Kuester, D. (2011). Avatars in a nonymous vs. anonymous online setting. In: *8th Conference of the Media Psychology Division (German Psychological Society)*. Bremen, Germany.
- Wang, C.; Huberman, B. A. (2012). Long trend dynamics in social media. *EPJ Data Science* **1(2)**, 8.
- Wang, F.; Yamasaki, K.; Havlin, S.; Stanley, H. E. (2006). Scaling and memory of intraday volatility return intervals in stock markets. *Physical Review E* **73(2)**, 1–8.
- Watson, D.; Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin* **98(2)**, 219–235.
- Wilson, T.; Wiebe, J.; Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In: *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT '05*. Morristown, NJ, USA: Association for Computational Linguistics, pp. 347–354.
- Wu, F.; Huberman, B. A. (2007). Novelty and collective attention. *Proceedings of the National Academy of Sciences of the United States of America* **104(45)**, 17599–601.
- Wu, Y.; Zhou, C.; Xiao, J.; Kurths, J.; Schellnhuber, H. J. (2010). Evidence for a bimodal distribution in human communication. *Proceedings of the National Academy of Sciences of the United States of America* **107(44)**, 18803–8.
- Yik, M.; Russell, J. A.; Barret, L. F. (1999). Structure of Self-reported Affect: Integration and Beyond. *Journal of Personal and Social Psychology* **77(3)**, 600–619.
- Yik, M.; Russell, J. A.; Steiger, J. H. (2011). A 12-point circumplex structure of core affect. *Emotion* **11(4)**, 705–731.
- Zanetti, M. S.; Schweitzer, F. (2012). A Network Perspective on Software Modularity. In: *Proceedings of the 1st International Workshop on Complex Sciences in the Engineering of Computing Systems*.
- Zhu, H.; Huberman, B.; Luon, Y. (2012). To switch or not to switch. In: *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. New York, New York, USA: ACM Press, p. 2257.

Zipf, G. K. (1935). *The Psycho-Biology of Language*. Oxford, England:: Houghton Mifflin, 336 pp.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. New York: Addison-Wesley.

Zizek, S. (1989). *The Sublime Object of Ideology*. Verso, 240 pp.

Zizek, S. (1998). Cyberspace, or, How to Traverse the Fantasy in the Age of the Retreat of the Big Other. *Public Culture* **10(3)**, 483–513.

