universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

'Text Mining and Computational Modeling of Collective Emotions in Online Discussions'

verfasst von / submitted by

Raffael Andre BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 013
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Middle European interdisciplinary master's programme in Cognitive Science (MEi:CogSci)
Betreut von / Supervisor:	Dr. David Garcia

Acknowledgements

I would like to thank Dr. David Garcia of the Complexity Science Hub Vienna for his supervision of this thesis, and his colleagues Max Pellert and Simon Schweighofer for their additional support. I am further grateful to Dr. Paolo Petta of the Austrian Research Institute for Artificial Intelligence for his invaluable guidance through the MEi:CogSci program. Lastly, I would like to thank all of my friends and colleagues who supported me during my years of study.

Contents

Introduction

Collective emotions are macro-level affective phenomena defined as emotional states shared and spread by a group of individuals through interaction (Goldenberg et al., 2017). Although it is possible to track these emotional dynamics individually, they may show emergent properties that cannot be captured on the micro-level, such as a decrease in variability. In order for these properties to emerge, individuals have to appraise the same event in a similar way. Hence, they can also be more formally described as "synchronous convergence in affective responding across individuals toward a specific event or object" (von Scheve and Salmela, 2014). Collective emotions appear online and were studied in the *Cyberemotions* project (Holyst, 2016). These affective processes occur in social networks (Garas et al., 2012) and involve online communication between individuals that are influenced by emotional states (Kappas, 2017). The increased use of social networks on a global scale raises questions about the influence of collective emotions on such interpersonal online interactions. It was argued that emotions cannot be studied without relying on a multidisciplinary research paradigm (Kappas, 2002). Cognitive science is an inherently interdisciplinary field of study that allows the combination of multiple disciplines, such as social science, psychology, computer science, and statistics, to study cognitive phenomena. Emotions are a central topic in cognitive science (Prinz, 2012). In the context of this master's thesis, we investigate collective emotions as cognitive phenomena that occur in online discussions. We use a computational approach to the human mind taken from cognitive science, and combine it with methods from social data science and computational social science, in order to study the effects of collective emotions on discussion dynamics and behaviors.

Online forums are core components of many social media platforms and provide a place for discussions to take place. In these communities, users can communicate with each other by expressing their opinions, thoughts, or emotions on a particular topic as written comments or posts, either in response to the topic itself, or to one of the other already published comments. Although such ordered collections of forum posts become less visible or harder to access over time when the discussed topic loses relevance, for example in discussions of news articles, they generally remain online indefinitely. These digital traces of online communication are thus pertinent means to investigate phenomena such as emotional contagion over time, because they allow for analysis on a much longer scale than traditional study environments (Aragón et al., 2017a). For this reason, this form of online discussion received increased attention in recent research in the field of computational social science (Chen, 2015; Aragón et al., 2017b). Online communities as a whole where shown to be driven by emotional expressions that create and module collective emotional states (Chmiel et al., 2011). This phenomenon of emotional contagion between users was formalized by the Cyberemotions agent-based modeling framework that describes the role of valence and arousal in the expression of emotions (Schweitzer and Garcia, 2010) based on a circumplex model of emotion (Russell, 1980).

Implementing the Cyberemotions framework, we model the emergence of collective emotions in online discussions using agent-based modeling, a method that abstracts a phenomenon in terms of agents and their interactions (Wilensky and Rand, 2015). It is a tool designed to conceptually bridge micro-level assumptions about individual behavior to observed behavioral patterns of groups, and established itself in interdisciplinary research paradigms that try to capture complex, dynamic, and interactive processes, such as social psychology (Smith and Conrey, 2007). For this reason, we believe it is especially well suited to model the emergence of collective emotions, because they are likewise complex and dynamic patterns observed on a macro-level that emerge from the actions and interactions of individuals. Agent-based modeling allows us to simulate individual agents, their emotional states, and their means of communication by expressing their emotions. Social simulations help us generate new insights into the dynamics of collective emotions that can be verified using real-world data.

In order to empirically validate our implemented model, we analyze the *One Million Posts* corpus, a data set of forum posts that was collected in the comment sections of derStandard.at, an Austrian newspaper website (Schabus et al., 2017). It contains the original news articles and their anonymized discussions as discussion threads, sets of comments ordered by their time and date of posting and information about the relationships between posts. Sentiment analysis serves as a quantitative text analysis tool to categorize the words of a text and is thus able to evaluate the emotional content of an online message based on the use of emotion words (Tausczik and Pennebaker, 2009). Applying sentiment analysis to the One Million Posts corpus, we aim to generate new insights about the emotion dynamics in online discussion forums, especially in context of upvote and downvote mechanisms that allow users to express agreement or disagreement with each other. Following the principle of synchronous convergence in affective responding, we not only expect to see correlations between sentiments of various posts in a discussion, but also correlations between the sentiment of a post and its polarization among other users measured as the relationship between upvotes and downvotes.

The sentiment analysis software LIWC2015 does not only provide insights about the emotional states of discussion participants, but also about their cognitive activity (Pennebaker et al., 2015). This allows us too compare our analysis results to predictions of dual process theories that postulate the respective use of two distinct cognitive systems. Initially described as hot and cold cognition (Abelson and Carroll, 1965) and popularized as system 1 and system 2 (Stanovich and West, 2000; Kahneman, 2011), one implies fast but emotional inference, whereas the other represents rational but slow thinking. Assuming users of online forums use either one or the other of the two when writing their posts, this dichotomy may be visible in online forum posts when we compare their emotional and cognitive content. In addition, our data set also contains a limited number of comment annotations by forum moderators that categorize the sentiment of the user who wrote the comment. We use these annotations for a validity check of our sentiment analysis data. Comparing them to the results of our sentiment analysis may verify its accuracy about how emotional posts are perceived by others.

The goal of our research is to contribute to the scientific understanding of online collective emotions by answering the following questions:

1. Can we replicate the properties of collective emotions in online discussions using an agent-based model of emotion dynamics and social interaction?

2. Can we predict the emotional content and affective reaction of a post based on the data of previous posts in the discussion?

3. Do the affective properties of a news article predict the collective emotional dynamics of its discussion?

4. Can we draw conclusions about the use of either system 1 or 2 by users based on the measured degree of cognitive complexity in their posts?

Our results and generated insights may then be of value to other researchers from various scientific fields studying online emotion dynamics and collective emotions, such as social psychology, computational social science, artificial intelligence, and of course cognitive science.

In the first two sections of this thesis after this introduction, we provide a theoretical background to the study of emotions in the interdisciplinary context of cognitive science, with an elaboration of our computational approach, social interactionism, the definition of collective emotions, and dual-process theories. We also summarize the Cyberemotions agent-based modeling framework that we base our model on. Following up on the theory part of our master's thesis, the next chapter outlines our computational model of collective emotions. It details agent-based modeling as a technique to generate hypotheses, and we provide an detailed description of our implementation of the Cyberemotions framework in the programming language Python (Van Rossum and Drake Jr, 1995). In the next chapter, we describe our dataset, The One Million Posts corpus (Schabus et al., 2017), and our method to analyze affect in online discussions using the quantitative text analysis tool LIWC (Pennebaker et al., 1999). Lastly, we present the simulation results collected by running our implemented model, and our analysis results after applying sentiment analysis on our dataset. Comparing the model data to real-world data allows us to draw conclusions about the validity of our model and to answer our research questions.

Cognitive science of emotions

Emotions are a central topic in cognitive science. They interact with basic cognitive systems such as perception and memory (LeBlanc et al., 2015), shape behavior (Baumeister et al., 2007), signal people's reaction to events in the world (Ellsworth and Scherer, 2003), and play a crucial role in decision-making processes (Schwarz, 2000; Loewenstein and Lerner, 2003; Lerner et al., 2015). While the cognitive science of emotions also encompasses non-cognitive approaches (Prinz, 2012), we base our research on cognitive theories of emotion and computational theories of the mind, because they constitute the established scientific basis for the practice of computational modeling in cognitive science and related disciplines, such as computational psychology (Sun, 2008a). The classic cognitivist approach presumes that the human mind consists of special kinds of cognitive processes and provided first outlines for modeling these processes computationally (Hebb, 1949; Turing, 1950). In this computational context "minds are what brains do" (Minsky, 1988) and brains can be considered to be information processors. In such an approach "emotions may be conceptualized as a dependent variable influenced by processes such as appraisal, or as an independent variable that influences information-processing" (Matthews, 1997). Potential substitutes to computational theory include 'systems design' that asks for the purpose of emotions in the cognitive system as a whole, or phenomenology of emotions which focuses on the subjective experience of individuals rather than on the material properties of events described as emotional. We believe neither provide suitable alternatives for this study. The former because we are interested in patterns of interactions and group behavior and not in individual cognitive systems, and the latter because it would require a qualitative rather than a quantitative research approach.

Generally, it can be differentiated between three cognitive approaches to emotions: 1. action-readiness, 2. core-affect, and 3. communicative theories of emotions (Oatley and Johnson-Laird, 2014). We consider these three to be principles that may be invoked individually or in combination when forming cognitive theories of emotion.

1. The *action-readiness* principle bases emotions on stimulus-response pairs that indicate states of 'readiness' for certain kinds of actions (Frijda and Parrott, 2011). As such, it is often used as basis for the design of computational models of emotions (Hudlicka, 2011).

2. The *core-affect* underlies all emotions as a state with two dimensions:

the degree of arousal and the degree of pleasure (Russell, 2003). Ascribed states such as anger or fear do not exist in this theory, but are rather socially constructed prototypes built on this continuous two-dimensional scale (Yik et al., 2011; Kuppens et al., 2013).

3. *Communicative* theories postulate that emotions are distinct adaptations that evolved in social mammals (Nesse and Ellsworth, 2009). In this context, emotions are cognitive signals used to appraise situations and communications to others, and produce empathy or maintain relationships.

In recent years it was argued that a focus on the functions of emotions in social relationships is a central task of future cognitive theories of emotions (Parkinson, 2013), and that emotions are of high importance to social cognition theories because they relate outer events and other people to inner concerns (Oatley and Johnson-Laird, 2014). Consequentially, emotions started to receive increased attention in the study of social cognition because of their central role in people's social lives (Ong et al., 2019). We continue this new trend in cognitive science by studying emotions with a social interactionist framework, highlighting patterns of 'collective emotions' that emerge when people interact by communicating their emotions to each other. Social interactionism was formulated specifically to bridge cognitive and social theories, tackling questions of emergence in the context of human social behavior, and studying the macro-level social outcomes of micro-level cognitive processes (Vlasceanu et al., 2018). It can be applied to any cognitive mechanisms that have effects in a social context, such as emotional contagion and the emergence of collective emotions. Tipping points (Granovetter, 1983) where communities suddenly start showing qualitatively different behavior after transitioning from one state to another, for example by changing the quantity of participants (Carneiro, 2000), are one proposed causal mechanisms in a social interactionist approach. We believe in this approach as suitable for the study of collective emotions, because they are likewise a pattern of group behavior that emerges from the interactions of many individuals, and their qualitative state shifts when their quantitative number of interactions changes.

As an inherently interdisciplinary research paradigm, one major experimental benefit of cognitive science is the ability to implement tools of various disciplines (Thagard, 2017), One of these tools is computational modeling, a method that established itself in computational social science for the study of social interactions using computer simulations (Cioffi-Revilla, 2014). In

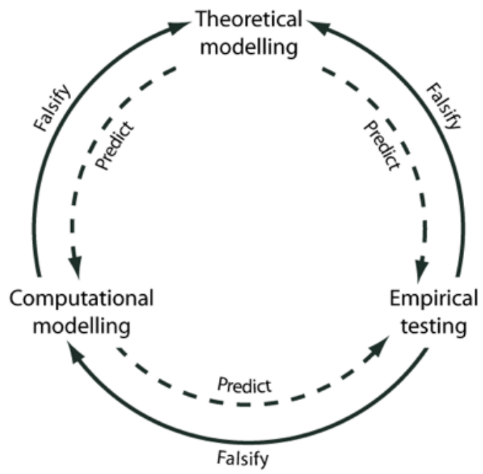

Figure 1: The relationships between theoretical modeling, empirical testing, and computational modeling as illustrated by Roesch et al. (2010). Model results are predicted by theories and can be falsified by empirical tests, while computational models are able to predict test results and falsify theories.

this context, we consider computational models to be those models that "can be, or have been, implemented on a computer" (Elkind et al., 2014). Computational models of the mind are a central topic in computational psychology (Sun, 2008a), with many including some form of affective cognition. For example, EMOCON (Thagard and Aubie, 2008) is a neural model of cognitive appraisal that includes an emotional consciousness computing valence using a simulated dopamine system. Figure 1 illustrates the role computational modeling can take in cognitive science by generating insights to falsify theories and predict empirical test results. A good computational model has the following properties (Fum et al., 2007):

- The model is a simpler and more abstract version of the system.
- Features that are considered essential are preserved.
- Features that are considered unnecessary are omitted.
- Results that are obtained by running the model can be applied to the original system.
- Examining the model increases the understanding of the modeled system.

The Cyberemotions framework (Schweitzer and Garcia, 2010) matches these criteria by providing a blueprint for the abstraction of collective emotions as a convergence in affective responding of agents, or more specifically as agents sharing information about their emotional states with each other in a similar manner. It relies on a circumplex model of affect that describes emotions as interrelated states with two dimensions: 1. arousal indicates the degree of activity induced by that particular state; and 2. valence which describes the degree of pleasure associated with the state (Russell, 1980). Figure 2 illustrates this model along the two pleasure-activity axes. Some of these emotional states may depend on contextual social relations. They can have different functions within social systems, such as motivating collective behavior and affective solidarity, or coordinating group activity, and they may be shared to a lesser or higher degree (Salmela, 2014). We distinguish two fundamentally different ways of looking at collective emotions: 1. with an aggregate account that assumes they are the sum of shared emotional states in individual group members; or 2. applying 'intergroup emotions theory' (Smith and Mackie, 2015) to distinguish between individual and group level emotions. The latter implies that collective emotions are more than just an aggregate of multiple individuals sharing the same emotional state. It was argued that in order for an emotion to be "genuinely" collective, it needs to generate group behavior that is specific to this state and that cannot be explained by referring to aggregate individual actions. Thus, in this context collective emotions can also be described as analogous "computations implemented in a distributed cognitive system that consists of multiple agents." (Huebner, 2011) They can be detected and modeled in online data (Hołyst et al., 2017), for example by measuring the mean or variance of emotional response of users to certain stimuli in online discussions, or burst of activity (Goldenberg et al., 2017).

Using the aforementioned definition and looking at collective emotions as cognitive computations distributed among multiple agents, we believe the agent-based approach that the Cyberemotions model provides is particularly applicable for abstracting collective emotion dynamics. Agent-based simulations have proven themselves as a novel method for studies in social psychology, among other disciplines, because of their ability to capture such complex, dynamic, and interactive social processes (Smith and Conrey, 2007).

Emotions also play a role in dual-process theories of cognition. They postulate two different modes of thought, initially coined as 'hot and cold cognition' (Abelson, 1963) and later popularized as 'system 1 and system 2' (Kahneman, 2011) that were partially based on conceptions of a supposed dichotomy between rationality and emotionality. Hot cognition was specifically described as "cognition colored by feeling." (Brand, 1985) System 1

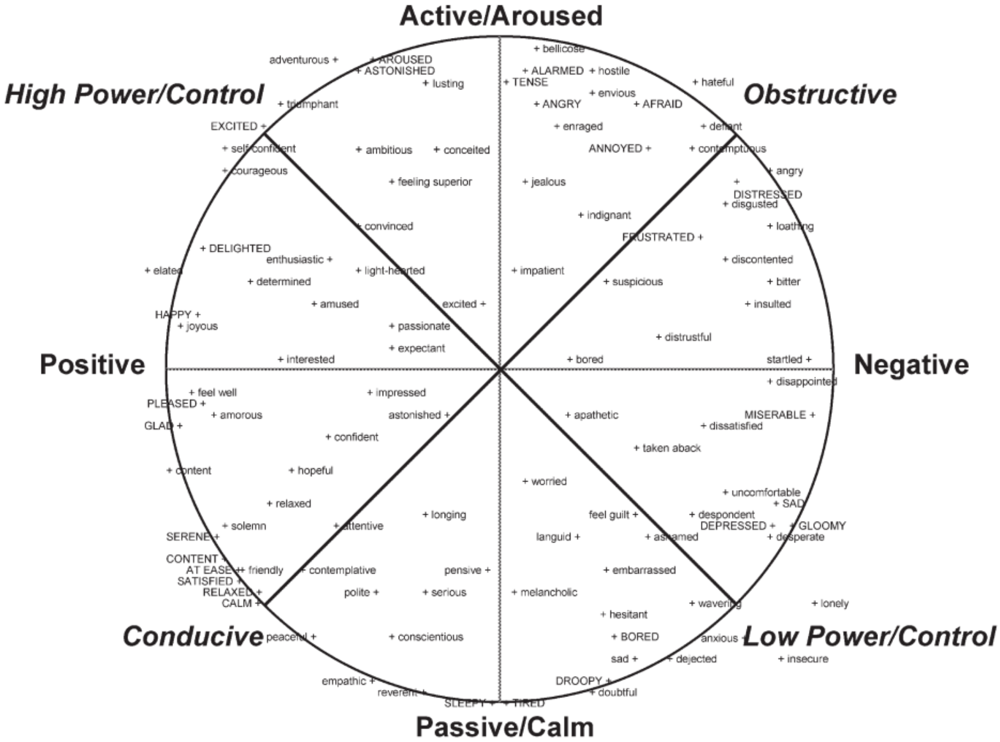

Figure 2: The circumplex model of affect as conceptualized by Russell (1980) and illustrated by Scherer (2005). It places emotional states along two axes, the degree of pleasure associated with an emotion and the level of activity it generates. For example, being relaxed is a pleasant state but does not usually induce much activation, while anger is generally an unpleasant feeling but comes with a high activity level.

was later defined as automatic, unconscious, computationally less demanding, and thus faster processing of information. System 2 on the other hand is slower but controlled conscious process that is more adept at representing information in terms of rules and underlying principles. The former implies intuitive and the latter analytical cognition. In addition, while system 1 personalizes, contextualizes, and socializes information, system 2 does not (Stanovich and West, 2000). Although the extent of its utilization may differ in different fields of research, and studies on the effects of emotions on cognitive control showing emotional influence on all modes of thinking (Rolls and Grabenhorst, 2008; Cohen et al., 2016), the general principle of a cognitive dualism, whereas one of two proposed systems of thought is more affected by emotions than the other, is still applied in cognitive sciences (Lodge and Taber, 2005; Anticevic and Corlett, 2012; Roiser and Sahakian, 2013).

One of the recent areas of dual-process application is political psychology and its 'affective intelligence' theory, which presumes that the degree of cognitive complexity involved in information processing depends on the emotional state of the individual (Brader and Marcus, 2013). Nonetheless, an opposition of reason and emotions is rejected in an affective intelligence approach (Zahn-Waxler et al., 2017). According to this theory, strong feelings such as enthusiasm and aversion lead to an increased use of low-complexity 'hot' system 1 processes. Using the circumplex model of emotions illustrated in Figure 2 we argue that this is well captured by emotional valence, the degree of pleasure associated with an emotion. Both enthusiasm and aversion can respectively be categorized as having very high and very low degrees of valence, and located on opposite sides of the valence scale. Besides enthusiasm and aversion, affective intelligence proponents focus on anxiety as a third indicator of the emotional state of an individual. A state of anxiety motivates the collection of new information and 'cold' high-complexity system 2 processing thereof, rather than relying on habits and intuitions. In the dimensional circumplex model, we suggest this can be abstracted by the arousal scale that describes the degree of activity induced by an emotion. Following this principle of categorizing emotions, it may be possible to draw conclusions about the current emotional state of an individual by measuring the complexity of their active cognitive processes.

The Cyberemotions framework

The Cyberemotions agent-based modeling framework describes collective emotions as a collective phenomenon "that arises from the interaction of many distributed system elements," whereas these elements are represented by agents (Schweitzer and Garcia, 2010). The framework is based on the concept of 'Brownian agents' that are described by a set of specific state variables and combine reflective and reactive components. Their states are determined by deterministic and stochastic factors (Schweitzer, 2007). This Brownian concept was defined specifically for modeling complex systems with multiple interacting agents whose group behavior cannot be inferred from the behavior of the individuals alone. Using a textbook classification, Cyberemotions agents are 'simple reflex agents' that take action based on programmed condition-action rules, but do not utilize internal goals or any models describing how their environment changes after taking an action (Russell and Norvig, 2016). This simplicity on the individual level is one of the strengths of agent-based modeling, because it emphasizes how complex patterns of interaction, such as collective emotions, can emerge from a few simple behavioral rules. This research strategy has proven itself useful for the study of online emotion dynamics, for example when simulating emotional persistence in chatting communities (Garas et al., 2012).

The framework is built on the premise that emotions are shared and spread by individuals as a result of external events and non-linear interactions between agents that change their emotional states. These emotional states are represented by the two variables *valence* and *arousal*. As the terms are commonly used in psychology and neuroscience as part of the circumplex model of affect (Russell, 1980; Posner et al., 2005), valence refers to the emotional value of an event that can be positive or negative, while arousal refers to the degree of activity associated with it.

In a Cyberemotions model, agents abstract human individuals that interact with each other using a communication field that represents an online discussion. The framework was specifically designed to capture the dynamics of online communities, because there they react on other time scales and act on different stimuli than offline (Schweitzer and Garcia, 2010). As mentioned above, agents are described by their state variables, namely 1. their valence which captures the degree of pleasure associated with their state, and 2. their arousal depicting the level of activity the state produces. An agent takes action and activates its *production rules* once its arousal passes a certain threshold by communicating its emotional state to the other agents,

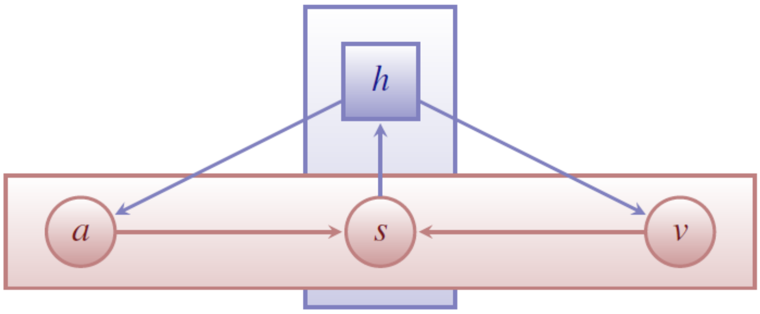

Figure 3: The emotion dynamics in a Cyberemotions model as illustrated by Garcia et al. (2016). The red box represents an agent and the blue box the communication field. Agent expression s depends on agent state variables arousal a and valence v, and affects the field charge h. The state of the field h, in turn, affects the agent by causing it to adjust a and v.

storing information about its valence in the communication field. The field collects this information of all agent states and provides feedback with its own state variable that takes a positive value when the overall simulated discussion is positively charged, and a negative one when it is negatively charged. At this point the feedback loop is active and *perception dynamics* cause agents to adjust their arousal and valence levels according to the field charge. If this again elevates the arousal beyond an agent threshold, this agent may also express its emotions to the rest, which in turn adjust their state variables. In a addition to a set of internal *eigendynamics*, specific rules that determine the development of agent states over time, there is a *feedback of expression* that adjusts the emotions of an agent right after it communicated its state to the others (Kennedy-Moore and Watson, 2001). Figure 3 illustrates the general outline of the model mechanics using the example of one agent perceiving the communication field and producing an emotional expression.

Expression rules and feedback, eigendynamics, and perception dynamics were refined based on three experimental studies measuring emotional responses of participants after reading comments in an online discussion (Garcia et al., 2016). The affective eigendynamics were further validated using data from social media (Pellert et al., 2020). We build on this research to build our own implementation of the Cyberemotions framework using Python. After confirming its accuracy, our model may then be used as a hypotheses generator, in order to predict patterns of collective behavior in real-world data.

Computational Modeling of Collective Emotions

Simulation is a theoretical way of generating knowledge and formulating hypotheses, and a fundamental part of synthetic approaches in cognitive science. It stands in contrast to analytic approaches that rely on empirical investigation as their primary source of scientific knowledge. Simulations of cognitive processes can be conducted using computational modeling (Sun, 2008b) and its use in cognitive science sharply increased in the last decade (Palminteri et al., 2017), spanning a wide area of applications. For example, computer simulations are used studying spatial attention, depression disorder, or cognitive control (Mozer and Sitton, 1998; Siegle et al., 2004; O'Reilly et al., 2010). Figure 1 illustrates the useful role computational modeling can take in the scientific method. It may be used to predict empirical test results, or to falsify scientific theories.

Computer simulations using cognitive models, and computer models of the mind, became novel methods for cognitive science with the rise of modern artificial intelligence research. The *General Problem Solver* model was used to illustrate how "the free behavior of a reasonably intelligent human can be understood as the product of a complex but finite and determinate set of laws" (Newell and Simon, 1963). Over time, more nuanced models of human cognition were developed, such as 4CAPS (short for: Cortical Capacity-Constrained Collaborative Activation-based Production Systems) that abstracts cortical functions of the human brain (Varma and Just, 2006). Cognitive architectures such as SOAR are attempts at creating a unified theory of cognition that allows the simulation of human-like intelligence, and designing a blueprint for 'general artificial intelligence' (Laird et al., 1986; Laird, 2008).

Computational modeling of emotions is a likewise prevalent approach to cognition research (Gratch and Marsella, 2005), although it was noted as lacking of consistent terminology and design guidelines (Hudlicka, 2011). One prominent example of an emotional model is EMOCON, an abstraction of human emotional consciousness that models cognitive appraisal by computing valence using a simulated dopamine system (Thagard and Aubie, 2008). Computational modeling of emotions was argued to be an interdisciplinary task that needs to bridge the gap between computer science and psychology (Reisenzein et al., 2013). In addition to input from these two fields, we add that it also requires input from social sciences, especially when simulating multi-agent models of emotional interaction.

The Cyberemotions approach (Schweitzer and Garcia, 2010; Garcia et al., 2016) targets a particular form of emotions, namely collective emotions that emerge from emotional contagion between multiple individual agents, as they manifest themselves online in social media. This requires a holistic stance that does not reduce collective behavior by the means of micro-level explanations, but instead focuses on social interactions between individuals that may give rise to novel group phenomena on the macro-level, such as collective emotions. Methodological holism has a strong tradition in social anthropology and sociology (Durkheim et al., 1938; Harkin, 2010), hence we believe that incorporating theory and methodology from social science may further the understanding of emotional interactions. For this reason we subscribe to the social interactionist approach to emergent phenomena that allows the study of cognition in a social context (Vlasceanu et al., 2018).

As a short summary of the Cyberemotions framework we detailed in the previous chapter of this thesis, what follows is the general outline of our computational model of collective emotions: Using the Cyberemotions framework and a 'core affect' approach, we model the dynamics of collective emotions online by simulating an environment where multiple agents may spread emotional information by storing it in a communication field. The emotional state of each agent is captured by the circumplex model in Figure 2 that defines valence as pleasure and arousal as activity associated with it. Both may increase or decrease based on perceived information in the field and due to stochastic factors. Once the arousal of an agent passes its internal threshold, the agent stores information about its valence in the communication field. The communication field is an abstraction of an online discussion thread that only stores information about whether the emotion expressed by an agent is either positive or negative. This field, in turn, affects other agents, who again may share their own emotional state, and so on. This way, we are able to simulate an online discussion between forum users, whereas the field represents an online platform, and the agents represent the discussion participants. Simulating agent interactions and the spread of emotional information between them, we expect to observe emergent patterns of collective emotions that manifest themselves in measurable group behavior.

One of the major challenges in affective modeling is the evaluation whether a model works as intended (Broekens et al., 2013). We address this challenge by comparing our model results to the collective behavior of forum users, relying on big data in order to assess the usefulness of our model for predicting macro-level dynamics of real-word online discussions.

Agent-based modeling

In artificial intelligence theory, an agent is an autonomous computational individual that has particular properties and actions, based on a set of behavioral rules. In the context of agent-based modeling, it can also be described as an identifiable, self-contained, discrete individual with a set of programmatic rules that determine its behavior. It is situated in an environment in which it may interact with other agents with the same characteristics (Macal and North, 2005). An agent-based model is "a form of computational modeling whereby a phenomenon is modeled in terms of agents and their interactions" (Wilensky and Rand, 2015). One primary advantage of such agent-based representations is that they are easier to understand than mathematical alternatives, because the latter require mathematical knowledge such as calculus, whereas agent-based models are more intuitive and similar to natural discourse by allowing the projection of human-like properties onto agents. This practice to anthropomorphize and ascribe agency has been extensively studied (Heider and Simmel, 1944; Guthrie, 1997; Dacey, 2017), and may be a reason why agent-based models appear to be more close to "natural" language and thinking.

The primary concept of agent-based modeling is that the dynamics of many phenomena can be described by agents in an environment taking actions and interacting with other agents by following a small set of simple rules. In addition to 'multi-agent' implementations, models may only include a single agent unit. While the agents in such a model are autonomous individual units, the environment is the computational space in which agents act. In a model, the environment is its own entity and can take different forms of shapes, for example it can be geometric or network-based. Following the principle of anthropomorphism, it immediately suggests itself that agents represent human individuals, and the environment characterizes their real-world circumstances. This does not necessarily have to be the case. Agents may represent groups or institutions, or in the case of cognitive modeling, agents may be used to describe multiple interacting modules of the mind. These are questions of 'granularity' and need to be asked when first conceptualizing a model (Guo and Tay, 2008).

Internal decision-making processes of agents can be described as 'agent cognition' and determine what actions an agent takes next (Wilensky and Rand, 2015). While agent cognition ideally follows a few simple rules, more complex decision-making processes can be implemented using machine learning techniques, such as neural networks (Liu et al., 2018), allowing agents to not only change their decisions, but their strategies of coming to conclusions. In contrast to common 'simple reflex agents', agents may also be implemented with specific internal goals, or their own models of 'how the world evolves' (Russell and Norvig, 2016).

Agent-based modeling is especially popular in the study of complex systems (Tisue and Wilensky, 2004), which are systems that (a) are structured by multiple interacting elements, (b) whereas the behavior of the system as whole is not predictable from the individuals alone, and (c) often contain probabilistic and random processes. Following the principles of good computational modeling practice (Fum et al., 2007), an agent-based model can usefully abstract the dynamics of complex systems or other phenomena in the following way:

- The agents and the environment provide abstract implementations of the system's entities that follow simple behavioral rules.
- Behavioral rules are based on the features of the system that are considered essential.
- Non-essential features of the system may be introduced as stochastic components.
- The behavior of the agents and their environment generates outcomes that can be applied to the original system.
- Examination of the behavioral outcomes provides insights about the system dynamics.

The simplicity on the individual level is one of the strengths of agent-based modeling, because it emphasizes how complex patterns of interaction can emerge from a few behavioral rules. For this reason, following the principle of *Occam's razor*, we believe agent-based modeling is a compelling method for the creation and validation of scientific theories. The use of agent-based modeling for theory building has also found support in social psychology, where "the goal of characterizing and theoretically understanding social and psychological phenomena requires a detailed understanding of such dynamic and interactive processes," because complex, dynamic, and interactive processes are especially important in the social world (Smith and Conrey, 2007). The reasons mentioned above, intuitive understanding and simple behavioral rules, also make it easier to verify the validity of agent-based models (Windrum et al., 2007).

Implementing the Cyberemotions framework

We implement the Cyberemotions framework using the object-oriented programming language Python.[1] (Van Rossum and Drake Jr, 1995) Agents are initialized as instances of a defined 'agent' object class, storing the emotional state variables arousal and valence. Likewise, the communication field is its own 'field' class and stores the field charge variable. The perception dynamics, production rules, feedback of expression, and eigendynamics of the agents are implemented as methods of their respective class. The field also stores information about the agent states when an emotional expression occurred and we capture this process in its own communication method. Mathematical formalizations of the model processes are directly taken from published Cyberemotions documentation (Schweitzer and Garcia, 2010; Garcia et al., 2016). Our model is an open-source implementation and its code is published as free software online under a GNU General Public Licence.[2] We hope this allows for easy replication and testing, as it is best-practice in agent-based modeling (Donkin et al., 2017).

Our addition to the framework is a formalization of the satiation of agents over time. Since online discussions do not sustain themselves forever as more and more people drop out, our model needs to accommodate this fact, otherwise a simulation would go on forever. We base our implementation of this process on 'optimal arousal' theories (Berlyne, 1967) that assume individual ranges of arousal levels at which a person experience satisfaction (Wirtz et al., 2000). However, the more often they experience this, the less satisfaction they get, leading to a state of satiation and inactivity (Poor et al., 2012). In our model, we assume that agents are increasingly likely to drop out of a discussion once their arousal becomes more negative, and that stochastic factors also play a role in agent satiation. This leads us to formalize agent satiation as a function of arousal.

The following section describes our implementation of the Cyberemotions framework using the updated ODD protocol (short for: Overview, Design concepts, and Details) for documenting agent-based models, especially in the use case of social simulation (Grimm et al., 2010, 2017). This protocol was conceptualized to standardize the published descriptions of agent-based models, by providing an outline for describing: 1. the purpose of the model; 2. its state variables and scales; 3. process overview and

[1] Version 3.7.

[2] `https://github.com/r-andre/abm-collective_emotions` (last accessed on 11.04.2020)

scheduling; 4. general design concepts; 5. how it is initialized; 6. its input; and 7. the submodels it relies on. Its goal is to ease understanding of model descriptions, make them more complete, and to allow rigorous evaluation (Grimm et al., 2006). Because of this, we believe the ODD protocol to be a convenient resource and rely on it to document our model implementation in a hopefully comprehensive and replicable manner.

Overview, design concepts, and details

Using the *Cyberemotions* framework (Schweitzer and Garcia, 2010; Garcia et al., 2016), this agent-based model abstracts the dynamics of collective emotions and the spread of emotional information online. Collective emotions are understood as emotional states shared and spread by a group of individuals in a similar manner, and defined as synchronous convergence in affective responding of agents (von Scheve and Salmela, 2014). As such, they may be studied as an emergent phenomenon using a complex systems approach. Following the principle of Brownian agents (Schweitzer, 2007), the agents of the Cyberemotions framework are described by a set of state variables that may change due to deterministic and stochastic influences. In the model, agents (representing human individuals) communicate their emotional states with each other using a communication field (representing aggregates of social media posts, such as threads in discussion forums). This field, in turn, provides feedback that affects the individual emotional states of the agents. On a group level, this leads to the spread of emotions from one agent to another and the emergence of collective emotions.

1. Purpose

This implementation of the Cyberemotions modeling framework was designed to investigate the dynamics of collective emotions in online discussion forums, where users may directly respond to the posts of others in the same discussion thread. The model and its results are intended to be used as a hypotheses generator and its predictions may be tested against real-world data of discussions in social media, and specifically of online forums.

2. Entities and state variables

The two types of entities of the model are 1. the agents representing forum users, and 2. the communication field that represents one specific thread in a forum. Both are treated as separate objects and defined as Python classes, with each instance of the respective class containing its own state variables. The state variables *valence* and *arousal* represent the emotional state of agents, based on a circumplex model of emotions (Posner et al., 2005; Scherer, 2005). They are numerical values that can be either positive or negative, with negative numbers indicating a negative emotional charge and positive numbers a positive emotional charge. As the number gets higher, the emotions it represents become more positive, and as it gets smaller, the more negative they become. A charge of 0, in that sense, is emotionally neutral and considered an equilibrium state.

While valence represents a quantification of the degree of pleasure associated with an emotion, arousal represents its degree of activity. The state variable *field charge* represents the emotional charge of the communication field and can likewise take positive and negative values, depending on communicated agent states.

3. Process overview and scheduling

For each simulation, the model initializes a set number of agents and the communication field. Once the initialization phase is over, it goes through the same following processes (abstracted as Python class methods) for every agent in the model and for every step in time. Over time, agents will enter a satiated state and drop out of the simulation. Once there are no more agents present, the simulation ends. Additionally, the model is initialized with a maximum number of time steps it may go through. Once this limit was reached, the simulation ends regardless of the number of agents left. The model processes that occur for each time step are scheduled as follows:

1. Every agent
    I. *perceives* the field and adjusts its state variables.
    II. may *express* its emotions.
    III. may enter a *satiated* state and drop out.
    IV. *relaxes* and adjusts its state variables.
2. The field enables *communication* by adjusting its state variable to the agent expressions.

4. Design concepts

The design of the model follows a social interactionist paradigm that highlights communicative interactions between individuals (Vlasceanu et al., 2018). These interactions may lead to unexpected patterns of behavior on the group level, and emergent phenomena. In this case, agents communicate their emotions (internal state variables) with each other by storing emotionally charged information (expression variables) in a field that is perceived by all agents. This expression of emotions, in turn, is triggered by the emotional charge of the field (field variable) itself. Once an agent has expressed its emotional state, it experiences a feedback effect that down-regulates its valence and arousal levels. Additionally, over time, valence and arousal and the field charge decrease towards their baseline values. With a decrease of arousal, the probability of the agent to drop out of the discussion altogether increases.

In principle, every agents has individual factors that determine what effect the emotional charge of the field has on its own emotional state (representing personality traits such as empathy or responsiveness), based on the general idea that agents already in a specific state may be more affected by particular emotions of others. For the purpose of this model, it is assumed that the impact of the emotional information that agents perceive depends on their emotional state in a non-linear manner. For example, the positive emotional state of an agent may become more positive when perceiving positive emotional information from other agents.

While the individual behavior of the agents follows a set of simple rules, new patterns of group behavior may emerge when emotional information spreads and emotional states are shared by multiple individuals. Both, all agents and the field, are treated as individual objects and instances of their respective class.

5. Initialization

The simulation is implemented as a function that initializes the model taking basic settings (number of agents and maximum number of time steps) as input arguments. This allows to easily run multiple instances of the model at various settings. At the beginning of each model run, the set number of agents and the field are initialized by creating instances of their respective class. The initial agent state variable baselines of valence and arousal, and their fixed arousal thresholds, are set using a standard normal distribution around a preset parameter, while the initial charge of the field is 0, assuming the discussion only starts after the first agent expressed its emotional state.

6. Input data

The only input data are the model parameters that determine the submodel processes, such as state variable baselines, and the basic simulation settings, number of agents and number of model runs.

7. Submodels

There are five major submodels that are all described as methods of their respective Python classes. I. *Perception*, II. *expression*, III. *satiation*, and IV. *relaxation* are methods of the agent class, while V. *communication* is a method of the field class. More detailed formalizations of submodels I., II., and IV., and an explanation of the equations they are based on, are part of the published Cyberemotions framework descriptions (Schweitzer and Garcia, 2010; Garcia et al., 2016). The satiation process described in submodel III. and its mathematical formalization constitutes our own addition to the framework.

<u>I. Perception</u>

This submodel describes the rate of change of the agent internal state variables *valence* and *arousal* through the effects of the *field variable*. It is based on an equation that takes deterministic influences (the communication field and emotional feedback of other agents) and stochastic factors (unexpected events represented by a random numbers taken at a given point in time from a distribution of stochastic shocks with a mean of 0) into account.

The perception function F describes how perceiving the emotional charge of the field, represented by the the field state variable h, changes valence v and arousal a of an agent i at time point t. The random number ξ represents stochastic components affecting v and a and the strength of its influence on an individual is determined by constant amplitudes A.

The equation

$$\dot{v}_i = F_{v_i}[h, v_i(t)] + A_{v_i}\xi_v(t)$$

describes the change of valence $\dot{v}$ of an agent i given the emotional charge of the communication field h, while

$$\dot{a}_i = F_{a_i}[h, a_i(t)] + A_{a_i}\xi_a(t)$$

describes the change in arousal $\dot{a}_i$ of the same agent. The perception functions F_v and F_a define the changes in v_i and a_i that are caused by different values of h and the sign of $|h|$. Furthermore, the unique stochastic components A and ξ are added that account for unpredictable factors in emotional changes of individuals (e.g. multiplicative noise).

The functions

$$F_v[h, v_i(t)] = h * [b_0 + b_1 v_i(t) + b_2 v_i(t)^2 + b_3 v_i(t)^3]$$

and

$$F_a[h, a_i(t)] = |h| * [d_0 + d_1 a_i(t) + d_2 a_i(t)^2 + d_3 a_i(t)^3]$$

define how the state variable v_i is affected by the the field variable h, while a_i is affected by the absolute value of the field variable $|h|$. Their change is determined by the constant coefficients b and d.

v valence

a arousal

i agent instance

F perception function

h field variable

t point in time

A amplitude

ξ stochasticity

b valence change coefficients

d arousal change coefficients

II. Expression

Submodel expression describes how agents express the valence of their emotions by creating an *expression variable* that is a function of the sign of its valence and activated when their arousal surpasses an individual *arousal threshold*. It is based on an equation that includes the assumption that agents do not communicate all details about their emotions, but only whether it is a positive or negative emotion. The expression of emotions also causes an immediate down-regulation of valence and arousal by a constant factor.

Expression variable s of agent i is a function of the sign of its valence v, but only stores a positive or negative value (1 or -1) if its arousal a is larger than the arousal threshold τ.

The equation

$$s_i(t) = sgn[v_i(t)]\Theta[a_i(t) - \tau_i]$$

describes the relationship between s_i and the sign of v_i as a Heaviside step function Θ. If a_i is larger that τ_i, then s_i stores a positive value when v_i is likewise positive, or a negative value when v_i is negative. Else, no expression is triggered when a_i does not pass threshold τ.

In case an emotion was triggered, the values a_i and v_i are instantly down-regulated using the formulas

$$v_i(t) \leftarrow [v_i(t) - b_i] * k + b_i$$

and

$$a_i(t) \leftarrow [a_i(t) - d_i] * k + d_i$$

that adjust the distance of v_i and a_i to their respective baselines b_i and d_i by a proportional value relying on the constant factor k.

s_i expression variable

Θ Heaviside step function

τ arousal threshold

b_i valence baseline

d_i arousal baseline

k down-regulation factor

III. Satiation

Satiation describes the process of determining if an agent drops out of the simulation due to reaching a satiated emotional state by creating a *satiation variable*. It is based on a quadratic *satiation function* of agent arousal and accounts for random influences on the emotional state of an agent. The agent may become satiated given a certain probability that is derived from the agent's state of arousal.

This submodel abstracts the decrease in activity in an online discussion over time on an individual level. We assume, as discussions go on, more and more people that were initially part of the interaction will leave the discussion due to personal factors. If no agent were to drop out, a simulated interaction between agents would go on indefinitely. We believe that modeling this process allows us to simulate the dynamics of online discussions in a more realistic manner.

The satiation variable n of agent i is assigned a value when the output of the quadratic function S is larger than stochastic component ξ (a random number between 0 and 1). The satiation function S describes the probability of the agent dropping out, which is determined by the absolute value of arousal a.

Satiation variable n_i only stores a value if function S is larger than ξ. This is formalized by

$$n_i(t) \leftrightarrow S[a_i(t)] > \xi$$

whereas $0 \leq \xi \leq 1$. Satiation function

$$S[a_i(t)] = |a_i(t)|^2 * c$$

squares the absolute value of a_i and returns a value representing the probability of an agent reaching a satiated state at time step t.

n satiation variable

S satiation function

c satiation coefficient

IV. Relaxation

The relaxation submodel describes the individual relaxation of the valence and arousal of agents towards their *baseline* values using specific *decay parameters* that are unique to every agent and to its respective state. It is based on an equation that assumes an exponential decay of valence and arousal in agents over time.

The state variables valence v and arousal a of agent i at time t are reduced by a combination of their initial values minus their baseline times the decay parameters γ (which vary for each agent and are different for valence and arousal) that define the time-scale of this decay. This decay represents the relaxation of emotional states towards the baseline values.

The functions

$$R[v_i(t)] = -\gamma_{v_i}(v_i(t) - b_i)$$

and

$$R[a_i(t)] = -\gamma_{a_i}(a_i(t) - d_i)$$

specify the relaxation of valence v_i and arousal a_i towards their respective values b and d given the decay parameters γ.

R relaxation function

γ decay parameter

V. Communication

This submodel describes how the emotional charge of the field changes when agents express their emotions. It is based on a function that calculates aggregates of all positive and negative *agent expressions*.

While N is the total number of agents with an either positive or negative expression variable, s is a fixed amount by which the field variable increases or decreases for every agents that expresses its emotions. In addition to agent expressions, the decay parameter γ represents a decrease in impact of emotional information on agents over time.

Equations

$$\dot{h}_+ = s_h N_+(t) - \gamma_+ h_+(t)$$

and

$$\dot{h}_- = s_h N_-(t) - \gamma_- h_-(t)$$

calculate the emotional charge of the respective positive or negative field variables h_+ or h_- using the total number of agents that contribute to the field N in combination with a fixed impact variable s. Additionally, the decay parameter γ_+ and γ_- account for a loss of emotional impact over time. While

$$h = h_+(t) - h_-(t)$$

creates a total of the emotional charge of the field that is either positive or negative, using the field variables h_+ and h_-.

h_+ (positive) field variable

h_- (negative) field variable

s_h impact coefficient

Data and Methods

derStandard.at is an Austrian news website and the official online presence of the daily newspaper *Der Standard*. It recorded 22,922,917 visits by 4,329,166 unique visitors in March 2016, making it the second most popular Austrian website overall and the most visited news website.[3] It also hosts its own discussion forums organized as comment sections in topics of published news articles. Registered users can post messages under articles and start discussions by directly replying to other users. In addition, the forums also utilize a voting mechanism that allows upvoting and downvoting other users posts as expressions of agreement or disagreement with its content. All posts in the forum contain a headline and a message with a maximum length of 500 words, although only of of the two is necessary for a comment to be posted. Figure 4 shows an exemplary post from the derStandard.at comments section, posted in March 2020.[4] While the website was redesigned in 2019, the forum user interface remains largely the same as in the decade before. The community guidelines define the rules for participating in the forums and are similar to other online discussion platforms, with active on-topic discussions using clear structured arguments being encouraged, and discriminating and illegal content being removed when discovered by professional moderators employed by website.[5] The official language of the site and the forums is German, but occasional messages in English can be found.

The One Million Posts corpus is a dataset of 1,011,773 forum posts, collected between 01.06.2015 and 31.05.2016 in the der.Standard.at comments section (Schabus et al., 2017). These comments were posted by 31,413 registered users in response to 12,087 articles. Basic statistics of the dataset

[3] Last accessed on 21.03.2020 using the internet archive: `https://web.archive.org/web/20160427122306/http://oewa.at/index.php?id=17198&sort=DESC&by=uc&cat=gesamt#ea`

[4] The name of the author was anonymized to protect their privacy. Searching for the whole comment or snippets of it using online search engines, or the search function of derStandard.at, will also not lead to the posted article and the username.

[5] `https://www.derstandard.at/communityrichtlinien` (last accessed on 21.03.2020)

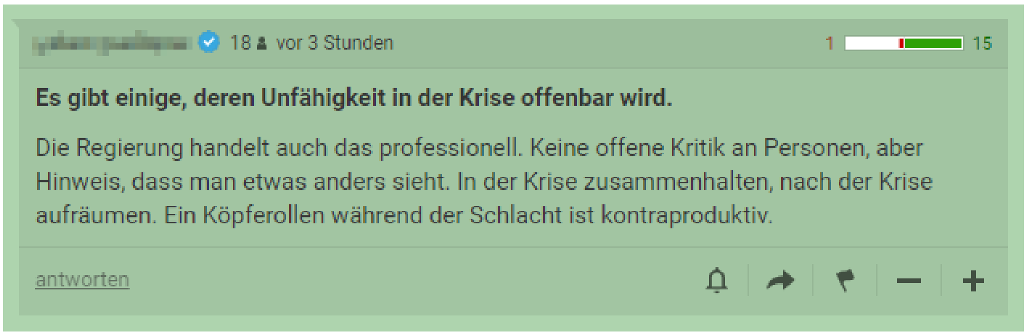

Figure 4: Exemplary comment on a news article in the derStandard.at forums. The top left shows the username of the author (in this case blurred) and a timestamp. In bold font is the headline of the post and the message body follows directly under it. Bottom left provides a link to write a reply to the post, while the + and − buttons on the right provide the option to upvote or downvote the comment as a means to signal agreement or disagreement with its content. The bar on the top right shows the current count of votes, with green numbers indicating upvotes, and the red number representing down votes. In this example, the post did not polarize and received a larger number of positive feedback, and only one negative vote.

are documented in Table 1. All posts in the corpus come with the following information: (1) unique identification numbers for the post itself, the user who wrote, the article discussion it is part of, the parent post it replies to; (2) a time stamp including time and date of posting; (3) the message text separated into headline and body; and (4) the numbers of upvotes and downvotes the post received. Likewise, information on the articles include: (a) its unique identification number; (b) the publishing date; (c) the thematic category it was published in, for example "International"; and (d) its text separated into headline and body.

In addition to posts and articles, the dataset also includes 11,773 annotations of posts by official moderators of the forum. Those annotations categorize posts based on: their broad category, such as "Offtopic"; their positive, neutral, or negative sentiment as perceived by the moderators; whether the authors provided clear and structured arguments for their statements; and if the content was inappropriate or discriminating. The annotation process purposely biased and primarily selected comments that were likely contain undesirable content, causing the majority of all sentiment annotations to document negative sentiment. Nonetheless, in addition to analyzing emotional content and vote polarization of posts, a comparison of the moderator annotations to our sentiment mining results might provide interesting insights.

Total number of posts	1,011,773
Number of unlabeled posts	1,000,000
Number of labeled posts	11,773
Number of category annotation decisions	58,568
Number of posts taken offline by moderators	62,320
Min/Median/Max post length (words)	0 / 21 / 500
Vocabulary size ($\geq$ 5 occurrences)	129,070
Number of articles	12,087
Number of article topics	1,229
Number of users	31,413
Min/Median/Max number of posts per article	1 / 22 / 3,656
Min/Median/Max number of posts per topic	1 / 142 / 44,329
Min/Median/Max number of posts per user	1 / 5 / 4,682
Min/Median/Max number of users per article	1 / 15 / 1,371
Min/Median/Max number of users per topic	1 / 78 / 6,874
Number of pos./neg. community votes	3,824,806 / 1,096,300

Table 1: Basic statistics of the One Million Posts corpus data set, as documented by Schabus et al. (2017). 'Labeled' refers to whether the post was annotated by moderators, and 'vocabulary size' reflects the number of words that occurred at least 5 times over all posts.

Text mining in online discussions

As a form of data mining, the term text mining generally refers "to the process of extracting interesting information and knowledge from unstructured text"(Hotho et al., 2005). The continuous increase of online textual content offers various possibilities to study linguistic phenomena (Mehl, 2006). Specifically, social media such as Facebook or Twitter offer major sources of data on social interactions and valuable modern opportunities for text mining (Salloum et al., 2017). Naturally, online discussion forums also provide large quantities of messages and user interactions for analysis. In comparison to status updates or short messages of other types of online social media platforms, discussion forums usually encourage the formulation of longer arguments and opinions, which allows for a more detailed and reliable analysis of individual messages.

Quantitative text analysis was broadly defined as a systematic reduction of a text to a set of statistically manipulable symbols that represent the presence, intensity, or frequency of relevant characteristics of the text (Shapiro and Markoff, 1997). In its simplest form, a quantitative approach implies the counting and categorization of words. In addition to quantitative approaches, text mining may also be use for qualitative text analysis (Wiedemann, 2016). However, in our thesis research we rely on a quantitative analysis method, which was noted to be a "powerful, efficient, and easy to use scientific method with a wide spectrum of applications" with a specific reference to its potential in psychological research (Mehl, 2006). We believe a quantitative approach is more suitable for our research questions, because it is established as a useful method to capture emotional content of online discussions, and can be easily compared to our simulation data. Quantitative analysis of text can vary along four different dimensions (Popping, 2000), which are mostly implicit applications of different concepts of analysis and neither specifically stated by the researchers, nor always clear in practice (Mehl, 2006):

1. *Representational* vs. *instrumental* aims: The goal of representational approaches are to decode the meaning of a text as accurately as possible and to represent the intention behind the message. The intent is of no relevance in more prominent instrumental approaches that focuses on latent content.

2. *Thematic* vs. *semantic* approaches: Thematic text analysis "maps the occurrence of a set of concepts in a text," (Mehl, 2006) while semantic

analysis extracts information on conversational meaning.

3. *Broad* vs *specific* bandwidth: The analysis may focus on few specific linguistic variables, such es emotion words only, or it can generate a broad linguistic profile of a text.

4. *Content* vs. *style* focus: A text can be analyzed based on those components that are transmitted on purpose, or on stylistic components that are unintended and automatic, often serving expressive rather than instrumental functions.

The study of psychometric properties is a primary application of quantitative text analysis. Studies show that the words in written language associated with basic psychometric properties are stable over time and across context (Gleser et al., 1959; Schnurr et al., 1986; Pennebaker et al., 2003). Nonetheless, there are several short comings of quantitative text analysis methods that are based on counting words. For example, they are unable to detect sarcasm or metaphoric language use, and they have trouble with detecting context in general (Mehl, 2006). One key premise of quantitative text analysis is that emotional states of authors can be detected by studying their use emotion words (Bestgen, 1994). This lead to the development of sentiment analysis and affective computing (Cambria et al., 2017), which capture the emotional content of texts and other characteristics, such as the cognitive complexity of text authors at the time of writing a text.

Text mining of online discussions is increasingly used to study the dynamics of online interactions, for example collective emotions (Chmiel et al., 2011), emotional persistence (Garas et al., 2012), integration and engagement in social movements (Alvarez et al., 2015), opinions (Chen, 2015), aggressive behavior (Ventirozos et al., 2017), or online deliberation (Aragón et al., 2017b). We follow this research trend by analyzing the One Million Posts dataset (Schabus et al., 2017) of online discussion posts for its emotional content, in order to gain insights about the dynamics of emotional contagion and collective emotions.

Measuring affective meaning in text

Sentiment analysis is a form of quantitative text analysis and based on the idea that the human mind conveys emotions through natural language and that these conveyed emotions can be analyzed using natural language processing (Cambria et al., 2017). Using large amounts of text, it is possible to link everyday language and word use with measures of personality, social behavior, and cognitive styles (Pennebaker et al., 2003). As mentioned in the last section, the analysis of social media comments, specifically for their emotional content, receives increasing interest in cognitive and psychological research. For example, sentiment analysis in online chatting communities was used to formulate an agent-based model of collective emotions (Garas et al., 2012). While there are multiple techniques for analyzing sentiment (Kharde et al., 2016), we focus on the LIWC (Linguistic Inquiry and Word Count) sentiment analysis software tool because of its ease of use and established efficiency when capturing the social and psychological states of people.

The LIWC software is an established tool in computerized text analysis and was validated as a method to capture emotional expressions in language (Kahn et al., 2007). It "reads a given text and counts the percentage of words that reflect different emotions, thinking styles, social concerns, and even parts of speech."[6] The program is based on a processing and a dictionary component, with the former processing text by going through it word for word. Each word is then compared to the dictionary and categorized as belonging to one or more specific types of word, such as emotion words. The dictionary is thus the key component to evaluate the sentiment of a text, because it provides the reference about when a word is considered to be emotional by the program. The categories of the dictionary were created based on human judgment (Pennebaker et al., 2015). The standard dictionary of the program is in English, although it comes with other supported languages, and additionally allows the upload and download of dictionaries made by the community.

Dictionaries were originally developed using human judges in two independent groups of three and any word is placed in a respective category when at least two judges agree on it. Likewise, when updating a dictionary, a word is removed from a category again when two of those three judges support the move. This process lead to a 93-100% agreement among judges after the second rating phase (Pennebaker et al., 2007). The increased use of "netspeak" language online lead to major updates in the 2015 version of the dictionary, allowing the capture of common abbreviations and emoticons. The preliminary list of emotion words was based on rating scales such as PANAS (Watson et al., 1988), which was specifically developed to measure positive and negative affect in language. This list was then again evaluated twice by a group of judges that decided about the inclusion and exclusion of words (Pennebaker et al., 2015).

[6] `https://liwc.wpengine.com/` (last accessed on 30.03.2020)

The official LIWC dictionaries broadly distinguish between content words and style words, the former indicating the content of a message for example nouns, verbs, and adjectives, whereas the latter are function words such as pronouns, prepositions, or articles. For example, the use of such function words was able to predict academic success in university admission essays (Pennebaker et al., 2014). All categorization of words is based on a probabilistic system and is supported by research such as studies on verb tense differences and psychological distance to events (Pasupathi, 2007), use of words in writing about emotional events (Kahn et al., 2007), or word choices when describing social processes, for example conveying status, lying, or romantic conversations (Semin and Fiedler, 1988; Sexton and Helmreich, 2000; Hancock and Curry, 2008; Simmons et al., 2008). In addition to information about sentiment, LIWC also captures information on how people process information and interpret it, and thus on cognitive mechanisms involved in writing a text. Cognitive complexity can be described as the manifestation of the following two processes: 1. the extent to which someone differentiates between multiple possible solutions; and 2. the extent to which someone integrates those solutions (Tetlock, 1980).

LIWC collects information on cognitive complexity using the cognitive mechanisms and six-letter words variables that are indicative of complex language. LIWC measures analytic thinking as rates of pronouns, auxiliary verbs, formal language style, propositions, or conjunction, based on studies showing their association with cognitive complexity (Pennebaker et al., 2014). Cognitive processes are measured using the number of function words such as "because" or "perhaps", while six-letter words is simply focused on words containing more than six letters. The emotional content of posts is aggregated as emotional tone, and individually as percentages of positive and negative emotions words, for example "nice" or "ugly". In total, the output of the program contains 90 variables, with a detailed description of each variable being part of the official LIWC2015 documentation (Pennebaker et al., 2015). The current implementation of the linguistic inquiry and word count is LIWC2015, named after its latest main release version. Because the One Million Posts corpus is a collection of German posts, the standard English dictionary cannot be used to conduct our analysis, since it does not contain the necessary reference words. Instead, we rely on the official German alternative DE-LIWC2015, or "LIWC auf Deutsch" (Meier et al., 2019).

The DE-LIWC2015 dictionary was assembled in a similar manner as that of the original, based on groups of judges deciding on the inclusion and exclusion of words in particular categories. These judges were native German speakers with backgrounds in psychology. Due to the grammatic difference between English and German, such as differences in personal pronouns, there were certain changes in the categories and structure of the dictionary, for example the creation of additional subcategories to capture formal 'you-talk'. All in all, the dictionary contains 18,711 words in the same 90 categories with a few additions that are unique to the German version because of a different use of personal pronouns, verbs, adjectives, and time focus. The general functionality remains the same to the standard version and it captures on average 83 percent of all words people use in written language (Meier et al., 2019).

The One Million Posts corpus

Using LIWC2015 and its official German dictionary DE-LIWC2015 (Meier et al., 2019), we conducted a sentiment analysis directly on the extracted, unaltered *Posts* and *Articles* tables of the One Million Posts corpus, which added extra columns for each of the LIWC variables to the respective data table. All $1,011,773$ posts and $12,087$ articles were processed this way. While the program provides over 90 output dimensions, it was not necessary to collect all of them for the purpose of this study. Therefore, we focused on variables that directly help answering our research questions and testing our hypotheses, namely those that indicate emotional states of users and the complexity of their active cognitive processes at the time of writing their posts, and the number of positive and negative votes their posts receive. In addition to the posts and articles, we also further analyzed the annotations table of the corpus, although no sentiment analysis was run on the annotations, because they do not include any comments of the moderators that could be categorised.

The following list describes all of the LIWC variables for the corpus data we collected and that were included in the processed data set that we used for further analysis. With the exception of a discrete word count, all other recorded dimensions are continuous variables that correspond to numbers between 0 and 99.99, whereas a score of 0 implies a complete lack of words of that category. Those numbers indicate the percentage of words in the text that fall into the respective word category. Only the variables 'emotional tone' and 'analytic thinking' are calculated as indices (Cohn et al., 2004).

1. *Word count*: As the name indicates, this is the total number of words in a post. It is used to provide general information about discussions and posts, and to provide an overview for us to evaluate how many posts contain enough words for their sentiment analysis results to be meaningful—since one-word or two-words comments do not provide much insights about the sentiment of the respective author.

2. *Positive emotions* and *negative emotions*: Both variables are percentages of words in their respective category. For example, a positive emotions score of 25.00 describes a fourth of all captured words in a comment inferring a positive emotional state of the author.

3. *Emotional tone*: This variable puts the two previous dimensions of emotionality into a single emotional positivity index (Cohn et al., 2004). The value 17.39 represents an emotional equilibrium at which

positive and negative sentiment are equally strong (the positive and negative emotions variables carrying the same value), and the higher the number, the more positive the emotion becomes.

4. *Cognitive processes*: The cognitive processing index indicates the degree of cognitive activity involved in organizing and formulating comments. It is a more narrow category of the former cognitive mechanisms category that incorporated words that depict causal thinking, including self-reflection, discrepancy, undoing, causation, and achievement (Berry et al., 1997). The newer dimension restricts constituent words to true markers of cognitive activity (Pennebaker et al., 2015). For these reasons, we use the cognitive processing index as the first of three indicators of the degree of cognitive complexity involved when writing a comment, and thus the use of either system 1 or system 2.

5. *Analytic thinking*: This dimensions is derived from function words and captures the usage of words that suggest formal, logical, and hierarchical thinking patterns. It is based on the categorical-dynamic index that contrasts analytic with narrative writing (Pennebaker et al., 2014). A higher score on the index implies the increased cognitive use of more complexly organized objects and concepts, and the larger the analytic thinking variable, the higher we consider the degree of cognitive complexity to be.

6. *Six-letter words*: As a simple percentage of words with at least six letters, it is one of the descriptive categories of LIWC that noted to capture the cognitive dimensions involved in writing a post (Pennebaker et al., 2015). We use it as the third indicator of the use of cognitive systems, and cognitive complexity. While the average number of letters in a word is higher in German than in English, the way these words differ from the rest of the German text correspond to the difference in an English sample (Meier et al., 2019).

In addition to the six recorded LIWC dimensions, the two variables of interest that we took directly from the corpus are:

7. *Positive votes* and *negative votes*: These are the discrete counts of upvotes and downvotes each post received, which are collected separately.

During our analysis we aggregated some of the above variables in order to compare discussions that develop around articles and other posts on a cumulative level. In this case, we either used the mathematical mean values for the respective variables (e.g. the mean value of emotional tone for all comments in a discussion), or their total sum (e.g. the sum of all upvotes for all comments in a discussion). For this aggregate-level analysis we relied on:

8. *Polarization*: The polarization index (Schweighofer, 2018) is based on the sum of all positive and negative votes of a discussion. It takes the value 0 when the percentage of positive votes is either 0% or 100%, indicating no polarization in the comments. A score of 1, on the other hand, represents an equal number of positive and negative votes. The lower the score, the higher the polarization. The index is based on the equation:

$$1 - 2 * |0.5 - \frac{u}{u+d}|$$

   wheras u is the number of upvotes and d the number of downvotes a post received. In addition to group level discussions, we also apply this formula to posts to calculate polarization on the individual level.

9. *Emotional charge*: This variable is a simple aggregate number of all emotion words in a post. In contrast to emotional tone, it does not indicate whether the emotions are positive or negative. The sum of positive and negative emotions is considered the emotional charge.

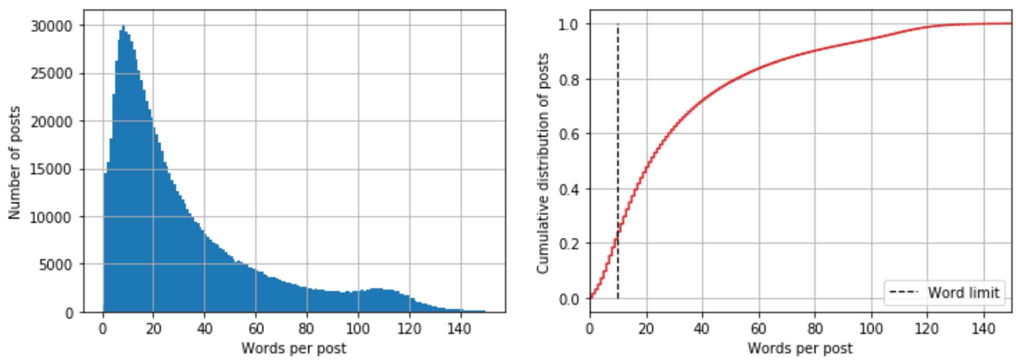

Figure 5: Distribution of individual posts in 150 bins by their number of words (left) and their cumulative distribution function (right). The dashed black line indicates the applied word limit and all posts left of it are removed from further analysis. With a mean of 32.6 (median 22.0) words per post, setting a limit of 10 words reduces the data set by 21.1% from 1,011,773 posts to 796,734 posts.

For the purpose of generating meaningful and reliable results, we restricted our main analysis to posts containing at least 10 words and to article discussions with at least 10 posts. According to the LIWC website, the analysis results of "any text with fewer than 50 words should be looked at with a certain degree of skepticism."[7] However, since setting a limit of 50 words per post would have reduced our dataset by roughly 80%, we opted for a 10 post limit that only removes about 20% of all posts. Figure 5 shows this distribution of posts by word count and illustrates the percentage of posts that is lost when applying a word limit. Figure 6 furthermore details the distributions of posts for their emotional tone and their vote polarization when setting different word limits. Articles themselves have an inherent 250 words minimum and are unaffected by such a limit. However, in order to increase statistical significance of our results, we limited our analysis to discussions containing at least 10 posts.

In the following section, we document descriptive statistical information on the results of our sentiment analysis that we conducted on the One Million Posts Corpus data set.

Statistical overview

A basic statistical documentation of the unaltered One Million Posts dataset, including information about number of posts, articles, or votes, can be found in Table 1. We organize our following visualized overview into three sections, targeting firstly individual posts and their characteristics, secondly discussions as groups of at least 10 posts in response to a news article, and thirdly moderator annotations of individual posts. More detailed statistics of posts and discussions are supplemented in Table 2 and Table 3.

The distribution of the 1,011,773 individual posts by their words per post is illustrated in Figure 5. It shows that with a minimum word limit of 10 words per post, the following analysis targets 796,734 posts, or 78.9% of all corpus posts. Figure 6 shows the distribution of posts by their emotional tone and polarization variables for different word limits. The patterns of distribution remain largely the same for a 10 word minimum as opposed to no word minimum, with a relatively even distribution of posts on both, the positivity and the polarization index. The gaps around the neutral point 17.39 of the positivity index are the result of the formula used to calculate the difference between positive and negative emotionality (Cohn et al., 2004).

Figure 7 shows the percentage of posts that received a certain number of upvotes and downvotes, and contain a particular percentage of emotions

[7] `https://liwc.wpengine.com/how-it-works/` (last accessed on 02.02.2020)

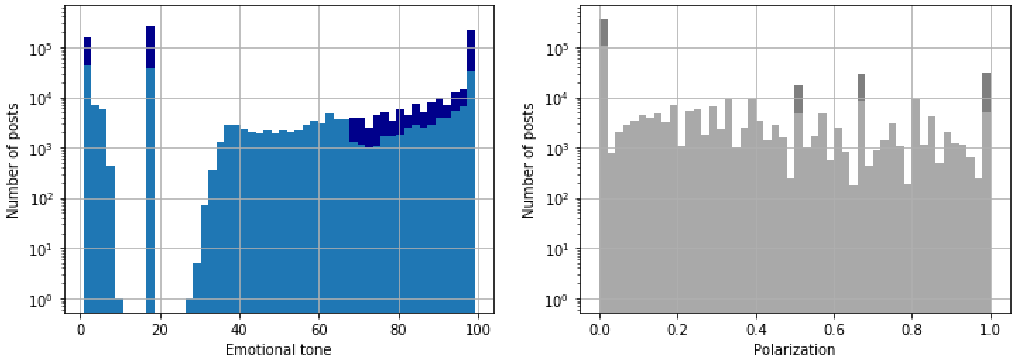

Figure 6: Distribution of posts in 50 bins by their emotional tone and vote polarizationon a log scale. For emotional tone, the darker color illustrates the distribution for all posts with the standard word limit (≥ 10 words), while the lighter color indicates posts with a more restrictive limit (≥ 50 words), showing no remarkable difference between both sets. In regards to polarization the darker color indicates the vote polarization of all posts, while the latter illustrates the distribution for a set arbitrary minimum of votes (≥ 5 votes), again showing no significant difference. Brighter colors illustrate the distribution when the set lower word limit is raised (to ≥ 50 words).

words, using a cumulative distribution function. In both cases, votes and emotions, positive evaluations are more pervasive than negative ones. Approximately 65% of all posts received one or more upvotes, as opposed to 30% with a minimum of one downvote. Similarly, roughly 62% of all posts contain positive emotions words and only 48% come with negative emotion words according to LIWC2015. This confirms the first indication provided in Table 1, the total number of upvotes across the whole data set being more than three times as high as downvotes, that there is a general prevalence of positivity in the voting system which, in turn, may be connected to the general positivity of positive emotional content.

An analysis of individual posts for their distribution when measuring cognitive complexity is illustrated by Figure 8. While the variables cognitive processes and six-letter words are normally distributed, with means lying within each others standard deviation as documented in Table 2, analytic thinking as a summary and index variable shows a U-shaped distribution with a very large number of posts that come with either the minimum or the maximum percentage of words that LIWC2015 categorizes as analytic. While the technical documentation of the program give more information as to which words are considered to indicate analytic thinking (Pennebaker et al., 2015), we randomly chose a sample of three posts from each end of the analytic thinking spectrum and record them below to provide an idea of why this strong differentiation may occur. The first three posts

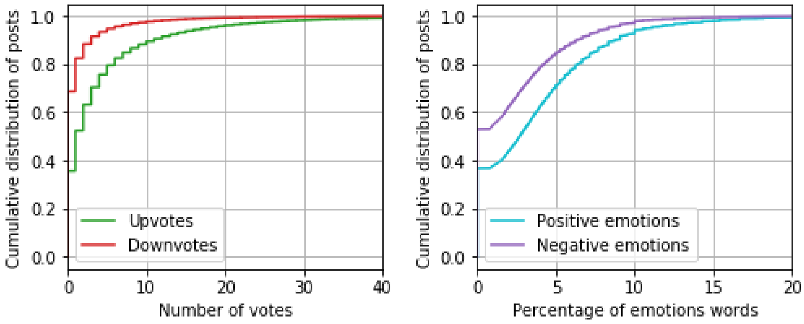

Figure 7: Distribution of posts (>10 words) for upvotes, downvotes, positive emotions, and negative emotions illustrated using cumulative distribution functions, showing a prevalence of positivity in both votes and emotional content of posts.

were all evaluated as having an analytic thinking variable of 99.0 (unaltered citations, including a timestamp and the category of the article it was posted under):[8]

> ”*Den Schwachsinn müssen wir wieder aus den USA importieren. Diesen ”regressive left” Wind.*” - Opinions, 02.04.2016

> ”*Praktizierbar wär's schon. Große, von EU/UNO verwaltete, Flüchtlingslager in den angrenzenden Staaten. Asylverfahren vor Ort sowie ein gerechter Aufteilungsschlüssel für die gesamte (!) EU. Jeder der mittels Schlepper ankommt wird sofort abgeschoben. Hohe Freiheitsstrafen für Schlepper, die im Moment ja nicht mal in U-Haft kommen. Nur wird das, aus welchen undurchsichtigen Gründen auch immer, nicht gewünscht!*”, Panorama, 20.08.2015

> ”*ich weiß, ein wenig illusorisch, aber halten wir einfach zu Montenegro. 3 Punkte für die wäre der Hammer - für UNS!*” - Sports, 14.06.2015

Likewise, the following three comments were evaluated with the smallest possible value of 1.0 for analytic thinking:

> ”*natürlich... nicht mal sein ziel hat rosetta erreicht. 46P/Wirtanen völlig verflogen und umkreist jetzt einen anderen kometen...*” - Science, 16.06.2015

[8]Usernames were not collected as part of the One Million Posts dataset to guarantee author privacy. Searching for the whole comment or snippets of it using online search engines, or the search function of derStandard.at, will not lead back to the posted article and the username.

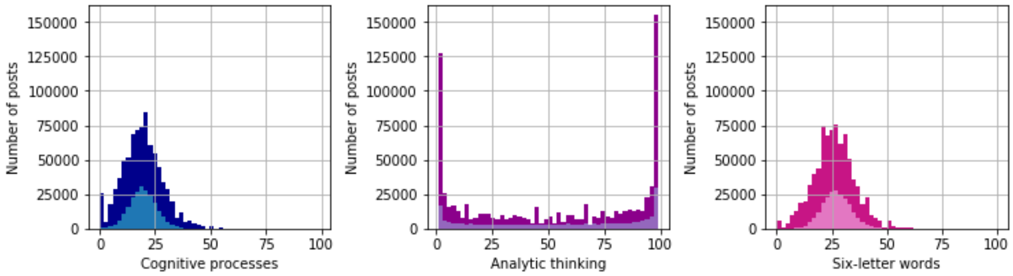

Figure 8: Distribution of posts for their measured cognitive complexity variables, cognitive processes, analytic thinking, and six-letter words. While the distributions of cognitive processes and six-letter words are unimodal and symmetric, analytic thinking shows large amounts of posts with either very high or very low values. The mean lies at 19.44 (standard deviation 8.50) for cognitive processes, 52.76 (56.23) for analytic thinking, and 26.40 (standard deviation 8.85) for six-letter words.

> "*Wir leiden all darunter, dass wir keine freie Sicht aufs Mittelmeer haben. Wenn ich mir wünschen darf, hätte ich lieber Triest, als Südtirol. Ist aber kein Wunschkonzert.*" - Panorama, 16.02.2016

> "*Mitnichten. Dann haben sie den Koran und das NT nicht gelesen.*" - International, 17.10.2015

Because there is no further information in the official documentation how analytic thinking is calculated other than being based on function words (Pennebaker et al., 2014, 2015) we do not have an explanation as to why these large numbers of posts with minimum and maximum values occur, other than it seemingly not being a calculation error resulting from our set word minimum.

Article discussions containing at least 10 posts with a 10 words minimum are illustrated in Figure 9, showing that the majority of discussions have less than 20 users who create less than 40 posts in total, with an average number of under 2 posts per user. Figure 10 illustrates how engaging these discussions are for users by the length of their arguments, measured as average and total number of words per post in the discussion. The bell-shaped distribution shows that most discussions fall around the average number words per post. A similar pattern is visible in Figure 11 in regards to upvotes and downvotes. All three emotion variables, positive and negative emotions, and emotional tone, are unimodal and symmetric, resembling a normal distribution and confirming a prevalence of positive emotions in discussions.

To get an overview of discussion dynamics, we also analyzed how the number of users affects voting behavior and emotion content of discussions.

Individual posts statistics							
	Mean	**Med.**	**Dev.**	**Var.**	**Err.**	**Min.**	**Max.**
Upvotes	4.07	1.00	8.09	65.45	0.01	0.0	455.00
Downvotes	1.19	0.00	3.78	14.32	0.00	0.0	233.00
Words	39.90	29.00	29.98	898.92	0.03	10.0	500.00
Emo. tone	45.74	17.39	40.73	1658.99	0.05	1.0	99.00
Positive emo.	3.57	2.70	4.10	16.84	0.00	0.0	100.00
Negative emo.	2.13	0.00	3.15	9.92	0.00	0.0	100.00
Cognitive p.	19.17	18.75	8.98	80.66	0.01	0.0	99.19
Analytic t.	52.85	56.84	38.27	1464.75	0.04	1.0	99.00
Six-letter w.	26.00	25.93	9.32	86.78	0.01	0.0	100.00
Polarization	0.20	0.00	0.31	0.10	0.00	0.0	1.00
Emo. charge	5.70	5.08	5.07	25.71	0.01	0.0	100.00

Table 2: Statistics of all posts containing at least 10 words, documenting mean, median, standard deviation, variance, standard error, minimum and maximum values. Emotional tone, analytic thinking, and polarization are index variables and described in the 'Methods' chapter. Positive emotions, negative emotions, cognitive processes, and six-letter words are percentages of words that fit the category. Emotional charge is the sum of positive and negative emotions.

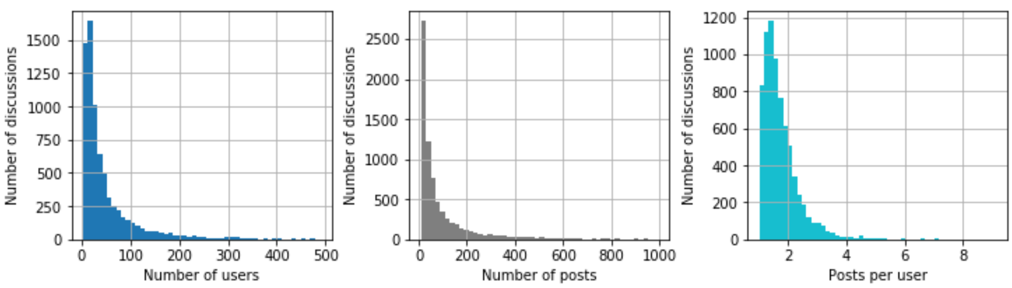

Figure 9: Distributions of article discussions in 50 bins each by their number of users, number of words, and the number of posts per user. The average discussion has a mean number of 53.61 (median 28.00) users, a mean number of 105.77 (44.00) posts, and the average user in a discussion writes 1.76 (1.58) posts. There are fewer discussions as either posts or users increase.

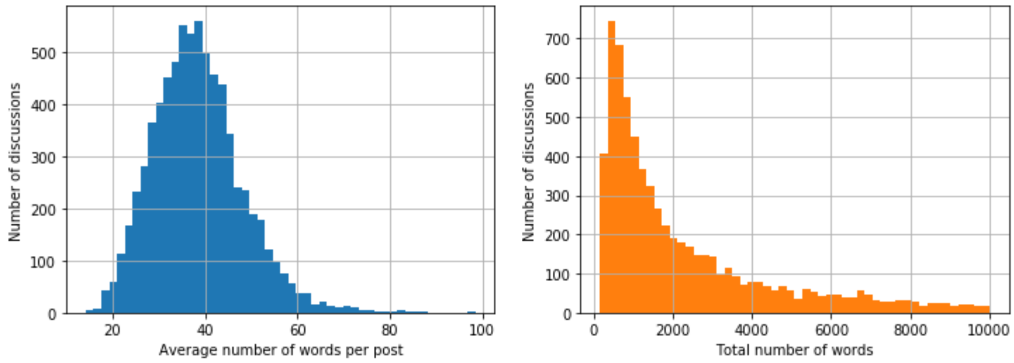

Figure 10: Distributions of article discussions for their average and total number of words per post. The mean number of the average words per post in a discussion is 38.48 (standard deviation 9.72), while the mean of the total number of posts lies at 4232.77 (median 1639.00). While the average number of posts is symmetric around a central mode, the number of discussions decrease as the total number of words increases.

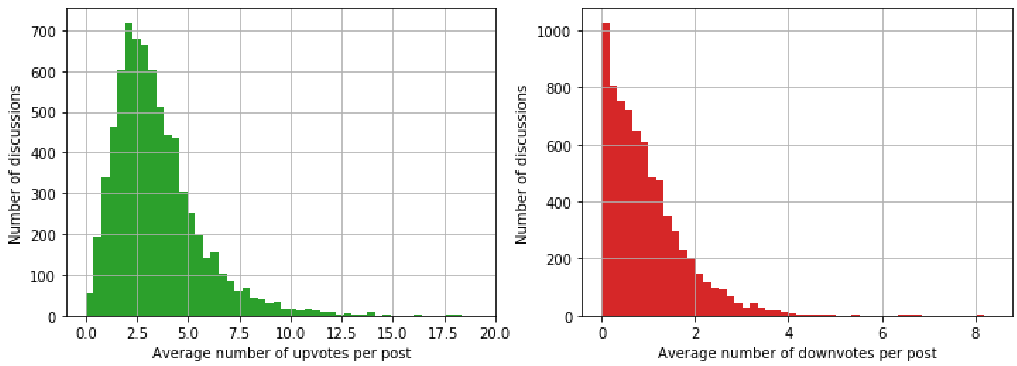

Figure 11: Distributions of article discussions for their average numbers of upvotes and downvotes per post. The distribution of discussions by their average number of upvotes per post resembles a normal distribution with a positive skew and a mean of 3.46 (standard deviation 2.13), while the distribution by average number of downvotes shows a clear prevalence of lower numbers of downvotes with a mean of 0.96 (median 0.76).

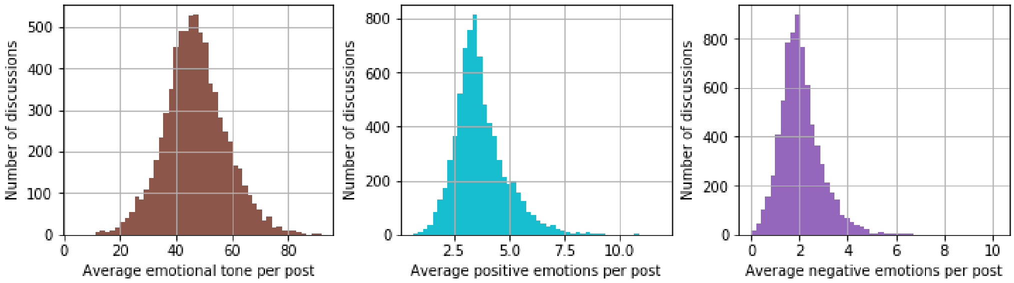

Figure 12: Distributions of discussions by the average emotional content of their posts. All three measured variables are bell-shaped distributions, with emotional tone having a mean of 46.92 (standard deviation 11.01), implying an overall positive tone of discussion. Positive emotions have a mean of 3.69 (1.17), and negative emotions 2.07 (0.87), confirming the prevalence of positive emotions over negative ones.

Figure 13 shows the relationships between number of users and (a) upvotes, (b) downvotes, (c) positive emotions, and (d) negative emotions. While there is a clear significant linear relationship for upvotes (Pearson coefficient 0.890) and downvotes (0.806; both p-values < 0.001) and a linear trend of both upvotes and downvotes increasing as more users participate, the same does not hold for emotional content. Neither positive (-0.051) nor negative emotions (0.038) correlate with the number of users in discussions significantly ($p < 0.001$). The relationship also does not translate to index variables emotional tone and polarization, as seen in Figure 14. This means that as the number of discussion participants rises, the emotional content of posts remains stable, and while the number of upvotes and downvotes increase individually, the polarization of post does not change on average.

In regards to annotations, only 2,951 of 11,773 annotated posts in the corpus contain information about sentiment. 1,492 posts were noted to contain a neutral sentiment, 1,430 a negative sentiment, and only 29 posts were annotated as positive. The other 8,822 annotations contain no information about emotional content and were hence disregarded for our analysis. This ratio is illustrated in Figure 15, showing that less than 1% of all annotated posts were described as having a positive sentiment, while the rest of the annotations are almost equally split between neutral and negative annotations. The smaller number of positive sentiment posts was caused by the selection process for comments to annotate, which specifically targeted those with a potentially negative sentiment (Schabus and Skowron, 2018).

Figure 16 compares the distribution of annotated posts for their emotional tone and polarization. There appear to be no differences between annotated posts when comparing their polarization. However, when com-

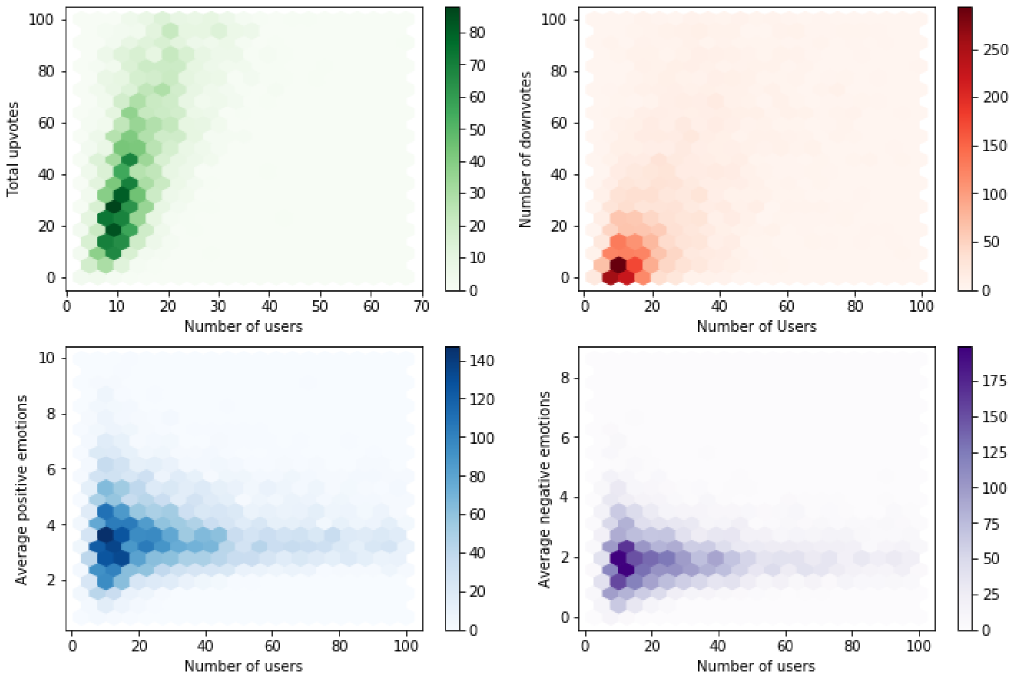

Figure 13: Distributions of discussions in grids of 20x20 bins, showing relationships of total votes and average emotions with the total number of users. Both upvotes and downvotes strongly increase with the total number of users in a discussion (Pearson coefficients 0.890 and 0.806; p-values < 0.001), but the average number for positive emotions (-0.051; < 0.001) and negative emotions (0.038; 0.001) remains stable.

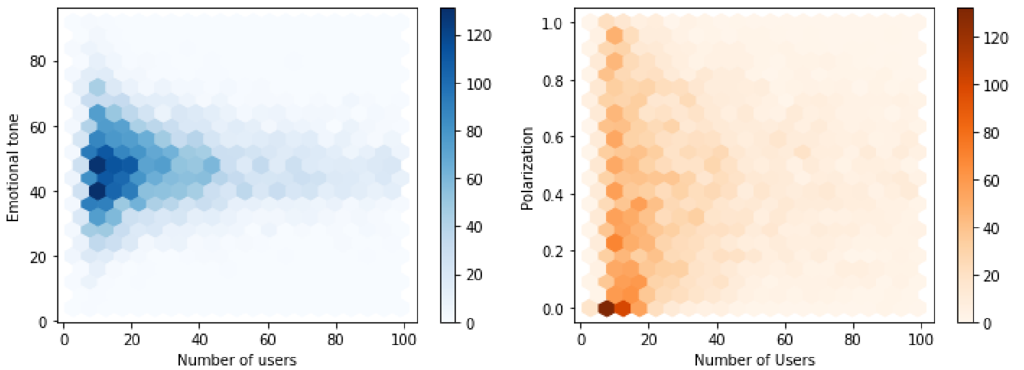

Figure 14: Distributions of discussions in grids of 20x20 bins for their relationship between their number of users and the average emotional tone and polarization of their posts. Neither show significant correlations with a Pearson coefficient of -0.053 (p-value < 0.001) for the relationship between the number of users and emotional tone, and 0.027 (0.018) for the number of users and polarization.

Aggregate discussions statistics							
Average	**Mean**	**Med.**	**Dev.**	**Var.**	**Err.**	**Min.**	**Max.**
Upvotes	3.46	3.04	2.13	4.54	0.02	0.00	19.09
Downvotes	0.96	0.76	0.86	0.75	0.01	0.00	17.73
Words	38.48	37.77	9.72	94.53	0.11	14.21	98.42
Emo. tone	46.92	46.49	11.01	121.20	0.13	4.14	91.68
Positive emo.	3.69	3.49	1.17	1.36	0.01	0.67	11.75
Negative emo.	2.07	1.95	0.87	0.76	0.01	0.00	10.17
Cognitive p.	18.77	18.83	2.31	5.32	0.03	4.31	32.60
Analytic t.	54.24	54.18	9.83	96.55	0.11	15.59	98.94
Six-letter w.	25.83	26.03	2.90	8.40	0.03	15.50	42.82
Polarization	0.42	0.41	0.25	0.06	0.00	0.00	1.00
Emo. charge	5.76	5.61	1.38	1.90	0.02	1.33	14.14
Total	**Mean**	**Med.**	**Dev.**	**Var.**	**Err.**	**Min.**	**Max.**
Words	4233	1639	7548	$10^{7.756}$	88	160	128307
Upvotes	435	138	978	$10^{5.98}$	11	0	27679
Downvotes	127	34	301	90637	4	0	6299
Users	54	28	77	5911	1	3	1131
Posts	106	44	184	33847	2	10	2992

Table 3: Statistics of article discussions as aggregates of at least 10 posts, documenting mean, median, standard deviation, variance, standard error, minimum and maximum values. The top part documents numbers for the average values across posts in a discussion (mean of all posts), while the bottom section shows total values in discussions (sum of all posts).

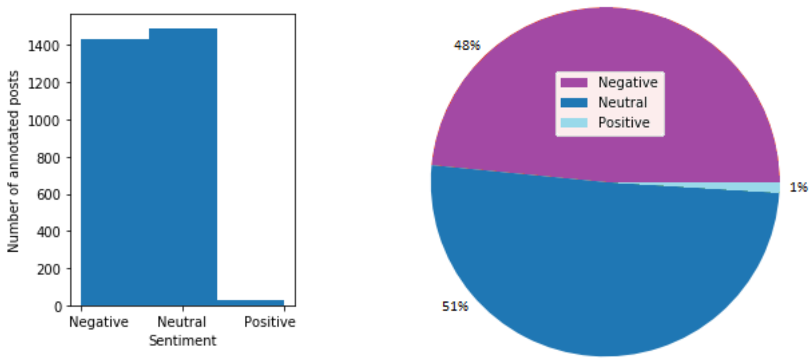

Figure 15: Visualization of the number of annotations, showing the prevalence of 1,492 posts that were annotated as having a neutral sentiment (50.56%) and 1,430 with a negative sentiment (48.45%), compared to 29 positive posts (0.98%).

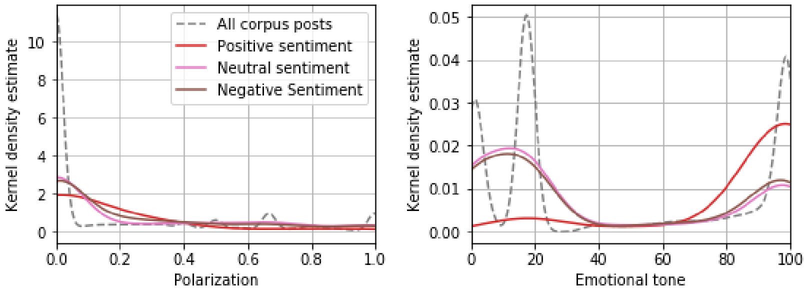

Figure 16: Distribution of posts by their emotional tone and polarization index variables, comparing the positively (red), neutrally (pink), and negatively (purple) annotated posts with the distributions of all posts (grey, dashed). While there appears to be no difference between annotated posts when it comes to their polarization, posts with a positive annotation also tend to have a higher emotional tone, indicating stronger positive content compared to those with neutral and negative annotation, which also show no difference between them.

paring positive, neutral, and negative sentiment posts for their emotional tone, there is a clear differentiation possible. Posts annotated as having a positive sentiment do, in fact, have an increased chance of having a high degree of emotional positivity in their textual content, as measured by the LIWC2015 emotional tone variable. Neutral and negative sentiment posts, in contrast, have a higher chance of being negative in tone. Interestingly, there appears to be no difference in the emotional tone between posts annotated as having a neutral sentiment, compared to those annotated as negative. We detail this observation in the next chapter of our thesis.

Results

The following chapter presents the results of our master's thesis in three parts, our simulation results obtained from running our implemented model of collective emotions and emotional expressions based on the Cyberemotions agent-based modeling framework (Schweitzer and Garcia, 2010), our sentiment analysis results generated using the quantitative text analysis tool LIWC2015 (Pennebaker et al., 2014) to measure the affective properties of online discussions of news articles of derStandard.at, and a comparison of our model data to the real-world data of the One Million Posts corpus (Schabus et al., 2017).

For the evaluation of linear relationhips between variables, we calculated Pearson correlation coefficients ϱ, rating the relationship with a number between -1 and 1, whereas a negative number indicates a negative relationship, a positive number a positive one, and a value of 0 implies no linear relationship. In this context, if $|\varrho| > 0.1$, we describe the relationship as weak; if $|\varrho| > 0.3$ as moderate; and if $|\varrho| > 0.5$ then we assume a strong correlation. We record a relationship as significant when the respective p-value is smaller than 0.05. In order to compare our simulation results to the One Million Posts data, we further use the least squares approach to calculate the slope of a linear regression with a 95% confidence interval.

Simulating online emotional expressions

We abstract emotional contagion (Aragón et al., 2017a) between individuals and the emergence of collective emotions using agent-based modeling, a method which is especially useful for emphasizing emerging patterns of group behavior that are the result of interactions between individuals (Wilensky and Rand, 2015). The Cyberemotions framework outlined in the previous chapter provides a blueprint for such a model, describing the dynamics of emotions in online interactions (Garcia et al., 2016). Using this framework, we implemented our model using version 3.7 of the programming language Python (Van Rossum and Drake Jr, 1995).

Our implementation of the Cyberemotions framework was designed to investigate the dynamics of collective emotions in online discussion forums, where users may directly respond to posts of others in the same thread. The model and its results are intended to be used as a hypotheses generator and its predictions may be tested against real-world data. In our model, agents represent human individuals who communicate their emotional states with each other using a communication field that may represent any thread in an online discussion forum. In turn, this communication field provides feedback that affects the individual emotional states of the agents. Adding to the mathematical formalization of the Cyberemotions framework (Schweitzer and Garcia, 2010; Garcia et al., 2016), our model also includes recent research into the temporal dynamics of affect (Pellert et al., 2020) that allows us to simulate the relaxation of valence and arousal towards realistic baseline values. Furthermore, our implementation of the satiation phase is the first attempt at modeling agent satiation as a quadratic function of arousal, to abstract the decrease of discussion activity over time in a realistic yet simple manner. The parameters used for our simulation are listed in Table 4.

Each model run goes through the following phases at each time step:

1. Perception: An agent perceives the communication field and adjusts its state variables valence and arousal according to the emotional charge of the field.

2. Expression: If the arousal of the agent passes an activity threshold, the agent expresses its valence to the communication field and experiences an immediate down-regulation of its emotions.

3. Relaxation: The agent relaxes and its emotional states variables valence and arousal regress towards their baseline values.

4. Satiation: If the arousal of the agent passes a satiation threshold, the agent drops out of the simulation and is no longer able to perceive the communication field or express its emotions.

5. Communication: The communication field adjusts its emotional charge according the the impact of all agent expressions.

Parameter	**Symbol**	**Value**
Valence baseline	b	0.056
Arousal baseline	d	-0.442
Decay factors	γ_v	0.367
	γ_a	0.414
Amplitudes	A_v	0.300
	A_a	0.300
Arousal threshold	τ	0.000
Down-regulation factors	k_v	0.380
	k_a	0.450
Satiation coefficient	c	1.000
Valence coefficients	b_0	0.140
	b_1	0.000
	b_2	0.057
	b_3	-0.047
Arousal coefficients	d_0	0.178
	d_1	0.145
	d_2	0.000
	d_3	0.000
Field charge baseline	h	0.000
Field decay	γ_h	0.700
Impact coefficient	s	$40/N$

Table 4: Set model parameters to generate our simulation results. The function I scales impact coefficient s linearly based on the number of simulated agents N. Parameters τ, c and h are free variables without published empirical estimates, while all others were derived from Schweitzer and Garcia (2010), Garcia et al. (2016), and Pellert et al. (2020).

Phases 1-4 occur once for every agent at every time step, while phase 5 happens once per time step after all agents took their actions. We formalize satiation as a function $S[a(t)] = |a(t)|^2 c$ of agent arousal a at time step t. Taking the squared absolute value of a and multiplying it with satiation factor c, it provides us with a value that can be compared to a stochastic component ξ, representing a random chance of an agent dropping out of a discussion. Hence, if $S[a(t)] > \xi$ an agent will drop out of a simulation and no longer be able to express its emotions to the communication field. Once all agents have dropped out, the simulation ends. We modeled satiation in such a way, because it allows us to:

- capture individual stochastic influences on agent participation;
- remove agents from the simulation over time in a realistic manner;
- keep the behavioral rules of the agent simple.

Test runs of our model showed the need to linearly scale impact parameter s to the total number of agents in the simulation, in order to observe the emergence of collective emotions. We formalized this using the function $I(N) = \frac{i}{N}$, whereas N constitutes the number of agents and i a preliminary arbitrary value. Setting $i = 15$ seemed to produce patterns of collective emotions among all numbers of agents, although not necessarily with realistic measurable outcomes.

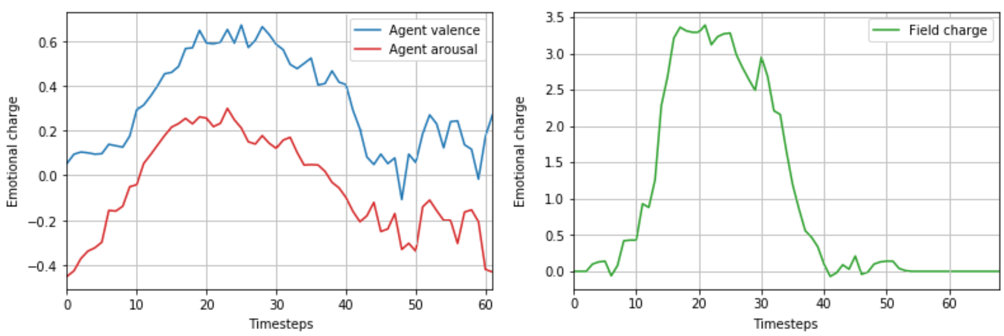

Figure 17: Depiction of agent and field variables over time during a Cyberemotions model run representing an online discussion, using mean values collected over 150 model runs with 150 agents each, and a time limit of 150 discrete time steps. Agent baselines were randomly drawn from a normal distribution with a standard deviation of 0.1 around baseline parameters (see Table 4). Impact factor s was set to 0.1 for the simulation, while c was 1.0. All model runs automatically ended when all the agents have dropped out.

Figure 17 illustrates the mean results of 150 test runs for 150 agents, assuming slightly positive valence baseline values for agents and a negative arousal baseline in accord with recent research into temporal emotion dynamics, and using an arousal threshold of 0. For this simulation we implemented a satiation factor of 1.0 and an impact parameter of 0.1 based on the formula outlined above, the latter defining the impact an agent expression has on the emotional charge of the communication field. It shows how a collective emotion appears when the emotional charge of a discussion rises, agent arousal thresholds are reached, and when agents express their emotions. After reaching an emotional high point, the collective emotion fades out again over time as agents drop out of the discussion after becoming satiated. This depiction of a collective emotion may not be a realistic representation of emotional content in an online discussion, but it shows the potential capabilities of our model to generate patterns of behavior that may be compared to real-world data, for example by looking at the number of expressions generated for a set number of agents.

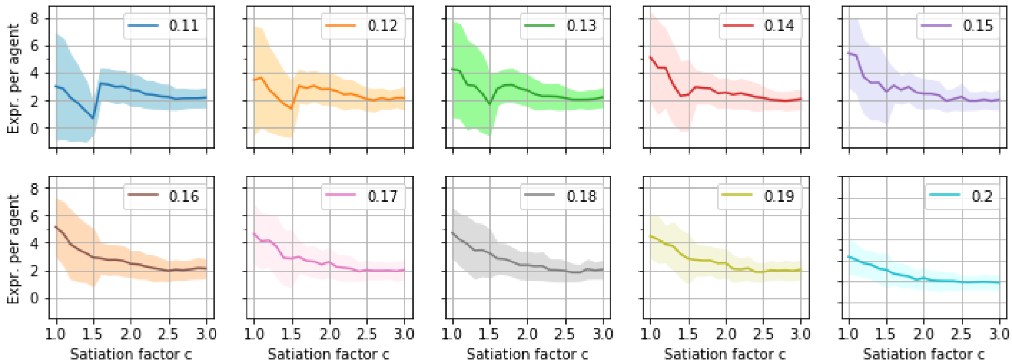

Figure 18: Relationship between the average number of expressions per agent in a simulation of 100 agents and the satiation factor c for different impact parameters s. Shown are mean values of the average number of (positive and negative) expressions per agent of 100 model runs per setting. Colored areas indicate their standard deviation. Other parameters are listed in Table 4.

As Figure 18 shows, using the average number of expressions per agent in a simulation as measure, the Cyberemotions model behavior can be controlled using satiation coefficient c and impact parameter s. An observed tipping point when $s < 0.11$ in the number of agent expressions fades away as s increases and only occurred in this form in simulations of 100 agents during test runs. For example, with 200 agents we instead observed a spike when $c = 2.5$ and $s > 0.11$, amplifying the expected number of expressions by a factor of approximately 2. Furthermore, the standard deviations of the average number of expressions per model run become smaller with an increase of s. While our implementation of the satiation function includes a stochastic component that takes individual influences on the dropout of agents into account and hence may remain constant across agents and model runs, we nonetheless have to first find a value that produces model results that accurately describe real-world data. Likewise, impact parameter s must be verified, potentially as a function of the total number of agents, because test runs showed that using the same value for s can produce vastly different outcomes for different number of agents. The collective emotion pattern visible in Figure 17, for example, does not emerge when running the model with 50 agents instead of 150 using $s = 0.1$ and $c = 1.0$, all other parameters being equal. For this reason, we suggest using a function of the number of agents simulated to scale the impact factor realistically, based on the assumption that the emotional impact of a single post on the overall sentiment of a discussion decreases as the number of participants in the same discussion increases.

Analyzing affective meaning in online discussions

After using LIWC2015 to evaluate our One Million Posts dataset, we primarily visualize the relationships between LIWC variables using linear regression and a 'least squares' approach. In order to analyze the cognitive complexity of discussion participants, we further relied on clustering using k-means vector quantization. For a comparison of post content and moderator annotations, we calculated a receiver operator characterstic and the area under its curve. We started our analysis of the One Million Posts corpus with four hypotheses:

1. The sentiment and polarization of a response post are correlated with the sentiment and polarization of its parent post.
2. The aggregate level of polarization in a discussion is correlated with the sentiment of its posts.
3. The sentiment of an article is correlated with the aggregate sentiment of its discussion.
4. The sentiment and cognitive complexity of posts are clustered in two modes that correspond to system 1 and system 2.

Our analysis results, including visualizations of the data, are publicly accessible online.[9] The following sections lay out each hypothesis one by one.

Hypothesis 1

The sentiment and polarization of a response post are correlated with the sentiment and polarization of its parent post.

Hypothesis 1 assumes a linear relationship for emotional charge and polarization between parent and response posts. This relationship is illustrated by Figure 19, showing a slight but significant increase in emotional charge of response posts as it increases in parent posts, for both positive emotions (Pearson correlation coefficient 0.175; p-value < 0.001), negative emotions

[9] `https://github.com/r-andre/sentiment_analysis-million_posts` (last accessed on 11.04.2020)

(0.244; $p < 0.001$), total charge (0.140; $p < 0.001$), and also vote polarization (0.182; $p < 0.001$). The same description applies to the relationship for individual upvotes in parent and response posts (0.301, $p < 0.001$). Only individual downvotes seem to share a marginal correlation (Pearson 0.087; p-value < 0.001). Considering these predictive relationships between parent and response posts for positive and negative emotional charge, upvotes, and overall polarization of votes, we conclude that hypothesis 1 is supported by our data. Although the correlation is weak, the sentiment and the polarization of a response post are correlated with the sentiment and polarization of their parent post.

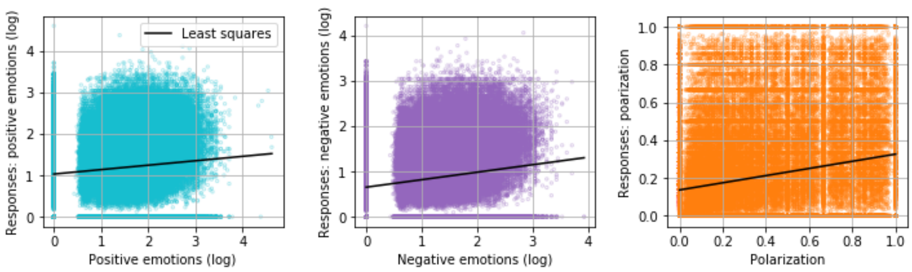

Figure 19: Relationship between parent and response posts for positive emotions (plotted: log transformed values), negative emotions (log values), and vote polarization. All three show a weak but significant correlation with the following Pearson coefficients: positive emotions 0.175; negative emotions 0.244; polarization 0.182 (all p-values < 0.001). These predictive relationships are illustrated by a linear least squares regression line.

Hypothesis 2

The aggregate level of polarization in a discussion is correlated with the sentiment of its posts.

Hypothesis 2 predicts a correlation between emotion variables and polarization in discussions. Figure 20 illustrates this using emotional tone and emotional charge of discussions, showing a weak but significant negative correlation (Pearson -0.175; $p < 0.000$) between emotional tone and polarization, but no correlation between total emotional charge and polarization (-0.015; $p < 0.196$). Further analysis of positive and negative emotions explains this discrepancy: while the correlation between positive emotions and polarization is negative (-0.140; $p < 0.000$), the relationship between negative emotions and polarization is positive (0.163; $p < 0.000$). For these reasons, we consider our hypothesis to be partially supported by the data. While the aggregate level of polarization in a discussion is correlated with

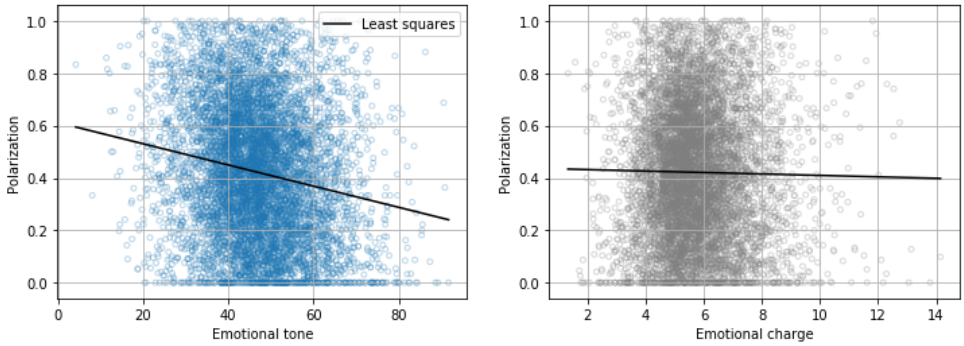

Figure 20: Relationship between discussion sentiment and polarization, the former measured as emotional tone and emotional charge. While emotional tone shows a weak but significant negative correlation with vote polarization (Pearson coefficient -0.119; p-value < 0.001), the latter indicates no linear correlation (0.002; $p = 0.841$). These predictive relationships are illustrated by a linear least squares regression line.

the emotional positivity of a discussion, there is no correlation with its overall emotional charge.

<u>Hypothesis 3</u>

The sentiment of an article is correlated with the aggregate sentiment of its discussion.

Hypothesis 3 concerns articles and discussion developing around them, predicting a correlation between the emotional content of an article and those of its posts. As shown in Figure 21, the spread of emotional tone in discussions is wide and there is no correlation with the tone of the corresponding article visible (Pearson -0.010; p-value 0.378). Likewise, there appears to be no correlation between the emotional charge of an article and the overall emotional charge of its discussion (-0.018; $p = 0.124$). On this basis, we reject hypothesis 3, because our analysis did not reveal a correlation between the sentiment of an article with the aggregate sentiment of its discussion.

<u>Hypothesis 4</u>

The sentiment and cognitive complexity of posts are clustered in two modes that correspond to system 1 and system 2.

Hypothesis 4 builds on dual-process theory and assumes clustering of posts in two clusters for the relationship of emotional charge with cognitive complexity, the latter being measured using cognitive processes, analytic

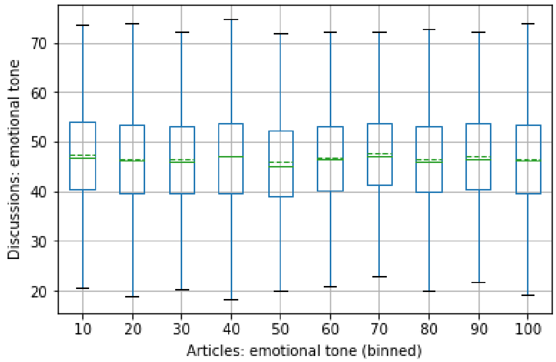

Figure 21: Relationship between the emotional tone of an article aggregated in 10 bins, and that of its discussion. It does not allow the prediction of the emotional tone of a discussion based on the emotional content of the article (Pearson -0.010; p-value 0.378). The plot shows feasible minimum and maximum values, first and third quartile, median (green), and mean of the variable (green, dashed). Outliers are not shown for transparency reasons due to their large number.

thinking, and six-letter words variables. This ideally implies clustering at 1. a high emotional charge and low cognitive complexity, and 2. a low emotional charge and high cognitive complexity. As shown in Figures 22 and 23, we did not observe such clustering in the posts of the corpus data set. Cluster analysis using k-means revealed two clusters at 1. moderate to high levels of cognitive complexity and moderate emotional charge levels, and 2. moderate to high cognitive complexity and low emotional charge for cognitive processes and six-letter words variables. For analytic thinking both clusters are located at moderate levels of emotional charge, with one being very high in analytic thinking, the other very low. Hence, we consider hypothesis 4 not to be supported by our data, since the sentiment and cognitive complexity of posts are not clustered in two modes that correspond to the two systems postulated by dual-process theory.

Analyzing annotated posts we did not discover visible patterns in their distribution for vote polarization. However, there are clear deviations in emotional tone for comments that were annotated as having a positive sentiment. These posts tend to also have a higher chance of being strongly positive in emotional tone. Figure 24 plots the relationships between annotated sentiment and the actual emotional content we measured using sentiment analysis. While there appears no difference in the distribution of negatively and neutrally annotated posts for positive or negative emotions and emotional tone, positive annotations seem to capture emotional content realistically. Posts that were annotated as having a positive sentiment are also high in positive and low in negative emotions, resulting in a very

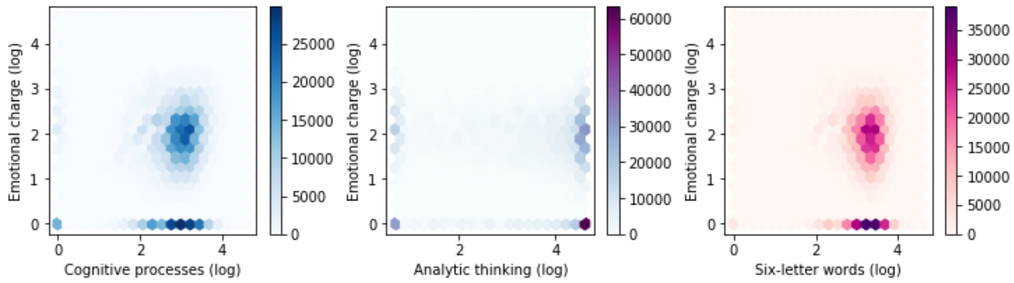

Figure 22: Distribution and clustering of individual posts for their cognitive complexity and emotional charge in a hexbin grid of 20x20 bins, using the variables cognitive processes, analytic thinking, and six-letter words as a measure of cognitive complexity (plotted: log-transformed values). Dense areas correspond to the locations of clusters generated by a k-means analysis.

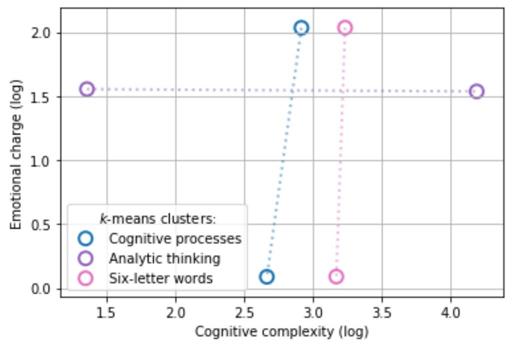

Figure 23: Clustering of individual posts for cognitive processes, analytic thinking, and six-letter words. Circles indicate the two k-means clusters of the nearest mean of each variable (using log-transformed values). Dotted lines indicate the two connected clusters of the same variable.

positive emotional tone.

Calculating the receiver operator characteristic and looking at the area under the curve, the performance for positive annotations is moderate (AUC score 0.840), while that of neutral and negative annotations is similarly low (neutral 0.479; negative 0.508), indicating that LIWC emotional variables may be reliable predictors of positive annotations, but not of neutral or negative ones. Nonetheless, the small number of such positive comments in our dataset leads us to caution in the interpretation of these results. As to why neutral and negative sentiment posts do not show more distinct patterns of distribution for emotional tone, or why they are seemingly hard to differentiate for moderators, an in-depth analysis of individual posts may lead to more meaningful conclusions.

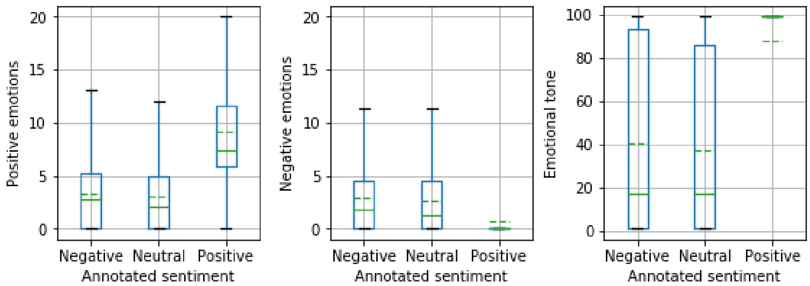

Figure 24: Relationships between annotated sentiment and actual emotional content as measured by LIWC2015, using the variables positive emotions, negative emotions, and emotional tone. While positively annotated posts have a higher degree of emotional positivity, there appears to be little to no difference between posts annotated as neutral or negative. The plot shows feasible minimum and maximum values, first and third quartile, median (green), and mean of the variable (green, dashed).

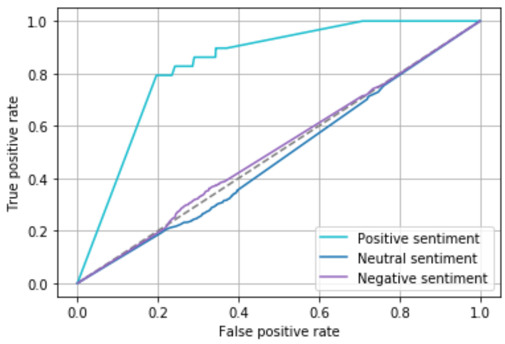

Figure 25: Ability to predict sentiment annotations of posts by moderators using LIWC measures of their emotion tone. Lines indicate the ROC (receiver operator characteristic) curves for posts annotated as having positive, neutral, or negative sentiments. Calculating the area under the curve confirms that positive annotations can be predicted with moderate accuracy, but neither neutral nor negative annotations are captured accurately (ROC-AUC scores: positive sentiment annotations 0.840, neutral 0.479, negative 0.508).

Comparing simulation results to empirical data

We compare the simulation results of our implementation of the Cyberemotions framework to the analysis results of the One Million Posts corpus using the level of activity in discussions. In the former case, we measure activity as sum of all emotional expressions by agents, and in the latter as total number of posts by users. In order to achieve comparable results, we ran our model 100 times each for 20 to 300 agents in intervals of 20 agents with the parameters listed in Table 4. A comparison of the number of expressions to the number of posts in corpus discussions with a similar number of users is illustrated in Figure 26, showing a increase in the number of expressions and posts as the number of discussion participants, agents or users, increases. Especially mean values of measured activity appear to scale in an analogous manner in both cases as the number of discussion participants increases, while the third quartile and feasible maximum values are higher in our simulation results beyond 100 agents.

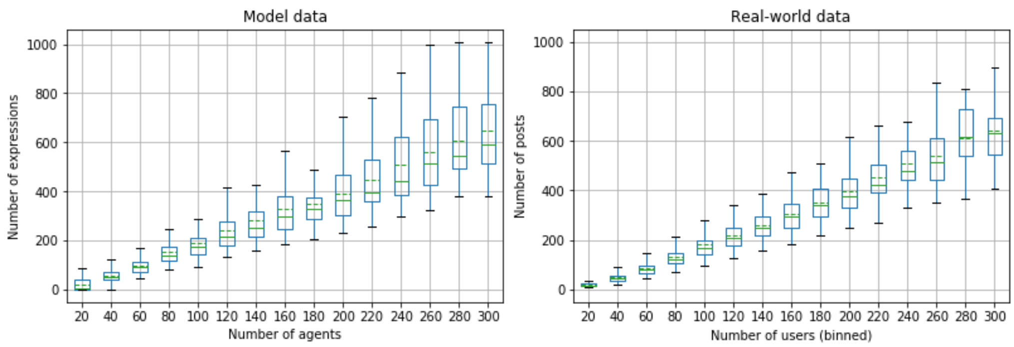

Figure 26: Comparison of the relationship between number of agents and agent expressions in simulations using our Cyberemotions model set to parameters listed in Table 4 (left), to number of users and posts in discussions of the One Million Posts corpus (right). While the model was run with the specific number of agents, corpus discussions were binned around intervals of 20 users (e.g. bin 50 includes all discussions with 41-60 users). In both cases, shown are feasible minimum and maximum values, first and third quartile, median (green), and mean of the variable (green, dashed).

We scale the impact parameter as a linear function of the number of agents $I(N) = \frac{i}{N}$. Based on simulation results, we find the values $i \approx 40$ to achieve the most realistic results when measuring discussion activity as number of agent expressions. In order to assess the ability of our model to capture activity in online discussions and to measure its accuracy, we used the logarithmic values of Figure 26 and calculated a linear regression using the least squares method. The regression results are documented in Table 5 and visualized in 27. Impact coefficient i, all other parameters being equal, controls the degree of the slope, with an increase of i also leading to an increase of exponent β. By adjusting i we are able to generate a value for the regression slope that closely resembles our real-world data set. In both cases, our model and the One Million Posts corpus, the number of posts in a discussion grows at an average superlinear rate β of 1.1 as the number of discussion participants increases (model expressions $\beta = 1.110$; corpus posts $\beta = 1.111$), with both exponents lying within each others 95% confidence intervals. The respective p-values signify that both relationships are significant.

Dependent variable	β	R^2	p	SE	95% CI
Model expressions	1.110	0.991	< 0.001	0.003	(1.026, 1.193)
Real-world posts	1.111	0.962	< 0.001	0.004	(1.103, 1.118)

Table 5: Regression table for the agents-expressions relationship (using $i = 40$ and parameters in Table 4) and the users-posts relationship illustrated in Figure 27, listing the degree of the slope of the regression line, its R-value, its p-value, its standard error, and the 95% confidence interval of the slope.

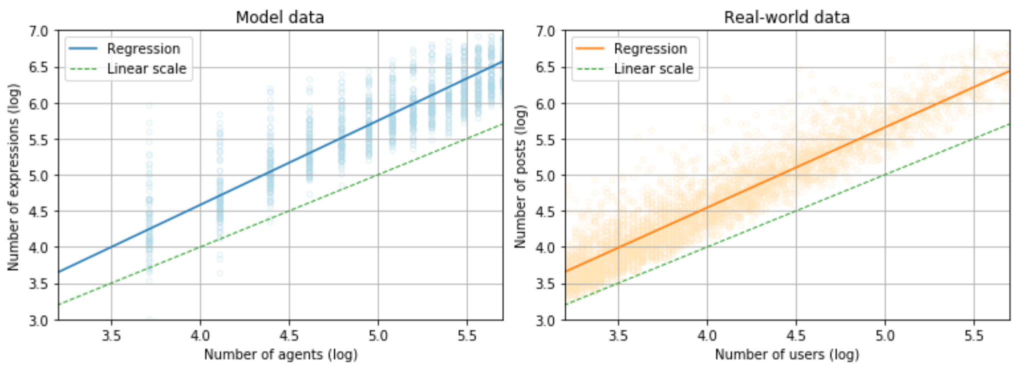

Figure 27: Comparison of the relationship between number of agents and agent expressions in simulations using our Cyberemotions model set to parameters listed in Table 4 (left; log-transformed), to number of users and posts in discussions of the One Million Posts corpus (right; log-transformed). The model plot markers indicate the individual results of all sets of model runs, while the real-world plot scatters all analyzed discussions. Straight lines indicate the linear least squares regression of the relationship. Dashed lines show linear growth rates with a scaling exponent $\beta = 1.0$ for comparison. Both model and real-world data scale on average at a similar superlinear rate (model $\beta = 1.110$; real-world $\beta = 1.111$).

Discussion

Our results of the One Million Posts corpus analysis demonstrate linear correlations between the emotional content of parent and response posts, and between their vote polarization. The simplicity of our method provides additional context when interpreting these results. They indicate that it is possible to predict the emotional content and affective reactions based on data of previous posts in a discussion using simple quantitative measures, such as counts of emotion words in a post and its upvotes and downvotes by other users. These correlations can further be interpreted as a indicators of collective emotion dynamics and the emergence of new qualities in emotional responses, because reduced variability on the collective level is an emergent property of collective emotions (von Scheve and Salmela, 2014; Goldenberg et al., 2017). Putting our analysis results in the context of collective emotion theory, they suggest that users influence each other's emotional states, leading to a convergence in the emotional content of their posts, and their evaluation of posts by other users. While the emotional tone of news articles influencing the tone of the discussion may also appear a logical explanation of reduced variability in affective responding, our analysis does not support this hypothesis. We did not find significant correlations when comparing the sentiment of news articles to the aggregated sentiment of its discussion, with articles generally containing less emotion words than forum posts. Nonetheless, a more elaborate analysis of articles collected in the One Million Posts corpus, for example by news category, may shed additional light on their relationship to the discussions developing around them and the emotional content of user reactions.

One limitation of using agent-based models to simulate human behavior is the lack of realistic input to generate applicable results. Social media data does not only provide means to verify and validate simulation results, but may also be used as input data to solve the aforementioned problem by providing empirical grounding for agent-based simulations (Padilla et al., 2014). Using data from experimental studies (Garcia et al., 2016) and text analysis of social media status updates (Pellert et al., 2020) as input for our model, our results illustrate how agent-based models can be successfully calibrated and validated using big data collected in social networks. Relying on the discussion dynamics of online forums as a reference, our adjusted model parameters were able to replicate real-world posting behavior on an aggregate level, by measuring the change in the number of emotional expressions as the number of discussion participants increases. Big data and data mining provide means to obtain information about human behavior and evaluate simulation results at a much larger scale than traditional approaches (Tolk, 2015). Specifically, social media captures patterns in the digital traces of human interactions (Padilla et al., 2014) that may otherwise remain unnoticed in traditional study environments. We therefore support the argument in favor of using big data as input to computational models of human behavior. This way, agent-based models may be increasingly used to approximate a realistic simulation of emotional contagion and expression, as is the goal with the Cyberemotions project and framework.

Using three LIWC counts of words that indicate cognitive activity and complexity, we were not able to find patterns of cognitive systems 1 and 2 in our data as we would suspect them. Initial formulations of 'hot' and 'cold' cognition systems suggest that hot system 1 processes are strongly influenced by emotional states, in contrast to the cold and complex system 2 (Abelson, 1963; Brand, 1985). This lead us to expect clustering of posts either (a) high in cognitive complexity and low in total emotional charge, or (b) low in cognitive complexity and high in emotional charge. Cluster analysis of the relationship between the aforementioned variables and positive and negative emotions, however, did not meet this expectation. Possible explanations are that the three variables we chose to indicate cognitive complexity may not be related to the cognitive processes described by system 1 and 2, or that these processes cannot be captured by analyzing the psychological meaning of words. Of course, recent theories on affective intelligence also highlight that systems 1 and 2 should not represent a dichotomy between emotion and reason, and that the influence of emotions on thinking is more nuanced and complex than a simple either-or differentiation between the two (Marcus, 2013; Zahn-Waxler et al., 2017). This would also mean that we need more sophisticated methods to detect these modes of thinking in online discussions, or that text is simply not the right medium to detect such subtle cognitive differences.

In our conceptual model, we used a simple linear function to scale the impact of agent expressions on the overall emotional charge of the discussion, based on the assumption that a single comment has more impact when there are fewer discussion participants. This may not be a realistic implementation and a more detailed formalization may lead to different results and other conclusions about the accuracy of the model. Likewise, the quadratic arousal function that we used to formalize emotional satiation and people dropping-out of online discussions may not capture the process as realistically as a more sophisticated model. It may also highlight the domain specificity of our results, because we believe the discussion environment can play a large role in satiation processes. Online environments such as Twitter can lead to a rapid development of large-scale discussions in a matter of minutes, while discussions of news articles of derStandard.at mostly appear to develop over the course of a couple of hours or a day. In our experience, users of traditional forums and message boards interact at an ever slower pace, with discussions being extended over weeks or even months. For these reasons, time scales may play a crucial role in satiation processes and users dropping in and out of discussions, and prove an interesting aspect to study and consider for future models of collective emotions and emotional contagion.

Comparing moderator annotations of post sentiment to emotional content measured by LIWC, we found that they do not correspond for neutral and negative sentiment. While posts annotated as positive tend contain a larger number of positive emotion words, we were not able to predict neutral and negative annotations using emotion variables provided by LIWC. The small sample size of positively annotated posts further leads us to be hesitant about drawing conclusions from these results. Additionally, the selection of posts for the annotation process was heavily guided towards a specific purpose, with a focus on posts with a negative sentiment (Schabus and Skowron, 2018), which indicates that the sample of annotated posts we analyzed may not be a representative sample of how moderators generally rate post sentiment. Our results in this regard could also be used as an example of the difference between qualitative and quantitative approaches to the study of affective meaning in text, with moderators potentially detecting subtleties in text and its context that are lost when simply counting words.

Although increasing in number in the recent years, there remain few studies on the effects of voting mechanisms in social media on individual users (Zhang et al., 2019). Despite their widespread use in online discussions, for example on Reddit (Van Mieghem, 2011), the role of upvotes and downvotes in discussion dynamics, their relationship with affective content, and their effects on the emotional well-being of users remains largely unexplored. While we did not detect significant relationships between positive or negative emotional content and upvotes or downvotes of individual posts, our analysis does show a moderate positive correlation in the polarization of votes between parent and response posts, and indicates that the degree of emotional positivity of a post impacts its level of polarization among other users. It also suggests that upvoted posts tend to generate responses that are likewise more likely to be upvoted, which does not hold true for downvotes. Interestingly, while the aggregate level of vote polarization in a discussion is correlated with the emotional tone of its posts and increasingly positive discussions tend to be more polarized than negative ones, there appears to be no relationship with total emotional charge. Because of these results, we speculate that the overall vote polarization of a discussion could be dependent on the kind of its emotional content, but not its degree. Our results may provide useful starting points for a more extensive study of voting mechanisms in online discussion forums and other social media platforms.

Scaling exponents in the range of ≈1.1-1.3 have been shown to be omnipresent in social scaling of quantities related to social currencies such as information, and used as a predictor for an increase of social pace as a population grows (Bettencourt et al., 2007). The similarity of this superlinear growth behavior across many different social phenomena was speculated to be caused by universal social dynamics underlying all of them. We used this premise to investigate the scaling of posting behavior as the number of participants in an online discussion increases. The almost analogous relationship between users and comments in our simulation and real-world data at a superlinear rate of 1.1 implies that we may replicate a crucial property of collective emotions in online discussions using our implementation of the Cyberemotions framework, namely the number of emotion expressions they induce. Our model may be extended in future studies to capture other features and emergent properties of collective emotions. Specifically, given that our model captures both arousal and valence of individual agents, it may be used to study the quality and magnitude of emotional responses (Goldenberg et al., 2017). In addition, our results fit into research of social scaling, potentially adding support to claims of universal mechanisms underlying societal growth. We also believe that the results of our model illustrate how agent-based modeling can be successfully used to capture and explore complex social and cognitive processes.

Conclusion

Building on collective emotions theory and using an interdisciplinary cognitive science approach, we studied collective emotions in online discussion forums employing agent-based modeling and text mining of online discussions. The goal of our thesis is to contribute to the scientific understanding of online collective emotions by answering our following four research questions:

1. Can we replicate the properties of collective emotions in online discussions using an agent-based model of emotion dynamics and social interaction?

2. Can we predict the emotional content and affective reaction of a post based on the data of previous posts in the discussion?

3. Do the affective properties of a news article predict the collective emotional dynamics of its discussion?

4. Can we draw conclusions about the use of either system 1 or 2 by users based on the measured degree of cognitive complexity in their posts?

1. Implementing an agent-based model of emotion dynamics and social interaction, based on the *Cyberemotions* framework, we were able to replicate the influence of collective emotions in online discussions on posting behavior, measured as expressions per agent. While agents represent forum users in our simulation, expressions represent posts of an online discussion. After calibrating the model and adjusting its parameters using big data as reference, our model output matches data collected in the forums section of an Austrian news website and should be able to predict the number of expressions in comparable online discussions for a certain number of discussion participants. This success in modeling emotion dynamics and social interaction in online discussions leads us to support recent arguments for using social media data as realistic input for agent-based simulations.

Our data consisted of comments originally posted in the discussion forum of derStandard.at and collected in the *One Million Posts* corpus. Measuring the affective meaning of these comments using the sentiment analysis tool LIWC2015, we analyzed forum posts for their emotion words to measure their sentiment, and indicators of cognitive complexity such as function words. Together with information on the number of upvotes and downvotes a post received, this allowed to investigate the dynamics of affective responding and the polarization of posts in online discussions.

2. Our results showed that it is possible to predict the emotional content and affective reaction of a post based on data of previous posts in the discussion. Both the emotional tone and the vote polarization of a response post were correlated with tone and polarization of the post it replies to. Furthermore, we detected a linear relationship between emotional content and vote polarization in individual posts. We interpret this decrease in variability of affective responding and convergence in emotional content on the collective level as an emergent property of collective emotions.

3. Contrary to our expectations, the affective properties of news articles, the main discussion points in our data set of online discussions, did not predict the emotion dynamics of their discussions. We did not find significant correlations between emotional content of articles and those of comments in their discussion thread, which also lends support to our theory that the convergence in affective responding in discussions is caused by the dynamics of collective emotions and not the discussion topic.

4. We also failed too draw conclusions about the use of 'hot' and 'cold' cognitive systems 1 and 2 by users based on the measured degree of cognitive complexity in their posts. Using the LIWC2015 variables analytical thinking, cognitive processes, and six-letter words, we were unable to show clustering of posts that would indicate an either-or use of emotional or rational cognitive systems and support dual-process theories in their most abstract form. This does not necessarily imply that the use of cognitive systems, as they are postulated by dual-process theories, cannot be detected using quantitative text analysis, only that the three variables we focused on did not provide us with sufficient information to come to a conclusion.

Comparing our simulation results to real-world data from the One Million Posts, we showed that in both cases the number of emotional expressions scale at a rate of 1.1 as the number of discussion participants increases. This fits into research on social scaling in complex societies, where similar exponents are responsible for the scaling of social quantities such as information. It may also indicate that there are universal social mechanisms at play that determine online discussion dynamics and which may be captured and explored using agent-based modeling. Our results lead us to conclude that agent-based modeling of collective emotions can not only provide insights into the cognitive mechanisms responsible for emotional contagion, but also the prevalent social dynamics that underlie such phenomena. Likewise, text mining has proven itself a useful tool for the study of emotions in cognitive science, providing an empirical foundation for computational models of the mind and a potential method for their calibration and validation.

In conclusion, we showed how digital traces of human interaction and communication are valuable resources for the study of human social behavior and cognition. Big data allows the calibration of computational models and the validation of simulation results. Its methods, such as text mining, provide means of generating insights into social processes and emotion dynamics on a group level that are lost in traditional study environments. Because of this, we predict that social media data will continue to grow in importance for social and cognitive scientists studying social cognition and emotional contagion between individuals. The variance of methods and theories involved in conceptualizing, implementing, and validating computational models of the mind also asks for the application of an interdisciplinary research framework. We therefore hope that our master's thesis not only contributes to the study and understanding of collective emotions in online discussions, but also highlights the value of an interdisciplinary cognitive science approach.

Bibliography

Abelson, R. P.
1963. Computer simulation of "hot cognition". *Computer simulation of personality*, Pp. 277–298.

Abelson, R. P. and J. D. Carroll
1965. Computer simulation of individual belief systems. *American Behavioral Scientist*, 8(9):24–30.

Alvarez, R., D. Garcia, Y. Moreno, and F. Schweitzer
2015. Sentiment cascades in the 15m movement. *EPJ Data Science*, 4(1).

Anticevic, A. and P. R. Corlett
2012. Cognition-emotion dysinteraction in schizophrenia. *Frontiers in psychology*, 3:392.

Aragón, P., V. Gómez, D. García, and A. Kaltenbrunner
2017a. Generative models of online discussion threads: state of the art and research challenges. *Journal of Internet Services and Applications*, 8(1).

Aragón, P., V. Gómez, and A. Kaltenbrunner
2017b. Detecting platform effects in online discussions. *Policy & Internet*, 9(4):420–443.

Baumeister, R. F., K. D. Vohs, C. N. DeWall, and L. Zhang
2007. How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review*, 11(2):167–203.

Berlyne, D. E.
1967. Arousal and reinforcement. In *Nebraska symposium on motivation.* University of Nebraska Press.

Berry, D. S., J. W. Pennebaker, J. S. Mueller, and W. S. Hiller
1997. Linguistic bases of social perception. *Personality and Social Psychology Bulletin*, 23(5):526–537.

Bestgen, Y.
1994. Can emotional valence in stories be determined from words? *Cognition & Emotion*, 8(1):21–36.

Bettencourt, L. M. A., J. Lobo, D. Helbing, C. Kuhnert, and G. B. West
2007. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, 104(17):7301–7306.

Brader, T. and G. E. Marcus
2013. Emotion and political psychology.

Brand, A. G.
1985. Hot cognition: Emotions and writing behavior. *Journal of advanced composition*, Pp. 5–15.

Broekens, J., T. Bosse, and S. C. Marsella
2013. Challenges in computational modeling of affective processes. *IEEE Transactions on Affective Computing*, 4(3):242–245.

Cambria, E., S. Poria, A. Gelbukh, and M. Thelwall
2017. Sentiment analysis is a big suitcase. *IEEE Intelligent Systems*, 32(6):74–80.

Carneiro, R. L.
2000. The transition from quantity to quality: A neglected causal mechanism in accounting for social evolution. *Proceedings of the National Academy of Sciences*, 97(23):12926–12931.

Chen, Y.
2015. Visual opinion analysis of threaded discussions. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. IEEE.

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

Cioffi-Revilla, C.
2014. Introduction to computational social science. *London and Heidelberg: Springer*.

Cohen, A. O., D. V. Dellarco, K. Breiner, C. Helion, A. S. Heller, A. Rahdar, G. Pedersen, J. Chein, J. P. Dyke, A. Galvan, et al.
2016. The impact of emotional states on cognitive control circuitry and function. *Journal of cognitive neuroscience*, 28(3):446–459.

Cohn, M. A., M. R. Mehl, and J. W. Pennebaker
2004. Linguistic markers of psychological change surrounding september 11, 2001. *Psychological Science*, 15(10):687–693.

Dacey, M.
2017. Anthropomorphism as cognitive bias. *Philosophy of Science*, 84(5):1152–1164.

Donkin, E., P. Dennis, A. Ustalakov, J. Warren, and A. Clare
2017. Replicating complex agent based models, a formidable task. *Environmental Modelling & Software*, 92:142–151.

Durkheim, E., G. E. G. Catlin, J. H. Mueller, and S. A. Solovay
1938. *The rules of sociological method*, volume 8. Free Press New York.

Elkind, J. I., S. K. Card, and J. Hochberg
2014. *Human performance models for computer-aided engineering*. Academic Press.

Ellsworth, P. C. and K. R. Scherer
2003. Appraisal processes in emotion. *Handbook of affective sciences*, 572:V595.

Frijda, N. H. and W. G. Parrott
2011. Basic emotions or ur-emotions? *Emotion Review*, 3(4):406–415.

Fum, D., F. Del Missier, and A. Stocco
2007. The cognitive modeling of human behavior: Why a model is (sometimes) better than 10,000 words. *Cognitive Systems Research*, 8(3):135–142.

Garas, A., D. Garcia, M. Skowron, and F. Schweitzer
2012. Emotional persistence in online chatting communities. *Scientific Reports*, 2(1).

Garcia, D., A. Kappas, D. Küster, and F. Schweitzer
2016. The dynamics of emotions in online interaction. *Royal Society Open Science*, 3(8):160059.

Gleser, G. C., L. A. Gottschalk, and W. John
1959. The relationship of sex and intelligence to choice of words: A normative study of verbal behavior. *Journal of Clinical Psychology*, 15(2):182–191.

Goldenberg, A., D. Garcia, E. Halperin, and J. Gross
2017. Collective emotions.

Granovetter, M.
1983. The strength of weak ties: A network theory revisited. *Sociological theory*, Pp. 201–233.

Gratch, J. and S. Marsella
2005. Evaluating a computational model of emotion. *Autonomous Agents and Multi-Agent Systems*, 11(1):23–43.

Grimm, V., U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S. K. Heinz, G. Huse, et al.
2006. A standard protocol for describing individual-based and agent-based models. *Ecological modelling*, 198(1-2):115–126.

Grimm, V., U. Berger, D. L. DeAngelis, J. G. Polhill, J. Giske, and S. F. Railsback
2010. The odd protocol: a review and first update. *Ecological modelling*, 221(23):2760–2768.

Grimm, V., G. Polhill, and J. Touza
2017. Documenting social simulation models: the odd protocol as a standard. In *Simulating social complexity*, Pp. 349–365. Springer.

Guo, Z. and J. C. Tay
2008. Granularity and the validation of agent-based models. In *Proceedings of the 2008 Spring simulation multiconference*, Pp. 153–161. Society for Computer Simulation International.

Guthrie, S. E.
1997. Anthropomorphism: A definition and a theory.

Hancock, J. and L. Curry
2008. S, goorha, & m. woodworth "on lying and being lied to: An automated linguistic analysis of deception in computermediated communication". *Discourse Processes*, 45:1–23.

Harkin, M. E.
2010. Uncommon ground: Holism and the future of anthropology. *Reviews in Anthropology*, 39(1):25–45.

Hebb, D. O.
1949. *The organization of behavior: a neuropsychological theory*. J. Wiley; Chapman & Hall.

Heider, F. and M. Simmel
1944. An experimental study of apparent behavior. *The American journal of psychology*, 57(2):243–259.

Holyst, J. A.
2016. *Cyberemotions: Collective emotions in cyberspace*. Springer.

Hołyst, J. A., A. Chmiel, and J. Sienkiewicz
2017. Detection and modeling of collective emotions in online data. In *Cyberemotions*, Pp. 137–158. Springer.

Hotho, A., A. Nürnberger, and G. Paaß
2005. A brief survey of text mining. In *Ldv Forum*, volume 20, Pp. 19–62. Citeseer.

Hudlicka, E.
2011. Guidelines for designing computational models of emotions. *International Journal of Synthetic Emotions (IJSE)*, 2(1):26–79.

Huebner, B.
2011. Genuinely collective emotions. *European Journal for Philosophy of Science*, 1(1):89–118.

Kahn, J. H., R. M. Tobin, A. E. Massey, and J. A. Anderson
2007. Measuring emotional expression with the linguistic inquiry and word count. *The American journal of psychology*, Pp. 263–286.

Kahneman, D.
2011. *Thinking, Fast and Slow*. Farrar Strauss & Giroux.

Kappas, A.
2002. The science of emotion as a multidisciplinary research paradigm. *Behavioural processes*, 60(2):85–98.

Kappas, A.
2017. The psychology of (cyber) emotions. In *Cyberemotions*, Pp. 37–52. Springer.

Kennedy-Moore, E. and J. C. Watson
2001. How and when does emotional expression help? *Review of General Psychology*, 5(3):187–212.

Kharde, V., P. Sonawane, et al.
2016. Sentiment analysis of twitter data: a survey of techniques. *arXiv preprint arXiv:1601.06971*.

Kuppens, P., F. Tuerlinckx, J. A. Russell, and L. F. Barrett
2013. The relation between valence and arousal in subjective experience. *Psychological bulletin*, 139(4):917.

Laird, J. E.
2008. Extending the soar cognitive architecture. *Frontiers in Artificial Intelligence and Applications*, 171:224.

Laird, J. E., A. Newell, and P. S. Rosenbloom
1986. Soar: An architecture for general intelligence. Technical report, STANFORD UNIV CA DEPT OF COMPUTER SCIENCE.

LeBlanc, V. R., M. M. McConnell, and S. D. Monteiro
2015. Predictable chaos: a review of the effects of emotions on attention, memory and decision making. *Advances in Health Sciences Education*, 20(1):265–282.

Lerner, J. S., Y. Li, P. Valdesolo, and K. S. Kassam
2015. Emotion and decision making. *Annual review of psychology*, 66:799–823.

Liu, C., E. Zhu, Q. Zhang, and X. Wei
2018. Modeling of agent cognition in extensive games via artificial neural networks. *IEEE transactions on neural networks and learning systems*, 29(10):4857–4868.

Lodge, M. and C. S. Taber
2005. The automaticity of affect for political leaders, groups, and issues: An experimental test of the hot cognition hypothesis. *Political Psychology*, 26(3):455–482.

Loewenstein, G. and J. S. Lerner
2003. The role of affect in decision making. *Handbook of affective science*, 619(642):3.

Macal, C. M. and M. J. North
2005. Tutorial on agent-based modeling and simulation. In *Proceedings of the Winter Simulation Conference, 2005.*, Pp. 14–pp. IEEE.

Marcus, G. E.
2013. The theory of affective intelligence and liberal politics. In *Emotions in Politics*, Pp. 17–38. Palgrave Macmillan UK.

Matthews, G.
1997. An introduction to the cognitive science of personality and emotion. *Advances in Psychology*, 124:3–30.

Mehl, M. R.
2006. Quantitative text analysis. In *Handbook of multimethod measurement in psychology.*, Pp. 141–156. American Psychological Association.

Meier, T., R. L. Boyd, J. W. Pennebaker, M. R. Mehl, M. Martin, M. Wolf, and A. B. Horn
2019. "LIWC auf deutsch": The development, psychometrics, and introduction of DE- LIWC2015.

Minsky, M.
1988. *Society of mind.* Simon and Schuster.

Mozer, M. C. and M. Sitton
1998. Computational modeling of spatial attention. *Attention*, 9:341–393.

Nesse, R. M. and P. C. Ellsworth
2009. Evolution, emotions, and emotional disorders. *American Psychologist*, 64(2):129.

Newell, A. and H. A. Simon
1963. Gps, a program that simulates human thought. *Computers and thought*, Pp. 279–293.

Oatley, K. and P. N. Johnson-Laird
2014. Cognitive approaches to emotions. *Trends in cognitive sciences*, 18(3):134–140.

Ong, D. C., J. Zaki, and N. D. Goodman
2019. Computational models of emotion inference in theory of mind: A review and roadmap. *Topics in cognitive science*, 11(2):338–357.

O'Reilly, R. C., S. A. Herd, and W. M. Pauli
2010. Computational models of cognitive control. *Current opinion in neurobiology*, 20(2):257–261.

Padilla, J. J., S. Y. Diallo, H. Kavak, O. Sahin, and B. Nicholson
2014. Leveraging social media data in agent-based simulations. In *Proceedings of the 2014 Annual Simulation Symposium*, Pp. 1–8.

Palminteri, S., V. Wyart, and E. Koechlin
2017. The importance of falsification in computational cognitive modeling. *Trends in cognitive sciences*, 21(6):425–433.

Parkinson, B.
2013. Comment: Journeys to the center of emotion. *Emotion Review*, 5(2):180–184.

Pasupathi, M.
2007. Telling and the remembered self: Linguistic differences in memories for previously disclosed and previously undisclosed events. *Memory*, 15(3):258–270.

Pellert, M., S. Schweighofer, and D. Garcia
2020. The individual dynamics of affective expression on social media. *EPJ Data Science*, 9(1).

Pennebaker, J., C. Chung, M. Ireland, A. Gonzales, and R. Booth
2007. The development and psychometric properties of liwc2007. austin, tx.

Pennebaker, J., M. Francis, and R. Booth
1999. Linguistic inquiry and word count (liwc).

Pennebaker, J. W., R. L. Boyd, K. N. Jordan, and K. Blackburn
2015. The development and psychometric properties of liwc2015.

Pennebaker, J. W., C. K. Chung, J. Frazee, G. M. Lavergne, and D. I. Beaver
2014. When small words foretell academic success: The case of college admissions essays. *PLoS ONE*, 9(12):e115844.

Pennebaker, J. W., M. R. Mehl, and K. G. Niederhoffer
2003. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, 54(1):547–577.

Poor, M., A. Duhachek, and S. Krishnan
2012. The moderating role of emotional differentiation on satiation. *Journal of Consumer Psychology*, 22(4):507–519.

Popping, R.
2000. *Computer-assisted text analysis*. Sage.

Posner, J., J. A. Russell, and B. S. Peterson
2005. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(03).

Prinz, J.
2012. Emotion. In *The Cambridge Handbook of Cognitive Science*, K. Frankish and W. Ramsey, eds. Cambridge University Pr.

Reisenzein, R., E. Hudlicka, M. Dastani, J. Gratch, K. Hindriks, E. Lorini, and J. Ch. Meyer
2013. Computational modeling of emotion: Toward improving the inter- and intradisciplinary exchange. *IEEE Transactions on Affective Computing*, 4(3):246–266.

Roesch, Eb, Korsten, N., Fragopanagos, Taylor, and Jg
2010. Emotions in artificial neural networks. In *A Blueprint for Affective Computing: A Sourcebook and Manual*, K. R. Scherer, T. Bänziger, and E. Roesch, eds. Oxford University Press.

Roiser, J. P. and B. J. Sahakian
2013. Hot and cold cognition in depression. *CNS Spectrums*, 18(3):139–149.

Rolls, E. T. and F. Grabenhorst
2008. The orbitofrontal cortex and beyond: from affect to decision-making. *Progress in neurobiology*, 86(3):216–244.

Russell, J.
1980. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178.

Russell, J. A.
2003. Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145.

Russell, S. and P. Norvig
2016. Artificial intelligence: A modern approach.

Salloum, S. A., M. Al-Emran, A. A. Monem, and K. Shaalan
2017. A survey of text mining in social media: facebook and twitter perspectives. *Adv. Sci. Technol. Eng. Syst. J*, 2(1):127–133.

Salmela, M.
2014. The functions of collective emotions in social groups. In *Institutions, emotions, and group agents*, Pp. 159–176. Springer.

Schabus, D. and M. Skowron
2018. Academic-industrial perspective on the development and deployment of a moderation system for a newspaper website. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).*

Schabus, D., M. Skowron, and M. Trapp
2017. One million posts. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR 17*. ACM Press.

Scherer, K. R.
2005. What are emotions? and how can they be measured? *Social science information*, 44(4):695–729.

Schnurr, P. P., S. D. Rosenberg, T. E. Oxman, and G. J. Tucker
1986. A methodological note on content analysis: Estimates of reliability. *Journal of personality assessment*, 50(4):601–609.

Schwarz, N.
2000. Emotion, cognition, and decision making. *Cognition & Emotion*, 14(4):433–440.

Schweighofer, S.
2018. *Affective, Cognitive, and Social Identity Related Factors of Political Polarization.* PhD thesis, ETH Zurich.

Schweitzer, F.
2007. *Brownian Agents and Active Particles.* Springer Berlin Heidelberg.

Schweitzer, F. and D. Garcia
2010. An agent-based model of collective emotions in online communities. *The European Physical Journal B*, 77(4):533–545.

Semin, G. R. and K. Fiedler
1988. The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of personality and Social Psychology*, 54(4):558.

Sexton, J. B. and R. L. Helmreich
2000. Analyzing cockpit communications: the links between language, performance, error, and workload. *Human Performance in Extreme Environments*, 5(1):63–68.

Shapiro, G. and J. Markoff
1997. A matter of definition. *Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts*, 1:9–34.

Siegle, G. J., S. R. Steinhauer, and M. E. Thase
2004. Pupillary assessment and computational modeling of the stroop task in depression. *International Journal of Psychophysiology*, 52(1):63–76.

Simmons, R. A., D. L. Chambless, and P. C. Gordon
2008. How do hostile and emotionally overinvolved relatives view relationships?: What relatives' pronoun use tells us. *Family Process*, 47(3):405–419.

Smith, E. R. and F. R. Conrey
2007. Agent-based modeling: A new approach for theory building in social psychology. *Personality and Social Psychology Review*, 11(1):87–104.

Smith, E. R. and D. M. Mackie
2015. Dynamics of group-based emotions: Insights from intergroup emotions theory. *Emotion Review*, 7(4):349–354.

Stanovich, K. E. and R. F. West
2000. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23(5):645–665.

Sun, R.
2008a. *The Cambridge handbook of computational psychology*. Cambridge University Press.

Sun, R.
2008b. Introduction to computational cognitive modeling. *Cambridge handbook of computational psychology*, Pp. 3–19.

Tausczik, Y. R. and J. W. Pennebaker
2009. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54.

Tetlock, P. E.
1980. Explaining teacher explanations of pupil performance: A self-presentation interpretation. *Social Psychology Quarterly*, Pp. 283–290.

Thagard, P.
2017. Cognitive science. In *The Oxford handbook of interdisciplinarity*, R. Frodeman, J. T. Klein, and R. C. D. S. Pacheco, eds.

Thagard, P. and B. Aubie
2008. Emotional consciousness: A neural model of how cognitive appraisal and somatic perception interact to produce qualitative experience. *Consciousness and cognition*, 17(3):811–834.

Tisue, S. and U. Wilensky
2004. Netlogo: A simple environment for modeling complexity. In *International conference on complex systems*, volume 21, Pp. 16–21. Boston, MA.

Tolk, A.
2015. The next generation of modeling & simulation: integrating big data and deep learning. In *Proceedings of the conference on summer computer simulation*, Pp. 1–8.

Turing, A.
1950. Computing machinery and intelligence. *Mind*, 59(236):433.

Van Mieghem, P.
2011. Human psychology of common appraisal: The reddit score. *IEEE Transactions on Multimedia*, 13(6):1404–1406.

Van Rossum, G. and F. L. Drake Jr
1995. *Python tutorial.* Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands.

Varma, S. and M. A. Just
2006. 4caps: An adaptive architecture for human information processing. In *AAAI Spring Symposium: Between a Rock and a Hard Place: Cognitive Science Principles Meet AI-Hard Problems*, Pp. 91–96.

Ventirozos, F. K., I. Varlamis, and G. Tsatsaronis
2017. Detecting aggressive behavior in discussion threads using text mining. In *International Conference on Computational Linguistics and Intelligent Text Processing*, Pp. 420–431. Springer.

Vlasceanu, M., K. Enz, and A. Coman
2018. Cognition in a social context: A social-interactionist approach to emergent phenomena. *Current Directions in Psychological Science*, 27(5):369–377.

von Scheve, C. and M. Salmela, eds.
2014. *Collective Emotions: Perspectives from Psychology, Philosophy, and Sociology.* Oxford University Press.

Watson, D., L. A. Clark, and A. Tellegen
1988. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology*, 54(6):1063.

Wiedemann, G.
2016. *Text mining for qualitative data analysis in the social sciences*, volume 1. Springer.

Wilensky, U. and W. Rand
2015. *An Introduction to Agent-Based Modeling.* MIT Press Ltd.

Windrum, P., G. Fagiolo, and A. Moneta
2007. Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation*, 10(2):8.

Wirtz, J., A. S. Mattila, and R. L. Tan
2000. The moderating role of target-arousal on the impact of affect on satisfaction—an examination in the context of service experiences. *Journal of retailing*, 76(3):347–365.

Yik, M., J. A. Russell, and J. H. Steiger
2011. A 12-point circumplex structure of core affect. *Emotion*, 11(4):705.

Zahn-Waxler, C., R. de Sousa, J. Campos, J. Averill, M. Nussbaum, J. Russell, J. Lerner, J. D'Arms, P. Ellsworth, K. Scherer, et al.
2017. How affective intelligence theory can help us understand politics.

Zhang, J., H. Labiod, and W. Meng
2019. A fistful of votes: Understanding the voting system in social question and answer websites. In *2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, Pp. 1–5. IEEE.

Abstract

We study collective emotions using agent-based modeling and sentiment analysis of online discussions. Simulation results using an implementation of the *Cyberemotions* framework show a scaling exponent in the relationship between the numbers of agents and emotional expressions that is almost analogous to that of real-world data. It corresponds to a range of values that is omnipresent in the social scaling of information, suggesting our implemented model may capture underlying social dynamics realistically. Applying the sentiment analysis tool LIWC on the *One Million Posts* corpus, a dataset containing over a million comments posted in the online forum of an Austrian news website, our results indicate that it may be possible to predict the emotional content and affective reaction of a post, measured as percentage of emotion words and upvotes, based on data of previous posts in the discussion. However, we failed to capture patterns of cognitive systems 1 and 2, as they are described by dual-process theories of the mind, using LIWC variables as indicators of cognitive complexity. Our results provide further insights into the relationship between the emotional content of a news article and its comments, and the ability of moderator annotations to predict sentiment.

Abstract in German

In dieser Studie untersuchen wir kollektive Emotionen in Online Diskussionen durch Anwendung von agentenbasierender Computermodellierung und Sentiment-Analyse. Simulationen unseres *Cyberemotions*-Modells zeigen einen Skalierungsexponenten in der Beziehung zwischen der Anzahl der Agenten und der generierten emotionalen Ausdrücke, der fast analog zu realen Werten steht. Der Wert fällt in einen Bereich, der in der sozialen Skalierung von Information omnipräsent ist, was suggeriert, dass unser Modell unterliegende soziale Dynamiken realistisch darstellen kann. Die Resultate unserer LIWC Sentiment-Analyse des *One-Million-Posts*-Korpus, ein Datensatz von über einer Million Kommentaren des Diskussionsforums einer österreichischen Nachrichten-Website, suggerieren die Möglichkeit den emotionalen Inhalt eines Kommentars und affektive Reaktionen darauf durch frühere Kommentare vorherzusagen. Während Ersteres durch die Prozentzahl emotionaler Wörter der Nachricht gemessen wird, betrifft Letzteres dessen Bewertung durch Upvotes und Downvotes Anderer. Allerdings scheint es durch Messen von LIWC-Variablen nicht möglich, in Kommentaren kognitive Muster zu entdecken die System 1 oder 2 entsprechen, wie es von kognitiven Dual-Prozess-Theorien postuliert wird. Unsere Ergebnisse geben weitere Einblicke in die Beziehung zwischen emotionalem Inhalt von Nachrichten-Artikeln und dessen Kommentaren und die Erfassung von Emotionen durch Forum-Moderatoren.