Dipl.-Ing. Ilire Hasani-Mavriqi, BSc

Consensus Dynamics in Online Collaboration Networks

DOCTORAL THESIS

to achieve the university degree of
Doktor der technischen Wissenschaften
submitted to

Graz University of Technology

Supervisor
Assoc.-Prof. Dipl-Ing. Dr.techn. Denis Helic

Additional Advisor
Ass.-Prof. Dipl-Ing. Dr.techn. Elisabeth Lex

Institute of Interactive Systems and Data Science
Faculty of Computer Science and Biomedical Engineering
Graz University of Technology
Graz, Austria

Graz, December 2018

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present doctoral thesis.

Date

Signature

Abstract

In many real-world settings, it is essential for a group of interacting individuals to reach shared understanding and consensus on a given issue. Consensus can strengthen groups and their impact on society. The popularity of online social and collaboration networks has influenced the way individuals interact with each other. Many collaboration sites (i.e., Stack-Exchange, Reddit or Wikipedia) enable users to exchange opinions, discuss certain topics and solve problems while interacting with other online users. This thesis aims at uncovering the dynamics of consensus building among users collaborating online. Consensus dynamics is closely related to the process of opinion dynamics. Opinion dynamics has been studied from the perspective of social sciences, physics, mathematics, complex system studies and network science. However, such studies often remain confined to these disciplines. Therefore, this thesis applies an interdisciplinary approach. It builds hypotheses based on social science theories, simulates opinion dynamics utilizing agent-based models from statistical physics and applies social network analysis on empirical datasets extracted from the Web. Methodologically, this thesis contributes a novel framework to study the role and interplay of some of the main factors in consensus building (i.e., users social status, network structure, users similarity and content creation). The presented method can be applied to run extensive simulations of opinion dynamics in arbitrary collaboration networks. The empirical findings of this thesis help draw recommendations on how to integrate the influence of user characteristics (e.g., social status) in opinion dynamics to optimize consensus building. Additionally, this thesis experimentally demonstrates how content dynamics drives the process of agreement and disagreement between users collaborating online. These results add to our understanding of the challenges of designing and implementing services that promote consensus building.

Kurzfassung

Für soziale Gruppen ist Konsens wesentlich zur Durchsetzung gemeinsamer (d.h., Gruppen-) Interessen. Nur dadurch gelingt gemeinsames Handeln, was den Zusammenhalt innerhalb der Gruppe stärkt. Die Verbreitung sozialer und kollaborativer Netzwerke im Web hat die Art und Weise beeinflusst, wie Individuen miteinander interagieren. Auf vielen kollaborativen Netzwerken im Web (d.h., StackExchange, Reddit oder Wikipedia) können Nutzer Meinungen austauschen, Themen diskutieren und Probleme lösen, indem sie mit anderen Nutzern dieser Netzwerke interagieren. Die vorliegende Arbeit zielt darauf ab, die Dynamiken der Konsensbildung zwischen Nutzern, die online kollaborieren, zu erforschen. Die Konsensdynamik hängt eng mit den Prozessen der Meinungsdynamik zusammen. Die Meinungsdynamik wurde bisher aus der Perspektive von Sozialwissenschaften, Physik, Mathematik, Komplexitätsforschung und Netzwerkwissenschaften untersucht. Diese Arbeiten blieben bisher jedoch häufig auf diese Disziplinen beschränkt. Aus diesem Grund verfolgt die vorliegende Dissertation einen interdisziplinären Ansatz. Zunächst wurden, basierend auf sozialwissenschaftlichen Theorien, Hypothesen entwickelt, um dann Meinungsdynamiken unter Verwendung agentenbasierter Modelle aus der statistischen Physik zu simulieren. Schließlich wurden Ansätze der Netzwerkanalyse auf aus dem Web extrahierte empirische Datensätze angewendet. In methodologischer Hinsicht entwickelt diese Dissertation ein neuartiges Framework, das es erlaubt, die Rolle und das Zusammenspiel einiger der wichtigsten Faktoren bei der Konsensbildung zu untersuchen. Mit der vorgestellten Methode können umfangreiche Simulationen der Meinungsdynamik in beliebigen Kollaborationsnetzwerken durchgeführt werden. Aufbauend auf den empirischen Ergebnisse dieser Arbeit können Empfehlungen erstellt werden, wie der Einfluss von Benutzereigenschaften in die Meinungsdynamik integriert werden kann, um die Konsensbildung zu optimieren. Zusätzlich weist diese Arbeit experimentell nach, wie die Inhaltsdynamik die Bildung von Konsens und Dissens zwischen Benutzern, die online kollaborieren, vorantreibt. Die daraus gewonnenen Ergebnisse führen zu praxisrelevanten Erkenntnissen für die Gestaltung und Implementierung von Tools zur Förderung der Konsensbildung.

Acknowledgments

This dissertation would have not been possible without the help and support of numerous people. First and foremost, I would like to express my special appreciation and thanks to my supervisor Denis Helic. I am grateful for his continuous support, patience and immense knowledge. His guidance played a crucial role in my research and writing of this thesis. In the same way, I am indebted to Elisabeth Lex, my additional adviser, for many fruitful discussions, her tremendous support and all great opportunities to work in interesting research projects. My gratitude extends to Tobias Ley, my external examiner, for his time, interest and helpful comments. Further, I would like to thank Stefanie Lindstaedt, head of Institute for Interactive Systems and Data Science and CEO & Scientific Director of Know-Center, for providing an inspiring working environment and for giving me the opportunity to write this thesis.

I would like to acknowledge my co-authors Florian Geigl, Dominik Kowald, Darko Stanisavljevic, Markus Strohmaier and Subhash Pujari for great collaboration and numerous discussions that resulted in successful publications. I would also like to thank my colleagues from the Social Computing team Sebastian Dennerlein, Tomislav Duricic, Leon Fadiljevic, Emanuel Lacic, Tony Ross-Hellauer, Stefan Reichmann, Dieter Theiler and Matthias Traub for their discussions and valuable feedback on my thesis. I would especially like to thank Lewis Marsh, who completed an internship at our institute, for his fruitful discussions and his enthusiasm for my research.

I owe my deepest gratitude to my parents, for their unconditional love and unwavering belief in me. Thank you for raising me well and for teaching me great values. A special thanks goes to my father for always challenging me with mathematical problems and for inspiring me to pursue a computer science career. I would also like to thank my parents-in-law for their encouragement and for taking care of our kids. Finally, more than anyone else, I would like to thank my husband Arian and my daughters Tiana and Hana for their love, patience and motivation in the last years. It has

not been easy, but you always stood by my side. Without your support, I would have not made it to this point. *Ju dua shumë!*

Institutional Acknowledgments. This thesis was partially funded by the EU funded projects Learning Layers (Grant Agreement 318209) and AFEL (Grant Agreement 687916). I would like to express my gratitude to Graz University of Technology, Institute for Interactive Systems and Data Science and Know-Center for giving me the opportunity to write this thesis. The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Ministry for Transport, Innovation and Technology, the Austrian Ministry for Digital and Economic Affairs and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency (FFG).

"Knowledge breeds knowledge, skill breeds skill, expertise breeds expertise. And each of these leads to success, which builds on itself."

Albert-László Barabási, "The Formula"

Contents

1 Introduction

1.1 Motivation

Consensus building among individuals is closely related to the process of opinion formation and dynamics of opinion exchange. Typically, individuals form their opinions in a complex social environment under the influence of exogenous factors (e.g., individual and group characteristics) [Friedkin and Johnsen, 1990]. This is also evident in our everyday life experiences. In the course of interactions with other individuals, trying, for example, to solve certain problems, our opinions and attitudes are shaped and modified by opinions or actions of others. In sociology and social psychology this phenomenon is known as social influence or interpersonal influence. Socialization, identity and decision making processes, taking place in a group of interacting individuals, are guided by the impact of interpersonal influence. As a result, a group of interacting individuals produces shared understandings and agreements that define the culture of the group and in turn the social control [Friedkin and Johnsen, 1999]. Even though not always plausible and maybe not even desirable, in many settings, it is essential for a group to reach shared understanding and consensus on an issue. Consensus makes a position of a group stronger and strengthens its impact on society [Castellano et al., 2009].

With the emergence of social media we are exposed to an intense social influence. Our interactions within our social environment are moving from offline to online settings. Online social network sites facilitate faster information flow and higher interconnections between individuals. In general, individuals tend to spread their opinions in online social networks by interacting with other online users. How opinions spread in online communities and which factors influence the dynamics of such process are pressing questions that our research community has already recognized.

Opinion dynamics is a typical representative of social dynamics that is elaborated with models from statistical physics [Castellano et al., 2009; Li et al., 2017]. The increasing number of research papers on this topic reveals that the statistical physics approach to opinion dynamics is currently attracting a great deal of interest [Castellano et al., 2009; Xia et al., 2011]. Statistical physics and agent-based models represent means to analytically capture complex phenomena with mathematical models. Such mathematical models define particular states of a population and rules for transitions between states. For example, in a typical opinion dynamics model, agents represent users that interact based on some predefined rules, opinions of which are represented with mathematical variables or dimensions. It is considered that consensus is reached among agents, if they unanimously adopt one single opinion.

It is questionable, if models from statistical physics can shed light on the process of opinion formation, having in mind that such process leads to the dynamics of consensus or disagreement between individuals. Indeed, the dynamics of consensus or disagreement between individuals is complex, because individuals are complex [Castellano et al., 2009]. Researchers agree that, despite the simplifications, agent-based models constitute important basics for the state-of-the-art research on opinion dynamics [Castellano et al., 2009; Fernández-Gracia et al., 2014; Xia et al., 2011].

The main contribution of today's streamline of research is to investigate opinion dynamics in pre-designed synthetic networks (i.e., the structure is known a priori) by neglecting structural changes of networks [Xia et al., 2011]. Pre-designed network structures do not reflect topologies arising from real user interactions. The structure and dynamics of real collaboration networks can be described by powerful means of network science [David and Jon, 2010; Newman, 2003, 2010; Palla et al., 2007]. Whereas studies on opinion dynamics in networks reveal valuable insights on *"dynamics on the networks"*, studies from the field of network science shed light on *"dynamics of the networks"* (i.e., evolution of networks over time) [Xia et al., 2011].

For many opinion dynamics models empirical validation is still outstanding [Baronchelli, 2018]. Also, there is very limited research on applying such models to analyze and categorize user behavior in online social and collaboration networks. Currently, it is not known if agent-based models are suitable to model consensus building in Web communities. There is also lack of evidence if social influence theories hold for groups interacting online. These are important questions to address, especially, since the Web has a strong impact on the processes such as opinion formation, political and cultural trends, globalization patterns, consumers behavior or marketing strategies [Castellano et al., 2009].

This thesis aims at tackling these open questions to uncover the dynamics of opinion spreading and consensus building in online collaboration networks, as a special case of online social networks. In tune with this aim, this thesis strives to understand the factors that govern consensus building processes, as well as mechanisms that may turn such processes into a success or into a failure.

In this regard, in Section 1.2 I discuss opinion and content dynamics in online collaboration networks. Section 1.3 gives an overview of the problem statement, objectives and general approach of this thesis. The following Section 1.4 states the research questions of this thesis. In Section 1.5 I list the main publications of this cumulative thesis and in Section 1.6 further publications that I contributed to during my PhD studies. Section 1.7 summarizes the main contributions and implications of this thesis and Section 1.8 gives an overview of the structure of this thesis.

1.2 Opinion and Content Dynamics in Online Collaboration Networks

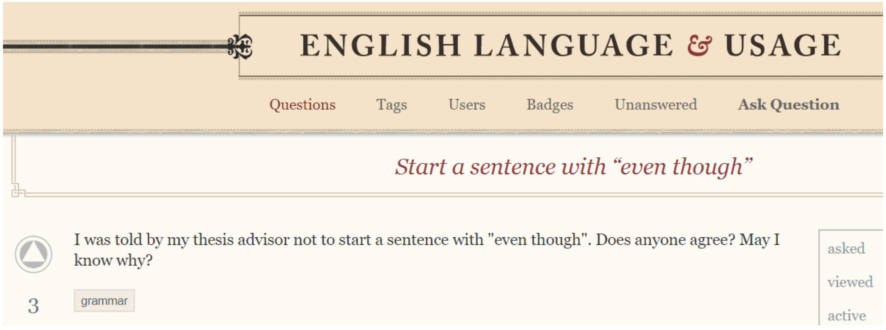
Online collaboration networks are considered a special case of online social networks. Typical examples include: question & answering (Q&A) sites, like StackExchange[1], discussion forums, like Reddit[2], online encyclopediae, like Wikipedia[3], or co-authorship networks[4]. Usually, users turn to online collaboration networks to seek for online help, share their knowledge and

[1] http://stackexchange.com/
[2] http://reddit.com/
[3] http://wikipedia.org/
[4] https://aminer.org

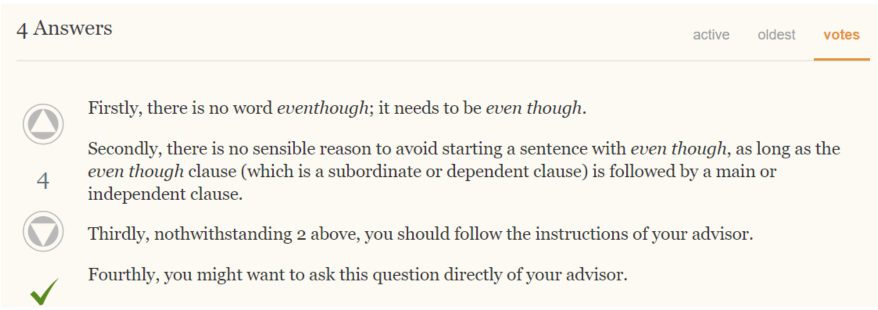
opinions, discuss problems and solutions, vote on each other's contributions or write and edit joint articles. Mostly, they engage in interactions with other users with a positive intent, for example, to provide solutions to problems by giving answers to questions in a Q&A site. A desired outcome would be that users involved in such iterative process come up with a shared common result. One of the questions addressed in this thesis is if users collaborating online manage to produce shared understanding and reach consensus on certain issues and if this process differs from offline settings? Realistically, users collaborating online do not always share common opinions, thus, consensus is not always reached and important topics or questions remain unresolved within online communities.

An illustrative example of typical user interactions in Q&A sites is presented in Figure 1.1. This example shows how users post questions and provide answers to problems related to English language and usage in the English edition of StackExchange. An example question is presented in Figure 1.1a and a sufficient answer to the question is showed in Figure 1.1b. Among other provided answers this is accepted as a best solution to the problem (cf. Figure 1.1c), however, a user expresses doubts on accepting only one answer as a valid solution (cf. Figure 1.1d). This implies that users do not always agree on the accepted answers as best solutions to certain problems.

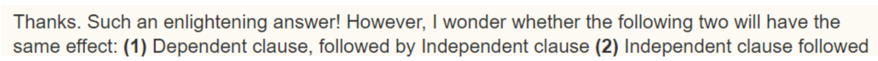
Such user interactions lead to relationships that can be studied in terms of a collaboration network. The StackExchange platform does not indicate relationships between users or friendship links. Nevertheless, co-posting activities of users can be analyzed to build relationships between them [Adamic and Adar, 2001; Halavais et al., 2014; Tang et al., 2012]. In Q&A sites, a co-posting activity between two users refers to a scenario under which two users contributed on the same post (i.e., by providing answers or comments to questions). Thus, if two users contributed in any way to a same post, they are connected via an edge in the collaboration network. Figure 1.2 shows a state of an example collaboration network after numerous interactions and exchange of opinions between users. After an iterative process (i.e., trying to reach consensus on a best solution to a given problem), blue nodes with a circle shape reached consensus on a



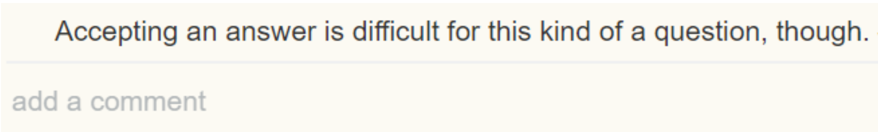
(a) Typical question



(b) Extensive Answer



(c) Accepting an answer



(d) Commenting answers

Figure 1.1: **Illustrative example of user interactions in StackExchange.** In the English edition of StackExchange users post questions and provide answers on different topics regarding the English language and usage. Figure 1.1a presents a typical question posted in Q&A sites like English StackExchange. In 1.1b a user provides an extensive answer to the posted question. This answer seems to be sufficient, so the user who posted the question thanks the user who provided the answer and continues the discussion by asking for a confirmation (cf. 1.1c). In respect to other provided answers and to the nature of the posted question, there are doubts as expressed in 1.1d, for example. Sometimes users do not agree on the accepted answer as a best solution to a given problem.

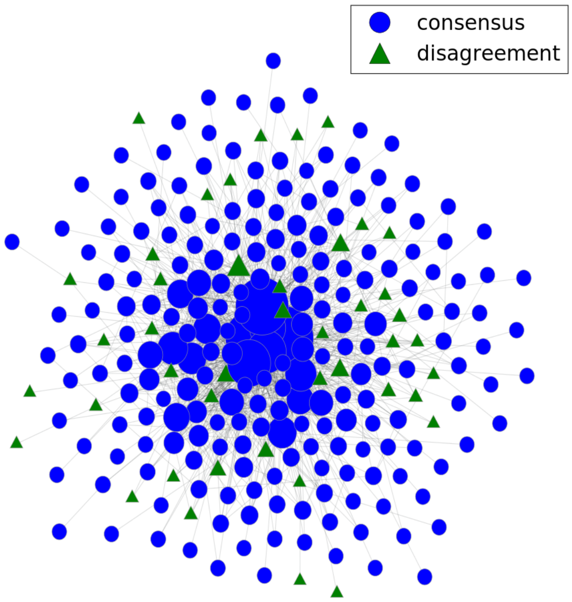


Figure 1.2: **Consensus building in a network.** Blue (circle) nodes reached consensus (have a single common opinion), whereas green (triangle) nodes did not adopt the common opinion (in a typical case they kept two or more opinions).

single common opinion, whereas green nodes with a triangle shape kept two ore more opinions (i.e., did not adopt the common opinion).

By design, online collaboration networks enable equal access to all their users. However, so called 'influencers' emerge very fast (i.e., due to heterogeneous user engagement) as opinion leaders in particular online communities. Such influencers induce (very often biased) opinions in the group of users interacting online and may act as a proxy to other resources [Garcia et al., 2017]. Online collaboration networks are relatively new (e.g., Wikipedia launched January 2001, StackExchange launched September 2009), so no specific theories were developed in social sciences to explain social influence in online settings. But, earlier established social influence theories [Latané, 1981] can be applied to explain the role of key players that influence opinion diffusion and consensus building among users collaborating online.

By considering individual user features (i.e., user social status or position in a network) and features of pairs of users (i.e., similarities between pairs of users), it is possible to uncover factors and mechanisms that drive opinion diffusion and consensus building in online communities. Incorporating these realistic features to agent-based models enables us to understand how microscopic behavior has macroscopic consequences in online collaboration networks [Baronchelli, 2018].

Online collaboration networks evolve with time, new users come, other users leave the sites. The content of such sites also evolves with time. To illustrate the point, let us consider Wikipedia as a wiki-based encyclopedia that enables its users (i.e., editors) to collaboratively create and edit content. The large-scale collaboration of Wikipedia volunteer-editors gives rise to knowledge construction and to communities with shared identity and practice [Iniguez et al., 2014]. Worth mentioning is the fact that there is no formal hierarchical organizational structure (i.e., central coordination) in Wikipedia. Even in the absence of a central coordination, the editing process in Wikipedia is usually characterized as peaceful and constructive [Liu and Ram, 2011]. But if, for example, editors have different point of views on some controversial topics, they might end up repeatedly overriding each other's contributions, making it harder to reach consensus on the content of articles [Iniguez et al., 2014; Tsvetkova et al., 2016]. Thus, content dynamics in Wikipedia provides great potential to study agreement and disagreement processes between users collaborating online.

1.3 Problem Statement, Objectives and General Approach

Problem Statement. With the emergence of social media our interactions within our social environment are moving from offline to online settings. We tend to form and spread our opinions in the course of interaction with other users in online communities. Online collaboration networks present an example, where users tend to seek for online help, express their opinions or experiences, provide solutions to given problems, rate each other's contributions or collaboratively write a joint article. Users engage in an iterative process trying to reach common understanding and agreement, for example, while trying to find the best solution to a given problem. However, that is not always the case due to differences in user opinions, experiences, backgrounds or point of views. So, users collaborating online do not always share common understanding, thus, consensus is not always reached and important topics or questions remain unresolved within online communities. Which are the factors that govern opinion dynamics and agreement or disagreement processes in online collaboration networks is still an open question for the research community.

Opinion dynamics is a complex endeavor to study because individuals are complex. As such, it is considered as an interdisciplinary field that requires an attention of social sciences, physics, mathematics and complex system studies, as well as network science [Xia et al., 2011]. It is difficult for our research community to conduct studies that cover each of the disciplines and aspects of opinion dynamics.

In terms of social sciences, well established theories from offline settings could be applied to investigate opinion dynamics and social influence in online settings. In regard to statistical physics and mathematical models, numerous opinion dynamics models have been already established. Overall, they are divided into discrete and continuous models. They consider (i) discrete opinions and proportional transition rules for opinion states (i.e., in the voter model [Clifford and Sudbury, 1973; Holley and Liggett, 1975] a randomly selected agent, possessing a binary state variable, takes the state of a randomly selected neighbor) or (ii) continuous opinions and transition rules based on some threshold (i.e., in the Deffuant model [Deffuant et al., 2000] agents will switch opinion states, only if two agent's opinions are within some predefined range). In general, such opinion dynamics models tend to answer questions whether agents unanimously adopt one single opinion (i.e., reach consensus) or more opinions coexist among agents and in which scaling time consensus is reached or not reached [Fernández-Gracia et al., 2014]. Such models have been rarely applied to investigate opinion dynamics in online collaboration networks. Also, they have been mainly used to simulate opinion diffusion in pre-designed synthetic networks, in which the structure of the relations between users is known a priori.

Objectives. To that end, this thesis aims to understand the general principles that govern opinion dynamics and consensus building in online collaboration networks. This step is a very important one for future research and more ambitious goals, such as predicting user behavior and designing and implementing services that best support users in online systems. Additionally, this thesis strives to fill the aforementioned research gap by contributing to the limited body of research on theoretical models from statistical physics and empirical data extracted from the Web. By applying agent-based models, such as the Naming Game model, on topologies from real collaboration networks (i.e., StackExchange, Reddit, co-authorship networks and Wikipedia), it is possible to understand the factors that drive consensus building processes in online collaboration networks, as well as mechanisms that may turn such processes into a success or into a failure. A very relevant goal of this thesis is to provide feedback (i.e., recommendations) to facilitate online system designers in developing tools to promote consensus building processes. To illustrate the point, connecting otherwise non-interacting users by recommendations may lead to discussions resolving issues that hinder consensus.

General Approach. This thesis takes an interdisciplinary approach by building hypotheses based on theories from social sciences, using and extending models from statistical physics and applying social network analysis on empirical datasets extracted from the Web. Early endeavors in social sciences have established social influence theories that are applied here to build hypotheses and to prove their plausibility in online collaboration networks. By applying network science means to empirical datasets this thesis extracts intrinsic properties and realistic network features. These realistic features are than incorporated to opinion dynamics models. This thesis takes a computational approach and analyzes consensus dynamics by simulating the diffusion of opinions in such models. Further, it studies the role of the following factors and their correlation in consensus dynamics, namely, users social status, underlying network structure and users similarity. Additionally, this thesis analyzes the dynamics of content creation in collaboration networks and investigates the development of consensus in such settings.

1.4 Research Questions

This thesis studies the process of opinion dynamics and consensus building in online collaboration networks. To that end, it investigates the influence of the following factors and their correlations on the speed towards consensus building. First, this thesis examines the influence of users social status in correlation with the underlying network structure on consensus building (RQ1). Second, it studies the role of the interplay between users similarities and users social status in consensus building (RQ2). Third, this thesis analyzes the development of consensus in collaborative content creation (RQ3). The following introduces the research questions answered in this thesis.

RQ1: What is the influence of social status on consensus in collaboration networks?

Problem. Each of us can recall situations, in which our opinions are influenced by the opinions of a person we trust, possessing a higher social status or reputation. Since our communications and interactions with our peers are taking place online, also the influence of social status on our opinion dynamics is moving from offline to online settings. So, the main focus of this research question is to investigate the relation between social status, underlying networks structure and dynamical processes that take place in online collaboration networks. In other words, this research question strives to find out (i) what are the dynamics of reaching a consensus in online communities with heterogeneous distribution of social status and (ii) how the underlying network structure affects this process.

The foundation for explaining such phenomena is given by theories from social sciences (i.e., social status, social impact and network exchange theories) [Friedkin and Johnsen, 1999; Latané, 1981; Markovsky et al., 1993; Walker et al., 2000; Willer, 1999]. Based on such theories, it is our natural predisposition to interact with people having higher social status in our community. Our behavior and our opinions, to some extent, are influenced from persons having a higher social status because they also have a higher convincing power. What is meant by social status in terms of this thesis? Without delving deeper into definitions from social status theory [Schooler, 2013; Tischler, 2010; Weber, 1964], it is considered that social status is related to a position individuals hold in a society or group, which is earned over their lifetime. It is also related to the reputation they gain, trust they earn, convincing power they possess, which in turn are a reflection of a professional expertise, education or a position in an hierarchy of an organization.

In online settings, some collaboration networks have established incentive systems to reward users for their contributions. Typical awards are badges or virtual points. While badges typically have no explicit value, they act as symbols of social status within an online community [Anderson et al., 2013; Immorlica et al., 2015].

Approach. To address this research question, this thesis takes a computational approach and analyzes opinion dynamics by simulating the diffusion of opinions in empirical collaboration networks with heterogeneous distribution of social status (i.e., datasets from StackExchange and co-authorship networks). To do so, in [Hasani-Mavriqi et al., 2016] we simulate scenarios of peer interactions in empirical datasets assuming that the status theory holds and observe the consequences. We use reputation scores (i.e., built-in incentive system of StackExchange) as a proxy for social status of users. We utilize agent-based models from statistical physics, namely, the Naming Game model [Baronchelli et al., 2006] and extend it by incorporating a mechanism to configure the degree of the influence of social status on the network dynamics. This mechanism is termed the *Probabilistic Meeting Rule*. With this extension we can model different scenarios of how social status may influence the opinion dynamics. For example, we can: (i) completely neglect the status by allowing any two individuals to exchange their opinions regardless of their social status (an *egalitarian* society) [Arneson, 2013], (ii) allow opinions flowing only in one direction – from individuals with a higher social status to those with a lower social status (a *stratified* society) [Weber, 1964], (iii) probabilistically model any situation in between these two extreme cases, that is, a case in which opinions are very likely to flow from individuals with a higher social status to those with a lower social status but with small probability they can also flow into the other direction (a *ranked* society) [Weber, 1964].

In addition, to investigate the role of network structure in consensus building, this thesis analyzes the impact of degree assortativity and the correlation between degree and social status. A degree assortativity is a graph metric. It represents the extent, to which nodes in a network associate with other nodes, being of similar sort or being of opposing sort, indicating that it can be positive or negative [Noldus and Van Mieghem, 2015]. It can be applied to other node characteristics other than the node degree (i.e., reputation or social status scores). In assortative networks, on average, nodes with similar degrees are connected together, which means physical connections between low and high are rare. Typically, co-authorship networks (also known as scientific collaboration networks) are assortative networks, which means that authors having similar social status in their community tend to publish an article together [Tang et al., 2008]. Disassortative networks have a negative degree assortativity coefficient. In disassortative networks, high-degree nodes are on average connected to nodes with low(er) degree. StackExchange networks are dissassortative. In other words, users with lower reputation scores are more likely to connect to users with higher reputation scores.

Based on the empirical networks, we generate specific synthetic networks (i.e., by constructing disassortative and assortative synthetic networks). In each of the constructed networks only one particular property of interest is preserved while others are eliminated. This way, in each experiment, the influence of a single property on the overall opinion dynamics process is assessed. Different configurations are created to analyze degree and status correlation of networks, for example, by randomly rewiring nodes, keeping the degree sequence, but randomly rewiring edges or shuffling social statuses (cf. Section 3.2.6 for an overview).

Findings and Contributions. Our initial hypothesis, based also on theories from social sciences, was that the influence of users social status will speed up consensus building among users collaborating online. However, our results only partially confirm our hypothesis, implicating that the social status of users influences consensus reaching in an intricate and complex way. Low values of the influence of social status tend to favor consensus reaching, whereas high values hinder this process. The findings reveal that the speed of reaching consensus among users collaborating online depends on the direction of the communication between low and high social status users. The best opinion convergence rates are achieved, when low social status users can communicate freely with each other and there is a barrier that prohibits them to inflict their opinions in high social status users. In contrary, the communication from high social status users to low social status users must not be disturbed. In this way, high social status users can pick up a user among low social status users, convey the message (opinion) and let the user spread it among other peers. Similarly, low social status users may delegate their opinions through one member. These results suggest that optimizing the process of consensus reaching is a tuning act of how to integrate social status in the opinion dynamics.

Investigations on the role of underlying network structure reveal which configurations benefit or hinder the process of consensus building. In general, hubs are crucial for the process of consensus building because they can distribute a single common opinion to a high number of other nodes. Further, our findings show that external interventions (i.e., Probabilistic Meeting Rule) have no effect if degree and status are not correlated or if a positive degree assortativity is evident in the network (i.e., co-authorship network). But, in disassortative networks, where degree strongly correlates with status (i.e., StackExchange empirical networks), this correlation delays consensus building, thus, it is necessary to apply external interventions to compensate this delay (i.e., insert a social barrier between low and high social status users). These findings can be applied to make recommendations on actions to be taken in specific settings and configurations to facilitate consensus building (cf. Table 3.2).

This research question makes a methodological and empirical contribution to the field of opinion dynamics. The presented model is flexible and can be easily extended to study the communication between users in different user groups. Overall, the methodology shows how agent-based models can provide more informative and illustrative analysis as compared to traditional regression approaches or black-box predictions. Empirically, this research question makes a necessary contribution to the limited body of research on Naming Game and empirical data. Further contribution is the simulation and evaluation framework provided as an open source project[5].

RQ2: How does consensus depend on user similarity and social status?

Problem. The previous research question considers user interactions that are visible or leave traces in online collaboration networks (i.e., two users providing answers to a question in StackExchange). In this line, in [Hasani-Mavriqi et al., 2016] we extracted such interactions and constructed networks, on which we simulated different scenarios of peer interactions. But in real-world online collaboration networks, there are user interactions that do not leave traces in the system logs, for example, many users turning to StackExchange or Reddit only read posts and do not leave any comments. This research question aims to capture such hidden interactions by adapting a model of interacting users, whose future interactions are not restricted to the edges of the observed interaction network. Rather, interactions are allowed between all pairs of users with varying preferences. Specifically, this research question investigates how the speed towards consensus building is governed by configurable influences of user similarity, user social status and a complex interplay between those two factors.

Approach. For tackling this research question, this thesis utilizes the Naming Game model and extends it to reflect (i) latent similarities between users and (ii) observed social status of users in online collaboration networks. Based on these two factors, this thesis provides a model which determines the likelihood of a future interaction between any two given users. To that end, in [Hasani-Mavriqi et al., 2018b], we study consensus building in different society forms, which are characterized according to user similarity into *open*, *modular* and *closed* societies and according to social status into *egalitarian*, *ranked* and *stratified* societies. Based on the influence of user similarity, open and closed societies [Watts, 1999,

[5] `https://git.know-center.tugraz.at/summary/?r=SocialNetworkAnalysis.git`

2004] represent two extreme cases. In an open society, any pairs of users can interact and exchange opinions with each other regardless of their similarity, whereas in a closed society only highly similar users are allowed to interact with each other. A modular society presents a case between these two, in which probability of users interaction is proportional to their similarity. The three society forms based on the influence of social status correspond to those described in RQ1.

With this model in place, we run experiments on 17 collaboration networks extracted from Reddit and StackExchange sites. To account for users social status we use built-in incentive systems of Reddit and StackExchange (i.e., karma and reputation scores). We determine similarities between users by calculating their regular equivalence. Regular equivalence is a measure from graph theory, based on which two nodes are regularly equivalent if their neighbors are themselves equivalent (i.e., they must not share the same neighbors, but they possess neighbors who are themselves similar) [Newman, 2010]. Thus, with this measure we capture not only observable similarities between users but also similarities between non-interacting users (i.e., latent similarities). We than simulate opinion dynamics in this setup and investigate how configurable influences of user similarity and user social status affect the speed of reaching consensus.

Findings and Contributions. The experimental results give interesting insights that do not fully support our expectations and hypothesis. Our initial hypothesis was that if users are driven by their similarities to exchange opinions between them, they will reach consensus faster. But, our findings show that user similarity and social status exhibit opposing forces with respect to consensus building in online collaboration networks. Specifically, an increase in the influence of user similarity delays the consensus building process, whereas a suitable increase of the influence of user social status compensates this delay. These results imply that the influence of these two factors should be carefully balanced to ensure a faster consensus.

RQ3: How does consensus develop in collaborative content creation?

Problem. The previous two research questions do not focus on the dynamics of content creation and evolution when studying consensus building in online collaboration networks. Previous research has shown that users engagement depends on the nature of content they interact with [Tsvetkova et al., 2016]. Some type of content provoke intensive interactions and conversations among users [Yasseri and Kertész, 2013]. Such interactions of users with the generated content and with other users give rise to user relationships in online communities. This research question studies the associations between content creation and consensus building in online collaboration networks. It aims to investigate the semantic stability process in collaboration networks, such as Wikipedia, that are driven based on policies, guidelines and community standards. In Wikipedia editors get actively involved in collaborative editing process while creating a vast amount of knowledge. In this regard, this research question strives to shed light on the dynamics of agreement and disagreement between editors on the content of Wikipedia articles by analyzing their fine grained edit actions.

Approach. To address this research question, this thesis elaborates collaborative editing in the context of Wikipedia, as a representative example of collaborative knowledge construction environments. Collaborative editing of volunteer editors drives the process of content and knowledge creation, which is reflected through numerous article revisions. First, this thesis assesses the semantic stability of Wikipedia articles by analyzing the evolution of article revisions over time. To do so, the Rank Biased Overlap method [Webber et al., 2010] is used to calculate the similarity between term vectors of consequent article revisions. This approach is evaluated in 10 Wikipedia language editions including the five largest language editions as well as five randomly selected small language editions (cf. Section 3.4). Second, this thesis performs a comprehensive analysis of granular edit actions for all articles in English Wikipedia to study editing dynamics and interactions between editors. In addition, it expresses a number of hypotheses on editing dynamics and utilizes the HypTrails framework [Singer et al., 2015] to find plausible explanations for the observed editing behavior (cf. Section 3.5).

Findings and Contributions. Findings gained in terms of this research question shed light on content-related processes that prompt or hinder Wikipedia communities to reach consensus on certain topics and content of Wikipedia articles. The results show that even in policy driven collaboration networks such as Wikipedia, semantic stability can be achieved, but with differences on the velocity of the semantic stability process between small and large Wikipedia editions. Small editions exhibit faster and higher semantic stability than large ones. In particular, in large Wikipedia editions, a higher number of successive revisions is needed in order to reach a certain semantic stability level, whereas in small Wikipedia editions, the number of needed successive revisions is much lower for the same level of semantic stability. The English Wikipedia, as the largest edition, represents a typical example, in which the editorial process of articles is more dynamic. Also the community contributing to English Wikipedia is much larger than in other editions and it is characterized with heterogeneous editors expertise, motivation and opinions. This implicates that it takes time until editors agree if sufficient and correct information is provided within an article. To further investigate editing dynamics in English Wikipedia this thesis conducts a comprehensive study of granular edit actions for all articles. The preliminary findings show that the editing behavior in English Wikipedia is mostly characterized with an intention to get involved in reducing, extending and again reducing content of Wikipedia articles (i.e., conflict-revenge actions [Tsvetkova et al., 2016]). The second most evident editing behavior is reflected with an intention to enhance the provided content of Wikipedia articles by updating or correcting information, deleting old content and inserting new information (i.e., identified as beneficial for the quality of Wikipedia articles [Liu and Ram, 2011]). Editing dynamics such as extending or reducing content consequently (i.e., gardening behavior) exhibits evidences in lower percentage of articles.

One of the contributions is the software solution provided as an open source project[6], which is highly modular, configurable and flexible and can be applied by anyone looking for an efficient way to analyze the semantics of natural language documents.

[6] https://doi.org/10.5281/zenodo.153891

1.5 Main Publications

This cumulative thesis consists of the following publications:

- **Journal 1:** [Hasani-Mavriqi et al., 2016] Hasani-Mavriqi, I., Geigl, F., Pujari, S. C., Lex, E., and Helic, D. (2016). The Influence of Social Status and Network Structure on Consensus Building in Collaboration Networks. *Social Network Analysis and Mining*, 6(1):1-17
- **Journal 2:** [Hasani-Mavriqi et al., 2018b] Hasani-Mavriqi, I., Kowald, D., Helic, D., and Lex, E. (2018). Consensus Dynamics in Online Collaboration Systems. *Computational Social Networks*, 5(1):24
- **Article 1:** [Stanisavljevic et al., 2016] Stanisavljevic, D., Hasani-Mavriqi, I., Lex, E., Strohmaier, M., and Helic, D. (2017). Semantic Stability in Wikipedia. *In Complex Networks & Their Applications V*, pages 379-390, Cham. Springer International Publishing.

In Section 3.5 I also include a work in progress [Hasani-Mavriqi et al., 2018a] that we will submit to an appropriate journal.

1.6 Further Publications

- **Journal 1:** [Seitlinger et al., 2017] Seitlinger, P. Ley, T., Kowald, D., Theiler, D., Hasani-Mavriqi, I., Dennerlein, S., Lex, E., and Albert, D. (2017). Balancing the fluency-consistency tradeoff in collaborative information search with a recommender approach. *International Journal of Human–Computer Interaction*, 34(6):557-575
- **Journal 2:** [Kopeinik et al., 2017] Kopeinik, S., Kowald, D., Hasani-Mavriqi, I., and Lex, E. (2017). Improving Collaborative Filtering Using a Cognitive Model of Human Category Learning. *The Journal of Web Science*, 2(4), 45-61
- **Article 1:** [Görögh et al., 2017] Görögh, E., Vignoli,M., Gauch, S., Blümel, C., Kraker, P., Hasani-Mavriqi, I., Luzi, D., Walker,

M., Toli, E., and Sifacaki, S. (2017). Opening up new channels for scholarly review, dissemination, and assessment. *In Proceedings of the 13th International Symposium on Open Collaboration (OpenSym '17)*, pages 6:1-6:11, New York, NY, USA. ACM

- **Workshop Article 1:** [d'Aquin et al., 2017] DÁquin, M., Adamou, A., Dietze, S., Fetahu, B., Gadiraju, U., Hasani-Mavriqi, I., Holtz, P., Kimmerle, J., Kowald, D., Lex, E., Lopez.Sola, S., Maturana, R., Sabol, V., Troullinou, P., and Veas, E. (2017). AFEL: Towards Measuring Online Activities Contributions to Self-directed Learning. *In Proceedings of Proceedings of the 7th Workshop on Awareness and Reflection in Technology Enhanced Learning (ARTEL) in conjunction with the 12th European Conference on Technology Enhanced Learning: Adaptive and Adaptable Learning (EC-TEL 2017)*

- **Article 2:** [Hasani-Mavriqi et al., 2015] Hasani-Mavriqi, I., Geigl, F., Pujari, S. C., Lex, E., and Helic, D. (2015). The Influence of Social Status on Consensus Building in Collaboration Networks. *In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, ASONAM'15*, pages 162-169, New York, NY, USA. ACM.

- **Article 3:** [Seitlinger et al., 2015] Seitlinger, P., Kowald, D., Kopeinik, S., Hasani-Mavriqi, I., Ley, T., and Lex, E. (2015). Attention Please! A Hybrid Resource Recommender Mimicking Attention-Interpretation Dynamics. *In Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion*, pages 339-345, New York, NY, USA. ACM

- **Article 4:** [Hasani-Mavriqi, 2011] Hasani-Mavriqi, I. (2011). Supporting creation of networked knowledge by automatically generated links. *In Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies, i-KNOW '11*, pages 14:1-14:8, New York, NY, USA. ACM.

- **Article 5:** [Helic et al., 2011] Helic, D., Hasani-Mavriqi, I., Wilhelm, S., and Strohmaier, M. (2011). The effects of navigation tools on the navigability of web-based information systems. *In Proceedings of* *the 11th International Conference on Knowledge Management and Knowledge Technologies, i-KNOW '11*, pages 16:1-16:8, New York, NY, USA. ACM.

- **Article 6:** [Hasani-Mavriqi et al., 2011] Hasani-Mavriqi, I., Leitner, H., Helic, D., and Maurer, H. (2011). Implementation of a wiki-based information and communication system for academia europaea. *In Proceedings of the ITI 2011, 33rd International Conference on Information Technology Interfaces*, pages 193-198. IEEE.

- **Article 7:** [Trattner et al., 2010] Trattner, C., Hasani-Mavriqi, I., Helic, D., and Leitner, H. (2010). The austrian way of wiki(pedia)!: Development of a structured wiki-based encyclopedia within a local austrian context. *In Proceedings of the 6th International Symposium on Wikis and Open Collaboration, WikiSym'10*, pages 9:1-9:10, New York, NY, USA. ACM.

1.7 Contributions and Implications

This thesis contributes methodologically and empirically to the fields of opinion dynamics, content dynamics, agent-based models, network science and computational social science. The main contributions can be summarized as follows:

- First, this thesis contributes with a data-driven model that formalizes the role of user social status and user similarity to explore how the underlying network structure in interplay with these two factors affects the speed towards consensus building. This model is flexible and can be easily extended to reflect various scenarios such as the emergence or disappearance of social classes in online collaboration networks. The methodology can be applied to run extensive simulations of opinion dynamics in arbitrary collaboration networks.

- Second, this thesis provides empirical results that reveal interesting facts on suitability of agent-based models to make opinion dynamics in online collaboration networks traceable for analytics. These findings suggest that experimental outcomes differ if the topology and the structure of the connections between users are known in advance (i.e., synthetic networks) or if empirical user connections are used. Empirical user connections are usually of intrinsically dynamic in nature. The empirical insights help to draw recommendations on how to integrate the influence of user characteristics (i.e., social status or user similarity) in opinion dynamics to optimize consensus building in any collaboration network.

- Third, this thesis experimentally demonstrates how content dynamics drives the process of agreement and disagreement between users collaborating online. The utilized methods assess the semantic stability of the co-created content and extensively evaluate editing dynamics and collaboration patterns among users.

The methods and empirical results provided in this thesis contribute as feedback to the research community on the potential of agent-based models to facilitate an informative and illustrative analysis of opinion dynamics and communication between users. The findings of this thesis showcase the usefulness of agent-based models to study complex social phenomena in online collaboration networks. Further, this thesis makes important recommendations on extending such models with empirical network features to reflect scenarios from real-world settings. Overall, this thesis provides a further step towards a more ambitious goal and larger challenge of developing tools that promote consensus building in online communities.

1.8 Structure of this Thesis

The reminder of this thesis is structured as follows. Chapter 2 gives an overview of the related work. Section 2.1 summarizes the main theories from sociology and social psychology relevant to this thesis. Agent-based models from statistical physics are outlined in Section 2.2. Section 2.3 presents network science methods applied to study dynamical processes in online collaboration networks.

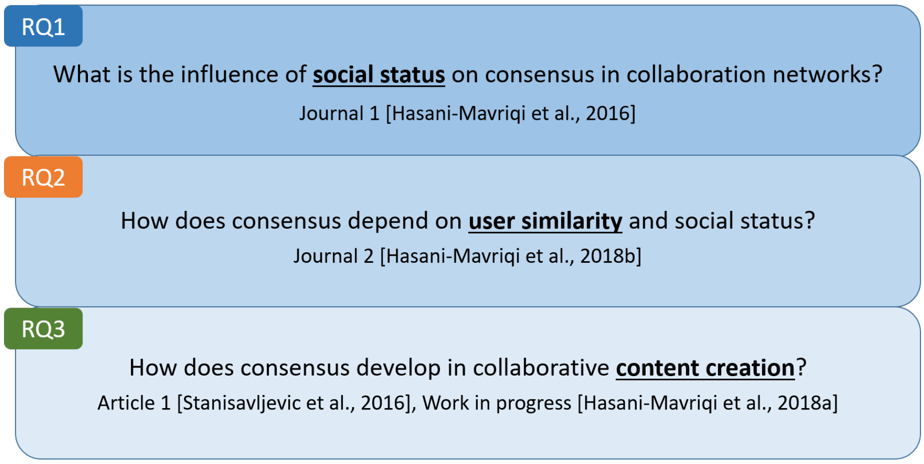


Figure 1.3: **Structural overview of the research questions.** This thesis strives to answer three research questions. In the first research question (RQ1) it investigates the influence of users social status on consensus building in online collaboration networks. The findings of this question open new directions for further investigations that are performed in terms of the second research question (RQ2), which studies the role of users similarity in correlation with users social status in consensus building. Finally, the third research question (RQ3) complements the results of previous two research questions by analyzing the development of consensus in collaborative content creation.

Chapter 3 includes the main publications of this cumulative thesis that address the research questions stated in Section 1.4. Figure 1.3 presents the structural overview of the research questions and the publications related with them. In Section 3.1 I describe my and all collaborators' contributions to these publications.

Chapter 4 concludes this thesis by giving a recap of research results and contributions in Section 4.1 and their implications in Section 4.2. Section 4.3 states the limitations of this thesis and Section 4.4 gives directions for future work.

2 Related Work

In this chapter, I provide an overview of the related work relevant to this thesis. Opinion dynamics and consensus building, being the main topics of this thesis, are studied from three different perspectives: social sciences, statistical physics and network science. Each of the following sections is dedicated to one of the perspectives. Section 2.1 summarizes the main theories from sociology and social psychology, on which this thesis bases the assumptions and hypotheses on the role of social influence in consensus building. Numerous agent-based models that have been developed in terms of statistical physics are presented in Section 2.2. These models are valuable not only for studying opinion dynamics but also other complex social phenomena such as cultural and language dynamics, formation of hierarchies and social spreading phenomena. Section 2.3 outlines the relevant literature, in which network science methods have been successfully applied to study dynamical processes in online collaboration networks.

2.1 Social Science Theories

The research field of opinion dynamics is considered as a sub-field of social dynamics and it originates from various disciplines in social sciences. Specifically, it has been studied in different social influence models developed in social psychology and sociology [Xia et al., 2011]. These studies reveal that in a group of interacting individuals, opinions and behavior of each of them are shaped by the group influence. This is also known as interpersonal influence or social influence [Friedkin and Johnsen, 1999] and is demonstrated by the tendency of interacting individuals to become more alike. The social influence process can also be beneficial for a group, by producing shared understandings and agreements between group members. Furthermore, based on Markovsky et al. [1993] and Weber [1964], individuals tend to create connections and interact with persons having a high social status in their group, who in turn may influence opinions and behaviors of others (i.e., due to their high convincing power). The next two sections outline the most relevant theories and studies to this thesis.

2.1.1 Social Influence

Previous research on social influence models has played an important role in the development of opinion dynamics [Xia et al., 2011]. One of the early endeavors of Katz and Lazarsfeld [1955] has introduced a so called "two-step flow" hypothesis to study the transmission of mass persuasion via the mass media. Based on this hypothesis, individuals that are more exposed to the mass media, pass on what they see, hear or read, to others, with whom they interact and whose mass media exposure is limited. This implies that there exist opinion leaders in society groups that induce norms or opinions (e.g., about fashion, marketing or movies) and fill the gap between media and the mass public. The hypothesis has potential to study recent observations of user behavior in online social networks. Despite the fact that online social networks are designed with the intention to enable equal access or exposure to all their users, so called 'influencers' emerge very fast as opinion leaders in particular online communities. These influencers induce (very often biased) opinions in the group of users interacting online and may act as a proxy to the mass media, even though the exposure to the mass media might not be limited at all.

Two-step flow hypothesis serves as a valuable background information for this thesis, especially, to identify the key players in empirical networks that influence the opinion diffusion and consensus building among users collaborating online.

Further contribution to social influence models and theories has been made by Latané with his social impact theory [Latané, 1981], which describes how individuals are influenced by their social environments. In general, the influence on an individual depends on the group size, the convincing and supportive power of other individuals (strength) and the distance from the subject (immediacy).

In respect to this thesis, the influence of users social status refers to the strength of the impact of other people (e.g., their authority or power of persuasion), whereas the user similarity is analogous to the immediacy of the others (e.g., their closeness in space or time) [Nowak et al., 1990]. Mathematically, the social impact felt by an individual, known also as a target, is a multiplicative function of the three features of a source person and is given in the following form: Impact $= f(S \cdot I \cdot N)$, where *Impact* is the social impact on the target person and S, I, and N, are the strength, immediacy and number of the source persons, respectively [Latané, 1981; Jackson, 1987]. The social impact function constitutes the theoretical basis for the agent-based model and multiplicative effects of the methodology of this thesis.

The social impact theory has been modeled by means of statistical physics in various ways. In [Nowak et al., 1990] researchers applied computer simulations to examine the extent to which group-level phenomena are driven by individual-level processes, establishing the first steps towards a dynamic social impact theory. In synthetic datasets that represent sets of individuals, they studied the attitude change of individuals and group polarization with respect to binary opinion states.

Similarly, this thesis applies agent-based modeling. However, the experiments are performed on empirical datasets from online collaboration networks and more than two opinion states are considered.

A following study by Latané [1996] complements previous studies of dynamic social impact theory by building a foundation for the current opinion dynamics research [Xia et al., 2011]. Social influence or interpersonal influence has been studied extensively by Friedkin and his colleagues [Friedkin and Johnsen, 1990; Friedkin, 1998; Friedkin and Johnsen, 1999]. Authors utilized a structural approach to investigate interpersonal influence within larger social networks. Their work contributed towards formalizing social influence and opinion evolution in social networks [Xia et al., 2011]. Friedkin's studies on consensus building within a group of interacting individuals highlight the importance of the group interdependence. Based on this a group trying to reach consensus on a matter can be seen as a "dynamical whole" [Friedkin, 1998].

2.1.2 Social status

According to definitions from social sciences [Schooler, 2013; Tischler, 2010; Weber, 1964], social status is a concept on individual users of a society that represents a degree of honor, reputation or prestige attached to a position of an individual. This position is earned over individuals' lifetime and is also related to the convincing power they posses (i.e., gained through individuals' expertise and education). Based on their convincing power, individuals may influence opinions and behaviors of others in accordance with their own intentions [Goldhamer and Shils, 1939]. Research on how the position and social status of an individual influence others (i.e., belonging to the same social environment or network) originates from network exchange theory [Markovsky et al., 1993; Walker et al., 2000; Willer, 1999].

Similarly, this thesis studies how the social status of a node in an interaction network affects the spread of opinion that leads to consensus building. Additionally, this thesis defines classes of nodes based on the social status and determines how their interaction affects the process of consensus building.

In online settings, some collaboration networks have established incentive systems to reward users for their contributions. Typical awards are badges or virtual points. While badges typically have no explicit value, they act as symbols of social status within an online community [Anderson et al., 2013; Immorlica et al., 2015].

This thesis also uses built-in incentive systems of StackExchange and Reddit (i.e., reputation and karma scores) as a proxy for social status of users to investigate how their social influence affects consensus building.

Aforementioned social science theories and studies constitute the theoretical foundation for experiments of this thesis on the role of social influence in consensus building. Typically, such studies are conducted in offline settings, nevertheless, this thesis builds hypotheses based on their outcomes to examine their plausibility in communities and groups interacting online.

2.2 Statistical physics and agent-based models

Agent-based models from statistical physics have become increasingly important tools and have proven valuable for studying dynamics of social behavior. The article by Castellano et al. [2009] gives an overview of the applications of statistical physics to phenomena in social networks. It introduces various areas of social sciences where models from statistical physics have been utilized, describes how they relate to each other and other areas of statistical physics, and also discusses shortcomings of such models. As both statistical physics and the study of social dynamics try to understand how the behavior of many agents on the micro scale affect the big picture on the macro scale, there has recently been a big interest in transferring models of statistical physics to social sciences. This massive increase of literature motivated the authors write a review of the state of the art, but at the same time forces them to admit that it is impossible to give a complete account of the matter.

In general, agent-based models represent simplifications that make complex problems of social dynamics tractable for analysis, by applying mathematical models and analytic approaches. For example, in an agent-based model used to study opinion dynamics in online social networks, agents would represent users interacting based on some predefined rules and mathematical variables or dimensions would represent the exchanged opinions. Simplifications of complex phenomena are often arguable. But, for example, based on the everyday life facts, we often only have a choice between a finite number of options when forming an opinion. Thus, scientists have a natural propensity to model these options as mathematical variables or dimensions [Castellano et al., 2009].

The approach of statistical physics to phenomena of crowds represents another example having the advantage that individuals can be viewed as adaptive rather than rational, yet fails to consider that humans cannot be described by a few variables as particles can be. Even though this is a well-grounded criticism, one may attempt to eradicate such imprecisions by introducing the physical concept of noise and in many situations the qualitative outcome does not depend on details in the micro level.

Numerous agent-based models exist on a wide range of topics, structured into the following main fields relevant to this thesis: opinion, cultural, and language dynamics, formation of hierarchies and social spreading [Castellano et al., 2009]. The following sections describe most prominent models in respective fields.

2.2.1 Opinion Dynamics

Opinion dynamics is a process characterized with a group of individuals reaching a consensus (i.e., the majority of a group share the same opinion). In opinion dynamics, the focus is on modeling the opinion state of an individual in particular and a population in general, as well as transitions between different opinion states [Castellano et al., 2009]. Several models have been invented to study the complex process of agreement or disagreement among individuals [Xia et al., 2011].

In the *voter model* [Clifford and Sudbury, 1973; Holley and Liggett, 1975], agents are arranged in a d-dimensional lattice and each agent has a binary state variable, only dependent on its neighbours' states. Hence, agents only adapt to pressure they feel from their most direct peers and bulk noise is not an issue for the most simple models. At each time step, a randomly selected agent takes the state of a randomly selected neighbor. While for a finite amount of agents a consensus will always be reached, for the infinite case the system asymptotes to an agreement if and only if $d \leq 2$. Castellano et al. [2009] discuss how soon consensus will be reached in the various cases and report that the runtime complexity differs significantly for $d \leq 2$ and $d > 2$. Following this, they discuss more sophisticated variations of this model, considering bulk and interfacial noise, as well as models allowing spontaneous opinion changes. Voter models can consider multitype agents [Sire and Majumdar, 1995], such as the existence of one or more zealots who do not change their opinions [Mobilia and Georgiev, 2005], several possible states, including intermediate states of agents [Vazquez et al., 2003], or models considering more than one neighbor in their decision forming process [Lambiotte and Redner, 2007].

Another type of model that is of interest and is considered akin to the voter model is the *majority rule model* [Castellano et al., 2009; Galam, 2000; Galam, S., 2002]. It is used to model the dynamics of group discussions and public debates. In the original majority rule model, agents with binary opinion variables form a complete graph. At each discrete time step, r agents are chosen at random (discussion group) and all agents of this discussion group take the opinion of the majority. In the event of a tie, they will always take a pre-selected preferred opinion. Thus, if r is odd, the threshold for the initial opinion ratio for the opinion that will eventually take over the whole system is $1/2$. Due to the biased choice in case of a tie, the threshold for even r is less than $1/2$. The majority rule model can also be applied to lattices, in which case the discussion groups are randomly chosen sites of constant size on the lattice [Chen and Redner, 2005b]. Furthermore, there exist variations allowing multi-state opinions, in which case consensus is reached in a similar way and time as in the original model, if a mean field limit is used [Chen and Redner, 2005a]. However, if the agents are arranged on a lattice no consensus will be reached if there are too many opinions available and a state of agreement will be reached by diffusive coarsening otherwise. In all of the mentioned above scenarios a worst case runtime complexity exists for a case in which meta stable states develop, which need a long time to dissolve again [Castellano et al., 2009]. Further possible modifications that have been previously studied include: agents that can move in space, heterogeneous node degree distributions [Lambiotte, 2007], varying probabilities of favoring opinions or the opinions of other users and inflexible agents. Modifications of decision rules are also possible, examples being: majority opinions only taking over a discussion with a fixed probability, models using the neighborhood of an agent as discussion group and agents remaining convinced of their own opinion until at least a certain fraction of the group endorses a different position [Castellano et al., 2009].

A further model in opinion dynamics is the so-called *Sznajd model* [Sznajd-Weron and Sznajd, 2000; Stauffer, 2002]. It is related to the voter model and can also be seen as a variation of the majority rule model. It follows the basic idea that one is easier to be convinced or influenced by two or more people who share the same opinion than by a single person [Xia et al., 2011]. In its most general and relevant variation, agents are arranged in a linear chain with binary opinions. At each time step, two neighbors get selected at random and if they have the same opinion, they impose it onto their other two neighbors. If they have opposed opinions, nothing happens (alternative rule: each agent imposes its opinion on the other agent's neighbor). This rule is supposed to model the fact that groups (i.e., pairs) are more convincing than individuals. In this variant of the Sznajd model consensus will always be reached. The model can be extended to a square lattice, in which case the consensus finding can be solved by the mean field limit. It can also be generalized to other graph topologies, including complete graphs, small world graphs and random graphs. The opinions can also be updated for all pairs of neighbors at once, with the possibility of one agent trying to get convinced by two opposing views, which can model frustration and lets the agent stick to its prior opinion [Castellano et al., 2009]. Applications of the Sznajd model include the modeling of competing products on a market or voting behavior [Sznajd-Weron, 2005].

Unlike the previously introduced opinion dynamics models, *bounded confidence models* describe opinions as a continuous variable rather than a discrete one. Communication will only take place if two agent's opinions are within some range ϵ (i.e., bound of confidence), in which case the opinions will move closer together. Possible asymptotical states are consensus, polarization or fragmentation. One example of such a model is the *Deffuant model* [Deffuant et al., 2000], in which a finite number of agents have individual opinions in the interval $[0, 1]$ (i.e., in contrast to discrete opinion models, here all agents usually start with different opinions [Castellano et al., 2009]) and agents are allowed to communicate if the difference between their opinions is lower than a given threshold, otherwise it does not make sense for them to exchange opinions. There exists a convergence parameter, scaling how close opinions will move together in case of a successful discussion. In the Deffuant model, clusters of opinions will emerge and in general small ϵ will produce a higher amount of clusters. The model can be discretized [Stauffer et al., 2004] and variations with individual ϵ, dynamic ϵ and spontaneous state changes exist [Castellano et al., 2009].

A very similar bounded confidence model is the *Hegelsmann-Krause* model [Hegselmann et al., 2002]. It mainly differs from the above by the decision rule, which interacts and takes an average with all neighbors at a given time step rather than just one. It is fully dependent on ϵ and does not require a convergence parameter. It is generally supposed to model formal meeting like interactions and develops its asymptotic states similar to the Deffuant model. Especially though, consensus will always be reached if ϵ is above a certain threshold. Variations are similar to the above model and it can also be modeled as an interactive Markov Chain. On top of these two introduced models, many other bounded confidence models have recently appeared, mainly trying to integrate randomness and irrationality as main features [Castellano et al., 2009; Xia et al., 2011].

To summarize, the main questions addressed by statistical physicists are: whether the opinion dynamics models reach consensus (i.e., ordered configuration), or coexistence (i.e., disordered configurations), which is the scaling of the consensus time and the characterization of the ordering process [Fernández-Gracia et al., 2014]. The development of opinion dynamics models has been uncoordinated and lacking of a general approach, implicating that there are evident similarities between models. Furthermore, there is very limited research on the extent that the rich set of aforementioned endeavors can be applied to explain social interactions and consensus building in real-life settings (e.g., online collaboration networks).

This thesis tackles this research question and contributes with empirical results on applying agent-based models to explain opinion dynamics and consensus building in online collaboration networks (i.e., Q&A sites as StackExchange, discussion forums as Reddit, co-author networks as DBLP, CiteSeer and Google Scholar, or online encyclopedias as Wikipedia).

2.2.2 Cultural Dynamics

Cultural dynamics is a field related to opinion dynamics with the main difference being that an opinion is usually thought of as a variable, while cultural traits of an individual are usually thought of as a set of variables (i.e., a vector). A prominent example of a cultural dynamics model is the Axelrod model [Axelrod, 1997], which manages to take both social influence and homophily (i.e., agents are more likely to be influenced by agents similar to them) into account. Agents are arranged on a regular lattice and are endowed with F integer variables (facets or cultural features), all in some range $0, ..., q-1$ (possible traits allowed per feature). Neighbors then interact depending on how similar they are to each other, with one neighbor updating a variable of the other agent in the interaction. In the long run, either all neighbors will have the same state or different cultural region coexist. Which of these states the system will converge to and how fast mainly depends on F and q parameters. Most analysis of the Axelrod model conducted so far is numerical and not analytical.

As the Axelrod model is highly general and manages to include the two features highly integral to social sciences above, there are numerous improvements and modifications of the traditional model. One of them, for example, introduces noise, discovering that a high amount of noise leads to a disordered steady state [Klemm et al., 2003]. Natural follow up research on this result is to find variations in which some kind of ordered steady state can coexist with noise, and proposed solutions to this question mainly include adaptions of the decision rules (e.g., a rule in which a trait is changed to that of the majority within the neighborhood of an agent instead of just that of a single interacting partner) [Kuperman, 2006]. Furthermore, there exist adjustments allowing random changes within traits, a minimum threshold for a possible interaction, models on small world and scale free graphs, modulations of mass media as an integral part of the network and decision rules considering the difference in trait-variables when changing them. On top of the various versions of the Axelrod model, there also exist vectorization of the Deffuant and Hegelsmann-Krause models [Castellano et al., 2009].

2.2.3 Language Dynamics

Language dynamics can be subdivided into two different approaches, namely sociobiological and sociocultural ones [Castellano et al., 2009].

Within the former fall the approaches of evolutionary game theory, which hypothesize that successful communicators, enjoying a selective advantage, are more likely to reproduce than worse communicators and this is transmitted genetically across generations. Generally, we have a population of agents living in an environment with n objects for which each of the agents can have up to m words [Nowak et al., 1999]. A language L consists of two Markov matrices: P, a $n \times m$ matrix whose p_{ij} entry represents the probability of an agent using word j to describe object i, and Q, a $m \times n$ matrix whose q_{ji} entry represents the probability of an agent associating sound j with object i. The definition of a successful communication between two agents using two different languages is trivial. The payoff, or fitness function, is defined as the mean of the probability of agent one understanding agent two and agent two understanding agent one when conversing about a randomly chosen object. Agents do not adapt their languages, but reproduce according to their fitness with the process of passing on a language to an offspring being well defined and allowing mutation. In this setting the system will converge to a common language, which, however, does not necessarily have to be optimal (i.e., possibility of synonyms and homonyms). This basic model assumes a fully connected network, which in real life settings is improbable, but models for different topologies do exist. Furthermore, there exist adaptions of this model taking errors of perception into account. This, however, gives the fitness function an upper bound, which also limits the number of objects that can be accurately described. The problem can be overcome by allowing the combination of signals into words, which can be infinitely long [Nowak and Krakauer, 1999]. Similar to the evolutionary game theory approach, but slightly more distinguished than its variations, is the Quasispecies-like approach [Castellano et al., 2009]. It modifies the inheritance process and the fitness function in a way such that they are analogous to Darwinian evolution. However, it is hard to obtain the required variables for this model from actual biological systems, which makes it hard to conduct quantitative predictions using this model.

In sociocultural approaches represented by semiotic dynamics, language is considered as a complex dynamical system that evolves and self-organizes, continuously shaped and reshaped by its users [Castellano et al., 2009].

Unlike the evolutionary game theory approach, in semiotic dynamics the space of agents is static while the language of agents is evolving. So, words can adopt new meanings; the relation between a word and its meaning or between a meaning and the world may shift. A well-known example of such a dynamic is the Naming Game model [Baronchelli et al., 2006] that is applied and extended in this thesis.

This thesis utilizes the Naming Game model to study opinion dynamics and consensus building in online collaboration networks because it reflects best dynamical processes taking place in online settings (i.e., collaboratively solving problems or co-creating content), which are mostly self-organized (i.e., no central control) and constantly shaped by users. The simplicity of the Naming Game model enables extensive computational simulations and analytical approaches and facilitates the implementation of a general framework to extend the model and compare it with other models. In the following section, I give an overview of the research conducted around the Naming Game model.

Naming Game

The concept of the Naming Game model was originally introduced in the work by Steels [1995] to explore the role of self-organization in the evolution of language. In this model each agent generates a vocabulary (word to meaning dictionary) at random. Throughout the game, agents interact (usually in pairs) and align their vocabulary in order to enjoy the benefits of successful communication. In most settings, a language will emerge in power-runtime. In general, the model can be amended in a way that it takes probabilities of an agent using or understanding a certain word for a given object into account, similar to the P and Q matrices in the evolutionary game theory model, with the difference that these probabilities are dynamic here [Ke et al., 2002].

The Naming Game model, in the simplest form, is presented by the work of Baronchelli et al. [2006], in which it is studied how large populations manage to reach consensus about the use of a single term for some object, relying only on self-coordination. Authors model such situations in discrete time (t) with simplified interactions between N users on a fully connected

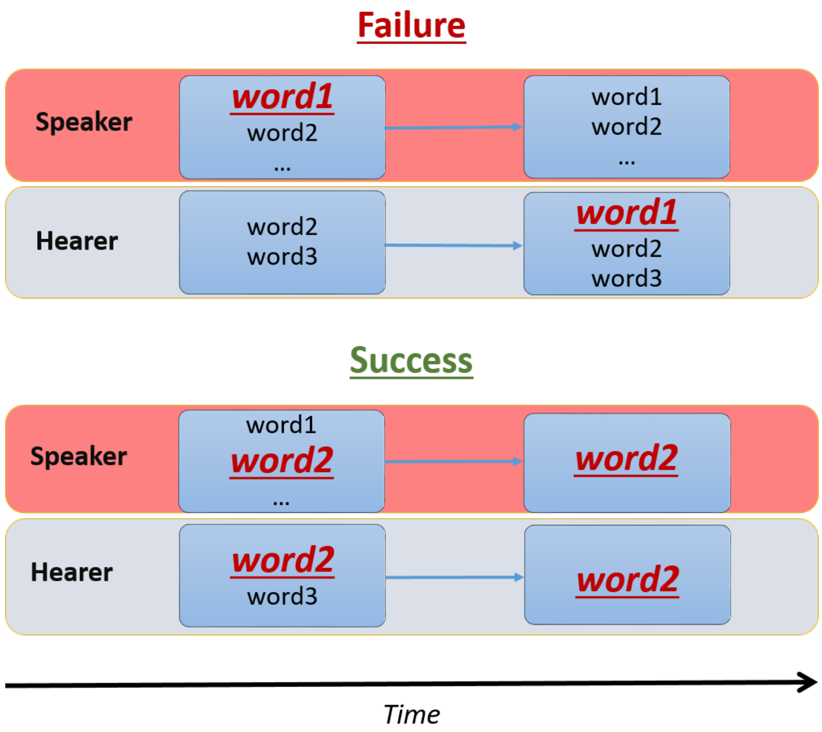


Figure 2.1: **Naming Game.** In the minimal naming game of Baronchelli et al. [2006] two scenarios are possible. 1) If the speaker transmits a word (red) that is unknown by the hearer, the hearer adds it to her inventory (*failure*). 2) If the word chosen by the speaker is also known to the hearer, they both agree on this word. In this case they both remove all other words from their inventories and keep only the transmitted one (*agreement*).

network. At each time step a "speaker" and a "hearer", both chosen at random, converse about an object from the games environment. The speaker picks a word for the object from its vocabulary (or makes one up if this is empty). In the case the hearer knows this term, they both agree on this word and delete all synonyms from their vocabulary. If not, the hearer adds the term to its vocabulary. This is illustrated in Figure 2.1.

That an asymptotic state will always be reached in such a model is yet to be formally proven, but boundedness and the average monotonicity of the system suggest so. After simulating the model for one object and various numbers of players, the researchers manage to identify three phases in the behavior of the system: In the early phase pairs of players play almost uncorrelated games which leads to a linear increase of words in the system. In the second stage, the success rate of games starts to increase approximately as $3t/N^2$, which they explain with the start of correlation building (i.e., multiple players have more than one word in common). In the last phase, a sudden convergence of the number of words in the system towards 1 occurs (i.e., consensus is reached). By evaluation on a log-log-scale authors show that the maximum of words in the system with respect to N (i.e., the point where invention peaks) is monomially distributed with exponent $3/2$. Furthermore, they analyze that the transition from disorder to order becomes steeper as N increases. Authors go on to do a similar analysis with the ratio of users which possess the R-ranked word and the development of the number of deleted words with respect to N and show that these are coherent with the previous findings.

Despite evaluations on fully connected network topology, there exists numerous papers on different topologies, such as low dimensional lattices, scale free networks, small world networks and some complex networks that discuss the effect of those topologies on the consensus finding process and its runtime [Castellano et al., 2009]. Some of them will be discussed in Section 2.3.

Another interesting version of the Naming Game is one introducing an "irresolution attitude" parameter [Baronchelli et al., 2007]; measuring how likely it is that two agents actually update their memory in the event of a successful interaction. In this model, an active stationary state characterized by polarization or fragmentation can evolve, besides a state of complete consensus.

The minimal Naming Game models how languages can develop from scratch in an isolated population. In the real world, however, populations are interconnected and languages already exist. Competition between languages in such a setting can be modeled on a macroscopic and a microscopic level [Castellano et al., 2009]. In the former, the most basic approach is to model the competition of two languages by a differential equation, taking in a fraction of speakers and a status associated with the language. In this case, except for an unstable stationary point in which both languages are perfectly balanced out, one language will always absorb the other [Abrams and Strogatz, 2003]. Empirical data shows that this model is suitable to predict the decrease of speakers of an already endangered language. More sophisticated variations allow the existence of bilingual speakers and to view status as a dynamic entity [Stauffer et al., 2007]. Especially in the former, minority languages are able to survive. Models allowing transmission of languages to offspring and strategies to defend endangered languages also leave a chance for minority languages to avoid extinction. Many microscopic models view a language as cultures in an Axelrod model. For example, the Schulze model [Schulze and Stauffer, 2005] uses the Axelrod model and transitions can happen in three ways: random changes in some parameter, transfer of one word of a language to another, and the adoption of a whole language. The probability for random changes gives a sharp transition between eventual consensus and fragmentation as steady states. Variations allowing reproduction of agents, considering different network topologies, and allowing bilingualism exist. A model proposed by de Oliveira et al. [2006] manages to describe the expansion of competing languages on a square lattice with very similar results to real life language distributions.

In summary, the minimal Naming Game model, even though it is very simplified, it manages to identify the main ingredients for how to describe how individuals develop a shared communication system and that the analysis on microscopic level allows also to draw conclusions for very large populations. Other models extending the minimal Naming Game deal with modifying communication rules, competing languages or different communication topologies.

The presented models perform evaluations only on synthetic data. Hence, this thesis utilizes and extends the minimal Naming Game model to account for social status of individuals and their similarity and evaluates the enhanced model on empirical data extracted from online collaboration networks such as, StackExchange, Reddit, co-authorship networks and Wikipedia.

2.2.4 Formation of Hierarchies

The principle idea of approaches modeling the evolution of hierarchies is that hierarchy is a result of fights and that individuals with a high rank in the hierarchy are more likely to win a fight than those with a low rank. Also, memory of agents should play a role [Castellano et al., 2009]. The first such model introduced is the Bonabeau model [Bonabeau et al., 1995]. Agents, endowed with a strength parameter, are arranged on a regular lattice and move around in random walks. If a lattice an agent wants to move to is already occupied a fight arises, with the outcome probabilities dependent on the force and the forces of the agents being updated according to the outcome. Memory is modeled within the latter process by a relaxation process. For simulations starting with an egalitarian society and a random initial distribution of positions on the lattice, there exists a critical density determining whether an egalitarian or authoritarian steady state will be reached. The model can be solved analytically in the mean field limit. The most popular modification of the model makes the outcome probability of a fight more dynamic [Stauffer, 2003]. This modification introduces a feedback mechanism between the running hierarchical structure of society and the dominance relationships between agents. Analytical research, using the mean field limit, shows that in this case the egalitarian fixed point is stable at all densities, which is seemingly in disagreement with numerical research suggesting the existence of a phase transition towards a hierarchical society. Both fixed points are stable and coexist as opposites. The initial conditions determine which state the system will asymptote to in this case. Further variations in which agents are more avoidant of fights have been studied as well.

This thesis also studies different society forms extracted from Web communities with different hierarchy distribution (i.e., social status distribution). However, real position of users within their communities are extracted from online collaboration networks. The focus of this thesis does not lie on investigating how users gained their positions, but rather on studying how these underlying hierarchical structures influence opinion dynamics and consensus building in online communities.

2.2.5 Social Spreading Phenomena

Social spreading phenomena are again related to opinion dynamics but yet have a very distinguishing feature. While in opinion dynamics it is usually assumed that two interacting agents can influence each other, social spreading phenomena deal with rumor-like opinions, for which the influence can only flow one direction: from the knowing to the unknowing agent. Due to the strong analogy, social spreading phenomena are often thought to be closer to epidemiology than to opinion dynamics. Hence, a natural question for this field to study is if there is an epidemic threshold or if the rumor will eventually reach everybody in the population. The currently most popular model proposed by Daley and Kendall [1964] has the important feature of separating the population into three groups: ignorants, spreaders, and stiflers. People will not be removed or recovered if they are 'infected', but will simply stop propagating a rumor as soon as everyone they want to inform is informed. Thus, the transition to the state in which all 'infected' agents are 'removed or recovered' is proportional to the density of spreaders in the network. An analytical study of the model shows that for homogeneous networks, no threshold as conjectured before exists. When complex or scale-free networks are studied, however, fewer people are reached. Remarkably, on the small-world Watts-Strogatz network, the epidemic transition depends on the rewiring parameter and if this is small enough, the rumor remains only local around its origin. Specialized models for the spreading of corruption and gossip, which unlike rumors in general are about a single person and thus local, exist and have been studied.

Social spreading phenomena is very present in online social networks. This thesis studies online collaboration networks as a special case of online social networks, thus, it's methodology benefits from the insights gained from such models.

To sum up, statistical physics methods constitute important basics for the state-of-the-art research of social dynamics, however, they narrow the scope of research down to theoretical models, which typically do not consider empirical data. As authors in [Castellano et al., 2009] point out, for many theoretical models empirical validation is still outstanding. Furthermore, there is limited research on using models from statistical physics to analyze and categorize behavior in online social and collaboration networks.

This thesis aims to fill this gap by contributing to the limited body of research on theoretical models from statistical physics and empirical data extracted from the Web. By applying agent-based models, such as the Naming Game model, on topologies from real collaboration networks (i.e., StackExchange, Reddit, co-authorship networks and Wikipedia), it is possible to shed light on how masses of users act on the Web and how users with specific intentions may affect the behavior of online communities. Experimental results of this thesis contribute as feedback to the research community on benefits of agent-based models on studying social dynamics processes in online collaboration networks. Such a study is considered a nontrivial endeavor due to intrinsic attributes of users (i.e., heterogeneity of individuals, their interests, behavior) and connections among them (i.e., vary in time). Findings of this thesis show that agent-based models are valuable tools to study complex social phenomena in online communities, however, to account for realistic scenarios from real-world settings they have to be extended with empirical features and they have to include combination of features across different models.

2.3 Network Science Methods

Understanding how crowds of online users collaborate on the Web and how users with specific intentions may affect the behavior of such a crowd is one of the goals of this thesis. Recently, network science has provided powerful means for describing the structure and dynamics of online communities and crowds [David and Jon, 2010; Newman, 2003, 2010; Palla et al., 2007]. Furthermore, methods of statistical physics when extended by network science means have proven to be very valuable for successfully studying opinion dynamics [Brigatti, 2008; Liu et al., 2011; Yang et al., 2008; Zhang et al., 2014].

The following sections give an overview of theoretical and empirical studies related to this thesis.

2.3.1 Theoretical evaluations

There is a considerable streamline of research contributing to evaluation of agent-based models in networks [Brigatti, 2008; Castellano, C. et al., 2003; Castellano, 2005; Li et al., 2017; Liu et al., 2011; Lu et al., 2009; Maity et al., 2013; Sood and Redner, 2005; Waagen et al., 2015; Yang et al., 2008; Zhang et al., 2014]. The relationships of agents in such models play a crucial role in social dynamics and their structure can be described by network science means. Opinion dynamics models are also extended to represent intrinsic properties of agents (nodes) and their relationships (edges). In the following, I describe research endeavors that are mostly relevant to this thesis.

The research paper by Brigatti [2008] develops a model based on the work of Baronchelli et al. [2006]. It points out that the model proposed by Baronchelli et al. lacks a notion of hierarchy amongst the players and only allows the invention of new words as a first step, something that is not coherent with real life social dynamics where new words may emerge at later stages as well. Therefore, each player gets equipped with a reputation variable that may change with time. At $t = 0$, these get randomly assigned by a normal distribution. In the case of a successful game, the reputation of the speaker increases by one. If a game fails, the hearer only amends his vocabulary if their reputation is less than that of the speaker. If not, the speaker invents a new word for the object that was subject of the conversation and their reputation decreases by one. Unlike in minimal Naming Game model, the number of different words does not reach its maximum proportionally to N, but to $N^{3/2}$, and also decreases more quickly after attaining the maximum, i.e., the graph does not show a plateau but develops similar to the graph for the total number of words in the system. It also analyzes that a higher variance in the distribution of reputations leads to a slower development of consensus. If the distribution is skewed towards low reputations (i.e., comparable to an authoritarian society), consensus is reached more quickly. Author also shows that the proportionality in which the time for reaching consensus develops changes at some critical point and becomes less time efficient for larger populations.

In contrast to this thesis, the research uses synthetic network data for simulations and creates synthetic reputation scores (i.e., assigned reputation scores are random numbers that change during iterations). This thesis employs empirical collaboration networks (i.e., real network structure) from StackExchange and Reddit with real reputation and karma scores that are assigned by the community.

Recent research [Waagen et al., 2015; Maity et al., 2013] utilizes the mean-field principle while using the Naming Game model for their experiments. For example, the work in [Maity et al., 2013] studies the impact of learning and the resistance towards learning (as two opposing factors) on consensus building among a population of agents. In [Waagen et al., 2015], authors consider the case of an arbitrary number of agent opinions and the presence of zealots in the Naming Game. Other examples for the Naming Game variations include the work of Liu et al. [2011], who studied the impact that spatial structures (e.g., geographical distances) have on meetings between individuals in a network, and Yang et al. [2008], who proposed a Naming Game that follows an asymmetric negotiation strategy and investigated the influence of hub effects on the agreement dynamics with specific focus on how quickly consensus could be achieved.

This thesis also investigates the role of hubs in opinion diffusion in empirical networks. It uses a parameterized probability function to define the probability of a meeting taking place between two nodes depending on their social status and latent similarities. In contrary, in this thesis agent selection is unbiased and empirical data with explicitly provided reputation scores are used.

Researchers in [Friedkin et al., 2016] give an algorithmic approach on how to describe how a network of individuals reach consensus, taking into account mutual trust amongst the individuals, openness towards interpersonal influences and logical constraints between believes. Each of these factors are modeled by matrices. A matrix representing the certainty of an individual i of some truth statement j in its ij-th entry is defined by a recursive matrix equation and an initial state at discrete time $k = 0$. They go on by simulating the dynamics of the certainty within the US population on whether the invasion of Iraq was just or not, as this was highly dependent on the certainty about the initial truth statement “Saddam Hussein has a stockpile of weapons of mass destruction”. The researchers thereby exhibit how logic constraints strongly affect the outcomes of simulations in their model, while openness towards interpersonal influences does not affect the type of consensus reached, but only the variance of the certainties about it within the population. Furthermore, they analyze how and under which conditions a subset of the population with the same beliefs can affect the outcome in favor of their position. Finally, they also conduct simulations under various logic constraints and competing logic constraints, showing that logic constraints affect the outcome consensus but also that competing logic constraints are often overridden by group dynamics. Even though real truth statements are used in this work, the simulation of interpersonal influences and consensus building is performed in a synthetic network of six individuals.

This thesis rather investigates the role of users social status and their latent similarities in the dynamics of consensus building by extracting empirical interaction networks from the Web.

The research paper by Rosenberg et al. [2009] studies topics in dynamic games with purely informational externalities in great generality and rigor. Their main objective is to show that if players do not act strategically, a consensus about the most successful strategy will be reached and players will eventually stop experimenting. The plays, including the exchange of information, are modeled as a very generalized space. Henceforth, the researchers derive sets of optimal actions and strategies from which they finally conclude that two players will have the same expected limit payoff, given they can observe each other directly or through various other players. With additional assumptions, it is even possible to show that the limit pay-off of two such players is the same. The paper is rounded off by several examples, highlighting the necessity of various assumptions and detailed mathematical proofs of the propositions and theorems that have been stated.

To sum up, the main contribution of the research streamline stated above is to investigate opinion dynamics in pre-designed synthetic networks the structure of which is known a priori (e.g., hierarchical, small-world or scale-free). Typically, intrinsic properties of agents (nodes) and their relationships (edges) are also created synthetically.

In contrary, this thesis applies social network analysis on empirical datasets extracted from the Web to determine intrinsic properties and realistic network features. Than these realistic features are incorporated to opinion dynamics models. This thesis takes a computational approach and analyzes opinion dynamics by simulating the diffusion of opinions in empirical collaboration networks.

2.3.2 Empirical evaluations

In this section, I provide an overview of the literature focused on explaining social dynamics in online social and collaboration networks, utilizing network science means to represent individual and relationship features of users [Akcora et al., 2013; Anderson et al., 2012; Burghardt et al., 2017; David and Jon, 2010; Garcia et al., 2017; Leskovec et al., 2010; Newman, 2003, 2010; Palla et al., 2007; Papadopoulos et al., 2012; Sekara et al., 2016; Xia et al., 2011].

The work by Burghardt et al. [2017] studies different factors that affect which answer is chosen as best to a given question on the online forum StackExchange. Throughout the paper, the researchers distinguish between different types of boards on the platform, assuming they may have different motivations and incentives. Dimensions they consider for their statistical evaluation are reputation of answering users, reputation increase, readability, number of hyperlinks used (i.e., documentation), tenure, eventual acceptance of answer, score of answer before vote, position on page, time since creation, word count and word proportion with respect to the given question. The model used by the researchers is logistic regression (LASSO penalized regression). In their analysis web page order and word share have the highest regression coefficients, implying that these are dominant factors, while tenure, readability and documentation do not seem to play a big role. The dependence on the chronological order and reputation of the answers is also relatively small. As the significance of the word share and web page order increases with the number of answers, the scientists argue that either the collective judgment changes with the size of the answer set or that this is affected by an unknown variable. The first of these hypotheses is based on the idea that if the answer set grows and becomes incomprehensible, users employ cognitive heuristics in attempt to filter out the best answer. Finally, the researchers use their regression models to make predictions and compare these with their data from StackExchange. This analysis suggests that the answer order is the most dominant factor and that especially for technical and meta boards, removing the word share dimension has little effect on the precision of the predictions. Amongst the different board types, tech boards are the most predictable. Overall askers are significantly more predictable in their behavior than voters. Authors explain this by the fact that askers usually have more predictable expectations towards answers and that there are less restrictions to being an asker than to being a voter, making them more likely to employ cognitive heuristics mentioned before to identify the best answer.

The content of questions and answers in StackExchange is not the focus of this thesis, but rather the dynamical processes taking place in such Q&A sites. The above described results could be considered for future work, for example, to perform opinion mining and sentiment analysis.

Recent work followed a theory-driven approach to conduct empirical analysis of Twitter data that supported the assumptions of the social impact theory [Garcia et al., 2017]. The paper represents a study of popularity, reputation and social influence on the Twitter network using a dataset of 40 million Twitter users in two snapshots in the interval of seven years. The authors studied the relationships between the popularity, reputation, social influence but also inactivity. It is important to note that the information flow on Twitter is asymmetric. The follower's network is presented as a directed graph where each node is a particular user and is connected with a directed link towards the nodes representing other users he or she follows. In-degree of a node is the number of links towards the node (number of followers the user has) and out-degree is the number of links pointing out from the node (number of users the user follows). The follower network from 2009 is the base for further studies, since it contains about 1.5 billion follower links. Overall, digital traces from year 2009 to 2016 are used. The authors also have recorded the date of the last tweet of each user. This is needed to calculate the state of activity/inactivity of each user. Popularity of a user is defined by the number of followers that a user has. For measuring the reputation, authors use D-core decomposition and focus on reputation as centrality measure, and define users with high reputation as nodes with high in-coreness. This means they have a large number of followers, but not the other way around. For the social influence measure, they take the retweet rate a user has. At last, every user is measured for their activity or inactivity by the time they last posted a tweet. Authors found that there is a range of values in which the risk of a user becoming inactive grows with popularity and reputation. Furthermore, their results revealed that social influence on Twitter is mainly related to popularity rather than reputation, but that this growth of influence with popularity is sublinear.

This thesis, however, studies the process of opinion dynamics in online collaboration networks, by applying a data-driven model as well as by simulating how opinions spread in those systems. It investigates the intrinsic interplay between local factors such as users social status and users similarity, and the global network structure on consensus building among users.

A framework for link prediction in evolving networks is presented in the work by Papadopoulos et al. [2012], where authors show that popularity is just one dimension of attractiveness, in the context of link creation, and another important dimension is the similarity between users. This indicates that user similarity and user popularity are two main forces that drive people to form links in various networks. User similarity in online social networks has also been studied in [Akcora et al., 2013], in which a method for evaluating social networks according to network connections and profile attributes is presented. In the work by Anderson et al. [2012], the effect of similarity (in terms of user characteristics) and social status, as well as their interplay is studied on online evaluations carried out among users. They found that when two users are similar social status plays less of a role when users evaluate each other. User actions are, for example, editing an article on Wikipedia, asking or answering a question on a Q&A site or rating a review on Epinions.

Major difference to our work is that the authors calculate user similarity as cosine similarity between user action vectors. In this thesis, however, user similarities are calculated by applying the regular equivalence that captures latent similarities even between non-interacting users and users who do not share common actions.

Similar work to [Anderson et al., 2012] is described in [Leskovec et al., 2010], with the difference that the authors consider only the relative social status between two users (i.e., their comparative levels of status in the group) when studying how users evaluate each other. The authors found that users with comparable status hesitate to give positive evaluations to each other.

This thesis also studies the effect of user similarity and user social status, but in relation to dynamical processes that take place in online collaboration networks.

In summary, research endeavors described in this section, combine network science means with regression approaches to analyze and predict user behavior and social influence in online social and collaboration networks.

In contrast, this thesis presents a data-driven model of consensus building in online collaboration networks. It applies empirical network data from Reddit, StackExchange, co-authorship networks and Wikipedia. The approach of this thesis formalizes the role of user status and similarity to explore how the underlying network structure in interplay with these two factors affects the speed towards consensus building. The analysis of communication between users of different status is informative and illustrative of how agent-based models can provide more than traditional regression approaches or black-box predictions.

Editing Dynamics in Wikipedia

The underlying dynamics of large-scale collaboration by volunteers provides great potential to study agreement and disagreement processes between users collaborating online. The repeated interactions of such volunteers give rise to knowledge construction and to communities with shared identity and practice [Iniguez et al., 2014]. Collaborative editing in Wikipedia is driven based on policies, guidelines and community standards. Based on these policies, both editors behavior and the process of article production is managed. Due to its popularity and accessibility of data, researchers from various disciplines focused on the analysis of Wikipedia in numerous studies. The following Wikipedia research topics are related to this thesis: (i) user behavior and interactions between Wikipedia editors [Adler and de Alfaro, 2007; Brandes et al., 2009; Ehmann et al., 2008; Flöck et al., 2017; Gandica et al., 2015; Kalyanasundaram et al., 2015; Liu and Ram, 2011; Pfeil et al., 2006; Sepehri-Rad and Barbosa, 2015], (ii) dynamics of collaboration processes and edit sequence analysis [Bochman, 2012; Keegan et al., 2016], and (iii) content-based analysis of revision history (i.e., detecting edit wars, conflict, disagreement and controversiality) [Gandica et al., 2014; Kalyanasundaram et al., 2015; Török et al., 2013; Yasseri and Kertész, 2013; Tsvetkova et al., 2016; Rudas et al., 2017].

The work by Bochman [2012] is a research paper motivated by Wikipedia's editor loss problem. It studies how peer deliberation in a computer-supported cooperative work (CSCW) environment can be modeled in game theoretic terms and how such models can be used to optimize protocols for consensus finding. Author points out that users inadvertently form hierarchies and social groups, which can make it hard to keep both newbies and veterans engaged. He then exhibits how he believes consensus can be modeled efficiently, taking the amount of available information and the level of socialization of the various participants into account. It is pointed out that such a newly found consensus may be fragile, or will even get blocked in the first place, due to factors such as formation of coalitions, cabal or an interest in the attenuation of other community members. Overall, the paper features several approaches to help understanding how different incentives affect strategies in CSCW disputes and how these strategies should be classified.

The aim of the work by Kalyanasundaram et al. [2015] is to model the process of edit wars and consensus reaching among Wikipedia editors, in order to study various factors that influence consensus formation and to predict the time needed to reach consensus. Their results showed that increasing the number of credible or trustworthy agents and agents with a neutral point of view decreases the time taken to reach consensus, whereas the duration is longest when agents with opposing views are in equal proportion.

The authors in [Iniguez et al., 2014] proposed a minimal model for a collaborative system like Wikipedia. The collaboration between editors in Wikipedia sometimes leads to conflicts, mostly about topics that are considered to be controversial. Conflicts that arise from disagreements on a topic lead to authors constantly overriding each other's edits. Authors model two types of conflict dynamics: agent-agent dynamics (i.e., discussions in talk pages) and agent-medium dynamics (editor's contribution to a Wikipedia page). The opinion formation process taking place in talk pages is modeled through the bounded confidence mechanism [Deffuant et al., 2000] (i.e., discussions would take place only if the opinions of the people involved are close to each other). Whereas the editing process of an controversial article is modeled through the 'inverse' bounded confidence process (i.e., editors change the document state of an article only if it differs too much from their point of view). Authors run experiments with a fixed pool of editors and a variable pool of editors. The results for a fixed agent pool show a rich phase diagram with several characteristic behaviors: (a) a stable article is constantly disputed by editors with extremist views with a slow convergence towards consensus, (b) the article oscillates between editors with extremist views, consensus is reached relatively fast at one of the extremes and (c) the extremist editors converge very fast to the mainstream opinion and the article has an fluctuating evolution. The results obtained for the variable pool of editors reveal that in such systems, there are peaceful and conflict periods that constantly change as different editors leave or enter the system. In the real Wikipedia system, the editors enter and leave the editing platform frequently which makes it more difficult for a consensus to be reached. Four regimes are identified in this case: eternal peace (the system reaches consensus fast and remains there), peace (in peace for great amount of time and sometimes interrupted), war (mostly in a state of disagreement) and perpetual war (no consensus can be reached ever). To detect conflicts in Wikipedia, authors focus on revert actions (i.e., an article has been completely undone by an editor and brought to the last version she or he wrote before someone changed it or updated it). Reverts are detected from the revision history containing two completely equal versions of an article. The latest edit is identified as a revert and the pair of editors involved as reverting and reverted editor. Reverts could happen for various reasons, sometimes conflict of opinions, sometimes vandalism, or correcting someone's mistakes. There also exist mutual reverts, where authors already enter into an edit war. The case much depends on whether the two editors are newbies, one is experienced editor with reputation and the other is a newbie, or both editors have editing reputation on Wikipedia. The last case is considered to be the most serious and the topic can be considered as a controversial topic. To determine the edit reputation of a certain person, the authors use weights which they assign to an editor. The weight is the sum of the edits performed by them. Further, by also weighting the mutual reverts, and summing them up, authors calculate the controversy level of the given article. They classified the controversial articles in three types, based on how they evolve over time: single war to consensus, multiple war-peace cycles and never ending wars.

This thesis aims at analyzing the whole editing process of Wikipedia to detect editing patterns not limited to reverts or disagreements between editors (i.e., reverts do not necessary mean negative social interactions). While previous studies investigate dynamics of collaboration processes in Wikipedia by selecting only samples of Wikipedia articles, this thesis contributes with an extensive evaluation of collaboration patterns in English Wikipedia by examining all articles in the corpus. I perform fine-grained investigations of the revision sequences of each article, by using word based differentiations of edit actions of Wikipedia editors. Furthermore, this thesis introduces a number of hypotheses that potentially explain editing dynamics and quantifies the evidence for those hypotheses in empirical data.

3 Publications

This chapter presents the main publications that constitute this cumulative thesis. Section 3.1 describes the contributions to these publications made by the author and all collaborators. The subsequent sections include the corresponding publications.

3.1 Contributions to the Publications

- **Journal 1:** [Hasani-Mavriqi et al., 2016] Hasani-Mavriqi, I., Geigl, F., Pujari, S. C., Lex, E., and Helic, D. (2016). The Influence of Social Status and Network Structure on Consensus Building in Collaboration Networks. *Social Network Analysis and Mining*, 6(1):1-17

The ideas for this work were developed and refined in discussions between Denis Helic, Elisabeth Lex, Florian Geigl and myself. I was responsible for further implementation of the framework[1] originally developed by Florian Geigl. I conducted a series of experiments, implemented the evaluation method and provided illustrations of the results. I led the writing of the paper. All authors were involved in the iterative process of developing the methodology, interpreting and discussing the results and writing the paper.

- **Journal 2:** [Hasani-Mavriqi et al., 2018b] Hasani-Mavriqi, I., Kowald, D., Helic, D., and Lex, E. (2018). Consensus Dynamics in Online Collaboration Systems. *Computational Social Networks*, 5(1):24

[1] https://git.know-center.tugraz.at/summary/?r=SocialNetworkAnalysis.git

The ideas for this article originated from the discussions with Denis Helic and Elisabeth Lex. I designed and implemented the entire approach, carried out the experiments, produced visual interpretations of the results and drafted the first version of the article. All authors aided in defining the methodology, interpreting the results and contributed intellectually to all research phases. All authors provided feedback and wrote the paper.

- **Article 1:** [Stanisavljevic et al., 2016] Stanisavljevic, D., Hasani-Mavriqi, I., Lex, E., Strohmaier, M., and Helic, D. (2017). Semantic Stability in Wikipedia. *In Complex Networks & Their Applications V*, pages 379-390, Cham. Springer International Publishing.

The ideas for this paper were developed in various discussions between the authors of this paper. Darko Stanisavljevic prepared the data and conducted the experiments. I led the writing process of this paper. All authors were involved in interpreting the results, giving feedback and writing the paper.

This thesis also includes a work in progress in Section 3.5. Ideas for this work in progress [Hasani-Mavriqi et al., 2018a] stem from the discussions between Denis Helic, Elisabeth Lex and me. I was responsible for implementing the code for the presented approach and evaluation methods. I performed the experiments and led the writing of the work. All authors contributed to interpretation and discussion of the results and to the writing process of this work.

3.2 The Influence of Social Status and Network Structure on Consensus Building in Collaboration Networks

This article tackles the first research questions of this thesis, namely what is the influence of users social status in correlation with the underlying network structure on consensus building. To answer this research question, this article takes an interdisciplinary approach. It builds hypotheses based on theories from social sciences, uses and extends agent-based models from statistical physics and applies social network analysis on empirical datasets extracted from StackExchange and co-authorship networks.

In this article we utilize the Naming Game model and extend it by incorporating a mechanism to configure the degree of the influence of users social status on opinion dynamics. We termed this mechanism the Probabilistic Meeting Rule. This rule enables us to study different interesting scenarios, such as the emergence or disappearance of social classes in collaboration networks. The invented extension is very flexible and can be used to control the opinion flow between different user groups by applying our computational approach for parameter estimation.

In addition, we analyze the influence of underlying network structure and the correlation between social status and network structure on consensus building. To that end, we investigate empirical assortative and dissasortative networks and synthetic networks constructed based on the empirical ones.

The results of this article show that low values of the influence of social status tend to favor consensus building. In such scenarios, consensus is always reached at a very fast convergence rate that is faster than in scenarios, in which social status does not play any role. However, if the influence of social status becomes too large, the consensus building process is hindered. Further, investigations on the role of underlying network structure reveal which configurations benefit or hinder the process of consensus building. Overall, the findings of this article suggest that optimizing the process of consensus building is a tuning act of how to integrate users social status in the opinion dynamics.

3.2.1 Abstract

In this paper, we analyze the influence of social status on opinion dynamics and consensus building in collaboration networks. To that end, we simulate the diffusion of opinions in empirical networks and take into account both the network structure and the individual differences of people reflected through their social status. For our simulations, we adapt a well-known Naming Game model and extend it with the *Probabilistic Meeting Rule* to account for the social status of individuals participating in a meeting. This mechanism is sufficiently flexible and allows us to model various society forms in collaboration networks, as well as the emergence or disappearance of social classes. In particular, we are interested in the way how these society forms facilitate opinion diffusion. Our experimental findings reveal that (i) opinion dynamics in collaboration networks is indeed affected by the individuals' social status and (ii) this effect is intricate and non-obvious. Our results suggest that in most of the networks the social status favors consensus building. However, relying on it too strongly can also slow down the opinion diffusion, indicating that there is a specific setting for an optimal benefit of social status on the consensus building. On the other hand, in networks where status does not correlate with degree or in networks with a positive degree assortativity consensus is always reached quickly regardless of the status.

3.2.2 Introduction

It is our natural predisposition to interact with people who have a high social status in our social communities. Customarily, our social interactions and, to some extent, our behavior are influenced by actions of individuals with a high social status. In the field of social psychology, the social status theory attempts to explain this phenomenon [Markovsky et al., 1993; Walker et al., 2000; Willer, 1999]. According to it, people tend to form their connections in a social network to maximize their perceived social benefits arising from the social status of their connections. Also, in the work of Guha et al. [2004] the authors relate social status to the mechanism of link formation in a social network, hypothesizing that people with a lower social status are more likely to create (directed) links with people of a higher social status.

In this paper, however, we are not interested in the relation between the social status and the process of link formation, but rather in the relation between social status and *dynamical processes* that may take place in a social or collaboration network (i.e., a special case of social network, in which users collaborate). One example of such dynamical process is a so-called opinion dynamics process. In our daily lives, we interact with our peers, discuss certain problems, exchange opinions and try to reach some kind of consensus. The question we want to answer in this paper is how social status influences such processes in a collaboration network. For example, in a university class there is a lively discussion between a student and her mentor regarding their newest experimental results and their interpretation. The mentor has a higher social status than the student, due to a superior education, a broader experience and a higher position in the organizational hierarchy. Undoubtedly, while trying to reach a consensus, the student will be influenced by opinions of her mentor because of the latter's convincing power [Castellano et al., 2009; Latané, 1981]. The literature [Castellano et al., 2009] identifies this process as dynamics of agreement/disagreement between persons belonging to a social group. For clarity, in this paper we will refer to it as opinion dynamics.

Problem. The aim of this work is to extend our previous investigations [Hasani-Mavriqi et al., 2015] in respect to the influence of social status on the process of reaching consensus within a social community that has a heterogeneous distribution of social status, by studying the underlying network structure. In particular, we investigate new empirical networks and construct synthetic networks to analyse the impact of degree assortativity and the correlation between degree and social status on opinion dynamics. While there is a substantial body of work on opinion dynamics (see Section 3.2.7) in general settings, we focus on a more specific and more realistic situation in which the dynamics are influenced not only by the network structure and the relevant parameters but also by the intrinsic properties of every single node in the network, such as for example, social status. In other words, we study the interplay between structure, dynamics and exogenous node characteristics and how these complex interactions influence the process of consensus building.

Approach & methods. In the field of statistical physics [Castellano et al., 2009], opinion dynamics are commonly studied by applying mathematical models and analytic approaches. To make these complex problems tractable for mathematical analysis, researchers make simplifications, such as presenting opinions as sets of numbers, ignoring the network structure (a typical approach from e.g., mean-field theory) and neglecting the individual differences between nodes. Simplifications narrow the scope of research down to theoretical models, which typically do not consider empirical data. Even so, statistical physics constitutes important basics for the state-of-the-art research on social dynamics in collaboration networks. In this paper, we build upon these basics.

We take a computational approach and analyze opinion dynamics by simulating the diffusion of opinions in empirical collaboration networks (specifically, we study datasets from a Q&A site StackExchange and a co-authorship dataset). In our simulations, we consider the network structure, apply a set of simple rules for opinion diffusion and take into account people's individual differences (e.g., their social status). In particular, we simulate scenarios of peer interactions in empirical datasets assuming that the status theory holds and observe the consequences. We model the dynamics of opinion spreading by adapting a well-known *Naming Game* model [Baronchelli et al., 2006] and extending it by incorporating a mechanism to configure the degree of the influence of social status on the network dynamics. We termed this mechanism the *Probabilistic Meeting Rule*. Through parametrization, we are able to explore various scenarios from the opposite sides of the spectrum: (i) we can completely neglect the status by allowing any two individuals to exchange their opinions regardless of their social status (an *egalitarian* society) [Arneson, 2013], (ii) we can have opinions flowing only in one direction – from individuals with a higher social status to those with a lower social status (a *stratified* society) [Weber, 1964], (iii) we can probabilistically model any situation in between these two extreme cases, that is, a case in which opinions are very likely to flow from individuals with a higher social status to those with a lower social status but with small probability they can also flow into the other direction (a *ranked* society) [Weber, 1964].

Contributions. The main contributions of our work are two-fold. Firstly, with our paper we contribute to the field of opinion dynamics *methodologically*. Secondly, with our work we also make an *empirical* contribution.

Our methodological contribution can be summarized as follows. To model various scenarios of how social status may influence the opinion dynamics, we have invented the Probabilistic Meeting Rule (see Section 3.2.3) and extended a standard Naming Game model with that rule. The extension is flexible and may reflect a variety of interesting scenarios, such as the emergence or disappearance of social classes in collaboration networks. Further, we provide an initial analysis on how this meeting rule may influence the consensus building process. This analysis allows us to obtain an intuition on the possible outcomes of our simulations. The opinion flow between different user groups can be easily controlled through our computational approach for parameter estimation (see Section 3.2.3). We also analyze the influence of network structure, particularly, the influence of degree assortativity and the correlation between degree and status on the process of consensus reaching in collaboration networks.

From the empirical point of view, we made a much-needed contribution to the limited body of research on Naming Game and empirical data [Gao et al., 2013] and obtained very interesting empirical experimental results. For example, based on the status theory it can be expected that consensus can be reached faster when social status plays a role. However, our results only partially confirm this expectation. In particular, if an opinion flows only in one high-to-low status direction, opinions do not converge at all since there are always a few people who do not adopt the common opinion from the network. However, with only a low influence of social status convergence is reached faster than with no status at all (as in a standard Naming Game). These results suggest that finding the optimal process of consensus reaching is a tuning act of how to integrate social status in the opinion dynamics. In addition, our investigations on the role of network structure in opinion dynamics, reveal that hubs are important factors for spreading a single common opinion among other nodes and that in networks with a positive assortativity degree or a degree sequence decorrelated to user's social status, the consensus is reached without external intervention.

The StackExchange empirical networks used in our previous work [Hasani-Mavriqi et al., 2015] are disassortative networks, i.e. they have a negative degree assortativity coefficient. In disassortative networks, high degree nodes are on average connected to nodes with low(er) degree [Noldus and Van Mieghem, 2015]. In this work, we extend our experiments with an additional type of empirical network, namely assortative networks, in which physical connections between low and high agents are very rare. We turn to co-authorship networks as an example of networks that exhibit a positive degree assortativity coefficient, indicating that, on average, nodes with similar degrees are connected together.

3.2.3 Methodology

Naming Game

Naming Game [Baronchelli, Andrea and Dall'Asta, Luca and Barrat, Alain and Loreto, Vittorio, 2006; Baronchelli et al., 2005, 2006; Dall'Asta et al., 2006a; L. Dall'Asta, A. Baronchelli, A. Barrat, and V. Loreto, 2006] is a networked agent-based topology, in which agent-to-agent interactions take place based on predefined gaming rules. In particular, agents exchange their opinions and try to reach a consensus about the name of an unknown object. When all agents in the network agree on the name, the network is considered to have established a common opinion.

Agents in the game are represented as nodes of a network and edges between two agents allow them to interact with each other. Names are represented with an inventory of words and each agent has her own inventory to store the words. Technically, an inventory is a set (i.e., a bag) of words. In the initial state, the inventories are empty. Two random adjacent agents are chosen in each simulation step to interact through a meeting, one agent is declared as a speaker and the other as a listener. In the course of the meeting, the speaker selects a word from her inventory and communicates it to the listener (note that if the speaker's inventory is empty, a new unique word is created and stored in the inventory). After communicating the word to the listener, two scenarios are possible (see Figure 3.1):

1. the word is not in the listener's inventory – the word is added to listener's inventory,
2. otherwise, both speaker and listener agree on that word and remove all other words from their inventories – they agree on the selected word.

Naming Game and social status

We modify the Naming Game to account for social status. As before, the agents are represented as network nodes, edges denote whether two agents can interact or not and names (opinions) are represented as word inventories.

The first difference between our model and a standard Naming Game is the simulation initialization. We initialize the inventories with a given number of selected words from a given vocabulary. The words are selected (with replacement uniformly) at random from the vocabulary. This results in an initial state where each opinion occurs with the same probability.

Secondly, we adopt the social status that governs how agent interactions are turned into meetings – not every agent interaction is turned into a meeting. During each interaction a random agent and a random neighbor are chosen to have a meeting. Then, the speaker and the listener are assigned randomly. Based on the difference between the speaker's and the listener's statuses, we randomly decide if the meeting occurs.

To decide if a meeting takes place, we introduce the Probabilistic Meeting Rule. Basically, the Probabilistic Meeting Rule is a function that takes the agents' social statuses as input and, based on the difference between the speaker's and listener's status, calculates the probability of the meeting taking place. The rule is defined by the following equation:

$$p_{sl} = min(1, e^{\beta \cdot (s_s - s_l)}), \tag{3.1}$$

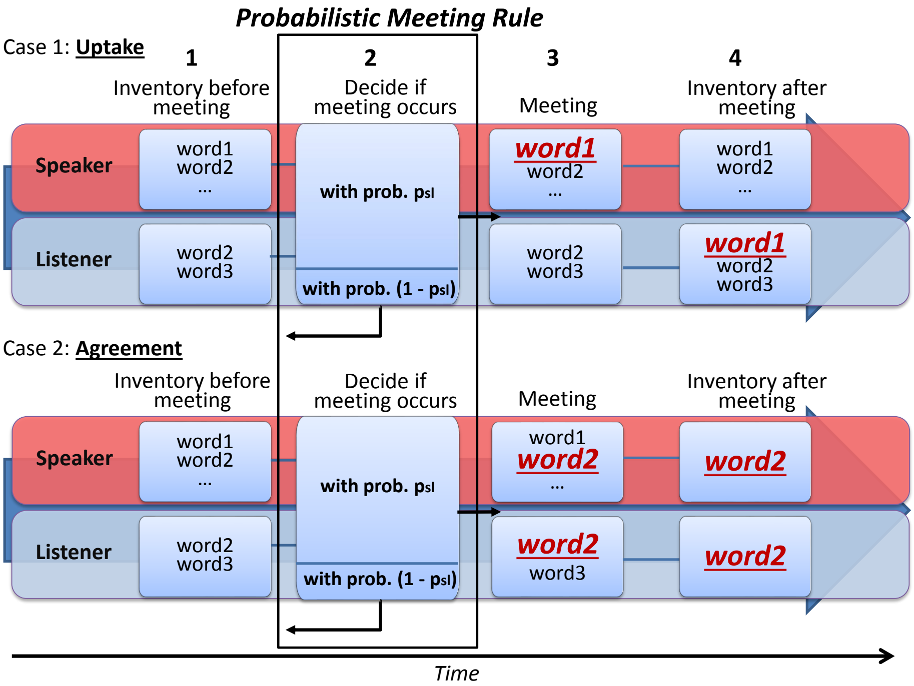


Figure 3.1: **Naming Game meeting.** The classical Naming Game consists of steps 1, 3 and 4, whereas our extension also includes the step 2. In step 2 we decide whether the meeting between two agents occurs by evaluating Probabilistic Meeting Rule (equation 3.1). For illustration, consider a ranked society with stratification factor $\beta = 0.0001$. *Example 1:* Speaker's status $s_s = 101$ and listener's status $s_l = 7967$. The meeting probability evaluates to $p_{sl} = 0.45$. We then draw a number from $[0, 1]$ uniformly at random (e.g., 0.93) and compare it with $p_s l$ – the meeting does not take place. *Example 2:* Let $s_s = 576$ and $s_l = 865$, which leads to the meeting probability $p_{sl} = 0.97$. We again draw a random number from $[0, 1]$ (e.g., 0.77) – in this case the meeting takes place. If the meeting takes place two scenarios are possible. 1) If the speaker transmits a word (red) that is unknown by the listener, the listener adds it to her inventory (*uptake*). 2) If the word chosen by the speaker is also known to the listener, they both agree on this word. In this case they both remove all other words from their inventories and keep only the transmitted one (*agreement*).

where s_s is the speaker's status, s_l is the listener's status and $\beta \geq 0$ is the *stratification factor*. The stratification factor β, which can be viewed as a measure of conformance to the agent's social status, is a tuning parameter in our model. The above equation results in the following probabilities. If the speaker's status is higher than the listener's status, p_{sl} has the value of 1, that is, such a meeting always takes a place. If the opposite is true, various scenarios are possible, depending on the value of the stratification factor. For example, $\beta = 0$, indicates an *egalitarian* society and p_{sl} is always equal to 1. However, if we slowly increase the stratification factor, p_{sl} will start to decay and in general will take a value between 0 and 1, which signifies a *ranked* society (see the running example in Figure 3.1). If we continue to increase β, we will soon (because of the exponential term in the equation) reach a situation where p_{sl} for all practical matters is equal to 0. In other words, we have reached a *stratified* society where meetings take place only if the speaker's status is higher than the listener's status but never in the opposite case.

The application of our Probabilistic Meeting Rule to our datasets is depicted in Figure 3.2. The probability of a meeting taking place is shown in correlation with the percentage of pairs of agents participating in that meeting. The above mentioned scenarios are represented as follows: *egalitarian* society (corresponds to $\beta = 0$) – green bar (circle texture), *ranked* society (e.g., $\beta = 0.0001$) – blue bar (line texture) and *stratified* society (e.g., $\beta = 1$) – red bar (star texture).

Estimating stratification factor

In this section, our primary goal is to investigate how the stratification factor β from Probabilistic Meeting Rule Equation 3.1 can be estimated such that the opinion flow between different classes of agents can be easily controlled. We first draw a line in the distribution of agents' statuses and separate the agents into two classes: high (agents with the status above 90^{th} percentile) and low (agents below 90^{th} percentile) class. Our focus lies on the estimation of the expected meeting probability between low and high status agents. Please note, however, that the methodology presented here can be applied also in a general setting to estimate, for

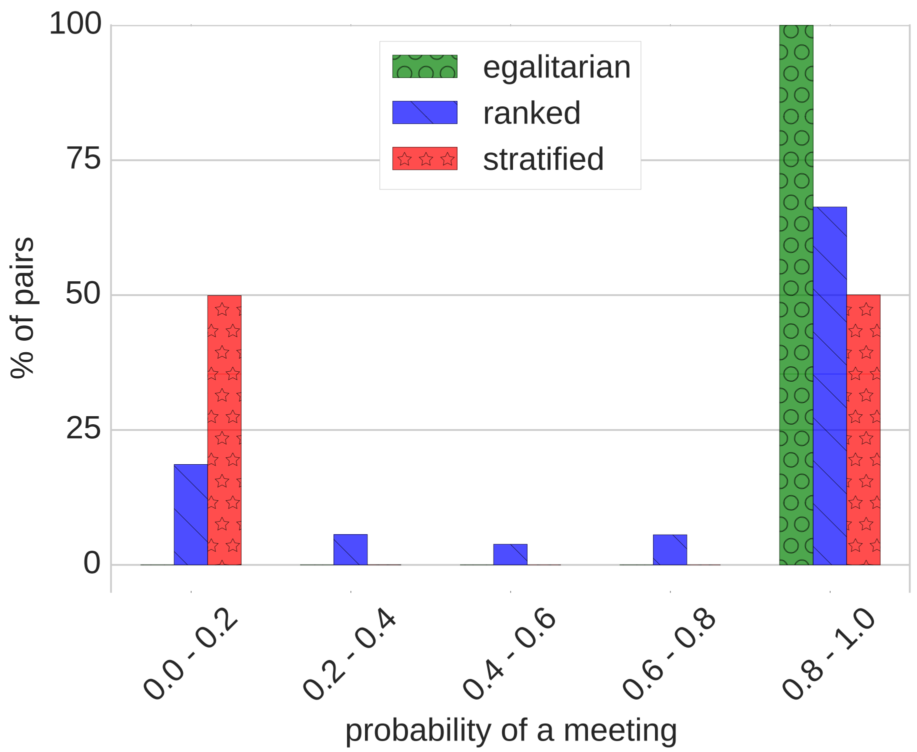


Figure 3.2: **Naming Game and social status**. The application of the *Probabilistic Meeting Rule* to our datasets and the emergence of social classes based on the stratification factor β are illustrated. The green bar with circle texture indicates an *egalitarian* society that corresponds to $\beta = 0$, in which each agent can meet every other agent. With an increase in β, our society becomes more conservative (as represented with the blue bars with line texture) and becomes a *ranked* society. In red bars with circle texture we observe a two-class society, that is, a *stratified* society.

example, expected meeting probability between low-to-low, or high-to-high agents.

The expected meeting probability depends on the differences between agents' social status, which in turn are random variables with unknown probability density functions. Formally, the problem is to calculate the expectations of a function (Probabilistic Meeting Rule) of a difference of two random variables, which are conditioned on their particular values, that is they are conditioned on either being a low or a high agent.

Let X be a random variable (r.v.) representing a user's social status. The probability density function (PDF) of the r.v. X is given with $p(x)$. We define now a new random variable conditioned on a specific value of that variable x_h, that is let us consider a random variable U for a low status agent and a random variable V for a high status agent. The PDF of U is then given by $p(u) = p(x|x \leq x_h)$ and PDF of V by $p(v) = p(x|x > x_h)$. Both of these PDFs can be obtained by normalizing with a cumulative and complementary cumulative distribution function evaluated at x_h.

To consider the differences between agents' social statuses, we would need to define a third r.v. $Z = U - V$, and under the assumption that the r.v. U and V are independent we could calculate the PDF of Z by calculating the convolution integral for U and V. Finally, we can define the expected value of Probabilistic Meeting Rule $e^{\beta \cdot z}$ as follows:

$$E[e^{\beta \cdot z}] = \int_{-\infty}^{\infty} e^{\beta \cdot z} \cdot p(z) dz \tag{3.2}$$

Since in practice none of these steps is tractable for the analytic solution, we resort to the empirical and approximative parameter estimation. To that end, we first create an empirical distribution for the random variable Z. First, we split agents into two classes: low and high defined by, for example, the 90^{th} percentile (although the choice for x_h is in fact arbitrary) in the distribution of agents' status values. Second, we iterate over all the links in the network and keep only low-to-high pairs to construct an empirical distribution of the differences between agents' statuses. Please note that the same procedure may be repeated for estimation of, for instance, the expected meeting probability of low-to-low or any other interesting pairs (instead of keeping low-to-high pairs we just need to keep the pairs in question). From this distribution, we then draw a random sample of size N and estimate the expectation value for $e^{\beta \cdot z}$ by applying the well-known Monte Carlo estimation [Metropolis and Ulam, 1949]:

$$E[e^{\beta \cdot z}] = \frac{1}{N} \sum_{i=1}^{N} e^{\beta \cdot z_i} \tag{3.3}$$

Our empirical solution is flexible and can be easily adapted to consider opinion flow in other agents' groups (e.g., high-to-high). By defining the percentage of allowed opinion flow between agents in different groups, we can determine β for networks of various structure and scope.

3.2.4 Datasets and Experiments

Datasets

In our experiments, we use two types of empirical datasets: (i) the first one is derived from a Q&A site (StackExchange[2]) and (ii) the second one is a co-authorship dataset introduced in [Tang et al., 2008].

In StackExchange users collaborate, ask questions and give answers on particular problems. After an iterative discussion process users exchange their opinions, find solutions to a problem and agree on the best suggested solutions [Tausczik et al., 2014]. Such Q&A sites have a reputation system which rewards users via reputation scores based on their contributions [Halavais et al., 2014; Movshovitz-Attias et al., 2013]. Based on the policies of this reputation system, users get appropriate reputation scores for giving good answers, asking good questions or for voting on questions/answers of other users. It is evident that high reputation users contribute high quality answers [Movshovitz-Attias et al., 2013]. We expect that high reputation users also demonstrate high convincing power during the agreement process, influencing opinions of other (low reputation) users. In our experiments, we apply reputation scores as a proxy for the social status and these two terms are used interchangeably throughout the paper. The StackExchange platform does not indicate associations between users or friendship links. For that reason, we turn our attention to collaboration networks which we extract by analyzing co-posting activities of users in order to have social ties between them [Adamic and Adar, 2001; Halavais et al., 2014; Tang et al., 2012]. In Q&A sites, a co-posting activity between two users refers to a scenario under which two users comment on the same post. Thus, if two users contributed in any way to a same post, they are connected via an edge in the collaboration network. We analyze the following StackExchange language datasets: French, Spanish, Chinese,

[2] http://stackexchange.com/

Japanese, German and English. They are available for downloading for research purpose from the StackExchange dataset archive.

We constructed our co-authorship network from the empirical dataset presented in [Tang et al., 2008] that is freely available under [3]. In this co-authorship dataset, publication data are combined from three different sources: DBLP, CiteSeer and Google Scholar and the problem of the author name disambiguation is addressed properly. Two authors are connected via an edge in the co-authorship network if they co-authored at least a publication together. The dataset provides citation counts for each author, which is used in our case as a proxy for the author's reputation.

Datasets statistics

The details of our empirical networks (derived from the above-mentioned datasets) and their properties are shown in Table 3.1, with the number of nodes (n), number of edges (m), mean (μ), median ($\mu_{1/2}$), standard deviation (σ) of the reputation scores, assortativity coefficient (r) and modularity (Q).

Among our StackExchange datasets, the English network is the largest one with $30,656$ nodes and $192,983$ edges, whereas the French is the smallest one with $1,478$ nodes and $6,668$ edges in the network. The German, Japanese, Chinese and Spanish networks lie in between the English and French networks in terms of network size. The co-authorship dataset is much larger in size compared to all StackExchange datasets, with $1,057,194$ nodes and $3,634,124$ edges it constitutes the largest dataset in our experiments.

The negative assortativity coefficient r in our StackExchange datasets indicates a negative correlation [Newman, 2003] between reputation scores over the network edges. In other words, users with lower reputation scores are more likely to connect to users with higher reputation scores. In particular, a typical post in our datasets has many users with low scores (e.g., who post a question) and only a few or even only a single user with a high score (e.g., who answers the question). This finding is in line with the

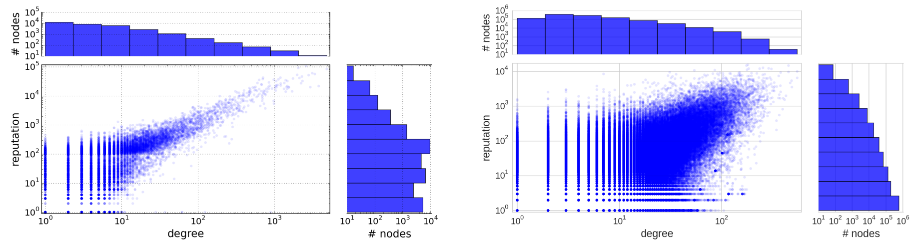
[3]https://aminer.org/DBLP_Citation

assumptions from the social status theory. The Chinese network has the lowest assortativity coefficient among our networks indicating that in this network there is a smaller chance of connection with a dissimilar reputation score. The Japanese and French networks have the highest absolute assortativity coefficient. The co-authorship dataset is characterized with a positive assortativity coefficient r, which is typical for co-authorship networks in general [Noldus and Van Mieghem, 2015], indicating that, on average, nodes with similar reputation scores are connected together. Particularly, this means that authors having similar social status in their community tend to publish an article together.

The modularity score is a measure of strength of the community structure in a network. A high modularity score indicates the existence of strong communities in the network, while a low modularity score means that the community structure is not that strong [Newman, 2006]. In our StackExchange networks, we observe low modularity values corresponding to a very weak or almost nonexistent community structure. As previously shown in a network without communities, in general Naming Game converges quickly to a single opinion [Baronchelli et al., 2006]. In contrary, our co-authorship network exhibits much higher modularity value, thus, the community structure in this network is stronger.

Table 3.1: **StackExchange and co-authorship datasets.** Description of StackExchange and co-authorship datasets with the number of nodes (n), number of edges (m), mean (μ), median ($\mu_{1/2}$) and standard deviation (σ) of the reputation scores, assortativity coefficient (r) and modularity (Q).

Dataset	Type	n	m	μ	$\mu_{1/2}$	σ	r	Q
StackExch.	French	1,478	6,668	298	111	1,273	−0.23	0.31
	Spanish	1,584	6,908	196	101	554	−0.19	0.38
	Chinese	1,985	8,556	160	61	477	−0.15	0.41
	Japanese	2,069	11,155	328	77	1,535	−0.23	0.34
	German	2,316	12,825	285	103	1,219	−0.16	0.32
	English	30,656	192,983	199	48	1,654	−0.19	0.33
Co-auth.	AMiner	1,057,194	3,634,124	20	2	138	0.15	0.67



(a) English StackExchange (disassortative) (b) Co-authorship (assortative)

Figure 3.3: **Distribution of reputation scores.** The correlation between the distribution of reputation scores and node degrees for the English StackExchange network (a) and co-authorship network (b). The subplots on the right show the heterogenous distribution of reputation scores in the both networks. The subplots on the top present the heterogenous distribution of node degrees. In the middle, the scatter plot of reputation scores vs. node degrees is shown. The Pearson correlation coefficient between the degree and the reputation score is 0.88 for the English StackExchange network. All other StackExchange datasets have comparable distributions and correlation coefficients. In the case of the co-authorship network, the Pearson correlation coefficient between the degree and the reputation score is 0.54. As it can be seen from the plot in (b), it is evident that some authors with a high citation count have a low degree (i.e., low number of co-authored publications), but there are also cases of authors with a low citation count and a high degree (i.e., they are active in scientific collaboration, but their publications have a low citation count).

The distribution of reputation scores and node degrees resemble a heterogenous distribution for all networks, which indicates that the majority of users in our collaboration networks have low reputation scores. Figure 3.3a shows the English StackExchange network, in which the correlation between the reputation scores and the node degrees is a linear correlation with a Pearson correlation coefficient of 0.88. All other StackExchange datasets have comparable properties. In the case of the co-authorship network shown (see Figure 3.3b), the Pearson correlation coefficient between the degree and the reputation score is 0.54. It is evident that there are cases of authors having a high citation count (used as a proxy for reputation) but low degree, which indicates that they possess a low number of co-authored publications that are frequently cited. For illustration purposes, we further investigated this property of our co-authorship dataset and retrieved the names of the authors having a low degree (lower than the $90th$ percentile) and a high citation count (higher than the $90th$ percentile). For example, the author Dennis M. Volpano [4] is characterized in our dataset with a degree of 6 and a citation count of 750. After checking the author's website and digital libraries such as IEEE Xplore, it is obvious that the author published most of his publications as a single author or in collaboration with other few authors, but his publications received a considerable attention from the community and are highly cited. The opposite scenarios are also possible, which correspond to authors being active in scientific collaboration (high degree), but their publications have a low citation count.

Simulations

In our experiments, we simulate Naming Game extended with the Probabilistic Meeting Rule. The simulation framework is provided as an open source project[5]. Our experiments consist of the following steps:

1. We calculate the stratification factor β using the approach from the Section 3.2.3, getting the values for the stratification factor that we need, to reflect a given situation. For all networks, we define five percentages, which correspond to the society forms defined earlier in this paper and control the opinion flow from low to high status agents (i.e., 100% - *egalitarian*, 75%, 50%, and 25% - *ranked*, and 0% - *stratified* society).

2. Each agent's inventory is initialized with a fixed number of three opinions (represented through numbers from 0 to 99). These opinions are selected uniformly at random from a bag of opinions to ensure that each opinion occurs with the same probability.

[4] http://faculty.nps.edu/volpano/
[5] https://git.know-center.tugraz.at/summary/?r=SocialNetworkAnalysis.git

3. We once create meeting sequences and apply the same sequences for the different values of stratification factors. Initialization of agent inventories differs for each meeting sequence, but same initializations are used for all β. Hence, it is ensured that the randomness between β is insignificant, due to the same meeting sequence and same initialization for different β.

4. For each meeting sequence, depending on the network size, we define the number of user interactions (iterations) for the simulations. We perform 4 million interactions for the largest StackExchange network (English), 1 million interactions for the five other StackExchange networks and 20 millions interactions for the co-authorship network.

5. We run 100 simulations per β and report the averaged simulation results to account for statistical fluctuations in the simulations.

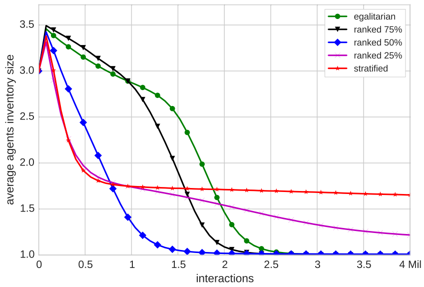
6. During the simulations, we store important information such as the appearance of agents as listeners/speakers, their participation in overall interactions versus successful meetings and the evolution of the agent's inventory size.

7. We modify the initialization of the agents' inventories to differentiate between opinions assigned to low and high status agents respectively, in order to evaluate the final agreement of agents.

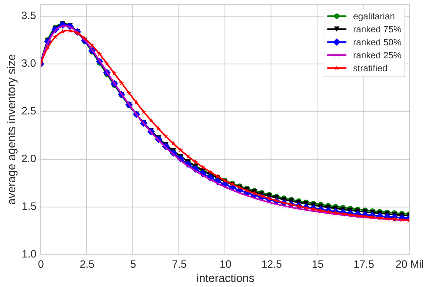
3.2.5 Results and Discussion

Figure 3.4 summarizes the results of our experiments by depicting the agent's inventory size as a function of the simulation progress for the (a) English StackExchange and (b) co-authorship networks.

Inventory size evolution of disassortative networks. The simulation results among all StackExchange networks are similar, thus, we show only the results of the largest StackExchange network (i.e., English in Figure 3.4a). In the case of *egalitarian* society ($\beta = 0$), the English network converges to a single opinion. This is in line with the previous experiments with the Naming Game – in networks without a strong community structure, we always reach a consensus. In the case of *stratified* society,



(a) English StackExchange (disassortative)



(b) Co-authorship (assortative)

Figure 3.4: **Inventory size evolution averaged over** 100 **runs per** β**.** Mean values of the agent's inventory size in relation to the number of interactions for English StackExchange (3.4a) and co-authorship (3.4b) networks. We compute five β for each network and control the opinion flow from low to high status agents. The green lines in the plots correspond to *egalitarian* societies (100% opinion flow), whereas the red lines represent the *stratified* societies (0% opinion flow). The lines in between (black, blue and magenta) depict the *ranked* societies, in which the opinion flow from low to high status agents is inhibited to 75%, 50% and 25% respectively. For readability reasons, error bars representing standard deviation of the mean agent's inventory size over 100 runs per β are not depicted in the plots. In the English StackExchange network (3.4a), in the case of an *egalitarian* society a common opinion is reached and the convergence rate is fast. In a *stratified* society, the opinions do not converge (the mean number of opinions lies between 1 and 2). *Ranked* societies also reach a common opinion with the highest convergence rate. Thus, for the English network, the consensus building depends on the status but in a non-obvious way, indicating that there is a specific setting at which the influence of the social status reaches the optimal state. In the case of co-authorship network in (3.4b), consensus is reached almost independently from β, so external interventions (such as our Probabilistic Meeting Rule) do not influence opinion convergence rates.

we do not observe convergence – consensus cannot be reached. This seems slightly counter-intuitive – an intuition would be that consensus building would benefit from the presence of agents with a high social status and their influence on agents with a lower social status.

Finding 1: Opinion dynamics in disassortative collaboration networks are affected by the individual's social status. If, due to the social status, opinions flow only in the high-to-low direction, the consensus building process is disturbed and consensus cannot be achieved, as opposed to when the status does not play any role at all.

The simulation results for *ranked* societies indicate that the impact of the social status on opinion dynamics is a complex one. In all our StackExchange networks, we observe the following situation. By starting at $\beta = 0$ and slowly increasing the stratification factor (note that higher values of stratification factor successively reduce percentages of meetings between low and high status agents), we are at first still able to reach consensus. Moreover, the convergence rate increases with a slightly increased stratification factor (cf. Figure 3.4a for e.g., *ranked* 75% – black line with triangle marker and *ranked* 50% – blue line with diamond marker). However, by further increasing the stratification factor, we reach a tipping point after which a further increase of the stratification factor results firstly in slower convergence rates before we again reach a state of no convergence at all (within e.g., *stratified* society).

Finding 2: The relation between the opinion dynamics and the stratification factor of a society in disassortative collaboration networks is intricate. Low values of stratification tend to favor consensus reaching – in such societies, consensus is always reached at a very fast convergence rate, which is higher than in *egalitarian* societies. However, if the stratification factor becomes too large, the consensus reaching process is hindered.

Inventory size evolution of assortative networks. Due to the large size of the co-authorship network, a much higher number of interactions is needed in order for all agent pairs to participate at least once in a meeting. For our experiments, we used 20 million interactions, but if the number of interactions is further increased the lines in Figure 3.4b will continue to drop towards 1. The co-authorship network is characterized with a positive assortativity coefficient that indicates that high status agents are, on average, connected to other high status agents, and low status agents are connected to other low status agents. The number of connections between low and high status agents is low, therefore, few meetings are taking place between these two classes. Consensus is reached almost independently from β (cf. Figure 3.4b), so our Probabilistic Meeting Rule does not benefit faster opinion convergence rates.

Finding 3: If a positive degree assortativity is evident in the network (e.g., co-authorship network), consensus is reached without external interventions.

Participation of agents in meetings across status groups. To further analyze these findings, let us investigate in more details the direction and intensity of opinions flow in our disassortative and assortative networks. To that end, we separate the agents into two classes: high (agents with the status above 90^{th} percentile) and low (agents below 90^{th} percentile) class. All reputation distributions are skewed to right and resemble a heterogenous distribution and the division into classes results in a reputation boundary of for example, 220 for English StackExchange network with all agents having reputation above 220 belonging to the high class and all agents below 220 belonging to the low class (for comparison the highest reputation score in English dataset is $105,678$). All other StackExchange networks are comparable to English and our analysis produces similar results. For that reason, we henceforth discuss only the English network as an example of our disassortative networks. In the case of our assortative network (i.e., co-authorship network), the highest reputation score is $15,758$ and the reputation boundary for the 90^{th} percentile is at 27, indicating that all low status agents have a reputation score below 27, while high status agents possess a reputation score above 27.

An important question is what happens when agents interact and how the Probabilistic Meeting Rule evaluates depending on the classes of agents participating in a meeting. In other words, we want to investigate the fraction of interactions that turn into a successful meeting (which consequently results in an opinion flow and increases the likelihood of two agents agreeing on a single word). We therefore classify each interaction according to the agent classes into four possible pairs: (i) low-to-low, (ii) low-to-high, (iii) high-to-low, (iv) and high-to-high where the first class corresponds to the speaker's class and the second corresponds to the listener class.

Figure 3.5 depicts the fractions of successful meetings among all interactions in the English StackExchange (3.5a) and co-authorship (3.5b) networks for three values of the stratification factor– *egalitarian* society (corresponds to $\beta = 0$), *ranked* society (up to 50% opinion flow is allowed between low and high status agents with optimal values $\beta = 0.0001$ for English and $\beta = 0.005$ for co-authorship network) and *stratified* society (e.g., $\beta = 1$ and $\beta = 5$). The only difference between plots in Figure 3.5a English (disassortative) and Figure 3.5b co-authorship (assortative) networks lies on the percentage of meetings taking place among low status agents and between low and high agents. As previously mentioned the number of physical connections between low and high agents in the co-authorship network is lower than in StackExchange networks and this results to the lower number of meetings taking place between these two classes. Since, in the co-authorship network agents belonging to the same classes tend to connect together, the number of meetings among low agents (low-to-low pairs) is much higher compared to StackExchange networks. The fraction of high status agents is equivalent for both networks, thus the number of meetings taking place between high status agents is almost the same.

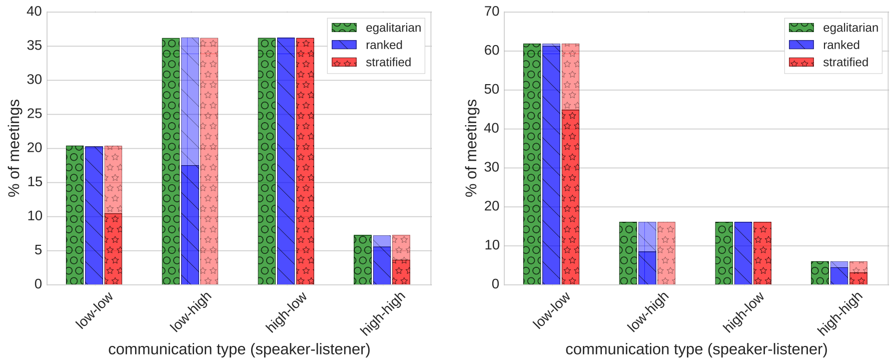
In the case of *stratified* society (red bars with star texture), opinions flow without restrictions only in high-to-low direction. Thus, the agents with a higher status can pass over their opinions to the agents with a lower status. The flow in the opposite direction is completely prohibited and therefore agents with a lower status cannot influence the opinions of the agents with a higher status. However, the Probabilistic Meeting Rule in this case is so strict and prohibitive that it greatly inhibits the opinion flow within the agents of the same status (i.e., high-to-high and low-to-low pairs). Because of the skewed nature of the reputation distributions, the inhibition in the low-to-low group (which is considerably larger than the high-high group) is more severe – the agents with a lower social status cannot efficiently exchange their opinions with each other and must rely on the agents with a higher social status to inject opinions into the low group by meeting each low agent separately. Since there are few high and many low status agents, consensus is never reached.

On the other hand, in the case of *egalitarian* society (green bars with circle texture), opinions flow without any restrictions in all directions. This results in the convergence of opinions and a rather fast convergence rate. However, the convergence rate is slightly slower as compared to the optimal case (*ranked* society). In our opinion, the explanation for this phenomenon lies in the dynamics of the low-to-high group meetings. Since everybody can impose her opinion onto everybody else, low status agents very often change the opinions of high status agents. Thus, low status agents increase the variance in the inventories of high status agents and they need additional meetings to eliminate these opinions. This results in slower convergence rates.

A particular dynamics of low-to-high meetings also explains faster convergence rates in *ranked* societies (blue bars with line texture). In this case, the opinion flow from the agents of low status to the agents of high status is strongly slowed down. Therefore, the disturbances in the opinions of high status agents are not substantial any more. On the other hand, as opposed to the *stratified* society, the opinion flow within the low-to-low group is not impaired at all. Thus, the injected opinions from the high status agents can be diffused among the low status agents themselves without need to address each low status agent separately. This, combined with the reduced disturbances flowing from low to high status agents, results in optimal opinion convergence rates.

Finding 4: The optimal convergence of opinions is achieved when low status agents can exchange their opinions among themselves without any restrictions. In addition, there must be a barrier that prohibits low status agents to inflict their opinions on high status agents so that disturbances in the opinions of high status agents are minimized.

Agents' final agreement. In order to gain insights into the final agreement of individuals we investigated each of the single opinions that agents agreed on. So, we modified the initialization of the agents' inventories to differentiate between opinions assigned to low and high status agents



(a) English StackExchange (disassortative)   (b) Co-authorship (assortative)

Figure 3.5: **Participation of agents in meetings across status groups.** The percentage of interactions resulting in meetings as a function of reputation classes in the English StackExchange (3.5a) and co-authorship network (3.5b). The high class comprises agents with the status above 90^{th} percentile and the low class all other agents. In the *stratified* society (red bars with star texture), a common opinion cannot be reached because the meeting rule is so strict that even communications between low agents (low-to-low pairs) are severely impaired. In the *egalitarian* society (green bars with circle texture), the convergence is slower because low status agents disturb high status agents by inflicting their opinion upon them (low-to-high pairs). In the *ranked* society (blue bars with line texture), the optimal convergence is achieved because low status agents can diffuse opinions among themselves (low-to-low pairs). At the same time, since the communications between low and high status agents are inhibited (low-to-high pairs), low status agents' opinions cannot disturb those of high status agents. The only difference between the plots in (3.5a) and (3.5b) lies on the percentage of meetings among low status agents and between low and high status agents. Since in the co-authorship (assortative) network, agents belonging to the similar classes tend to connect together, the number of meetings between low-to-low pairs is higher than the number of meetings between low-to-high and high-to-low pairs.

respectively. After rerunning the experiments and evaluating the results, we found out that for very low stratification factor (correspond to higher percentages of meetings taking place between low and high status agents, e.g., *egalitarian*, *ranked* 75% and *ranked* 50% in Figure 3.4) the final agreement of agents is mostly on the opinion of a low status agent, whereas for higher stratification factor (e.g., *ranked* 25% and *stratified*in Figure 3.4) the opinion on which all agents agreed on is usually one of a high status agent. This is in line with the fact that for very low stratification factor the intensity of the communication from low to high status agents is high, so the probability that an opinion of a low status agent is the final opinion on which all agents agreed on is high. By increasing beta we decrease the probability of a communication taking place between low and high agents. Thus, the final agreement is mostly on the opinion of a high status agent.

Finding 5: The final agreement of agents is mostly on the opinion of a low status agent, if the opinion flow from low to high agents is not disturbed at all, or if it is disturbed up to 50%. By further prohibiting the opinion flow from low to high status agents, the winning opinion, on which all agents agree on, is usually one of a high status agent.

3.2.6 Analysis of network correlations

In this section, we study how network structure and, in particular, the correlation of structure and status affect the process of consensus reaching in collaboration networks by constructing disassortative and assortative synthetic networks.

Decorrelating networks

Our aim is to study in detail how the network structure and, in particular, the correlation of structure and status affect the process of consensus reaching in our networks. Obviously, the connections between hubs and other nodes play a crucial role, as well as the distribution of degree sequence and the position of high reputation nodes in the network. For this study, we generated specific synthetic networks, whereas in each case, only one particular property of interest is preserved while others are eliminated. This way, in each experiment, we can assess the influence of a single property on the overall opinion dynamics process.

Degree and status correlation. In order to analyze the role of network structure and especially the role of the degree assortativity in the process of opinion spreading, we generate three synthetic networks based on the original collaboration networks introduced in Section 3.2.4. All synthetic networks have the same number of nodes n and edges m as the empirical networks but we modify the connections between nodes and the correlation between degree and reputation as follows:

- *Random network.* Here, we rewire the edges uniformly at random. This means that all nodes have equal probability of getting selected for creating an edge. The resulting network corresponds to the Erdős–Rényi model proposed in [Erdös and Rényi, 1959] and its node degree distribution follows a homogeneous Poisson distribution. With this network, we eliminate the degree sequence and the correlation with reputations.
- *Configuration model.* In this case, the edges from the original network are randomly rewired, but the degree sequence remains the same [Bender and Canfield, 1978; Molloy and Reed, 1995]. An uncorrelated rewiring minimizes the bias for connections in a network as all nodes are randomly rewired to different nodes than in the original network. Since the degree sequence is not modified, this results in a heterogeneous degree distribution with the same slope as in the original network. With this network, we eliminate the correlation between nodes over the edges, for example, we eliminate the correlations caused by the friendship relations.
- *Shuffled reputations.* Finally, we do not modify the network structure itself, but shuffle the reputation of nodes randomly. In the resulting network, the node degrees are decorrelated with reputations.

For all the experiments in the synthetic networks, we use as basis the English StackExchange and the co-authorship datasets and we follow the experimental setup described in Section 3.2.4.

Results of decorrelated networks

Our experimental results reveal some interesting insights. In Figure 3.6, we show the evolution of agent's inventory size during the interactions, averaged over 100 runs. To better understand the variation of the stochastic processes performed throughout our simulations, we calculated standard deviations over 100 runs per β, but for readability reasons, we removed error bars from the plots. Typical standard deviation values range between 0.48 (e.g., English shuffled reputations network) and 0.66 (English Erdős–Rényi network).

Disassortative networks. We recall the results of the original English network once more for an easier comparison with the results with synthetic networks. The simulation results with the English StackExchange original networks show that the *ranked* societies reach a common opinion with the highest convergence rate, higher than in *egalitarian* societies (e.g., *ranked* 50% compared to *egalitarian* in Figure 3.4), whereas in a *stratified* society consensus is not reached at all.

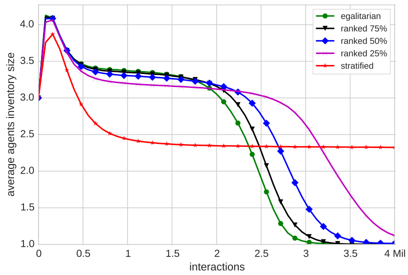
The simulation results for the English Erdős–Rényi network differ from the original network (see Figure 3.6a). Except for *stratified* society, for which consensus is not reached within the limit of interactions, for other societies, the process of consensus reaching is slowed down. The fastest convergence is achieved with $\beta = 0$, respectively in *egalitarian* societies. This result shows that the convergence rate is highly dependent on the existence of hubs in a network. In an Erdős–Rényi network, the high status agents are not hubs any more since their degrees are much smaller and therefore, they can not spread their opinions to low status agents as quickly as in the original network.

We find a further evidence for this behavior in the English configuration model in which the calculated stratification factors and the evolution of agent's inventory size are identical to the original network, thus, the figure presenting the results is not included. In this example, we keep the same degree sequence but rewire the edges in the English StackExchange network. Since we now keep the hubs and the degree-status correlation, we do not disturb the consensus reaching process. We simply reconnect the low degree/status agents to different high degree/status agents. This result also shows that additional external correlations such as friendship/-collaboration correlations do not influence the consensus reaching. Mainly, it is the degree/status correlation that provides support for achieving the consensus.

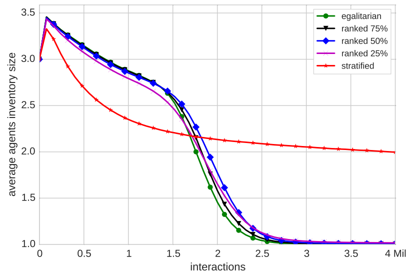
In the English network with shuffled reputations (Figure 3.6b), the estimated stratification factors that define the five societies are identical to the English Erdős–Rényi network. However, agents agree to a common opinion almost independently from the society form, except for the *stratified* society. The convergence rate is faster than in English Erdős–Rényi network. This outcome indicates that in networks with heterogenous degree distribution and uncorrelated reputations of users, consensus is reached automatically without need for external interventions. Since, however, in most of empirical collaboration networks degree strongly correlates with user reputation we need another mechanism that can positively influence opinion dynamics. That mechanism includes controlling the communication between low and high status nodes through the stratification factor.

Assortative networks. Figure 3.6c shows the simulation results of the co-authorship Erdős–Rényi network, in which the hubs are removed. The calculated β differ from the empirical co-authorship network and the consensus reaching process is slowed down in this case. This outcome confirms once more that the presence of hubs is crucial for the consensus reaching process.

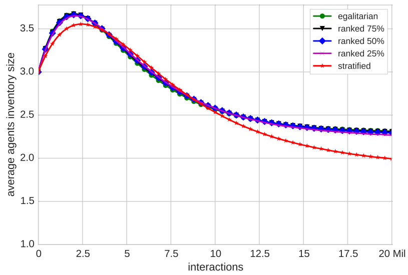
Applying the configuration model to the co-authorship network while keeping the degree sequence changes the connection patterns between nodes. So, rewiring the edges reduces the number of high-to-high and low-to-low connections, simultaneously increasing the number of high-to-low links. This results in a decreased assortativity. In fact, in the configuration model, we measure the assortativity coefficient of 0.0001, whereas in the original co-authorship network that factor is 0.15. This can be seen also in Figure 3.6d, where the opinion convergence rates are slowed down.



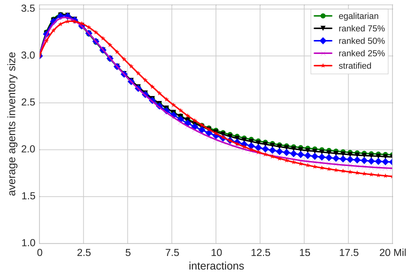
(a) English Erdős–Rényi



(b) English shuffled reputations



(c) Co-authorship Erdős–Rényi



(d) Co-authorship configuration model

Figure 3.6: **Decorrelating networks.** Mean values of the agent's inventory size in relation to the number of interactions for English Erdős–Rényi (3.6a), English shuffled reputations (3.6b), co-authorship Erdős–Rényi (3.6c) and co-authorship configuration model (3.6d) networks. The process of consensus building varies among networks. In the English Erdős–Rényi network, the process of consensus reaching is slowed down, whereas in the English shuffled reputations, the opinion convergence rate is faster (agents agree to a common opinion almost independently from β). In the English configuration model opinions converge with the highest rates in the case of *ranked* societies (e.g., *ranked* 50%), which corresponds to the English original network, thus the plot is omitted. In the co-authorship Erdős–Rényi and configuration model, the consensus building process is slowed down compared to the co-authorship original network. The simulation results of the co-authorship shuffled reputations network are identical with the original co-authorship network, consequently it is not included in the figure.

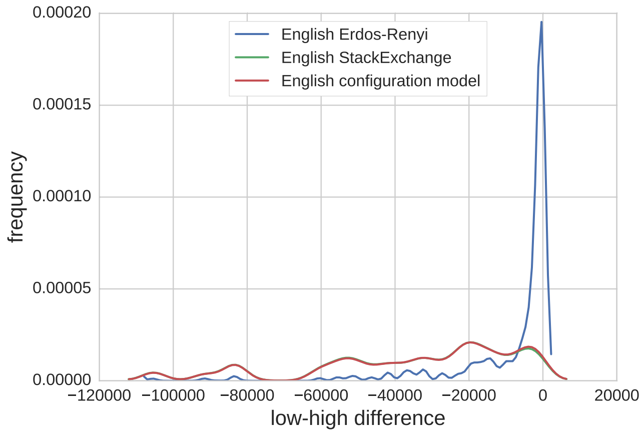
Shuffling the reputations in the co-authorship network does not impact the simulation results as they are identical with the empirical co-authorship network. Thus, the respective plot is omitted from Figure 3.6.

Distribution of status differences. To further quantify our findings, we investigated the distribution of status differences between two connected nodes in our networks. The differences are calculated for two neighboring nodes if one of the nodes is a low and the other one is a high status node (defined by the 90^{th} percentile). The results for disassortative and assortative networks are depicted in Figure 3.7.

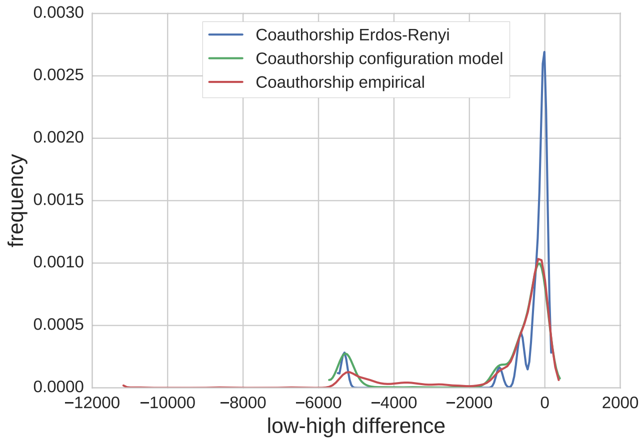
In the networks with a heterogenous degree distribution, a negative degree assortativity and a strong correlation between degree and status (red and green lines in Figure 3.7a), there are many connections from low to high status nodes and therefore we frequently observe high negative differences. In other words, there are many potential meetings between low and high agents that given that they take place often can disturb the high status agents and consequently the consensus reaching process. Thus, to reduce the number of meetings that take place we need to apply a mechanism such as our Probabilistic Meeting Rule and inhibit the opinion flow in the low-to-high direction.

In the case of the English Erdős–Rényi network (blue line in Figure 3.7a), there are lower differences between low and high status agents (the majority of differences is close to 0), due to the lower number of connections between these two groups of agents. Thus, not many of the meetings that take place are high-to-low agent meetings and additionally with our Probabilistic Meeting Rule , we are also prohibiting the opinion flow from low to low status agents. Consequently, this slows down the consensus reaching process.

In the English shuffled reputation network, the number of connections between low and high status agents is the same as in the original network, but the differences between agents' statuses are lower (with only one peak close to 0, thus, it is omitted in Figure 3.7a), which speeds up the consensus reaching even without external interventions such as Probabilistic Meeting Rule.



(a) English StackExchange (disassortative)



(b) Co-authorship (assortative)

Figure 3.7: **Kernel density estimation of the distribution of status differences between low and high agents.** Disassortative networks are shown in (a) and assortative networks in (b). The distribution of agents' status differences in the English StackExchange and configuration model networks in (a) are almost identical, thus the blue and the red lines overlap. Due to many connections from low to high status agents, we frequently see high negative differences. In the English Erdős–Rényi network (blue line), the majority of differences between low and high status agents is close to 0, because of the lower number of connections between these two groups of agents. The English shuffled reputations network is not shown in the plot, because of very low status differences with only one peak around 0. In (b) are shown lower differences between agents' statuses in the co-authorship empirical network and synthetic networks.

In Figure 3.7b, it is shown that, in general, there are lower differences between agents' statuses in the co-authorship empirical network and synthetic networks derived from it, which explains the fact that in co-authorship original network consensus is reached fast and independent from β. The opinion convergence rates are slowed down only if the presence of hubs is lower or if the degree assortativity is decreased.

Finding 6: A common opinion is adopted in collaboration networks with heterogenous degree distribution. Hubs are key to reaching consensus since they can distribute a single common opinion to a high number of other nodes. If degree and status are not correlated or if a positive degree assortativity is evident in the network (co-authorship network), consensus is reached quickly and without external interventions. In disassortative networks, where degree strongly correlates with status (StackExchange empirical networks), this correlation slows down the convergence rate, making it necessary to take actions such as applying the Probabilistic Meeting Rule to insert a social barrier between low and high status agents.

Table 3.2 summarizes the results of our experiments both on empirical and decorrelated networks.

Table 3.2: **Summary of our findings.** This table summarizes the results of our work.

Network	Type	*egalitarian*	*ranked* 75%	*ranked* 50%	*ranked* 25%	*stratified*
Disassortative	Empirical	converge	converge	fastest convergence	no converge	no converge
English StackExch.	Erdős–Rényi	fastest convergence	slowed down	slowed down	slowed down	no converge
	Configuration model	converge	converge	fastest convergence	no converge	no converge
	Shuffled reputations	converge	converge	converge	converge	no converge
Assortative	Empirical	converge	converge	converge	converge	converge
Co-Authorship	Erdős–Rényi	slowed down	slowed down	slowed down	slowed down	slowed down
	Configuration model	slowed down	slowed down	slowed down	slowed down	slowed down
	Shuffled reputations	converge	converge	converge	converge	converge

3.2.7 Related Work

At present, we identify three main lines of research related to our work: opinion dynamics, social status theory and naming game.

Opinion dynamics

Opinion dynamics is a process characterized with a group of individuals reaching a consensus (i.e., the majority of a group share the same opinion). In opinion dynamics, the focus is on modeling the opinion state of an individual in particular and a population in general. Opinion dynamics has been tackled in the past in the context of statistical physics [Castellano et al., 2009; Iniguez et al., 2014]. As discussed in [Castellano et al., 2009], if opinion dynamics is viewed from a perspective of statistical physics, an individual is analogous to a particle with properties that may or may not change over a period of time. Thus, the social process of interaction among individuals can be designed as a mathematical model that represents a change in the local and global state of an individual and a group. One of the examples of such a process is the Naming Game model, a variant of which we are using in our work, that models how individuals behave during a meeting and exchange their opinions. In our experiments, the meeting process is further enhanced by taking reputation scores of individuals into account. Constraining the system to favor high reputation nodes resulted in reaching consensus later as compared to an unconstrained model.

In a different context opinion dynamics is studied in [Blondel et al., 2010; Hegselmann et al., 2002; Krause, 2011; Lorenz, 2007; Muller, 2006], where an opinion is represented as a real number and a classical approach of individual opinion formation involves averaging over opinions of other agents in the system. In such a setting a consensus is considered to be reached if all the agents in the system agree to the same value of opinion. The process of opinion dynamics is studied in [Krause, 2011] both from the local and global perspective. They defined the opinion formation process as local when a user takes into account only the nearest neighbors, whereas in the case of global opinion formation the user takes into account all other agents in the network. The process of opinion formation is studied in [Blondel et al., 2010] by means of a continuous time multi-agent system.

In their work they proved that opinion converge to a set of clusters, where agents in each of the cluster share a common value. The work of Lorenz [2007] studied a continuous model of opinion dynamics under bounded condition. The bounded condition restricts users to interact with their peers only if they are close to each other. Such a process of opinion dynamics leads to formation of clusters with characteristic location and size patterns. They found the drifting phenomenon in composition of cluster in case of heterogeneous bounds. In [Muller, 2006] authors studied the process of internal organisation within communities of practice and how such a process leads to some members obtaining a leadership status. They developed a model to depict the self-organising process and found that leaders are the members who correspond to higher level of activity in the community.

Social status theory

Research on how the position and status of a node influence a network is mostly carried out in the context of network exchange theory [Markovsky et al., 1993; Walker et al., 2000; Willer, 1999]. This theory states that connections and a position in a network lead to a power condition that is based on how the nodes are connected and which position they take in the network [Walker et al., 2000]. For example, in [Markovsky et al., 1993], researchers differentiate between weak and strong powers network in terms of node positions and network properties. The authors give a theoretical extension to the network exchange theory to explain why in sparsely connected networks a stronger power effect is observed than in densely connected networks. They found that in densely connected networks, weak position nodes have an advantage since they have a higher connectivity, which enables them to short circuit the structural advantages of strong position nodes. This is related to our work, as we concentrate on investigating how the reputation of a node in a network affects the spread of opinion that leads to establishing consensus in the network. Also, we define various classes of nodes based on reputation and determined how their interaction affects their overall process of consensus building.

Naming Game

The Naming Game has been introduced in the context of linguistics [Dall'Asta et al., 2006b] and the emergence of a shared vocabulary among agents [Baronchelli et al., 2006] with the aim to demonstrate how autonomous agents can achieve a global agreement through pair-wise communications without central coordination [Zhang et al., 2014]. With that regard, we present a selection of variations of the Naming Game that are relevant to our work.

Similarly to our approach, the work in [Brigatti, 2008] describes a variation of the Naming Game that incorporates the agents' reputation scores. In the beginning, reputation is randomly distributed (Gaussian distribution) among the agents. Successful communication increases the agents' reputation and during each iteration, the agent with a higher reputation score acts as a teacher and the one with the lower score as a learner. The main difference from our work is that in [Brigatti, 2008], they use synthetic data for the simulations and that the assigned reputation scores are random numbers that change during iterations. In our work, we employed empirical collaboration networks from StackExchange with reputation scores that were assigned by the community. As opposed to the work of Brigatti [2008] where there is an open-ended game with unlimited number of words, the inventory of our agents consists of predefined sets of three opinions.

Other examples for the Naming Game variations include the work of Liu et al. [2011] who studied the impact of spatial structures (e.g., geographical distances) have on meetings between individuals in a network, and Yang et al. [2008], who proposed a Naming Game that follows an asymmetric negotiation strategy and investigated the influence of hub effects on the agreement dynamics with specific focus on how quickly consensus could be achieved. Each agent in the network is assigned a weight defined by the agent's degree and a tunable parameter α. During iteration, two nodes are randomly selected and based on their degree and the configuration of the parameter α, they are either the speaker or the listener (i.e., if $\alpha > 0$, high degree agents have more chances to be speakers and vice versa). This way, the dynamics of the game can be investigated in light of the varying influence of high degree agents. Our work is somewhat related as we also use a parameterized probability function to define the probability of a meeting taking place between two nodes, in our case depending on their reputation score. The main difference to our work is that agents' selection is unbiased and empirical data with explicitly provided reputation scores are used.

The diffusion of opinions across networks and the potential of reaching consensus are strongly influenced by the availability of communities and, specifically, by the presence of strong community boundaries [Lu et al., 2009]. To investigate this effect, authors in [Lu et al., 2009] assigned a group of nodes in a network as a committed fraction, that is, nodes that are not influenced by other nodes in a network and don't ever change their opinion. In our dataset, however, no strong community structures are present.

3.2.8 Conclusion and Future Work

Understanding opinion dynamics and how consensus is reached in social networks has been an open and complex challenge in our community for years. In this work, we addressed a sub-problem related to this challenge by investigating a specific case of collaboration networks in which individual nodes have a certain social status.

To that end, we presented an extension (Probabilistic Meeting Rule) to the standard Naming Game model of opinion dynamics. We evaluated our approach on six large empirical collaboration networks, as well as on three specifically created synthetic networks, which reflected the characteristics of the empiric networks. In this work, we provided a computational approach for the general estimation of the stratification factorof our Probabilistic Meeting Rule and we analyzed the role network structure plays in the process of consensus building. These studies constitute the methodological contribution of our work to the field of opinion dynamics. Additionally, we investigated various real-world scenarios such as the emergence and disappearance of social classes in collaboration networks. From the empirical point of view, our investigations revealed insights about the influence of social status on the diffusion of opinions. Our main finding indicates that social status strongly influences the opinion dynamics in a complex and intricate way. More specifically, weakly stratified societies reach consensus at the highest convergence rate, whereas completely stratified societies do not reach consensus at all. The most important issue in this process is related to low status agents and how their communication is controlled. In particular, the optimal convergence is achieved when (i) low status agents are allowed to freely exchange opinions between themselves (since this reduces the need for high status agents to interact with low status agents) and (ii) simultaneously there is a communication barrier reducing the number of interactions of low status agents towards high status agents (since this reduces the variance in opinions of high status agents). Furthermore, our investigations on the role of the network structure reveal that hubs are in general crucial to reach consensus, since they can spread a single common opinion to a high number of nodes. In assortative networks, in which connections between low and high agents are very rare, external interventions do not benefit faster convergence rates. A similar situation is observed in disassortative networks when degree is not correlated with a user's status. If there is a strong correlation between status and degree in a disassortative network, this slows down the convergence rate, making it necessary to take actions such as applying the Probabilistic Meeting Rule to disturb the communication between low and high status users.

Limitations. In our opinion, our work has the following limitations. Firstly, we represent social status with a single number – for certain scenarios this representation may be too simplistic. For example, people often play different roles in social networks and a non-simple interplay between the roles and status may exist. Secondly, a more finely grained classification of agents into various groups (e.g., low, mid and high groups or even finer divisions) may shed more light on the opinion dynamics. Finally, in our work, we consider only static snapshots of networks and reputation scores. However, not only opinions but also networks are dynamic, as new agents may arrive to the network, new edges may form and inactive edges may disappear from the network. Moreover, reputation itself is very dynamic and depends on the agent's activity and the current perception of an agent by her peers.

Future work. In our future work, we plan to address some of the limitations of our current work and extend our approach and experiments to other scenarios. For example, one interesting avenue for further research are the networks with a strong community structure. As communities tend to slow down the consensus reaching process, it would be interesting to investigate how status and/or network structure can be adjusted to support the process. Apart from social status, the influence of trust is of utmost importance in various social systems and in particular in social media. Thus, adapting the presented approach to analyzing how trust relates to opinion dynamics is another promising research direction for the future.

3.2.9 Acknowledgments

This work is supported by the Know-Center, the EU funded project Learning Layers (Grant Agreement 318209) and the Austrian Science Fund (FWF) research project "Navigability of Decentralized Information Networks" (P 24866-N15). The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Ministry of Transport, Innovation and Technology, the Austrian Ministry of Economics and Labor and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency (FFG).

3.3 Consensus Dynamics in Online Collaboration Systems

This article addresses the second research question of this thesis, that is, what is the role of users similarity in correlation with users social status in consensus building. In detail, this article investigates how the process of consensus building is affected by configurable influences of user similarity, user social status and a complex interplay between those two factors.

The presented approach enables us to capture user interactions that are not restricted to the edges of the observed interaction network (i.e., user interactions that do not necessary leave traces in system logs). This article applies the Naming Game model and extends it to reflect (i) latent similarities between users and (ii) observed social status of users in online collaboration networks. Based on these two factors, we provide a model which determines the likelihood of a future interaction between any two given users. With this model in place, we run experiments on 17 collaboration networks extracted from Reddit and StackExchange sites.

Our results reveal that an increase in the influence of user similarity delays the consensus building process (i.e., when users are guided by their similarity to communicate with other users). This delay can be compensated by a suitable increase in the influence of user social status. The results implicate that user similarity and user social status exhibit opposing forces and their reciprocal influence should be carefully balanced to ensure a faster consensus.

3.3.1 Abstract

In this paper, we study the process of opinion dynamics and consensus building in online collaboration systems, in which users interact with each other following their common interests and their social profiles. Specifically, we are interested in how users similarity and their social status in the community, as well as the interplay of those two factors influence the process of consensus dynamics. For our study, we simulate the diffusion of opinions in collaboration systems using the well-known Naming Game model, which we extend by incorporating an interaction mechanism based on user similarity and user social status. We conduct our experiments on collaborative datasets extracted from the Web. Our findings reveal that when users are guided by their similarity to other users, the process of consensus building in online collaboration systems is delayed. A suitable increase of influence of user social status on their actions can in turn facilitate this process. In summary, our results suggest that achieving an optimal consensus building process in collaboration systems requires an appropriate balance between those two factors.

3.3.2 Background

In this work, we study opinion dynamics and consensus building in online collaboration systems. In collaboration systems such as online encyclopediae, question & answering (Q&A) sites or discussion forums, users engage in complex interactions with others' to reach a common goal, such as to write an article or to answer a difficult question. Often, this is a long-lasting iterative process, in which users share their knowledge and opinions, discuss problems and solutions, write and edit joint articles, or vote on each other's contributions. Ideally, this process converges to a shared common result. However, many times, consensus cannot be reached and a given topic or question remains unresolved within the community.

Understanding the factors, which govern a consensus building process in online collaboration systems, as well as mechanisms that may turn such a process into a success or failure is one of the pressing questions that our research community has already recognized. In many studies, researchers analyzed the underlying dynamics of opinion formation to identify key factors that contribute to consensus building (see, e.g., [Castellano et al., 2009] for an overview). Such studies may be seen as a first step towards a more ambitious goal of developing tools that promote consensus building processes in online communities. For example, connecting otherwise non-interacting users by recommendations may lead to discussions resolving issues that hinder consensus.

To study consensus building processes, researchers frequently apply agent-based models. In an agent-based model, opinions of individual agents are represented as states and agents interact with each other following a set of predefined interaction rules. Through such interactions, agents alter their states until some criteria are met or the system reaches an equilibrium state. The interactions between agents give rise to a particular behavior of the whole population. The Naming Game model [Baronchelli et al., 2006] is among the most prominent agent-based models for studying opinion dynamics and consensus building in groups of interacting agents. Often, such studies simulate opinion dynamics on synthetic networks, see for example [Brigatti, 2008; Dall'Asta et al., 2006a; Li et al., 2013; Liu et al., 2011; Lu et al., 2009; Gao et al., 2013; Waagen et al., 2015; Wang et al., 2007].

In one of our previous works [Hasani-Mavriqi et al., 2016], we studied the influence of social status on consensus building in online collaboration systems. In that study, we assumed that the underlying network of previous interactions determines the constraints on the possible future interactions. In other words, only users who have already interacted with each other in the past were allowed to interact in the future. For example, user interactions on Reddit include users writing comments or voting on postings of other users. Such interactions allow us to extract user interaction networks from the system logs. In such networks, users are nodes and two users are connected by a link if they interacted in the past. However, in real-world online collaboration systems, there are certain user actions and interactions, which leave no or inconclusive traces in the system logs. For example, when users on Reddit simply read submissions but never leave comments or votes, in general we do not know which particular comments and postings these users actually have read.

Moreover, many real-world datasets contain inaccuracies and are therefore inherently uncertain [Martin et al., 2016].

In this paper, we set out to study consensus building by adopting a model of interacting agents, whose future interactions are not restricted to the edges of the observed interaction network. Rather, we allow interactions between all pairs of users with varying preferences. In particular, we apply the Naming Game model and extend it to reflect (i) latent similarities between users and (ii) observed social status of users in real-world systems. Technically, with those two factors, we parametrize a probability distribution over pairs of users, which determines the likelihood of a future interaction between any two given users. This results in a flexible approach that enables us to explore and analyze various interesting and realistic configurations as opposed to restricting interactions to the edges of the observed network, which fixes the interaction probabilities to zero for previously non-interacting users.

To that end, we investigate consensus building within different society forms, which we characterize according to user similarity into *open*, *modular* and *closed* societies and according to social status into *egalitarian*, *ranked* and *stratified* societies. *Open* and *closed* societies represent two extreme cases based on the influence of user similarity: in an open society, any pairs of users can interact and exchange opinions with each other regardless of their similarity, whereas in a closed society only highly similar users interact with each other. Between these two society forms we define a *modular* society, in which probability of users interaction is proportional to their similarity. Similarly, *egalitarian* and *stratified* societies represent two extreme cases governed by configuring the influence of social status: in an egalitarian society, the influence of social status is neglected, indicating that users can interact and exchange opinions with each other regardless of their social status, whereas in a stratified society, opinions can flow only from users with a higher social status to those with a lower social status. Between these two extreme cases, we can model different situations (*ranked societies*) by tuning the influence of social status so that opinions are very likely to flow from individuals with a higher social status to those with a lower social status, but with small probability they can also flow into the other direction.

For our experiments, we extract 17 collaboration networks from the real-world systems Reddit and StackExchange. For each of these networks, we first determine user similarity and their social status. We determine user similarity by calculating their regular equivalence [Newman, 2010]. With regular equivalence we capture global user similarities between non-interacting users as opposed to local similarity measures, which take into account only the immediate neighbors of a node. To determine social status of users we use the built-in scoring schemes of Reddit and StackExchange. With these networks in place, we simulate opinion spreading among users to study how the process of consensus building is governed by configurable influences of user similarity, user social status and a complex interplay between those two factors.

The contributions of our work are twofold. First, we extend the Naming Game model with an interaction mechanism that is based on user similarities and their social status. With this extension we conduct experiments on empirical collaboration networks and contribute in this way to the limited line of research on opinion dynamics in empirical networks. Second, our experimental results reveal interesting and non-trivial findings, namely, that user similarity and user social status are opposing forces with respect to consensus building. Whereas user social status may speed up the emergence of consensus, user similarities typically hinder that process. Thus, for an efficient consensus building the negative effect of similarity needs to be carefully compensated by the positive effect of social status.

3.3.3 Related Work

At present, we identify three main lines of research related to our work: (i) social impact theory, (ii) works that study the interplay between user similarity and social status and its impact on user behavior in online systems, and (iii) opinion dynamics in interaction networks.

Social impact theory

In the field of social psychology, the social impact theory of Latané [Latané, 1981] attempts to explain how individuals are influenced by their social environments. According to it, the social impact felt by individuals can be explained in terms of social forces, to which they are exposed [Jackson, 1987; Pettijohn, 1998]. Latané [1981] compares these social forces to physical forces, such as electromagnetic forces or forces that govern the transmission of light, sound and gravity [Jackson, 1987]. In this analogy, social forces felt by individuals are moderated by the strength, immediacy and number of other people present in their social environment. In relation to our work, the influence of users social status in our experiments refers to the strength of the impact of other people (e.g., their authority or power of persuasion), whereas the user similarity is analogous to the immediacy of the others (e.g., their closeness in space or time) [Nowak et al., 1990]. Mathematically, the social impact felt by an individual, known also as a target, is a multiplicative function of the three features of a source person and is given in the following form: Impact $= f(S \cdot I \cdot N)$, where $Impact$ is the social impact on the target person and S, I, and N, are the strength, immediacy and number of the source persons, respectively [Latané, 1981; Jackson, 1987]. The social impact function constitutes the theoretical basis for our agent-based model and its multiplicative effects.

Connecting the social impact theory with agent-based modeling has been also the aim of previous research [Nowak et al., 1990], in which researchers applied computer simulations to examine the extent to which group-level phenomena are driven by individual-level processes. In synthetic datasets that represent sets of individuals, they studied the attitude change of individuals and group polarization with respect to binary opinion states. Similarly, in our work we apply agent-based modeling. However, we perform experiments on empirical datasets from online collaboration systems and consider more than two opinion states.

Recent work followed a theory-driven approach to conduct empirical analysis of Twitter data that supported the assumptions of the social impact theory [Garcia et al., 2017]. In our work, however, we study the process of opinion dynamics in online collaboration systems, by applying a data-driven model as well as by simulating how opinions spread in those systems.

Cultural dynamics in society classes and their role in the adaption of fashion are the focus of the work of the sociologist Georg Simmel [Simmel, 1957]. According to Simmel's theory the latest fashion is defined by the higher society classes and the lower ones imitate and copy the fashion from them. As soon as this happens, higher classes move from the current fashion and adopt a new style to differentiate them from the masses. Similarly, in our analysis, we define higher and lower social status classes and analyze the opinion flow between them. The effect of lower status agents inflicting opinions to the higher ones, observed in our experiments, is analogues to the phenomenon of imitation, whereas the effect of limiting the communication from low-status agents to high-status agents reflects the phenomenon of differentiation.

The work presented in [Pedone and Conte, 2001] applies an agent-based model to simulate the effects of Simmel's theory by exploring its spatial dimension. While the authors use synthetic data and synthetic agents' social statuses, we use empirical datasets from Reddit and StackExchange and apply the empirical reputation scores provided by both systems as a proxy for social status.

Research on how the position and social status of a node influence the network originates from network exchange theory [Markovsky et al., 1993; Walker et al., 2000; Willer, 1999]. Similarly, we study how the social status of a node in an interaction network affects the spread of opinion that leads to consensus building. Additionally, in our work we define classes of nodes based on the social status and determine how their interaction affects the process of consensus building.

The influence of the interplay between user similarity and social status on user behavior in online systems

In our previous work Hasani-Mavriqi et al. [2016], we studied the impact of social status on opinion dynamics and consensus building in online collaboration systems. In contrast, in the present work, we study how latent user similarity and the interplay between the user similarity and user social status impact the process of consensus building.

In Papadopoulos et al. [2012] the authors present a framework for link prediction in evolving networks and show that popularity is just one dimension of attractiveness, in the context of link creation, and another important dimension is the similarity between users. In other words, user similarity and user popularity are two main forces that drive people to form links in various networks. In our work, we also study the effect of user similarity and user social status, but in relation to dynamical processes that take place in online collaboration systems.

User similarity in online social networks has also been studied in Akcora et al. [2013]. Here, the authors present a method for evaluating social networks according to network connections and profile attributes. In Anderson et al. [2012], the effect of similarity (in terms of user characteristics) and social status, as well as their interplay is studied on online evaluations carried out among users. They found that when two users are similar social status plays less of a role when users evaluate each other. Major difference to our work is that the authors calculate user similarity as cosine similarity between user action vectors. User actions are, for example, editing an article on Wikipedia, asking or answering a question on a Q&A site or rating a review on Epinions. In our work, we calculate user similarities by applying the regular equivalence that captures latent similarities even between non-interacting users and users who do not share common actions. Similar work to Anderson et al. [2012] is described in Leskovec et al. [2010], with the difference that the authors consider only the relative social status between two users (i.e., their comparative levels of status in the group) when studying how users evaluate each other. The authors found that users with comparable status hesitate to give positive evaluations to each other.

Opinion dynamics in interaction networks

Research on opinion dynamics in interaction networks builds upon insights from the field of statistical physics Castellano et al. [2009]; Iniguez et al. [2014]. In this field, social processes of interaction among individuals are modeled mathematically by representing how changes in the local and global state of an individual and a group take place. A well-known model following this approach, the Naming Game, has been introduced in the context of linguistics Baronchelli et al. [2006]; Dall'Asta et al. [2006b] with the aim to demonstrate how autonomous agents can achieve a global agreement through pairwise communications without central coordination Zhang et al. [2014].

Recent research Waagen et al. [2015]; Maity et al. [2013] applies the mean-field principle while using the Naming Game model for their experiments. For example, the work in Maity et al. [2013] studies the impact of learning and the resistance towards learning (as two opposing factors) on consensus building among a population of agents. In Waagen et al. [2015], the authors consider the case of an arbitrary number of agent opinions and the presence of zealots in the Naming Game. They provide a methodology to numerically calculate critical points in two special cases: the case in which there exist zealots of only one type and the case in which there are an equal number of zealots for each opinion. Similarly to our approach, the work of Brigatti et al. Brigatti [2008] describes a variation of the Naming Game that incorporates the agents' social status scores. In the beginning, social status is randomly distributed among the agents via a Gaussian distribution. Successful communication increases the agents' social status and during each iteration, the agent with the higher social status acts as a teacher and the one with the lower status as a learner. In contrast to our work, the dynamic social status scores are synthetically created, whereas we adopt empirical status scores.

3.3.4 Methodology

We base our model on the Naming Game Baronchelli et al. [2006]; Dall'Asta et al. [2006a]; Baronchelli, Andrea and Dall'Asta, Luca and Barrat, Alain and Loreto, Vittorio [2006]; Baronchelli et al. [2005]; L. Dall'Asta, A. Baronchelli, A. Barrat, and V. Loreto [2006]. The Naming Game is an agent-based model, in which agents are represented as nodes in a network. Agents interact with each other by following a set of predefined rules, with the aim of giving a name to a single unknown object. Consensus is reached when all agents agree on a single name for the object.

Each agent possesses an inventory of names or words (i.e., opinions), which is initially empty. At each interaction step, two agents are randomly chosen to meet (i.e., to communicate), where one of them is designated the role of the speaker while the other one is the listener. If the speaker's inventory is empty, a word is invented and it is communicated to the listener, or otherwise the speaker selects randomly a word from her inventory and communicates it to the listener. If the communicated word is unknown to the listener (i.e., it does not exist in the listener's inventory), the listener adds this word to her inventory. Contrarily, if the communicated word is known to the listener, both speaker and listener agree on that word and drop all other words from their inventories.

In this work, we extend the Naming Game with an interaction mechanism that accounts for latent user similarities and social status. In Papadopoulos et al. [2012], the authors have identified user similarity and user popularity as two main forces that drive people to form links in various networks. User similarity is a property of pairs of users, whereas social status is a property of individual users. In general, in collaboration systems, users tend to connect with similar users, i.e., with those sharing similar interests, tastes or social backgrounds, as well as with users of a higher social status or a higher popularity Scholz [2010].

Regular equivalence

To calculate the user similarity, we apply similarity measures from graph theory and social network analysis. In these fields, there are two main types of similarity: (i) *structural similarity*, and (ii) *regular equivalence* Newman [2010]. In particular, two nodes in a network are structurally similar if they share many common neighbors. On the other hand, two nodes are regularly equivalent if they have common neighbors that are themselves similar even if they do not share the same neighbors. Thus, regular equivalence quantifies not only observable but also latent similarities.

With regular equivalence, the basic idea is to define a similarity score σ_{ij} between nodes i and j, such that i and j are similar if i has a neighbor k

that is similar to j Newman [2010]:

$$\sigma_{ij} = \alpha \sum_k A_{ik}\sigma_{kj} + \delta_{ij}, \tag{3.4}$$

where α is a constant known as damping factor, A_{ik} are elements of the adjacency matrix $\mathbf{A}$ (with $A_{ij} > 0$ if i and j are connected by an edge with a positive weight and $A_{ij} = 0$ otherwise), σ_{kj} is the similarity score between k and j, and δ_{ij} is the Kronecker delta function, which is 1 for $i = j$ and 0 otherwise. The damping factor α should satisfy $\alpha < 1/\kappa_1$ in order for similarity scores to converge, where κ_1 is the largest eigenvalue of the adjacency matrix. The recursive calculation of the regular equivalence counts paths of all lengths between pairs of nodes. It assigns high similarity values to nodes that either share many common neighbors or to nodes that are connected by many longer paths, or both. By choosing α closer to $1/\kappa_1$, we assign more weight to longer paths, whereas smaller α values prefer shorter paths. Since we want to capture as much of latent similarities as possible, we set $\alpha = 0.9/\kappa_1$.

The formula for similarity scores tends to give higher similarity to high degree nodes due to their many neighbors Newman [2010]. A standard approach to remedy this situation is to normalize the scores by dividing them with the node degree.

Furthermore, we once more normalize the similarity values by subtracting for each user the minimum similarity of her direct neighbors. This lets us take into account the social adaptation of individual agents to the local norms induced by their neighbors Sayama and Sinatra [2015]. As a result, we have positive similarity values only for the direct neighbors, as well as for all other users that are more similar than the direct neighbors (see Figure 3.8 for an example of regular equivalence calculation).

Probabilistic meeting rule

Algorithm 1 describes the procedure of our extension of the Naming Game. In particular, we modify the meeting rule between two agents and replace it with our Probabilistic Meeting Rule, which defines the probability of a meeting taking place:

$$p_{sl} = \underbrace{\min(1, e^{\gamma\sigma_{sl}})}_{\text{similarity}} \cdot \underbrace{\min(1, e^{\beta(s_s - s_l)})}_{\text{social status}}. \tag{3.5}$$

Here, σ_{sl} is the similarity score between speaker s and listener l, s_s is the speaker's social status, s_l is the listener's social status, $\gamma \geq 0$ is the *closeness factor* and $\beta \geq 0$ is the *stratification factor*. Note that those two factors are the controlling parameters in our model.

The Probabilistic Meeting Rule is a flexible rule that enables us to model various scenarios and society forms. The first term in the equation $(\min(1, e^{\gamma\sigma_{sl}}))$ controls the degree of openness of a society. It evaluates to 1 for $\gamma = 0$. We refer to this scenario as *open* society, in which any pair of users (mean field approach) can interact independently of how similar are they to each other. In other words, in an open society, the similarity between users does not play a role and everybody can interact with everyone else. Open society thus reflects the Solaria world introduced by Watts Watts [1999, 2004].

By increasing γ, the influence of the user similarity becomes stronger indicating a so called *modular* society (i.e., the first term in the Probabilistic Meeting Rule takes on a value between 0 and 1). In this scenario, highly similar users interact with each other with a high probability, whereas less similar users still interact with each other but with a smaller probability than highly similar users. By further increasing the closeness factor we arrive at a *closed* society, in which users interact only with other highly similar users and the interaction probability between less similar users evaluates to a value close to 0. This scenario is analogous to the Watts' caveman world, in which users who live in "caves" (i.e., closed communities) interact with each other but they never or rarely interact with users from other "caves" Watts [1999, 2004].

Similarly, with the stratification factor, we can configure the level of influence of the users social status on the probabilities of their interactions. In particular, if the speaker's social status is higher than the listener's social status, the second term $(\min(1, e^{\beta(s_s - s_l)}))$ in Eq. 3.5 takes the value of 1. This means that a meeting between a speaker with a social

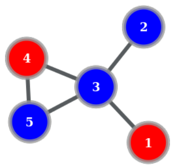
Algorithm 1: The Naming Game model with Probabilistic Meeting Rule.

variables:
σ_{sl} - similarity score between speaker and listener;
s_s - speaker's social status;
s_l - listener's social status;
γ - closeness factor;
β - stratification factor;
p_{sl} - probability of a meeting between speaker s and listener l;
x - random number $\in [0, 1]$;
input: initial inventory, interaction network, γ, β, σ s ;
for *each iteration* **do**
  randomly select two users;
  randomly select a speaker;
  calculate $p_{sl} = \min(1, e^{\gamma\sigma_{sl}}) \cdot \min(1, e^{\beta(s_s - s_l)})$;
  if $p_{sl} > x$ **then**
    speaker selects randomly one word from her inventory;
    speaker communicates the selected word to the listener;
    if *the word is not in the listener's inventory* **then**
      uptake: listener adds the word to her inventory;
    end
    else
      agreement: both speaker and listener keep only the communicated word and delete all others;
    end
  end
end

status higher than the listener's always takes place. When the listener has a higher social status than the speaker, several scenarios are possible, depending on the value of the stratification factor.

For example, for $\beta = 0$, the second term evaluates to 1 and we call this scenario *egalitarian* society. In an egalitarian society, everyone can talk to everyone else independently of their social status. If we increase the stratification factor, the second term starts to decay and in general, takes a value between 0 and 1. We refer to this situation as a *ranked* society, in which opinions always flow from individuals with a higher social status to those with a lower social status. Flow into the other direction is also

Inter. Network



$$s_1 = \mathbf{1} \quad s_2 = 2 \quad s_3 = 3 \quad s_4 = \mathbf{4} \quad s_5 = 5$$

$$\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 1 & \mathbf{0} & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ \mathbf{0} & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Similarity

$$\boldsymbol{\sigma} = \begin{pmatrix} -2.1 & -3.1 & 0 & \mathbf{-2.3} & -2.3 \\ -3.1 & -2.1 & 0 & -2.3 & -2.3 \\ 0 & 0 & 3.4 & 0.8 & 0.8 \\ \mathbf{-1.2} & -1.2 & 1.6 & 0.7 & 0 \\ -1.2 & -1.2 & 1.6 & 0 & 0.7 \end{pmatrix} \qquad \boldsymbol{P_\sigma} = \begin{pmatrix} 0 & 0.21 & 1 & \mathbf{0.31} & 0.31 \\ 0.21 & 0 & 1 & 0.31 & 0.31 \\ 1 & 1 & 0 & 1 & 1 \\ \mathbf{0.54} & 0.54 & 1 & 0 & 1 \\ 0.54 & 0.54 & 1 & 1 & 0 \end{pmatrix}$$

Social Status

$$\boldsymbol{S} = \begin{pmatrix} 0 & -1 & -2 & \mathbf{-3} & -4 \\ 1 & 0 & -1 & -2 & -3 \\ 2 & 1 & 0 & -1 & -2 \\ \mathbf{3} & 2 & 1 & 0 & -1 \\ 4 & 3 & 2 & 1 & 0 \end{pmatrix} \qquad \boldsymbol{P_S} = \begin{pmatrix} 0 & 0.61 & 0.37 & \mathbf{0.22} & 0.14 \\ 1 & 0 & 0.61 & 0.37 & 0.22 \\ 1 & 1 & 0 & 0.61 & 0.37 \\ \mathbf{1} & 1 & 1 & 0 & 0.61 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Figure 3.8: **Probabilistic Meeting Rule - illustrative example.** *Top row:* we depict an interaction network with five users, the social status of users (s_1 to s_5) and the adjacency matrix $\boldsymbol{A}$. All edge weights in $\boldsymbol{A}$ are 1, indicating that the corresponding users interacted only once with each other in the past. If we restrict meetings to the edges of the interaction network, the meeting probabilities are symmetric and equal to the entries of $\boldsymbol{A}$. Thus, the users 1 and 4 cannot participate in a meeting since $p_{14} = p_{41} = 0$ (elements marked in red in $\boldsymbol{A}$). The average meeting probability p_m corresponds to the network density and evaluates to 0.5. *Middle row:* we calculate the regular equivalence matrix $\boldsymbol{\sigma}$ and normalize it with the degrees and the minimal neighbor similarity (normalization results in asymmetric similarities). We set closeness factor $\gamma = 1/2$ (modular society) and calculate the matrix of meeting probabilities $\boldsymbol{P_\sigma}$ (we set zeros on the diagonal since each meeting requires two users). The rows correspond to the meeting probabilities of a user acting as the speaker. We observe now non-zero probabilities between users who are not connected by an edge. For example, for users 1 and 4 (cf. red-marked elements in $\boldsymbol{P_\sigma}$), the meeting probability is $p_{14} = 0.31$ (user 1 acts as the speaker) and $p_{41} = 0.54$ (user 4 acts as the speaker). In this setting, the average meeting probability is significantly higher than previously $p_m = 0.69$. *Bottom row:* the matrix $\boldsymbol{S}$ keeps the (asymmetric) social status differences between users. Again, the rows correspond to users acting as the speaker in a meeting. We set stratification factor $\beta = 1/2$ (ranked society) and calculate the matrix of the meeting probabilities $\boldsymbol{P_S}$. The social status mechanism results in non-zero probabilities between all pairs of users. For example, for users 1 and 4 (cf. red-marked elements in $\boldsymbol{P_S}$), the meeting probability is $p_{14} = 0.22$ (user 1 is the speaker) and $p_{41} = 1$ (user 4 is the speaker). The average meeting probability for this configuration is $p_m = 0.71$. Finally, if similarity as well as social status rules apply, the final meeting probabilities are calculated by element-wise multiplication of $\boldsymbol{P_\sigma}$ and $\boldsymbol{P_S}$.

possible, however only with small probability. By further increasing β, we reach a situation where the second term always evaluates to a value close to 0 if the speaker's social status is smaller than the listener's social status. In other words, we have reached what we call a *stratified* society where meetings take place only if the speaker's social status is higher than the listener's social status but never in the opposite case. Thus, with varying configurations of both terms, we can explore nine different combinations of the above-mentioned scenarios.

In Figure 3.8, we show an illustrative example for the calculation of the meeting probabilities for a modular, ranked society. In general we observe two effects of our approach: (i) the meeting probabilities increase as compared to a model which restricts interaction to the edges of the interaction network, and (ii) the meeting probabilities are asymmetric.

3.3.5 Datasets and Experiments

In our experiments, we use 17 empirical datasets from Reddit and StackExchange by selecting them randomly to ensure a broad coverage of different topics.

Extracting interaction networks

In Reddit, registered users post new submissions (typically links or texts), comment and discuss existing submissions, or create new communities (so-called subreddits), which revolve around a specific topic. For our experiments, we parsed the dumps of 16 different subreddits from the year 2014, which belong to four main categories[6]: *Movies* (*Documentaries*, *True film*, *Movie details* and *Harry Potter*), *Politics* (*Political discussion*, *Political humor*, *Neutral politics* and *World politics*), *Programming* (*Julia*, *Python*, *Ruby* and *Compsci*) and *Sports* (*Skiing*, *Tennis*, *Badminton* and *Volleyball*). To construct the Reddit interaction network, we extract the

[6] https://www.reddit.com/r/ListOfSubreddits/wiki/listofsubreddits/

users contributions from the submission[7] and from the comment[8] dumps. We then create an interaction network, in which users are represented as nodes and two users are connected by an edge if one user commented on the submission of another one, or if they both commented on the same submission of a third user. For each edge, we set a weight, which corresponds to the number of interactions between two given users.

StackExchange[9] is a Q&A site, where users collaboratively solve problems through asking and answering questions in posts. Similarly to the Reddit networks, we construct the StackExchange interaction networks to represent co-posting activities. Specifically, two nodes (i.e., users) are connected via a weighted edge if the users contributed to the same question. Correspondingly, the edge weight encodes the number of common contributions. We use the following StackExchange editions covering different topics for our experiments: *English, Cooking, Academia, Movies, Politics, Music, German, Japanese, History, Chinese, Spanish, French, Sports*[10].

Finally, in all networks, we extract the largest connected component and perform all experiments on that component. We give the basic statistics of our empirical datasets such as the number of nodes (n) and edges (m), as well as average node degree (d), average social status (s), average edge weight (e) and density (ρ) in Table 3.3. The network density ρ calculated as $2m/(n(n-1))$ is defined as the fraction of all possible edges that are present in a network. In the context of our model, density can be interpreted as an average meeting probability if meetings are restricted only to the edges of the network. In other words, the probability that a randomly selected pair of users participates in a joint meeting equals, on average, to the network density. In practice, the majority of social and other networks such as interaction networks are extremely sparse networks with densities that lay way beyond 1%. Thus, our empirical interaction networks indeed constitute a very rigid constraint on any possible interactions.

[7] https://www.reddit.com/r/datasets/comments/3mg812/full_reddit_submission_corpus_now_available_2006/
[8] https://www.reddit.com/r/datasets/comments/3bxlg7/i_have_every_publicly_available_reddit_comment/
[9] https://stackexchange.com/
[10] https://archive.org/details/stackexchange

Determining social status

To determine the social status scores for users, we exploit the built-in user rewarding system of Reddit and StackExchange. In Reddit, users can accumulate so-called "karma" scores that rise if their posts receive good ratings from other users. Thus, karma scores represent the reflection of the user "vibes" in the community and we apply it as a proxy for social status. Since karma scores are not included in the publicly available Reddit dumps, we crawled those scores using the public API[11] and the python-based PRAW API wrapper[12]. On the other hand, in StackExchange users are rewarded by the community with reputation scores for providing not only valuable answers but also valuable questions. As shown in Movshovitz-Attias et al. [2013], the scores given by this user-rewarding system highly correlate with the quality of the user content and thus, we assume that a high-reputation user contributes with a high-quality content to the

Table 3.3: **Dataset characteristics.** This table shows the number of nodes (n), number of edges (m), average node degree (d), average social status (s), average edge weight (e) and density (ρ) of our networks.

Dataset	Type	n	m	d	s	e	ρ
Reddit	Movies	38,006	138,907	7.3	6	1.1	0.00019
	Politics	25,946	92,285	7.1	8.2	1.2	0.00027
	Programming	23,074	70,232	6.1	5.4	1.1	0.0003
	Sports	18,441	97,073	10.5	10.2	1.2	0.00057
StackExchange	English	30,656	192,983	12.5	199	1.7	0.0004
	Cooking	9,637	40,437	8.4	175	1.6	0.0009
	Academia	5,098	26,805	10.5	312	1.7	0.002
	Movies	4,425	13,952	6.3	194	1.8	0.0014
	Politics	4,349	21,428	9.8	229	2	0.002
	Music	3,699	15,750	8.5	213	1.7	0.0023
	German	2,316	12,825	11	285	2.2	0.0048
	Japanese	2,069	11,155	10.7	328	2.5	0.0052
	History	2,054	11,048	10.7	271	2.2	0.0052
	Chinese	1,985	8,556	8.6	160	1.8	0.0043
	Spanish	1,584	6,908	8.7	196	1.9	0.0055
	French	1,478	6,668	9	298	2.1	0.0061
	Sports	1,276	3,513	5.5	178	1.8	0.0043

[11] https://www.reddit.com/dev/api/

[12] https://praw.readthedocs.io/en/stable/

community. Reputation scores are provided in the dataset dumps and we use them as a proxy for social status in StachExchange systems. This setup allows us to investigate the effect of social status from two points of view: in Reddit, the social status is a reflection of how other persons experience a given user in the society (i.e., charisma) and in StackExchange, social status is more related to a position that users earn in a society based on the quality of their work (i.e., reputation).

Experiments

Our experiments consist of six steps. First, for each interaction network, we construct a weighted adjacency matrix $\mathbf{A}$ by setting A_{ij} to the edge weight between users i and j, if they are connected or to 0 otherwise. Second, we compute the matrix of similarity scores using the methodology described in Methodology section.

Third, we compute the closeness factor γ and the stratification factor β by estimating the expected meeting probability in our networks using a standard Monte Carlo method Hasani-Mavriqi et al. [2016]. This enables us to control the communication intensity between users. For the closeness factor, we determine two parameter values to depict modular and close societies by controlling the percentages of successful meetings induced by the first similarity factor in our multiplicative Probabilistic Meeting Rule: (i) for the modular society, we determine γ such that approximately 75% of all possible meetings (up to the statistical fluctuations) take place, (ii) for the closed society, we determine γ for which approximately half of all meetings are successful on average. In addition, for the open society, in which all meetings take place, we set $\gamma = 0$.

Average meeting probability of 50% is 2 orders of magnitude higher than the average network density of our empirical interaction networks (0.27%) (cf. Table 3.3). Thus, even though our model biases the user interactions towards more similar users, it is substantially less restrictive than an alternative model solely based on the interaction network. Another (simpler) alternative to avoid the restrictions of the interaction networks would be to, for example, allow for each second interaction to take place between arbitrary pairs of (non-adjacent) users. Nevertheless, this approach would miss the possibility to induce similarity or social status biases.

Similarly to the closeness factor, we also estimate two values for the stratification factor β that correspond to the ranked and stratified society forms. Here, we control the opinion flow from low to high social status users and set β such that on average, 50% of meetings take place (ranked society) and so that none of the meetings takes place (stratified society). Again, we only control the second social status factor in the multiplicative meeting rule. In addition, by setting $\beta = 0$ we achieve the egalitarian society, in which all meetings take place. Note that we define high social status users as users with a social status above the 90^{th} percentile, whereas low social status users have a social status below the 90^{th} percentile.

Fourth, we initialize agents inventories by randomly selecting three words from a set of unique words for each agent. Fifth, we create a set of meetings, i.e., randomly selected pairs of users. From this set, we generate meeting sequences by picking meetings at random without repetition for each possible combination of closeness factor and stratification factor. This ensures that the random factor due to the meeting sequence remains insignificant for various values of γ and β. We determine the length of the meeting sequence (c) (i.e., maximum number of user interactions) based on the number of users in a given dataset. The length of the meeting sequence c is two orders of magnitude higher than the number of users n. For each configuration, we simulate the meetings 100 times and report the averaged simulation results.

Finally, we store the state of the agents network for each $c/100$ interaction of our 100 runs as well as for all values of closeness factor and stratification factor. This enables us to investigate the distinct number of overall opinions adopted by each agent during the interactions. Additionally, we can derive the percentages of agents that have reached consensus on a particular opinion.

Source code. To ensure the reproducibility of our experiments, we provide our simulation framework as an open source project. The source code can be downloaded from our Git repository[13].

3.3.6 Results and Discussion

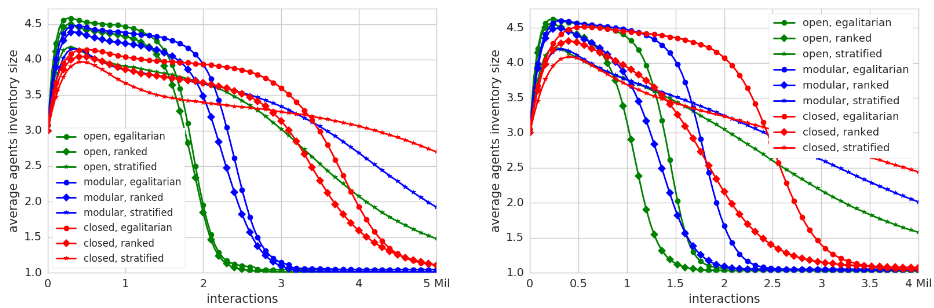
The influence of user similarity and social status on consensus dynamics

We show our simulation results in Figure 3.9. The plots in Figures 3.9a and 3.9b depict the evolution of the agents inventory mean size (over 100 runs) as a function of the simulation progress for the Reddit Movies and StackExchange English datasets, respectively. All other empirical datasets exhibit comparable results, so we omit them from Figure 3.9; but we provide them in Appendix 3.3.9 in Figure 3.12. Each line in the plots corresponds to the results obtained using one particular closeness factor and stratification factor. Line colors depict different values of closeness factor, whereas line markers illustrate values of stratification factor.

Due to our Probabilistic Meeting Rule, whenever we set one of the factors to 0, we can study the impact of the other factor on the process of consensus building. Thus, by analyzing society forms with $\beta = 0$ (*egalitarian*) and varying closeness factor, we can investigate the effect of user similarity on the consensus building process. Our results reveal that in *(modular, egalitarian)* and *(closed, egalitarian)* societies (cf. blue and red lines with circle markers in Figure 3.9) consensus is slowed down as compared to *(open, egalitarian)*, which represents a society where all meetings take place. Thus, as soon as user similarity starts to exhibit influence on the meeting probabilities the consensus building process is delayed. Although the average meeting probability in *modular* society forms is still very high, even this slight preference towards meeting with more similar users is able to slow down the spread of opinions.

On the other hand, a weak increase in the influence of the user social status is beneficial for the consensus. In *(modular, ranked)* and *(closed, ranked)* societies (cf. blue and red lines with diamond markers in Figure 3.9) we

[13] https://git.know-center.tugraz.at/summary/?r=SocialNetworkAnalysis.git



(a) Reddit Movies (b) StackExchange English

Figure 3.9: **The influence of user similarity and social status on consensus dynamics.** The plots show the mean size (100 runs) of the agents inventories (y-axes) in relation to the number of interactions (x-axes) for Reddit Movies (left) and Stack-Exchange English (right) datasets. Each line represents results for one particular γ and β. The line colors represent three values of γ and line markers three different values of β. We notice that in (*modular, egalitarian*) and (*closed, egalitarian*) societies (marked with blue and red lines with circle markers), opinion convergence rates are slower than in (*open, egalitarian*) society marked with green and circle markers. This indicates that as soon as user similarity plays a role, consensus building is delayed. However, in (*modular, ranked*) and (*closed, ranked*) societies (blue and red lines with diamond markers) we observe faster consensus building. This means that by increasing the effect of social status, we are able to partially compensate the negative effect of similarity. By further increasing the impact of the social status through the stratification factor, the positive effect of social status dissolves. This is visible in the green, blue and red lines with star markers corresponding to (*open, stratified*), (*modular, stratified*) and (*closed, stratified*) societies. Thus, for a faster consensus building, a careful balancing between the influence of similarity and social status is needed.

observe faster consensus building. Thus, by increasing the effect of social status, we can compensate the initial negative effect of similarity.

Nevertheless, the positive effect of social status diminishes quickly. In *(modular, stratified)* and *(closed, stratified)* societies (cf. blue and red lines with star markers in Figure 3.9) the convergence rate again slows down. Thus, an initially positive effect of social status in ranked society forms undergoes a phase transition towards a negative effect in stratified societies.

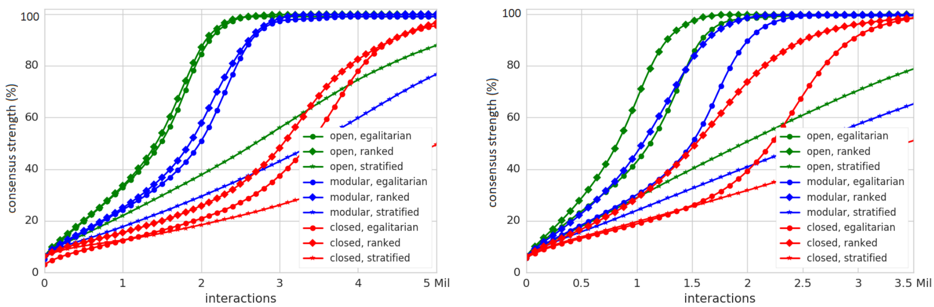
Findings. Our simulation results indicate that user similarity and social status exhibit opposing forces with respect to consensus building in online collaborative systems. While an increase in the influence of user similarity has a negative effect, the social status exhibits both the phase of a positive effect as well as the phase of a negative effect. Consequently, an optimal configuration for a faster consensus requires a careful balance between those two factors.

Coarse analysis

We consider the average inventory size of agents equalling 1 as a first criterion for reached consensus among agents (cf. Figure 3.9). Further, we aim to determine the distinct number of opinions present in the agents network and the consensus strength during the interactions. We define the consensus strength as percentages of agents having one single opinion in their inventories over the course of simulations. The consensus strength reaches its maximum when all agents unanimously agree on one particular opinion.

Figure 3.10 shows consensus strength (averaged over 100 runs) over the number of interactions for the Reddit Movies (Figure 3.10a) and Stack-Exchange English (Figure 3.10b) datasets. Analogous to Figure 3.9, each line represents results for one particular γ and β. The line colors represent three values of γ and line markers three different values of β.

For almost all societies except for *(open, stratified)*, *(modular, stratified)* and *(closed, stratified)* (cf. green, blue and red lines with star markers in Figure 3.10), there is a saturation of the consensus strength visible in the plots. The growth curves resemble logistic growth curves with the phases of quick initial growth and a saturation phase as the process reaches its equilibrium. The growth rates of the consensus strength lines determine how quickly agents reach consensus. The growth rates are faster for *(open, ranked)*, *(modular, ranked)* and *(closed, ranked)* (cf. green, blue and red



(a) Reddit Movies (b) StackExchange English

Figure 3.10: **Coarse analysis.** Percentages of consensus strength (averaged over 100 runs) reached among agents on one particular opinion (y-axes) are shown as a function of the number of interactions (x-axes) for different values of γ and β. The line colors illustrate three different values of γ and line markers three values of β. The plot in the left illustrates Reddit Movies results, whereas the plot in the right presents the results of StackExchange English dataset. Each line represents results for one particular configuration of γ and β. We consider that the consensus strength reaches its maximum when all agents unanimously agree on one particular opinion. With each interaction, agents exchange opinions and the consensus strength increases, but with different growth rates for different configurations of γ and β. The growth rates of the consensus strength lines determine how quickly agents reach consensus. A saturation of consensus strength lines is visible for almost all society forms except for *(open, stratified)*, *(modular, stratified)* and *(closed, stratified)* (cf. green, blue and red lines with star markers). The growth rates are faster for *(open, ranked)*, *(modular, ranked)* and *(closed, ranked)* (cf. green, blue and red lines with diamond markers) compared to *(open, egalitarian)*, *(modular, egalitarian)* and *(closed, egalitarian)* societies (cf. green, blue and red lines with circle markers). These results complement our previous findings presented in Fig. 3.9 and reveal that the appropriate balance between user similarity and social status enables faster consensus strength growth rates in online collaborative systems.

lines with diamond markers) compared to *(open, egalitarian)*, *(modular, egalitarian)* and *(closed, egalitarian)* societies (cf. green, blue and red lines with circle markers). These results complement our findings presented in the previous subsection, namely, with the increase of the influence of user similarity on the meeting probabilities, consensus building among agents is delayed. This negative effect is compensated to some extent with the increase of the influence of social status (*ranked societies*). A further increase of the influence of social status yet hinders consensus building, which means that no saturation state can be observed in case of *stratified* societies (at least not in the number of interactions that we simulate).

Findings. Our coarse analysis reveals that the optimal balance between user similarity and social status enables faster growth rates towards consensus building in our datasets.

Communication intensity between social classes

Now we are interested in identifying causes of these observed effects. For this, we investigate the communication intensity (i.e., the number of successful meetings) across user social classes that we introduced earlier, namely high social status class with users above the 90^{th} percentile and low social status class with all other users.

In our previous study Hasani-Mavriqi et al. [2016], we found that the direction of opinion flow impacts how fast opinions converge. Specifically, the flow from low social status to high social status users, as well as from low social status users to low social status users, is crucial. As described in Leskovec et al. [2007], high social status users are typically able to impose their opinions to other users in a system. Thus, whenever the opinions of these high social status users frequently change the system as a whole experiences oscillatory behavior and cannot reach consensus. Due to the heterogeneous distributions of user social status in many systems, the number of low social status users is substantially higher than the number of high social status users. Therefore, whenever the communication intensity in the direction from low social status users to high social status users is high, low social status users are able to cause oscillations in the opinions of high social status users and the consensus building process is delayed.

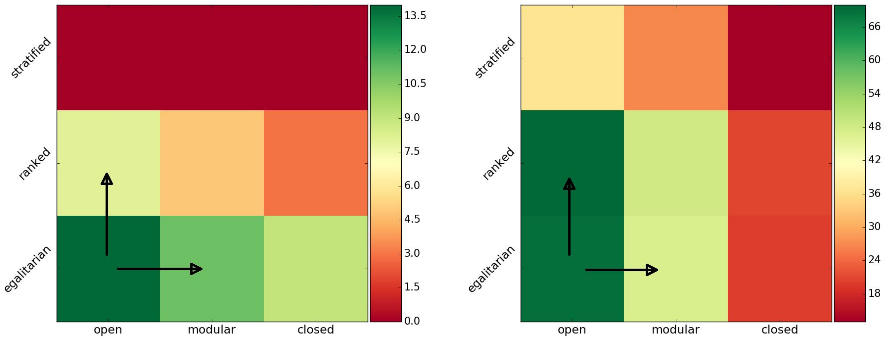
On the other hand, it is important that communication direction from low social status users to other low social status users remains unhindered. Due to the high number of low social status users, they have to be able to intensely communicate among themselves to spread opinions. Low social status users cannot rely on a small number of high social status users to reach many low social status users and distribute opinions. In other words, the process of consensus building among low social status users is a two phase process. First, high social status users impose their opinions onto a small fraction of low social status users, and second, those opinions are subsequently spread among low social status users themselves.

These mechanisms can potentially explain the results of our experiments. For example, due to their numerous previous interactions high social status users are on average more similar to other users than low social status users. Therefore, whenever user similarity is the driving force behind meetings taking place we expect users with high social status to participate in a large number of meetings.

On the other hand, the number of low social status users is high and the second meeting participant is very likely a low social status user. Thus, our expectation is that we will observe many meetings with one high social status and one low social status user. This increases the communication intensity between low and high social status users and this leads to increased opinion fluctuations for high social status users. This in turn can slow down the consensus building process.

To further investigate this hypothesis, we analyze the percentages of users interactions that turn into successful meetings after applying our Probabilistic Meeting Rule. Specifically, we analyze two important communication directions and their intensities: (i) low-to-high and (ii) low-to-low, where the first term refers to the speaker and the second to the listener.

In Figure 3.11 we show a heatmap with communication intensities between social classes for all nine combinations of society forms for the StackExchange English dataset. Again, here we only present the results for this dataset, since in all other datasets we obtain comparable results; we provide them in Appendix 3.3.9 in Figures 3.14 and 3.15. The heatmap in Figure 3.11a depicts the percentages of successful meetings in the low-



(a) Low-to-High

(b) Low-to-Low

Figure 3.11: **Heatmaps of the communication intensity in low-to-high (3.11a) and low-to-low (3.11b) social status classes of users for the StackExchange English dataset.** The columns represent three society forms based on similarity: open, modular and closed, whereas rows show three social status society forms: egalitarian, ranked and stratified. The colors depict the intensity of the communication between users (i.e., percentages of the successful meetings taking place). In the plot in Figure 3.11a, we notice that the communication intensity from low- to high-status users is decreased by increasing either the influence of user similarity (switch from open to modular society) or the social status (switch from egalitarian to ranked society). In Figure 3.11b, we see that by switching from an open to a modular society the communication intensity from low- to low-status users is decreased. But, for optimal consensus building, the communication in this class of users should not be disturbed. When we switch from an egalitarian to a ranked society, the intensity of the communication between users in the low-to-low class remains unchanged. This is one of the factors that in the ranked societies we observe fast opinion convergence rates. To summarize, through the increase in similarity the communication channel from low status users to other low status users is being closed and this slows down the consensus building process.

to-high class of users, whereas the heatmap in Figure 3.11b depicts the percentages of successful meetings taking place in the low-to-low class.

Columns of the heatmaps show the society forms based on similarity (i.e., open, modular and closed) and rows show the social status society forms (i.e., egalitarian, ranked and stratified).

The communication intensity from low to high social status users (cf. Figure 3.11a) is decreased when either the influence of user similarity (switch from open to modular society) or social status (switch from egalitarian to ranked society) is increased. In the ranked society, we observe a slightly higher reduction in the opinion flow from low to high social status users than in the modular society. Thus, even though high social status users are on average more similar to other users, increase in the influence of similarity reduces the opinion flow from low social status to high social status users. Since this is a desired behavior there seems to be some other mechanism causing the delay in the opinions convergence.

Therefore, we turn our attention now on the low-to-low communication direction. By switching from an open to a modular society, we observe a decreasing opinion flow from low to low social status users (cf. Figure 3.11b). However, for optimal consensus building, the communication in this class of users should not be disturbed. On the other hand, when we switch from an egalitarian to a ranked society, the intensity of the communication between users in the low-to-low class remains unchanged and we observe fast convergence rates. Thus, through the increase in similarity the communication channel from low social status users to other low social status users is being closed and this slows down the consensus building process. Similar behavior can be also observed for the social status when we switch from ranked to stratified society form. Thus, a balanced influence of social status improves convergence rates, whereas even a low influence of similarity hinders the process.

Findings. Our analysis indicates that the increased influence of similarity reduces the communication intensity between both low social status users and high social status users, as well as between low social status users and other low social status users. While the former has a positive effect on the spreading of opinions the latter hinders that process and causes the delay in consensus. Meetings governed by similarity are locally contained to smaller groups of users and the communication between different users groups is less intensive.

3.3.7 Conclusion and Future Work

In this paper, we studied the process of opinion dynamics and consensus building in online collaboration systems. Specifically, we adopted a model of interacting agents, in which we allow interactions between all pairs of users with varying preferences beyond the observed interaction network. To that end, we presented an extension to the Naming Game model, i.e., the Probabilistic Meeting Rule that reflects (i) latent similarities between users and (ii) observed social status of users in real-world systems. We conducted our experiments on 17 empirical datasets from Reddit and StackExchange.

Our experimental results revealed that user similarity and social status exhibit opposing forces with respect to consensus building in online collaborative systems. Our main finding indicates that while an increase in the influence of user similarity has a negative effect, social status exhibits both the phase of a positive effect as well as the phase of a negative effect. Consequently, for a faster consensus, a careful balance between those two factors is required.

To explain our results, we further investigated the communication intensity (i.e., the number of successful meetings) between the social classes we defined. Our findings showed that the increased influence of similarity reduces the communication intensity between both low-status users and high-status users, as well as between low-status users and other low-status users. While the former has a positive effect on the spreading of opinions the latter hinders that process and causes the delay in consensus.

Limitations. In our opinion, our work has the following limitations. First, we neglected any dynamic changes of user similarity and social status and the networks as such. In reality, social networks constantly change as users may leave the system while others join. We could gain more realistic insights by comparing results of dataset snapshots between different points in time. Second, we used a simplification for opinions of users exchanged in online collaboration networks by presenting them as a set of numbers. An alternative would be to use the real content exchanged among users.

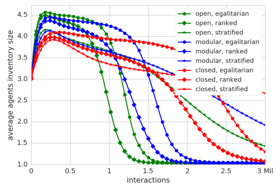
Future work. For future work, we plan to use our insights to design personalized user recommendation algorithms. Thus, by identifying the factors that lead to barriers and conflicts in collaborations, we plan to design meaningful interventions by suggesting possible collaborators with the goal to create network structures, in which consensus building is supported (i.e., recommending experts or high social status users as possible collaborators with the goal to speed up the process of consensus building). We also plan to verify our findings in other types of empirical networks, for example, gathered from the collaborative editing system Wikipedia, where we will investigate the dynamics of the editing process.

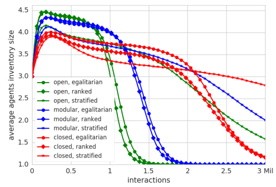
3.3.8 Acknowledgments

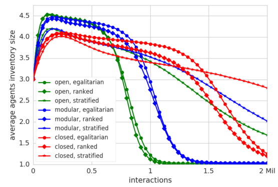
This work is supported by the Know-Center Graz and the AFEL project funded from the European Union's Horizon 2020 research and innovation programme under grant agreement No 687916. The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Ministry of Transport, Innovation and Technology, the Austrian Ministry of Economics and Labor and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency (FFG).

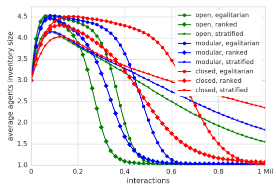
3.3.9 Appendix

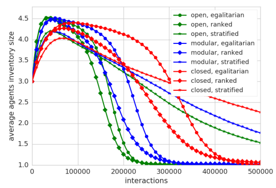
See Figures 3.12, 3.13, 3.14, 3.15.

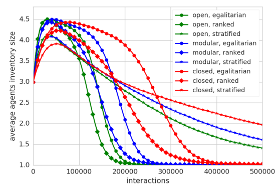

(a) Reddit Politics

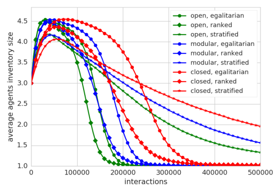

(b) Reddit Programming

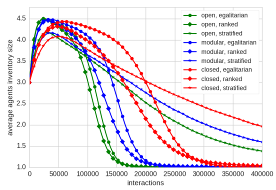

(c) Reddit Sports

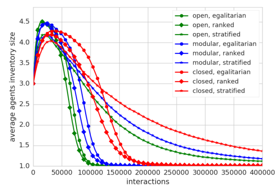

(d) StackExchange Cooking

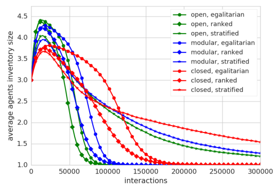

(e) StackExchange Academia

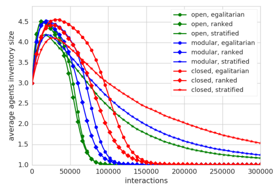

(f) StackExchange Movies

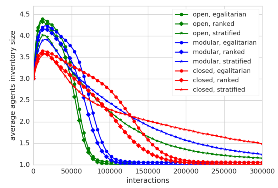

(g) StackExchange Politics

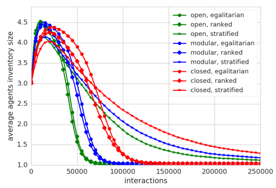

(h) StackExchange Music

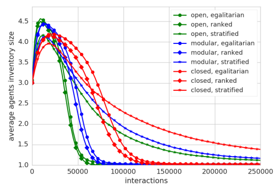

(i) StackExchange German


(j) StackExchange Japanese


(k) StackExchange History


(l) StackExchange Chinese


(m) StackExchange Spanish


(n) Stackexchange French

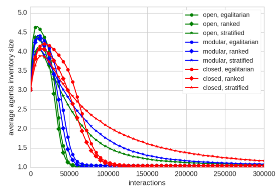
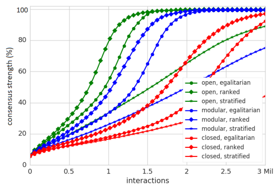
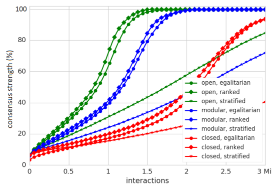

(o) StackExchange Sports

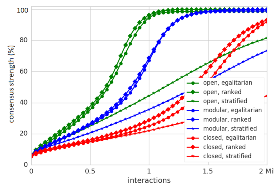
Figure 3.12: **The influence of user similarity and social status on consensus dynamics.** The plots show the mean size (100 runs) of the agents' inventories (y-axes) in relation to the number of interactions (x-axes) for all datasets from Table 3.3 not included in Figure 3.9. The simulation results are similar to the ones presented in Figure 3.9.



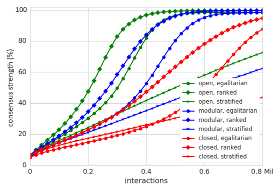
(a) Reddit Politics



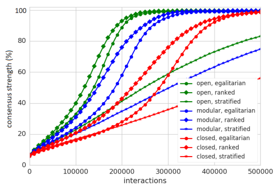
(b) Reddit Programming



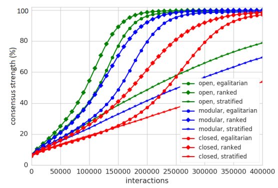
(c) Reddit Sports



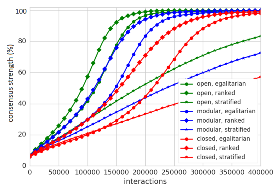
(d) StackExchange Cooking



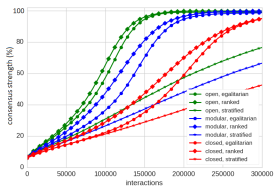
(e) StackExchange Academia



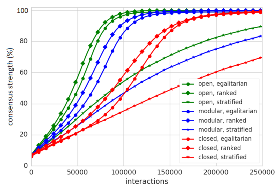
(f) StackExchange Movies



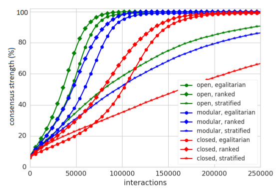
(g) StackExchange Politics



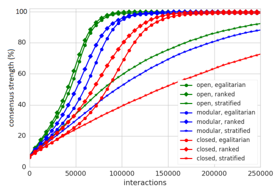
(h) StackExchange Music



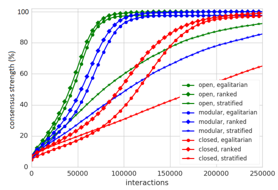
(i) StackExchange German



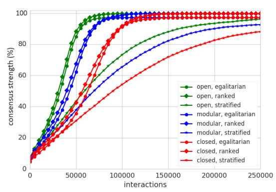
(j) StackExchange Japanese



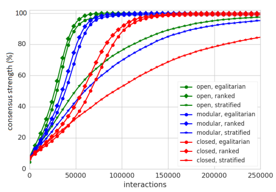
(k) StackExchange History



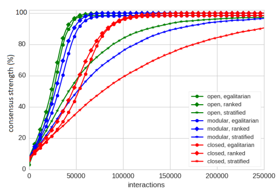
(l) StackExchange Chinese



(m) StackExchange Spanish

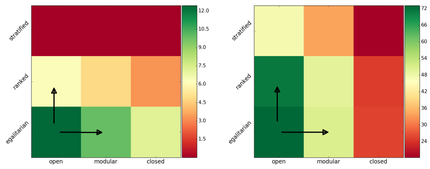


(n) Stackexchange French

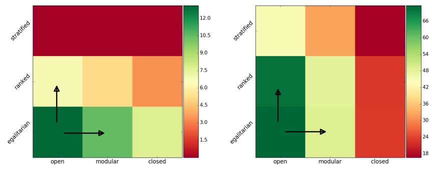


(o) StackExchange Sports

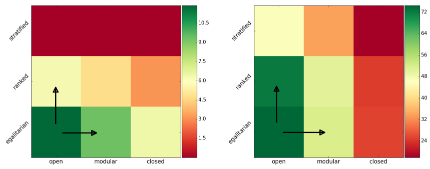
Figure 3.13: **Coarse analysis.** Percentages of consensus strength (averaged over 100 runs) reached among agents on one particular opinion (y-axes) are shown as a function of the number of interactions (x-axes) for different values of γ and β for all datasets from Table 3.3 not included in Figure 3.10. Coarse analysis results are similar to the ones presented in Figure 3.10.



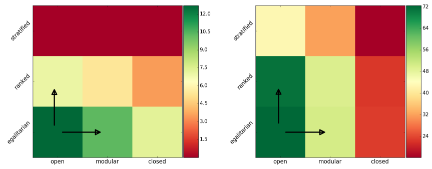
(a) Reddit Movies



(b) Reddit Politics

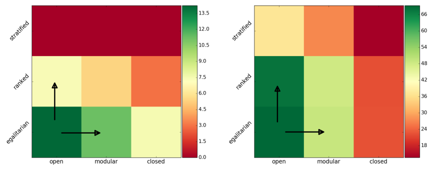


(c) Reddit Programming

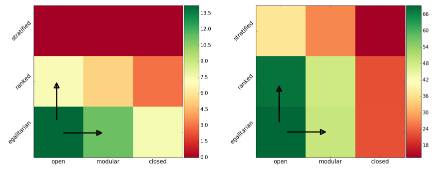


(d) Reddit Sports

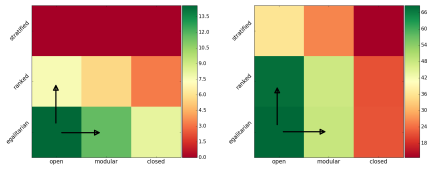
Figure 3.14: **Communication intensity between social classes.** Heatmaps of the communication intensity in low-to-high (left) and low-to-low (right) social status classes of users for all Reddit datasets from Table 3.3 not included in Figure 3.11. The results are very similar to those presented in Figure 3.11.



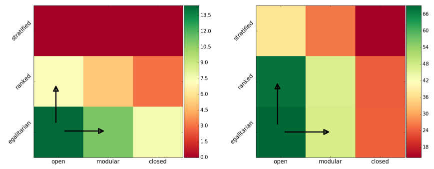
(a) StackExchange Cooking



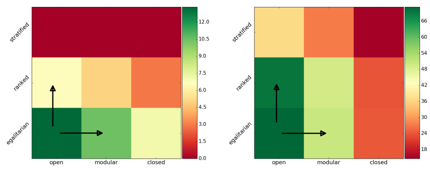
(b) StackExchange Academia



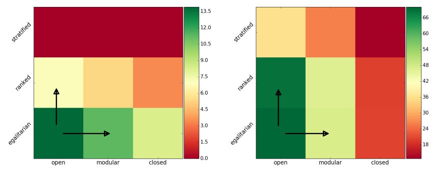
(c) StackExchange Movies



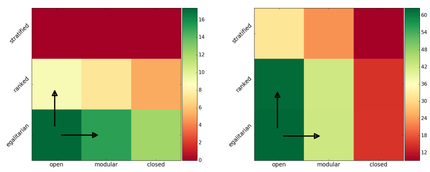
(d) StackExchange Politics



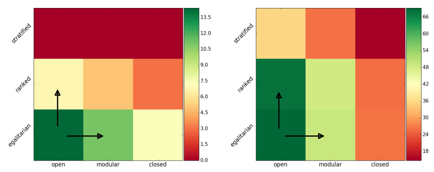
(e) StackExchange Music



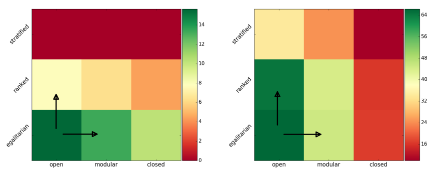
(f) StackExchange German



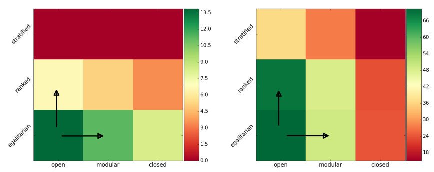
(g) StackExchange Japanese



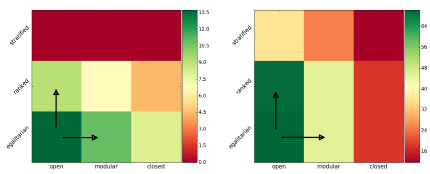
(h) StackExchange History



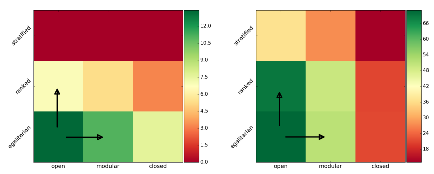
(i) StackExchange Chinese



(j) StackExchange Spanish



(k) Stackexchange French



(l) StackExchange Sports

Figure 3.15: **Communication intensity between social classes.** Heatmaps of the communication intensity in low-to-high (left) and low-to-low (right) social status classes of users for StackExchange datasets from Table 3.3 not included in Figure 3.11. The results are very similar to those presented in Figure 3.11.

3.4 Semantic Stability in Wikipedia

This article tackles the third research question concerning the development of consensus in collaborative content creation. To study the associations between content creation and consensus building in online collaboration networks, this article elaborates the semantic stability of articles in Wikipedia (i.e., as a typical example of collaborative knowledge construction sites). In this regard, we assess the evolution of article revisions over time by calculating the semantic similarity of consecutive revisions of Wikipedia articles. This approach is evaluated in 10 Wikipedia language editions including the five largest language editions as well as five (randomly selected) small language editions.

The outcomes of this article reveal that in analyzed Wikipedia editions semantic stability can be achieved. But, there are evidences of different velocities on the semantic stability process between small and large editions. Small editions exhibit faster and higher semantic stability than large ones. Specifically, in large Wikipedia editions, a higher number of successive revisions is needed in order to reach a certain semantic stability level. On the other hand, the number of needed successive revisions in small Wikipedia editions is much lower for the same level of semantic stability. These findings are in line with the fact that communities contributing to large editions such as English Wikipedia are much larger compared to small editions. Such communities are characterized with heterogeneous editors expertise, motivation, opinions and points of view, implicating that editors need longer time to agree if sufficient and correct information is provided within an article.

3.4.1 Abstract

In this paper we assess the semantic stability of Wikipedia by investigating the dynamics of Wikipedia articles' revisions over time. In a semantically stable system, articles are infrequently edited, whereas in unstable systems, article content changes more frequently. In other words, in a stable system, the Wikipedia community has reached consensus on the majority of articles. In our work, we measure semantic stability using the Rank Biased Overlap method. To that end, we preprocess Wikipedia dumps to obtain a sequence of plain-text article revisions, whereas each revision is represented as a TF-IDF vector. To measure the similarity between consequent article revisions, we calculate Rank Biased Overlap on subsequent term vectors. We evaluate our approach on 10 Wikipedia language editions including the five largest language editions as well as five randomly selected small language editions. Our experimental results reveal that even in policy driven collaboration networks such as Wikipedia, semantic stability can be achieved. However, there are differences on the velocity of the semantic stability process between small and large Wikipedia editions. Small editions exhibit faster and higher semantic stability than large ones. In particular, in large Wikipedia editions, a higher number of successive revisions is needed in order to reach a certain semantic stability level, whereas in small Wikipedia editions, the number of needed successive revisions is much lower for the same level of semantic stability.

3.4.2 Introduction

Wikipedia is one of the largest, freely accessible web-based encyclopedias and its content is open for editing by users. Wikipedia articles are mainly a contribution of volunteer editors who collaboratively create and manage the largest repository of human knowledge. This way, different editors can contribute with their expertise, ideas and opinions. Wikipedia contributors, however, may have different motivations and opinions, for example, it may take some time for them to agree if sufficient and correct information is provided within an article. If editors have different point of views on a particular topic, especially on controversial topics, they might end up overwriting each other's content such that articles cannot become semantically stable. These are also known as edit wars [DeDeo, 2016; Kalyanasundaram et al., 2015; Török et al., 2013; Yasseri and Kertész, 2013]. On the contrary, if Wikipedia editors achieve consensus on the content, implicitly, articles become semantically stable.

Problem & objectives. The goal of this paper is to investigate the semantic stability process in collaboration networks, such as Wikipedia, that are driven based on policies, guidelines and community standards. Based on these policies, both editors' behavior and the process of article production is managed [Osman, 2013].

Approach & methodology. In order to assess the semantic stability of Wikipedia, we turn to semantic similarity of consecutive revisions of Wikipedia articles. Semantic similarity of two textual documents expresses the extent to which two documents deal with semantically similar topics or content. This concept is key to understanding the comparison of documents written in natural language. Typically, semantic similarity is calculated by means of document statistics. An advantage of statistical approach is that it does not require predefined models, which describe the meaning of particular words (terms). The method applied in this work, i.e., Rank Biased Overlap, is also a statistical method and it is first introduced in [Webber et al., 2010]. The basic procedure carried out during the calculation of the semantic similarity is the modeling of the semantic space in accordance with the term distribution in a corpus of documents. In such a space, each document is represented by a vector and semantic similarity is calculated by performing vector operations on those vectors. This approach is based on the distributional hypothesis, according to which the terms with similar meanings show tendency to appear in similar contexts [Sahlgren, 2005].

The concept of semantic stability applied in our paper is based on the work presented in [Wagner et al., 2014], which studies the semantic stability of social tagging systems. In our work, we are interested in the semantic stability of Wikipedia. Thus, we take a Wikipedia corpus of documents that contains the complete edit history for each article and which includes all existing article revisions. The following Wikipedia language editions are used: English, German, French, Spanish, Italian, Czech, Finnish (Suomi), Danish, Greek and Swedish. The intention behind the choice of these particular languages is to have five Wikipedia editions with a large number of articles and five smaller editions. This enables us to study the relation between semantic stability and corpus size. Our long term goal is to investigate the consensus building process in Wikipedia based on the semantic stability. The work of Wagner et al. [2014] states that semantic stability implies implicit consensus on the description of a resource in a social tagging system.

Findings & contributions. One of the contributions of our work is the software solution that we provide as an open source project[14], which is highly modular, configurable and flexible and can be applied by anyone looking for an efficient way to analyze the semantics of natural language documents contained, for example, in the Wikipedia XML dump files. From the empirical point of view, we conduct experiments in 10 different Wikipedia language editions and discuss the experimental results and their implications. Our experimental results reveal that the mean semantic stability of large Wikipedia editions is significantly lower compared to the mean semantic stability of small Wikipedia editions. In particular, in large Wikipedia editions, a higher number of successive revisions is needed in order to reach a certain semantic stability level, whereas in small Wikipedia editions for the same level of semantic stability, the number of successive revisions needed, is much lower.

3.4.3 Technical Approach

Preliminaries

Particularly important for this paper is the theory describing: (i) evaluation of importance of terms in a single document or in a corpus of documents and their representation in a form of matrix - TF-IDF (Term Frequency - Inverse Document Frequency), (ii) calculation of semantic similarity measure and (iii) calculation of semantic stability over time.

[14] https://doi.org/10.5281/zenodo.153891

We represent each revision of the parsed Wikipedia articles as a TF-IDF vector. *Term Frequency - Inverse Document Frequency* is one of the methods in the theory of Information Search and Retrieval used to represent the relevance of terms in a document belonging to a collection of documents - *corpus* [Debole and Sebastiani, 2003; Salton and Buckley, 1988; Turney and Pantel, 2010; Zaman, 2012].

The comparison of the TF-IDF vectors is performed using a modified version of RBO (Rank Biased Overlap) method as in [Wagner et al., 2014]. However, our approach is flexible and can be extended to include additional similarity measures. The RBO method is used to calculate the similarity measure of two given vectors, each of them representing the rankings of terms contained in a single Wikipedia article. Its main characteristic is that it takes the cumulative overlap of the given rankings as a measure for similarity. It is represented with the following mathematical equation:

$$RBO(\sigma_1, \sigma_2, p) = (1-p)\sum_{d=1}^{\infty} \frac{2 * \sigma 1_{l:d} \cap \sigma 2_{l:d}}{|\sigma 1_{l:d} + \sigma 2_{l:d}|} p^{(d-1)} \qquad (3.6)$$

where $\sigma 1$ and $\sigma 2$ are not necessarily conjoint lists of ranking and $\sigma 1_{l:d}$ and $\sigma 2_{l:d}$ are ranked lists at depth d. RBO evaluates to a value in the range $[0, 1]$, where 0 means disjoint and 1 means identical. The parameter p defines the steepness of the weights and takes a value in interval $(0 \leq p < 1)$. When $p = 0$, RBO considers only the top ranked item of the lists and its value is either 0 or 1. When p is arbitrarily close to 1 the weights are almost the same for all depths and the analysis is arbitrarily deep.

The similarity measure described in Equation 3.6 is used as basis for determining the semantic stability over time. Based on the work of Wagner et al. [2014], for a given value of RBO threshold k, an article is semantically stable if its RBO value at the point of time t is equal or higher than the threshold k. A rather simple mathematical formulation of this method for inspection of stabilization process in a given data set is as following:

$$f(t,k) = \frac{1}{n}\sum_{t=1}^{n}\begin{cases} 1, & \text{if } RBO(\sigma_{t-1}, \sigma_t, p) \geq k \\ 0, & \text{otherwise} \end{cases} \tag{3.7}$$

Based on the Equation 3.7, for each article in a Wikipedia corpus, the rank-biased overlap similarity measure is calculated. Inputs are the revisions before and after the time point t as well as the parameter p. If the calculated similarity is equal or greater than the threshold k, 1 is added to the sum, otherwise 0 is added. With no more articles in corpus to iterate, the sum is divided by the total number of iterated articles from the Wikipedia corpus. Thus, the result will be the percentage of the stable articles at time-point t for a predefined threshold value k.

For our experiments, the rank-biased overlap similarity measure algorithm is parametrized with the $p = 0.9$ which means that the first ten ranks of the ranking list have 86% of the weight of the evaluation as stated in [Wagner et al., 2014]. Empirically, we also find that $p = 0.9$ is appropriate because of the value of parameter d (depth of evaluation) chosen for rank-biased overlap. This means that the TF-IDF vectors will be checked for similarity only up to the depth of 20. Of course, one can take a much higher depth, but that will increase the computation time as well as the storage space. Namely, the TF-IDF vector representing a single revision of an arbitrary article can have several thousands of values, but not all of those values are stored. Only the values up to the depth needed for rank-biased overlap calculation are stored. So, if 20 elements are used for rank-biased overlap measure, the first 10 elements of the ranking weight 86% of the evaluation and the other 10 elements weight only 14%. It is exactly because of this fact that there is no need to do the similarity calculation for much higher depths as those are not regarded as very important. In every case, the top 20 (most-weighted) elements of the TF-IDF vector are more than enough to precisely describe the semantics of the article revision they represent.

3.4.4 Experimental Setup

We study two different aspects of the stabilization process: (i) semantic stabilization of the Wikipedia corpus over a predefined period of time and (ii) semantic stabilization of the Wikipedia corpus after a number of successive revisions. The idea behind the examination of the Wikipedia corpus stabilization over the time is to choose a point in time t and count the number of articles existing at that point in time and the number of articles existing at that point in time that are also semantically stable. This is possible because of the fact that every article revision is uniquely identified in the database by the compound key consisting of the article ID and the revision timestamp.

Another way to inspect the stabilization process of the document corpus is to find out how many successive revisions are required before a percentage of the available articles becomes stable (in reference to the stability threshold). The idea is very similar to the previously discussed one, but now it is assumed that all articles have the first revisions starting at the same date and time. The timestamp information is now completely neglected and only the number of revisions per article is important. So, at the beginning, the first value of the similarity vectors of all articles is examined. The stability threshold takes the maximal value at the beginning of the calculation, 1. If the desired percentage of the articles is stable, the next value of the similarity vector is inspected. If not, the threshold is decreased and the calculation is repeated until the value of the stability threshold, for which the desired percentage of articles is stable, is found. Analysing the semantic stability from two different point of views, provides more useful insights about the examined corpus.

Dataset Preprocessing. The Wikimedia[15] provides XML dumps of all active Wikipedia projects. The basic building block of all Wikipedia editions is a page. Every page represents an article and every article has at least one, but usually more than one, revision. There are articles in bigger Wikipedia editions which have tens of thousands of revisions.

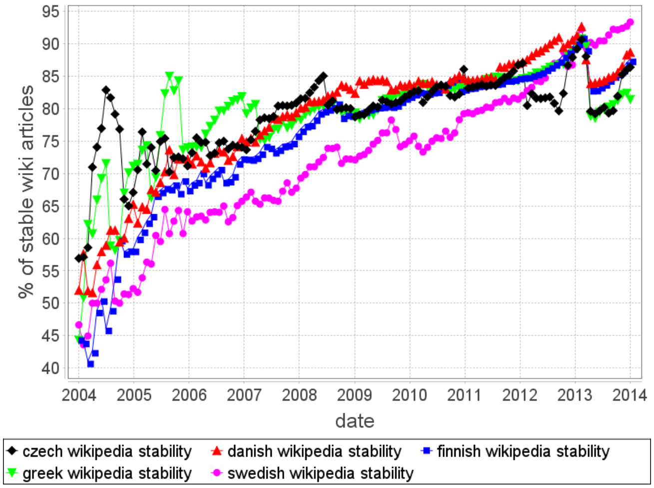
[15] https://dumps.wikimedia.org/

We analyze 10 Wikipedia language editions, five of which are (randomly selected) small language editions and the remaining five are the largest language editions. Our goal is not to analyze the full Wikipedia corpus of the large editions, thus, the sampled data of 10 thousand randomly selected articles with their complete revision history is used for 8 out of 10 Wikipedia editions. Only Czech and Finnish Wikipedia corpus is fully analyzed.

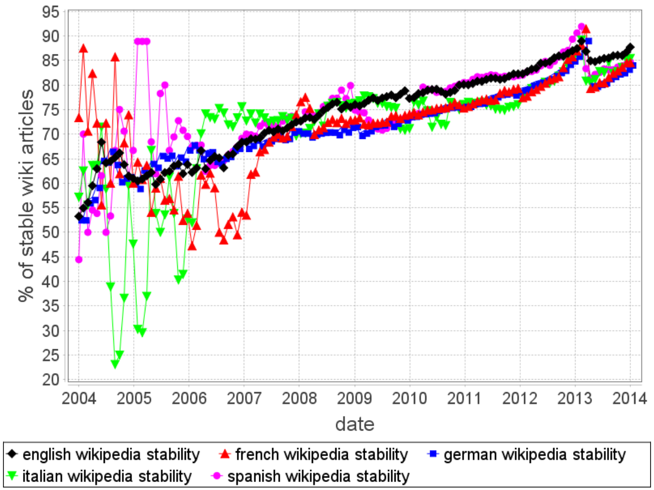
3.4.5 Results and Discussion

Figure 3.16 compares the stabilization process between small and large Wikipedia language editions over a period of time. A portion of the stable articles (in percentages) is shown for a chosen point in time t, in order to spot periods of increased stability or instability of an article corpus. The plots in Figure 3.16 correspond to the RBO threshold $k = 0.8$. We run experiments with two other values: $k = 0.4$ and $k = 0.6$, to investigate the role of the threshold parameter k in the stability calculation method proposed in [Wagner et al., 2014]. Once the similarities of all revisions of a single Wikipedia article are calculated, the value representing the similarity in a given moment of time t is taken and compared to the value of the parameter k. Our intuitive assumption is that, for a low value of RBO threshold k, there are a lot of articles in the examined corpus, whose stability value in a given instant of time is higher than the chosen threshold. Our results are consistent with our initial assumptions. Thus, as the value of the RBO threshold increases, the number of stable articles decreases. The document corpus stability is inversely proportional to the value of parameter k. However, the steepness of the stabilization curves remains the same over different parameters k, thus, we include plots for only $k = 0.8$ to show the least stability.

From the plot in Figure 3.16a, it is noticeable that all small Wikipedia editions exhibit semantic stability variations in almost the same range (with a deviation $\pm 2\%$ from the average). The only exception to this is the case of Swedish Wikipedia that has the semantic stability well below the average semantic stability of the other four small Wikipedia editions.



(a) Small Wikipedia editions



(b) Five largest Wikipedia editions

Figure 3.16: **Semantic stabilization of the Wikipedia corpus over a period of time.** Percentages of stable articles (y-axis) are shown in relation to a predefined period of time (x-axis) for **(a)** small and **(b)** large Wikipedia editions. Semantic stability curves shown, correspond to the RBO threshold $k = 0.8$ and steepness parameter $p = 0.9$. For illustration, consider the plot in (a), for a chosen point in time, (e.g.,) year 2008, in (e.g.,) Czech edition, is indicated that 70% of articles have reached a semantic stability equal or higher than 0.8. The steepness of the stabilization curves remains the same over different parameters k, however, the percentage of stable articles decreases with increasing k. Comparing plots in (a) and (b), one can see that the mean semantic stability of small Wikipedia editions is significantly higher in contrast to large ones. This is in line with the fact that small Wikipedia editions contain large portions of articles simply translated from the English Wikipedia, for example. Such articles are usually rarely changed substantially and they increase the overall stability of small editions. In contrary, the editorial process in large editions is much more dynamic.

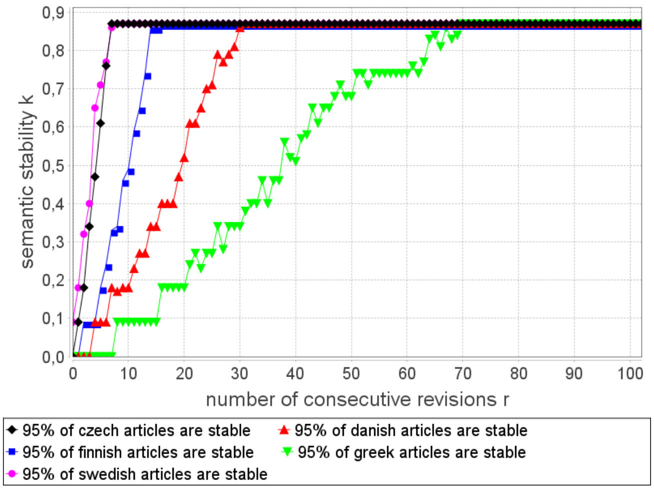
Figure 3.16b shows that in large Wikipedia editions, semantic stabilization curves oscillate more at the beginning of the editorial process compared to small editions. Thus, they are, on average, more unstable than the small Wikipedia editions. Our explanation for this is that the small Wikipedia editions consist mainly of articles which are the translated versions of the articles from the main Wikipedia editions (for example from the English Wikipedia). Once translated and created, such articles are rarely edited a lot. Whereas in large editions such as in the English one, a higher number of new articles that are authored from scratch is present. Of course, the editorial process of such articles is more dynamic.

We observe a very interesting phenomenon in both plots in Figure 3.16, namely, in both small and large Wikipedia editions, a sudden increase of the semantic stability is noted, with a peak around year 2013. Right after this point of time, the stability decreases for all Wikipedia editions and than continues to increase again. We wanted to find an explanation for this observation by contacting the Wikipedia community by writing several posts in the *Wikimedia.org*[16]mailing list, but we did not receive any plausible answer. Some of the assumptions are that: some of the Wikipedia servers were down for a short maintenance, or some of the Wikipedia maintenance bots were active and editing Wikipedia contents was shortly blocked or malfunctioning of Wikipedia servers was induced by malicious software or hacker attacks. But, the temporary peak in semantic stability in year 2013 could also be seen as a consequence of a change in Wikipedia policies of how to handle edit wars (e.g, the introduction of a new rule such as the three-revert rule). Still, no hard evidence was brought into light.

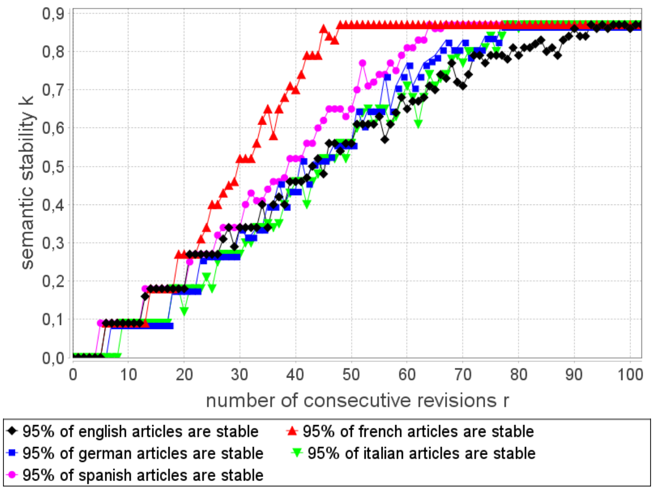
Figure 3.17 visualizes the number of consecutive revisions per article needed to achieve the stability of 95% in both small and large Wikipedia editions. This means that 95% of articles in a corpus become semantically stable, evaluated based on different RBO (for $p = 0.9$) thresholds k (y-axis in Figure 3.17), after r consecutive revisions (x-axis).

In Figure 3.17a, 95% of stable articles is reached after, for example, 70 revisions for the Greek Wikipedia and 30 or less revisions for all other

[16] https://lists.wikimedia.org/mailman/listinfo/wiki-research-l



(a) Small Wikipedia editions



(b) Five largest Wikipedia editions

Figure 3.17: **Semantic stabilization of the Wikipedia corpus after a number of successive revisions.** 95% of articles in a corpus become semantically stable, evaluated based on different RBO thresholds k (y-axis), after r consecutive revisions (x-axis). The plot in (a) illustrates that almost all small editions exhibit, at the beginning, a fast increase of the stabilization curves, which remain relatively stable after few successive revisions. An exception presents the Greek edition, which is the most frequently edited among the small ones. The plot in (b) depicts that the stabilization process in large editions is delayed. This indicates that in large editions a higher number of successive revisions is needed in order to reach the same semantic stability level as in small Wikipedia editions. These results are consistent with the fact that the size of the community contributing to the large editions, such as English, can not be compared to the small ones. Large communities are characterized with heterogeneous contributors' expertise, motivation and opinions, which implicates that it takes time until contributors agree if sufficient and correct information is provided within an article.

small Wikipedia editions. It can be seen that for the Greek Wikipedia, 95% of the articles has the stability of 0.5 or higher after almost 35 revisions, where $k = 0.5$ is considered as a medium stability Wagner et al. [2014]. From this fact one can conclude that the Greek Wikipedia edition is the most frequently edited one amongst the analyzed small editions. The Czech and Swedish editions are showing much more semantic stability. 95% of the article corpus of this two editions has the semantic stability of 0.5 or higher after only about 5 revisions.

Figure 3.17b shows the stabilization process of large Wikipedia editions where the achieved stability is 95%. This time, as expected, the English Wikipedia is the most unstable one. Almost the complete corpus of analyzed articles becomes stable after almost 95 revisions of each article. The medium semantic stability of the corpus that is defined by the value of parameter $k = 0.5$ is, in the case of English Wikipedia, reached after about 45 revisions, and in the case of the French one (the most stable one) after about 30 revisions.

These results are in line with the fact that larger communities contribute to the largest Wikipedia editions (e.g., English, German or French), in comparison to the communities editing the small Wikipedia editions, written in languages, which are only used by a very small percent of the world population. Large authoring community indicates a heterogeneous community based on authors' expertise, ideas and opinions, which in turn implies that the contributed content is more colorful. If content contributors have different point of views on a particular topic, especially on controversial topics, they might end up overwriting each other's content such that articles cannot become semantically stable. Thus, in large Wikipedia editions a higher number of revisions is needed until contributors agree if sufficient and correct information is provided within an article.

Key findings. Our findings can be summarized as follows: even in policy driven collaboration networks such as Wikipedia, semantic stability can be achieved. However, there are differences on the velocity of the semantic stability process between small and large Wikipedia editions. In large Wikipedia editions, semantic stability curves oscillate more at the beginning of the editorial process compared to small editions. Thus, the mean semantic stability of large Wikipedia editions is significantly lower in contrast to small Wikipedia editions. In other words, small Wikipedia editions stabilize faster and achieve higher levels of semantic stability.

3.4.6 Related work

The process of consensus reaching among Wikipedia editors has been on the focus of many recent studies [Biancani, 2014; DeDeo, 2016; Kalyanasundaram et al., 2015; Müller-Birn et al., 2013; Osman, 2013; Török et al., 2013; Yasseri and Kertész, 2013]. Authors in [Kalyanasundaram et al., 2015] study the problem of edit wars in Wikipedia and model this phenomenon using agent-based systems, based on theories of group stability and reinforcement learning. Authors show that consensus is reached faster if the number of credible or trustworthy agents and agents with a neutral point of view is increased. In the contrary, consensus is hindered when agents with opposing views are in equal proportion. Similarly, authors in [Török et al., 2013] apply also an agent-based model to emulate conflict scenarios in edit wars and validate their model by empirical Wikipedia data. Recently published work of DeDeo [2016] uses hidden Markov models to approximate and characterize the computational structure of conflicts in Wikipedia.

The work presented in [Osman, 2013] investigates the role of conflict in the editorial process in Wikipedia by studying talk pages. Experimental results reveal that conflict is central to the editorial processes of Wikipedia; it is a generative friction that is used by Wikipedia editors as part of a coordinated effort within the community to improve the quality of articles.

There are several research approaches published in the field of semantic similarity measurements [Hajian and White, 2011; Shirakawa et al., 2013; Stefanescu et al., 2014; Takale and Nandgaonkar, 2010]. Hajian and White [2011] propose a multi-tree similarity algorithm as a non-linear technique for measuring similarity based on hierarchical relations which exist between attributes of entities in an ontology. This method compensates for the lack of semantic relatedness among features using taxonomic relations that exist among the features of two entities. In [Shirakawa et al., 2013] authors implement a probabilistic method of measuring semantic similarity for real-world noisy short texts like microblog posts. Their method adds related Wikipedia entities to a short text as its semantic representation and uses the vector of entities for computing semantic similarity. The work presented in [Stefanescu et al., 2014] shows that the combination of knowledge and corpus-based word-to-word similarity measures can produce higher agreement with human judgment than any of the individual measures. Authors in [Takale and Nandgaonkar, 2010] present an approach for measuring semantic similarity between words using the snippets returned by Wikipedia and the five different similarity measures of association. Their results demonstrate that the snippets in Wikipedia have a significant influence on the accuracy of semantic similarity measure between words.

The Rank Biased Overlap or shortly RBO method is introduced in [Webber et al., 2010]. Our study is based on the scientific work of Wagner et al. [2014], in which a modified version of RBO is applied to investigate the semantic stability of social tagging systems. However, in our work we assess the semantic stability of Wikipedia articles.

3.4.7 Conclusion and Future Work

In this work, we study the semantic stabilization of Wikipedia with a focus on the dynamics of Wikipedia articles' revisions over time. Our experimental results reveal that: (i) the analyzed Wikipedia language editions show medium semantic stability and (ii) large Wikipedia editions exhibit a significantly lower mean semantic stability value compared to the small Wikipedia editions.

Our first findings are in line with the research results of the work presented in [Wagner et al., 2014], in which authors state that natural languages are semantically stable in their nature. In our case, all the analyzed datasets have at least medium semantic stability.

Our second experimental results indicate that the large Wikipedia editions, which were utilized for the purpose of this paper are semantically less stable than the small ones. This observation can be logically explained by the fact that large Wikipedia editions have much more contributors than the small ones. The sheer size of the community supporting and developing the English Wikipedia edition cannot be compared to e.g., the size of community working on the Czech Wikipedia edition. Having many more users contributing to the content means that higher semantic instability is brought to the system. The users of English Wikipedia are changing the content of the articles much more than the users of small Wikipedia editions. Additionally, many articles available in small Wikipedia editions are simply translations of the articles found in the English Wikipedia. Once translated, such articles are rarely changed significantly, which contributes to a higher semantic stability of the small Wikipedia editions.

One of the limitations of our work is that we evaluated only sampled data for the large Wikipedia editions. However, our software solution is flexible and could be easily extended to analyze the full Wikipedia corpus of the large editions.

For future work, we plan to investigate the consensus building among editors in different Wikipedia categories, in order to find out if there are categories that are unstable. We also want to specifically study the semantic stability of articles marked as controversial. One of our future plans is to combine the content based approach introduced in this work with a network based approach. Vandalism detection is also a topic that could benefit from our work.

Acknowledgments. This work is supported by the Know-Center, the VIRTUAL VEHICLE Research Center and the AFEL project funded from the European Union's Horizon 2020 research and innovation programme under grant agreement No 687916. The Know-Center and the VIRTUAL VEHICLE Research Center are funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Ministry of Transport, Innovation and Technology, the Austrian Ministry of Economics and Labor and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency (FFG).

3.5 Empirical Analysis of Editing Dynamics in English Wikipedia

This work in progress provides further answers to the third research question by delving deeper into the editing dynamics in English Wikipedia. To that end, we perform a comprehensive analysis of editing actions, on the word level, for each article in English Wikipedia. We first categorize each revision of articles based on the editors contribution (i.e., number of added, deleted and reinserted words). For each article we construct an ordered sequence of categorized revisions. We apply these sequences to further investigate the transitions between subsequent revisions and underlying dynamics of editing behavior. To explain the observed editing dynamics, we build a number of hypotheses (i.e., cooperative, gardening and combative) and utilize the HypTrails framework to quantify the evidence for those hypotheses in empirical data.

Our preliminary findings showcase that editing dynamics in English Wikipedia is mostly characterized with an intention to get involved in reducing, extending and again reducing content of Wikipedia articles (i.e., combative behavior). The second most evident editing dynamics is reflected with an intention to enhance the provided content of Wikipedia articles by updating or correcting information, deleting old content and inserting new information (i.e., cooperative behavior). Editing dynamics such as extending or reducing content consequently (i.e., gardening behavior) exhibits evidences in lower percentage of articles. We plan to further differentiate our preliminary results by comparing in more detail the evidences of the tested hypotheses.

3.5.1 Abstract

In this paper, we study interaction patterns that drive editing dynamics in Wikipedia. To that end, we conduct a comprehensive analysis of all edit actions in each individual article from English Wikipedia. In particular, we start off by quantifying and categorizing article revisions based on the edit actions performed on the word level. From this we construct an ordered sequence of categorized revisions for each article. We use these sequences to elaborate the transitions between subsequent revisions and editing dynamics. In the next step, we introduce a number of hypotheses that potentially explain editing dynamics and utilize the HypTrails framework to quantify the evidence for those hypotheses in empirical data. Our preliminary results on hypotheses evaluation show that Wikipedia editors very often get involved in combative dynamics by reducing, extending and again reducing content of articles. However, they also tend to enhance the provided content of Wikipedia articles by updating or correcting information, featuring a cooperative behavior.

3.5.2 Introduction

Wikipedia constitutes the largest encyclopedia of human knowledge produced by huge numbers of volunteer editors writing the majority of Wikipedia articles. The main characteristic of the editorial process in Wikipedia is that there is no formal hierarchical organizational structure. Volunteering editors collaboratively create and edit content. They contribute with their expertise, ideas and opinions, however, occasionally they have different motivations. For example, it often takes a long period of time for them to agree if an article provides sufficient and correct information (e.g., it takes in average 2000 days for an article in Wikipedia from its creation to the date it is assigned one of the quality grades: Featured Articles, A-class, Good Articles, B-class or C-class [Liu and Ram, 2011]). Even in the absence of a central coordination, the editing process in Wikipedia is usually characterized as peaceful and constructive [Liu and Ram, 2011]. But, the work of Jurgens and Lu [2012] describes editor motives in Wikipedia as either being cooperative or combative. Especially, if editors have different point of views on some controversial topics, they might end up repeatedly overriding each other's contributions, making it harder to reach consensus on the content of articles [Iniguez et al., 2014; Sumi et al., 2011; Tsvetkova et al., 2017, 2016; Yasseri et al., 2012]. For example, there is a seemingly endless debate on Nikola Tesla's nationality in Wikipedia[17] and consequently, proponents of all sides regularly edit the article[18].

Objectives. In our work, we aim to quantify and understand editing dynamics in Wikipedia and to explain the dynamics we observe. Our work is motivated by the fact that the editing dynamics impact how content is created and maintained [Jurgens and Lu, 2012], as well as which information prevails in the end. The awareness of how editing processes function helps us better understand and shape Wikipedia as a collaborative system. To find plausible explanations for our observations, we evaluate a number of hypotheses in the English-language edition of Wikipedia by applying a sound Bayesian statistical framework.

Approach & methodology. In this work, we follow a data-driven approach. To that end, we first quantify and categorize article revisions. For each revision of an article, we generate a sequence of edit actions on word level, similar to the work of Flöck et al. [2017]. Then, we categorize the revision as either *extension* or *reduction* of the original article based on the number of added, deleted, and reinserted words. This allows us to generate an ordered sequence of categorized revisions for each article.

We then use these sequences as a basis to investigate the transitions between subsequent revisions. For this, we first construct a first-order Markov chain based on the sequences, following the approach presented in [Singer et al., 2014] where the states of the chain correspond to the revision categories (extension or reduction), and the sequence order to the state transitions. Then, we evaluate the transitions between subsequent revisions using the following hypotheses: cooperative, gardening and combative.

To evaluate our hypotheses, we use the HypTrails framework [Singer et al., 2015]. HypTrails is a Bayesian framework that is based on a first-order Markov chain model and the conjugate prior for the Markov models.

[17] https://en.wikipedia.org/w/index.php?title=Nikola_Tesla&action=history
[18] https://www.pcworld.com/article/236486/wild_wars_of_wikipedia.html#slide2

Hypotheses, which represent our belief in parameters of the model are expressed as prior distributions over the model parameters. Consequently, HypTrails incorporates these hypotheses as elicited Dirichlet priors into a Bayesian inference process. The relative plausibility of hypotheses to each other is determined by their marginal likelihoods and Bayes factors.

Note that while previous work [Iniguez et al., 2014; Kalyanasundaram et al., 2015; Liu and Ram, 2011] has studied editing behavior and collaboration processes in Wikipedia on *samples* of Wikipedia articles, in this paper, we test our hypotheses on the whole corpus of the English-language edition.

Findings & contributions. We contribute with a comprehensive evaluation of granular edit actions and editor contributions of all articles in English Wikipedia. In our experiments, we find that the editing dynamics of Wikipedia editors is mostly characterized with a combative behavior and an intention to get involved in reducing, extending and again reducing content of Wikipedia articles (i.e., conflict-revenge actions Tsvetkova et al. [2016]). The second most evident editing behavior is the cooperative behavior reflected with an intention to enhance the provided content of Wikipedia articles by updating or correcting information, deleting old content and inserting new information (i.e., identified as beneficial for the quality of Wikipedia articles Liu and Ram [2011]). Editing dynamics such as extending or reducing content consequently (i.e., gardening behavior) exhibit evidences in lower percentage of articles.

3.5.3 Related Work

At present, we identify three main lines of research related to our work: (i) the study of user behavior and interactions between Wikipedia editors, (ii) the analysis of edit sequences and dynamics of collaboration and (iii) content-based analysis of revision histories to detect conflicts, controversiality and edit wars.

User behavior and interactions between Wikipedia editors. A number of studies investigate user interactions, respectively, user edit actions in Wikipedia and define a taxonomy of actions in this regard

[Adler and de Alfaro, 2007; Brandes et al., 2009; Ehmann et al., 2008; Flöck et al., 2017; Gandica et al., 2015; Kalyanasundaram et al., 2015; Liu and Ram, 2011; Pfeil et al., 2006; Sepehri-Rad and Barbosa, 2015]. In [Kalyanasundaram et al., 2015], authors define two basic actions: a commit operation, which is adding or editing content in an article and a revert operation, which is restoring an article to a previous version. In [Pfeil et al., 2006], for example, authors define a broader taxonomy of editor actions in Wikipedia: add information, clarify information, delete information, add link, fix link, delete link, format, grammar, mark-up language, style/typography, spelling, reversion, and vandalism. We use the taxonomy defined by Flöck et al. [2017] comprising of three main edit actions: (i) adding a completely new word, (ii) deleting a single word and (iii) performing the reinsertion of a single word, to conduct an edit sequence analysis for all articles and their corresponding revisions in English Wikipedia. We build a number of hypotheses and evaluate their evidences on empirical data to better understand the mechanisms producing editing dynamics in Wikipedia.

Analysis of edit sequences and dynamics of collaboration. The work in [Keegan et al., 2016] conducts a sequence analysis of editorial process in Wikipedia, as a case study for analyzing organizational routines in online knowledge collaborations. Authors argue for a conceptual reorientation towards sequences as a fundamental unit of analysis for understanding work routines in online knowledge collaboration. Their results show that four distinct key patterns are identified while grouping motifs that represent similar edit sequence characteristics (i.e., solo contributor, reactive contributing, inactive contributor and distinctive motifs). Despite the interesting findings, this work considers only a sample of Wikipedia articles. In our work, we investigate editing sequences of all articles in English Wikipedia and contribute with an extensive analysis of collaboration dynamics in the whole corpus. Moreover, we build a number of hypotheses on editing dynamics and test their evidences on empirical data.

Content-based analysis of revision history. Phenomena such as conflict, disagreement and controversiality in Wikipedia are studied in depth from the research community. The research ranges from theoretical models (i.e., agent based) [Gandica et al., 2014; Kalyanasundaram et al., 2015; Török et al., 2013; Yasseri and Kertész, 2013], and calculations of controversial measures or number of reverts and detecting disagreement (empirically) [Sumi et al., 2011; Tsvetkova et al., 2017, 2016; Yasseri et al., 2012] to evaluating theoretical models using empirical data [Iniguez et al., 2014; Rudas et al., 2017].

Extensive work on agent based modeling and edit wars in Wikipedia is conducted in [Iniguez et al., 2014]. The presented model illustrates different scenarios, for example, simulating new coming editors and fluctuations that they cause to the editing process. Authors analyzed the dynamics of an empirical controversial measure over time for sample articles (i.e., top-10 most controversial articles from 13 different Wikipedia editions) and they identified three types of consensus reaching processes: single war to consensus, multiple war-peace cycles and never ending wars.

In [Tsvetkova et al., 2016] authors aim to study negative interactions in online collaboration networks that are not always explicitly declared (i.e., not visible). This related work analyzes the timing and configuration of sequences of contributions in Wikipedia to identify patterns of negative social interactions among users. General approach to detect disagreement, is to detect reverts. Researchers analyzed different patterns (motifs) of reverts and named them accordingly (e.g., 'repeated attack', 'self-defense', 'third-party defense', or 'pay forward'). Their results showed that, in general, more active and experienced editors do more reverts and hence, the reverters tend to have higher status than the reverted user. However, reverts do not necessary mean negative social interactions, thus, in our work we analyze the whole editing process of Wikipedia with the aim to detect editing patterns not limited to reverts or disagreements between editors. We perform fine-grained investigations of the revision sequences of each article, by using word based differentiations of edit actions of Wikipedia editors. Furthermore, we introduce a number of hypotheses that potentially explain editing dynamics and apply the HypTrails framework to quantify the evidence for those hypotheses in empirical data.

3.5.4 Methodology

We investigate the dynamics of interactions between editors during the editing process in Wikipedia by analyzing the edit sequence of granular edit actions (i.e., word-based) extracted for all articles and their corresponding revisions in English Wikipedia. Each article revision consists of a sequence of words and corresponding edit actions (i.e., *add* a word, *delete* a word, and *reinsert* a word deleted in one of the previous revisions). For each article, we extract all revisions and their word editing sequences from the TokTrack dataset [Flöck et al., 2017].

In the next step, we categorize each revision as either an *extension* (E) or as a *reduction* (R) by calculating the difference in word count Δw between the given and the previous revision [Brandes et al., 2009]:

$$\Delta w = a + r - d, \tag{3.8}$$

where a, r, and d are respectively numbers of additions, reinsertions, and deletions in a given article revision. Whenever $\Delta w > 0$ we categorize the revision as an *extension* otherwise as a *reduction*. We argue that for revisions with $\Delta w > 0$ the intention of a given editor is to extend the text of the given article by either adding new words or by reinserting deleted words. On contrary, whenever $\Delta w \leq 0$ the intention is to reduce the text of an article by deleting words Brandes et al. [2009].

Thus, for each article i, we obtain an ordered sequence of revision categories, which we denote with A_i. We analyze article sequences by constructing a first-order Markov chain [Singer et al., 2014] for each sequence. In each Markov chain states correspond to revision categories (i.e., to *extension* or *reduction*), and the sequence order to the transitions between the states. First-order Markov chains are memoryless, which means that the next state in the sequence only depends on the current one and not on the history of preceding states. To determine the transition probabilities we resort to maximum likelihood estimation (MLE), which is simply given by the proportions of transitions from one state to the other:

$$p_{ij} = \frac{n_{ij}}{\sum_j n_{ij}}, \tag{3.9}$$

where n_{ij} is the number of observed transitions from state i to state j. Typically, all transition probabilities p_{ij} are collected in transition matrix $\mathbf{P}$.

We are interested in better understanding the mechanisms producing dynamics of extending and reducing content of articles. To that end, we build a number of hypotheses on editing dynamics, by defining functions that assign values to the fields of the state transition matrix and compare their plausibility relatively.

We utilize the HypTrails framework [Singer et al., 2015] to test evidences of our hypotheses and compare them with empirical data. In HypTrails, hypotheses are represented as matrices $\mathbf{Q}$ with q_{ij} elements expressing the beliefs about transitions between states i and j (i.e., higher probability values correspond to higher beliefs). The defined hypotheses are then incorporated as Dirichlet priors into the Bayesian inference process:

$$P(\mathbf{P}|\alpha) = \prod_i \frac{\Gamma(\sum_j \alpha_{ij})}{\prod_j \Gamma(\alpha_{ij})} \prod_j p_{ij}^{\alpha_{ij}-1} = \prod_i Dir(\boldsymbol{\alpha}_i), \tag{3.10}$$

where $\alpha_{ij} > 0$ are so-called pseudo-observations of transitions from state i to state j and Γ is the gamma function, Dir is the Dirichlet distribution and $\boldsymbol{\alpha}_i$ is the vector that collects all pseudo-observations for the state i, and $\boldsymbol{\alpha}$ is a matrix collecting all α_{ij}. In particular, higher values of α_{ij} represent that we expect more frequent transitions from state i to state j. The exact values for α_{ij} are calculated from the matrix $\mathbf{Q}$ and the weighting parameter k, which reflects the overall strength of our belief in a hypothesis. The technical details on the exact elicitation of values α_{ij} can be found in the original HypTrails paper [Singer et al., 2015].

Using this conjugate prior together with the observed data one easily arrives at closed-form expression for the posterior distribution over $\mathbf{P}$, which we write here in the abbreviated form:

$$P(\mathbf{P}|D, \boldsymbol{\alpha}) = \prod_i Dir(\mathbf{n}_i + \boldsymbol{\alpha}_i), \tag{3.11}$$

where D represents all observed transitions, $\boldsymbol{\alpha}_i$ is as before and $\mathbf{n}_i$ collects all transitions from state i into a single vector.

$\log_{10}(B_{10})$	B_{10}	Evidence against H_0
0 to 1/2	1 to 3.2	Not worth more than a bare mention
1/2 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
>2	>100	Decisive

Figure 3.18: **Interpretation table of log-Bayes factors according to Kass and Raftery [1995].**

Using expressions for prior, likelihood and posterior we can easily arrive at the expression for the marginal likelihood:

$$P(D|\boldsymbol{\alpha}) = \prod_i \frac{\Gamma(\sum_j \alpha_{ij})}{\prod_j \Gamma(\alpha_{ij})} \frac{\prod_j \Gamma(n_{ij} + \alpha_{ij})}{\Gamma(\sum_j (n_{ij} + \alpha_{ij}))}. \tag{3.12}$$

The relative plausibility of hypotheses is determined by their relative marginal likelihoods and Bayes factors [Kass and Raftery, 1995; Singer et al., 2015].

$$B_{1,2} = \frac{P(D|\boldsymbol{\alpha}_1)}{P(D|\boldsymbol{\alpha}_2)}, \tag{3.13}$$

where $\boldsymbol{\alpha}_1$ are α_{ij}s for the first hypothesis and $\boldsymbol{\alpha}_2$ for the second one.

In Figure 3.18, we provide the interpretation table of log-Bayes factors according to Kass and Raftery [1995]. To determine the strength of the Bayes factors we resort to this interpretation table. For example, when comparing more than two hypotheses with a data hypothesis, we calculate the corresponding differences of evidences (i.e., marginal likelihoods) and consult the interpretation table. If all differences are decisive, we focus on the scale of each of the evidences. The larger the evidence for a given hypothesis, the more plausible it is compared to other tested hypotheses. Otherwise, if the significance is not present, the compared hypotheses are considered as being equal.

Our methodology is illustrated in Figure 3.19. In the first row, we consider five editors working on a Wikipedia article in a sequence of five article revisions. Based on the number of *added*, *deleted*, and *reinserted* words in each revision, we determine the sequence of revision categories. For example, the corresponding sequence of revision categories for the example

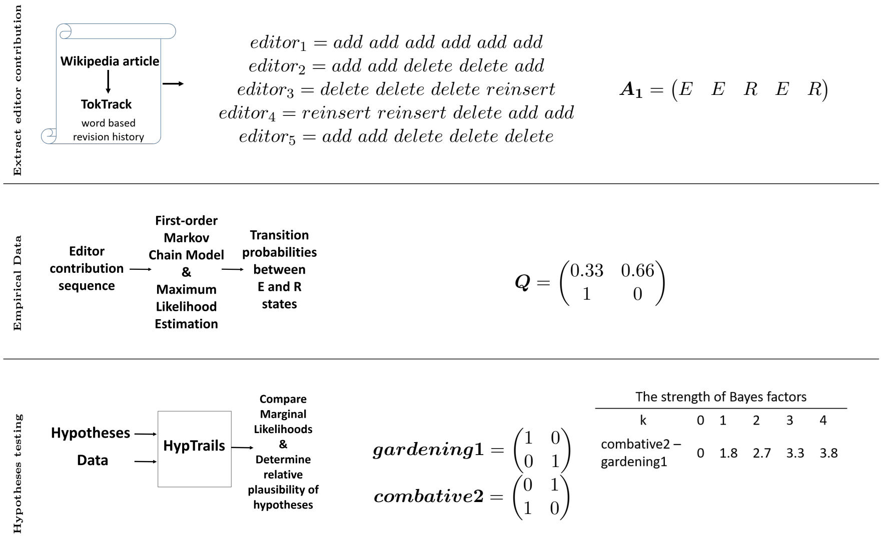


Figure 3.19: **Illustrative example.** *First row* represents five editors working on a Wikipedia article in a sequence of five article revisions. Each revision consists of a sequence of words and corresponding edit actions (*add* a word, *delete* a word, and *reinsert* a word deleted in one of the previous revisions). In the next step, we categorize each revision as either an *extension* (E) or as a *reduction* (R) by calculating the difference in word count of that particular revision. Further, we obtain the following ordered sequence of revision categories $A_1 = (EERER)$. We analyze this sequence by constructing a first-order Markov chain, where states correspond to revision categories, and the sequence order to the transition between the states. In the *second row* we show the analysis of four different transitions (EE, ER, RE and RR) between two states (E and R). The matrix Q shows the following transition probabilities: $p_{EE} = 0.33$, $p_{ER} = 0.66$, $p_{RE} = 1$, $p_{RR} = 0$. Finally, we introduce a number of hypotheses trying to explain the editing behavior. For example, in the *third row* we depict gardening1 and combative2 hypotheses and show the calculated Bayes factors for five values of weighting factor k. For $k = 0$ both hypotheses are considered as being equal. The combative2 hypothesis exhibits a strong Bayes factor (cf. Figure 3.18) for weighting factor $k = 1$ and decisive ones for higher k, implying that it is more plausible than the gardening1 hypothesis.

article looks as follows: $A_1 = (EERER)$. In the second row of Figure 3.19, we show the analysis of four different transitions (EE, ER, RE and RR) between two states (E and R). The example matrix Q has the following transition probabilities: $p_{EE} = 0.33$, $p_{EE} = 0.66$, $p_{RE} = 1$, $p_{RR} = 0$. In the third row of Figure 3.19, we illustrate two hypotheses gardening1 and combative2 and show how their evidences are calculated and compared. The table on the right shows the differences of the evidences (i.e., Bayes factors) between combative2 and gardening1 hypotheses for five values of weighting factor k. For $k = 0$ both hypotheses are considered as being equal. The combative2 hypothesis exhibits a strong Bayes factor (cf. Figure 3.18) for weighting factor $k = 1$ and decisive Bayes factors for higher k. This implies that the combative2 hypothesis is more plausible than the gardening1 hypothesis.

3.5.5 Hypotheses

In our work, we build three different hypotheses to characterize editing behavior of Wikipedia editors and test their evidences against empirical findings. Our hypotheses are explained below.

Cooperative. In this hypothesis, we express our intuition that editors intend to extend the text of articles and also enhance the provided content by updating or correcting information, for example, deleting old content and inserting new information. We base our hypothesis on previous works [Jurgens and Lu, 2012; Liu and Ram, 2011], which found that Wikipedia editors engaging in a variety of actions, such as sentence creations, modifications, and deletions, contribute to high quality articles (i.e., featured and good articles). Thus, in a cooperative hypothesis editors are likely to extend a given article after another extension and they are also likely to reduce or extend the given article after another reduction action, reflecting our intuition that editors first collaboratively extend content, which they then polish, proof-read and correct during several rounds of rewriting. To define the cooperative hypothesis, we construct two matrix variants and assign the following probabilities: *cooperative1*: $p_{EE} = 1$, $p_{ER} = 0$, $p_{RE} = 1$, $p_{RR} = 0$ and *cooperative2*: $p_{EE} = 1$, $p_{ER} = 0$, $p_{RE} = 0.5$, $p_{RR} = 0.5$.

Gardening. This hypothesis has the intuition that Wikipedia editors perform consequent extension and reduction actions with high probability. Authors in [Liu and Ram, 2011] report such editor behavior and describe two editor roles corresponding to our hypothesis: content justifiers (i.e., focusing on sentence creations, link creations and reference creations) and cleaners (i.e., focusing on removing sentences, references and links). We characterize this behavior as gardening contribution, which implicates that editors tend to extend or reduce the content of Wikipedia articles after already extending it and they also tend to reduce the content after they already performed a reduce action. We present two variants of the gardening hypothesis with the following probabilities: *gardening1*: $p_{EE} = 1$, $p_{ER} = 0$, $p_{RE} = 0$, $p_{RR} = 1$ and *gardening2*: $p_{EE} = 0$, $p_{ER} = 1$, $p_{RE} = 0$, $p_{RR} = 1$.

Combative. With this hypothesis, we express our belief that in controversial scenarios, editors end up in a circle of reducing, extending and again reducing content of Wikipedia articles. Our intuition is that editors tend to write content on certain topics based on their point of view. If a topic is controversial and disagreements are evident, they override those parts that differ from their own viewpoints. This is in line with related work [Jurgens and Lu, 2012; Tsvetkova et al., 2016], which has investigated editing behavior in controversial Wikipedia articles. In [Tsvetkova et al., 2016], authors report on different patterns of editor actions: revenge/self-defense, generalized revenge (e.g., after A reverts B (AB), B reverts A back and B reverts C) or repeated-attacks (e.g., after A reverts B (AB), A reverts B again). In other words, editors with a combative behavior are likely to reduce the content of an article right after an extension. Also, they are likely to extend or reduce the given article after another reduction action. We describe this behavior with the following transition probabilities: *combative1*: $p_{EE} = 0$, $p_{ER} = 1$, $p_{RE} = 0.5$, $p_{RR} = 0.5$ and *combative2*: $p_{EE} = 0$, $p_{ER} = 1$, $p_{RE} = 1$, $p_{RR} = 0$.

Empirical data. We test evidences of our hypotheses against empirical data that we consider as an upper limit [Singer et al., 2015].

3.5.6 Datasets and Experiments

We use the TokTrack datasets established in the work of Flöck et al. [2017] that is publicly available[19][20]. TokTrack tracks the origin and changes of all words in the articles of the English Wikipedia until October 2016. From the TokTrack dataset we exclude stubs and disambiguation pages from articles and bots and scripts from editors. We also exclude articles having less than 30 revisions and 2 editors. Overall, we analyze $2,428,580$ articles, $399,593,420$ revisions and $40,965,833$ editors (registered and unregistered)[21].

First, we extract all revisions and their word editing sequences from each article. Second, we calculate the difference in word count of a particular revision [Brandes et al., 2009] and construct sequences of revision categories (i.e., E for extend or R for reducing text) for each Wikipedia article and store them. Third, we define hypotheses to characterize editing dynamics of editors, test their evidences using the HypTrails framework and store the results. Fourth, we compute the differences between evidences of hypotheses, averaged over the hypothesis weighting factor k, and determine their plausibility for Wikipedia articles in our dataset. Fifth, we investigate more in depth articles that exhibit higher plausibility of particular hypothesis by examining dynamics of editing behavior. Finally, we run statistical hypothesis tests to quantify the independence of transitions between revision categories E and R.

To ensure the reproducibility of our work, we provide our experimental framework as an open-source project. The source code can be downloaded from our Git repository[22].

3.5.7 Results and Discussion

We compare evidences of defined hypotheses for all Wikipedia articles in our dataset. Since the empirical data hypothesis expresses the real data

[19] https://zenodo.org/record/834557
[20] https://zenodo.org/record/439699
[21] *The English Wikipedia currently has* $34,541,561$ *users who have registered a username, of which* $128,442$ *are actively editing*, source: https://en.wikipedia.org/wiki/Wikipedia:Wikipedians
[22] https://git.know-center.tugraz.at/summary/?r=SocialNetworkAnalysis.git

(i.e., is the most plausible one), we investigate the plausibility of other three hypotheses by comparing their evidences against empirical data and averaging the results over values of the weighting factor k.

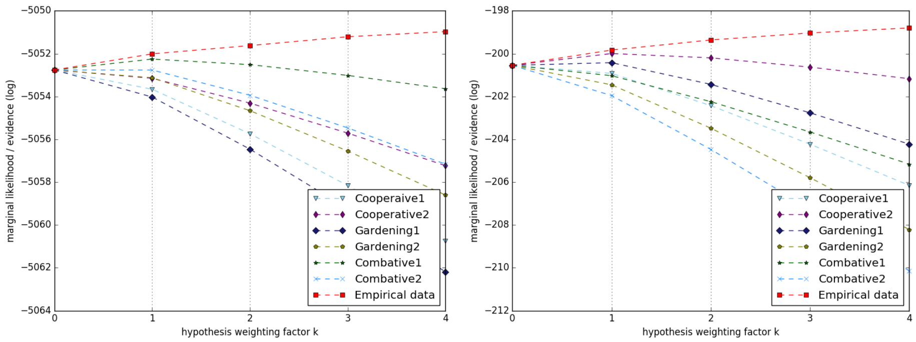
Our experimental results for the whole corpus (Table 3.4) show that in 54.9% of examined Wikipedia articles the combative hypothesis is most plausible, followed by the cooperative hypothesis with a plausibility of 28.5% and gardening hypothesis, plausible in around 16.6% of articles.

Our main results indicate that the intention of editors to get involved in reducing, extending and again reducing content of Wikipedia articles (i.e., conflict-revenge Tsvetkova et al. [2016] actions reflected in our combative hypothesis) is strongly evident. This is followed by the intention of a progressive and cooperative behavior of editors to enhance the provided content of Wikipedia articles by updating or correcting information (i.e., contributions to the article quality [Liu and Ram, 2011]). Editing dynamics such as extending or reducing content consequently (gardening hypothesis) exhibit evidences in lower percentage of articles.

We further delve into the nature and characteristics of articles corresponding to stronger evidences of combative and cooperative hypotheses. For illustration, we present hypotheses testing for the complete revision history of two Wikipedia articles exhibiting high and low conflict measures in Flöck et al. [2017], respectively.

Table 3.4: **Summary of our preliminary results.** This table shows statistics for article groups, in which corresponding hypothesis is the most plausible: percentages of articles ($A(\%)$), number of unique editors (e), number of unique revisions (r) in the whole group, mean editors ($\bar{e}$), mean revisions ($\bar{r}$), mean edit actions ($\bar{a}$), standard deviation editors (s_e), standard deviation revisions (s_r), and standard deviation edit actions (s_a) per article.

Hypothesis	$A(\%)$	e	r	$\bar{e}$	$\bar{r}$	$\bar{a}$	s_e	s_r	s_a
cooperative	28.5	$12,285,114$	$103,014,945$	75.5	163.1	6398.3	118.7	310.6	27113.2
gardening	16.6	$1,324,572$	$15,161,720$	27.2	65.1	1603	22.1	76.4	4071.4
combative	54.9	$33,148,476$	$281,416,755$	98.7	207	8226.4	237.8	561.1	57354.4



(a) Mustafa Kemal Atatürk article (b) Hank Jones article

Figure 3.20: **Hypotheses testing for the complete revision history of Wikipedia articles.** In both plots of example articles, the x-axis depicts the weighting factor k [Singer et al., 2015], while the y-axis shows the corresponding evidence (marginal likelihood) value. We compare hypotheses with each other by comparing the evidence values for the corresponding values of k. Higher values mean higher plausibility. Empirical data hypothesis is the most plausible in both Wikipedia articles as it expresses exactly the real data. In Mustafa Kemal Atatürk article (a) the combative1 hypothesis is more plausible than other five. In Hank Jones article (b) a more peaceful editing behavior is present making a good prior for the cooperative2 hypothesis to be evaluated as the most plausible one.

Specifically, we show evidence testing of hypotheses for the following Wikipedia articles: Mustafa Kemal Atatürk article (8034 revisions and 2800 editors) in Figure 3.20a and Hank Jones article (434 revisions and 166 editors) in Figure 3.20b. In both plots, the x-axis represent the weighting factor k and on the y-axis the corresponding evidence (i.e., marginal likelihood). Hypotheses are marked with different colors and markers. For $k = 0$ all hypotheses have the same evidence and we determine their plausibility by comparing the evidence values for the respective values of k. Higher evidence values indicate higher plausibility of a particular hypothesis [Singer et al., 2015].

Not surprisingly, the empirical data hypothesis is the most plausible in both examples, as it expresses exactly the real data and is considered as an upper limit. The Mustafa Kemal Atatürk article is identified as the second most conflicted article in English Wikipedia in [Flöck et al., 2017] with a conflict measure $cbTime = 4397110$. In this line, our experiments show that the most plausible hypothesis among defined hypotheses for the Mustafa Kemal Atatürk article is the combative1 hypothesis, followed by combative2 , cooperative2 and gardening2 hypotheses. Cooperative1 and gardening1 hypotheses are evaluated as weak evidences. Our findings illustrate a typical scenario identified in previous work [Tsvetkova et al., 2016] on controversial topics in Wikipedia, where editors aim to provide sufficient information and facts, but in the same time they tend to correct facts and arguments presenting different views than theirs.

In the Hank Jones article, the conflict measure determined in [Flöck et al., 2017] is much lower $cbTime = 17.9$, which is indicated also in our experimental results, exhibiting stronger evidence for cooperative2 hypothesis. These results implicate a more peaceful editing behavior in the case of Hank Jones Wikipedia article.

Validation of results. To investigate the implications of our findings we focus on the data matrices representing empirical observations of transitions between *extension (E)* and *reduction (R)* states. We aim to verify if probabilities in data matrices could also be derived by chance alone. Specifically, we determine how likely it is that any observed transition between the states arose by chance. So, we calculate the Pearson's chi-square test of independence [Pearson, 1900] for the matrices containing

Table 3.5: **Summary of statistics for articles exhibiting significant results at all usual levels.** This table shows statistics for article groups, in which corresponding hypothesis is the most plausible: percentages of articles ($A(\%)$), number of unique editors (e), number of unique revisions (r) in the whole group, mean editors ($\bar{e}$), mean revisions ($\bar{r}$), mean edit actions ($\bar{a}$), standard deviation editors (s_e), standard deviation revisions (s_r), and standard deviation edit actions (s_a) per article.

Hypothesis	$A(\%)$	e	r	$\bar{e}$	$\bar{r}$	$\bar{a}$	s_e	s_r	s_a
cooperative	23.8	$2,433,175$	$16,325,370$	170.5	524.5	27552	261.4	775.4	79658
gardening	14.5	$287,619$	$2,628,044$	42.5	137.2	4308	49.6	192.2	9651
combative	61.7	$18,160,796$	$103,735,491$	582.8	1257.9	71185	726.2	1764.6	215469

the frequencies (i.e., counts) of transitions between E and R states. This enables us to compare the observed frequencies of transitions between these states with the frequencies that one would expect to get by chance. The Pearson's chi-square test returns the calculated statistic and p-value for interpretation. Depending on the threshold for the p-value or the significance level - α (i.e., typically $\alpha = 0.01$ for significant results at all usual levels [Urdan, 2010]), the test is significant or it rejects the null hypothesis (H0) if the given p-value is below the threshold. In our case 6% of data matrices or $145,801$ articles exhibit significant results at all usual levels, which means they reject the null hypothesis (H0).

In Table 3.5 we provide the same statistics as in Table 3.4, but for the group of articles exhibiting significant results at all usual levels in our Pearson's chi-square test of independence. It can be seen that the plausibility of three main groups of hypotheses remains almost the same. The combative hypothesis is again the most plausible one, exhibiting higher evidences in 61.7% of articles followed by the cooperative hypothesis with a plausibility of 23.8% and gardening hypothesis, plausible in around 14.5% of articles. Noticeable is the fact that in this group of articles the mean values of editors, revisions and actions per article are much higher than in Table 3.4, implicating an evidence of a more intensive editing behavior.

We further investigate the top 10 articles with the lowest p-values in the Pearson's chi-square test of independence. In all 10 articles combative1 and combative2 hypotheses are evaluated as the most plausible ones. The same articles show high conflict measures in [Flöck et al., 2017]. For example, articles Wikipedia ($cbTime = 3768268$) and Doctor Who ($cbTime = 2497633$) are both in top 10 with lowest p-values and combative editing behavior and in most conflicted articles in [Flöck et al., 2017].

3.5.8 Conclusion and Future Work

In this paper, we studied editing dynamics and interactions between editors in Wikipedia, by performing a comprehensive analysis of granular edit actions for all articles in English Wikipedia. We applied editor contribution sequence analysis on empirical data to capture interactions between

Wikipedia editors. Additionally, we expressed a number of hypotheses on editing dynamics and utilized the HypTrails framework [Singer et al., 2015] to test evidences of our hypotheses and compare them with empirical data.

Our preliminary findings revealed that hypotheses characterizing a combative behavior of editors (i.e., reducing, extending and again reducing content of Wikipedia) exhibit strong evidences in 54.9% of the analyzed articles. This observed editing dynamics is very often characterized with propagated conflict between editors, for example, when editors deal with a controversial topic and try to bring facts from opposing views. The evidences of a cooperative behavior of editors (i.e., intention to enhance the provided content of Wikipedia articles by updating or correcting information, deleting old content and inserting new information) are strong in 28.5% of articles. Other tested hypotheses exhibit plausibility in lower percentage of articles. Our empirical results present the first step towards our long term goal of designing concepts for Wikipedia tools to intervene, for example, when edit wars are evident.

For future work we plan to further differentiate our preliminary results by comparing the strength of the Bayes factors between all tested hypotheses. We will than determine the overlap between these results and the results from Pearson's chi-square test of independence. We also intend to analyze other Wikipedia editions and compare the results with English Wikipedia and our previous work on semantic stability of Wikipedia [Stanisavljevic et al., 2016]. Furthermore, we want to tackle particular categories or topics and build and test hypotheses for sets of articles contained in such categories. Similarly, we will test hypotheses on different communities detected on collaboration network of editors constructed for the whole Wikipedia corpus. Interesting to investigate would be to consider editors social status or similarity when constructing hypothesis.

4 Conclusions

The popularity of online social and collaboration networks has affected the way we interact with our social environment. Online social and collaboration sites enable faster information flow and higher interconnections between individuals. Individuals turn to collaboration networks (i.e., special case of social networks) to express their opinions, discuss and solve certain problems or collaboratively write joint articles. How opinions spread in such sites and which factors influence the dynamics of consensus or disagreement between individuals are pressing questions that our research community has already recognized. Opinion dynamics and consensus building are complex phenomena to study, because individuals themselves are complex [Castellano et al., 2009]. As such, they have been studied from point of views of social sciences, physics, mathematics, complex system studies and network science. It has been challenging for our research community to conduct studies that cover each of the disciplines and their aspects. This thesis aims to tackle this challenge by combining aspects of aforementioned disciplines to investigate the factors and mechanisms that drive consensus dynamics in online collaboration networks. To that end, this thesis contributes with a methodology and a framework to study the role of some of the main factors (i.e., users social status, network structure, users similarity and content creation) and their interplay in consensus building.

This chapter concludes this thesis by summarizing results and contributions in Section 4.1, discussing the implications in Section 4.2, stating the limitations in Section 4.3 and the potential future work in Section 4.4.

4.1 Results and Contributions

In this section I give answers to the research questions introduced in Section 1.4.

RQ1: What is the influence of social status on consensus in collaboration networks?

Consensus building among users collaborating online is closely related to opinion formation and opinion dynamics processes. Opinion dynamics has been studied in the field of statistical physics by utilizing agent-based and mathematical models [Castellano et al., 2009]. Agent-based models represent simplifications that make complex problems, such as opinion dynamics, tractable for analysis. Such simplifications narrow the scope of research down to theoretical models that typically do not consider empirical data. The main contribution of today's streamline of research is to investigate opinion dynamics in pre-designed synthetic networks, in which the structure of the relations between users is known in advance [Xia et al., 2011]. Pre-designed network structures do not reflect topologies arising from real user interactions and also neglect real user characteristics (i.e., social status). Thus, the focus of the first research question has been to find out how users social status in correlation with the underlying network structure influences consensus building in empirical datasets extracted from the Web. Section 3.2 presents the work that has been conducted to answer this research question. This work contributes with a framework that facilitates the simulation of opinion dynamics and consensus building in arbitrary empirical networks with heterogeneous distribution of users social status. The results of this work have shown that a moderate influence of users social status speeds up the process of consensus building among users collaborating online, whereas a high influence of users social status hinders this process. The findings regarding the correlation with the network structure have demonstrated that some network configurations benefit consensus building and no external interventions are needed (i.e., controlling opinion flow between users). In general, the results suggest that to optimize the process of consensus building it is necessary to accommodate users social status in opinion dynamics.

RQ2: How does consensus depend on user similarity and social status?

The first research question has considered user interactions that are visible or leave traces in online collaboration networks (i.e., two users providing answers to a Q&A site). However, some user interactions do not leave any traces in the system logs. For example, many users turn to such sites only to read posts or articles and do not leave any comments or perform contributions of any kind. The second research question has aimed to investigate consensus building by capturing such hidden user interactions. In this regard, Section 3.3 presents the work that has provided answers to this research question. This work applies a model of interacting users, whose future interactions are not restricted to the edges of the observed interaction network. Rather, interactions are allowed between all pairs of users with varying preferences. In detail, in this work we have studied how the speed towards consensus building is affected by configurable influences of users latent similarities, users social status and the interplay between those two factors. The findings showcase that an increase in the influence of user similarity delays the consensus building process, whereas a suitable increase of the influence of user social status compensates this delay. These results indicate that user similarity and social status exhibit opposing forces with respect to consensus building in online collaboration networks and their influence should be carefully balanced to ensure a faster consensus. The evaluation framework is provided as an open source project[1].

RQ3: How does consensus develop in collaborative content creation?

The third research question has addressed the development of consensus in collaborative content creation. Experimental results of this research question have complemented the findings of previous two research questions that do not consider aspects of content creation and content evolution. Previous research has shown that some type of content provoke intensive interactions and conversations among users [Iniguez et al., 2014; Tsvetkova

[1] https://git.know-center.tugraz.at/summary/?r=SocialNetworkAnalysis.git

et al., 2016]. Typically, such interactions are evident when users collaborating online do not agree on the provided content (i.e., editing articles in Wikipedia). Sections 3.4 and 3.5 present the works that study the associations between content creation and consensus building in online collaboration networks. First, this research question has aimed to investigate the semantic stability process in collaboration networks, such as Wikipedia, that are driven based on policies, guidelines and community standards. Second, it has studied the dynamics of agreement and disagreement between editors on the content of English Wikipedia. The results demonstrate that larger Wikipedia editions (i.e., characterized with larger contributing communities) exhibit slower semantic stability compared to smaller editions. Furthermore, preliminary results of the analysis of fine grained edit actions in English Wikipedia (i.e., representative example of large editions) show that editing behavior is mostly characterized with an intention to get involved in reducing, extending and again reducing content of Wikipedia articles. Nevertheless, the second most evident editing behavior is reflected with an intention to enhance the provided content of Wikipedia articles by updating or correcting information, deleting old content and inserting new information. One of the contributions is the software solution provided as an open source project[2].

4.2 Implications and Potential Applications

Understanding the factors that influence opinion dynamics and consensus building among users is crucial for designing and implementing services that best support users in online systems. In addition, studying the mechanisms that may turn such processes into a success or into a failure is necessary to facilitate online system designers in developing tools to promote consensus building. This thesis provides a further step towards this ambiguous goal. I am confident that future research can benefit from the methodology, valuable insights and knowledge gained from this thesis. The reminder of this section provides some implications and potential applications.

[2] https://doi.org/10.5281/zenodo.153891

Benefits of an interdisciplinary approach. This thesis has shed light on the question whether theoretical models are actually eligible to study consensus dynamics in real-world collaboration networks. It has demonstrated that the synergy of approaches from social sciences, statistical physics and network science proves valuable for studying consensus dynamics. In this thesis, I show how existing agent-based models can be enhanced through realistic network features and intrinsic user attributes. This approach has revealed non-trivial results and has opened up new research questions and opportunities for real-world applications. In the first place, this approach is applicable in settings that support unanimous consensus as an outcome of a certain process. However, due to it's flexibility, it can be applied in scenarios, the outcome of which is not necessarily an unanimous consensus. For example, our model could facilitate the coordination between arbitrary players or entities. In the following, I provide some examples of potential applications.

Promote consensus. One of the possible applications is enhancing user recommendation services. To illustrate the point, connecting otherwise non-interacting users by recommendations may lead to discussions resolving issues that hinder consensus. This thesis provides very valuable insights that can be applied to design personalized user recommendation algorithms. For example, identifying the most influential users and recommending them to other users could enhance the collaboration process. The most influential users or users with a high social status contribute with high quality content when collaborating with their peers. They are also characterized with a high convincing power, so they can influence the opinions of others and speed up the process of consensus building. Such meaningful interventions create network structures that support consensus building. Another application scenario would be social-economic networks. For example, data on purchasing habits of customers, their demographic features and social status could be used to study relations between social and consuming behavior. Insights gained in this thesis can be applied to aid resource allocation, marketing and recommendation system design.

Support coordination. The approach presented in this thesis could be applied to support the coordination of different entities, for example, in an organization where some hierarchy is in place. A typical scenario would be enforcing new policies that are beneficial for an organization but not so smoothly adopted from all the entities. Our model could facilitate the communication between different levels of the hierarchy to benefit the coordination between them (i.e., which person from a particular level should speak with which person from another level in a hierarchy).

Prevent edit-wars. In platforms such as Wikipedia, where creation of content relies on its users, it is important to uncover how users interact with content and with each other. The results of this thesis imply that identifying the dynamics of such interactions help us understand mechanisms that prompt or hinder interests of Wikipedia communities. Edit-wars in Wikipedia present an example of heated arguments between editors over certain issues or topics. When content creation is the only form of expressing own arguments and opinions, such edit-wars become a threat to the growth and stability of online encyclopedia as Wikipedia. The knowledge obtained from this thesis could be used to foster the design of improved guidelines, policies and tools to prevent edit-wars.

4.3 Limitations

In this section I discuss the limitations of this thesis.

Dataset restrictions. This thesis considers snapshots of datasets (i.e., StackExchange and Reddit) when constructing empirical collaboration networks. In reality, collaboration networks change over time, new users may join, others may leave the site. Opinion and consensus dynamics are affected by the existing network structures. But, new connections between users may emerge or existing ones may disappear as a result of opinion and consensus dynamics. One possible solution to gain insights on the dynamics of networks is to compare results of dataset snapshots between different points in time. Our work on the development of consensus in collaborative content creation makes the first steps in this direction. We analyze editing behavior in Wikipedia over a period of 15 years, however, we do not investigate how collaboration networks of editors change over time. This is the very next step planned for our future work. In this line, this thesis also considers static reputation scores as a proxy for social status. Reputation scores are dynamic and they depend on users activities and the perception by their peers. By comparing dataset snapshots of different points in time, we could find out how user reputation scores change over time and how this relates to consensus dynamics.

Methodological restrictions. Methodologically, this thesis applies some simplifications that may be seen as limitations. First, it uses a simplification for opinions exchanged among users, by presenting them as a fixed set of numbers. Such simplifications are indeed arguable, but very often not too far from reality. Very often we have a choice between a fixed number of options when forming an opinion. An alternative would be to use the real content of opinions exchanged among users. Similarly, social status is also presented with a single number (i.e., reputation scores) that for certain scenarios may be too simplistic. For example, users often play different roles in collaboration networks and a non-trivial interplay between the roles and status may exist. Even though, such reputation scores typically have no explicit value, they still act as symbols of social status within an online community.

Empirical evidence of consensus. This thesis lacks of validation of consensus by means of it's notion. Extracting the notion of consensus from real-world collaboration networks is a non-trivial task. The notion of consensus itself depends on the context of collaboration networks. For example, in StackExchange consensus would mean that a conversation is over? Or some people leave the conversation and others agree on an answer? One possibility would be to observe and follow a thread in a StackExchange site and find out which answer was accepted as the best one. The author of that answer could be tracked and marked as the most trusted one, opinions of which should win in our Naming Game experiments as well. In a co-authorship network the notion of consensus is totally different and much more extensive to track. Aiming to validate the model with a notion of consensus from the real world, might compromise the generalizability of the corresponding approach.

Generality of findings. Even though this thesis studies four types of collaboration networks, the generality of the findings is restricted to the evaluated datasets. The characteristics of empirical observations are guided by specific collaboration networks studied. However, this thesis aims at providing a general solution to enhance the consensus building when, for example, a collaboration network is known and its degree distribution and social status distribution are given. Thus, the presented methodology is comprehensive and can be easily adapted for arbitrary collaboration networks.

4.4 Future Work

Finally, I state some potential future works that are also implicated from the limitations of this thesis.

Extending the analysis on opinion dynamics. The evaluation framework presented in this thesis has great potential to be extended with other opinion dynamics models. For example, the evaluation of opinions possessed by two agents could be extended to determine not only if two opinions are the same or not, but, also how similar they are. This implies that the framework could be enhanced with the implementation and extension of existing continuous models. Furthermore, we plan to extend our framework to account for the content of opinions by utilizing text mining and sentiment analysis approaches.

Extending the analysis on network dynamics. To account not only for the role of network structure in opinion dynamics, but also for the role of opinion dynamics in the adjustments of the network structure, it is necessary to study the evolution of collaboration networks over time. An important analysis would be the appearance and disappearance of individual nodes and edges as well as the evolution of network metrics over time. This would give insights into network dynamics and would establish a next step towards a general framework for both opinion dynamics and network dynamics.

Implementing a generic model. One of the next future works is to extend the study on development of consensus in collaborative content creation. So, we plan to establish an agent-based model that mimics the editing process in Wikipedia. We would run experiments with synthetic editors and synthetic parameters and than exploit the results presented in this thesis to construct empirical networks. As the next step, would be to evaluate and calibrate the parameters based on empirical data.

User recommendations. An interesting new direction for future work would be to incorporate the results of this thesis in user recommender services. For example, services of the recommender framework known as ScaR (Scalable Recommendation-as-a-service) [Lacic et al., 2014, 2017] could be enhanced with the insights obtained from this thesis regarding the role of users social status and users similarities.

Opinion dynamics and consensus building will continue to be on the focus of future research, especially, considering the mass exposure to social media, user engagement in online systems and the speed of information and opinion spread in such systems. With this thesis, I hope to have provided incentives for future research to study consensus dynamics from different perspectives and aspects. I believe, that the research community will benefit from the framework and methodology developed in this thesis.

List of Figures

List of Algorithms

List of Tables

Bibliography

Abrams, D. M. and Strogatz, S. H. (2003). Linguistics: Modelling the dynamics of language death. *Nature*, 424:900–900.

Adamic, L. and Adar, E. (2001). Friends and neighbors on the web. *Social Networks*, 25:211–230.

Adler, B. T. and de Alfaro, L. (2007). A content-driven reputation system for the wikipedia. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 261–270, New York, NY, USA. ACM.

Akcora, C. G., Carminati, B., and Ferrari, E. (2013). User similarities on social networks. *Social Network Analysis and Mining*, 3(3):475–495.

Anderson, A., Huttenlocher, D., Kleinberg, J., and Leskovec, J. (2012). Effects of user similarity in social media. In *WSDM '12*, pages 703–712, New York, NY, USA. ACM.

Anderson, A., Huttenlocher, D., Kleinberg, J., and Leskovec, J. (2013). Steering user behavior with badges. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 95–106, New York, NY, USA. ACM.

Arneson, R. (2013). Egalitarianism. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Summer 2013 edition.

Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. 41:203–226.

Baronchelli, A. (2018). The emergence of consensus: a primer. *Royal Society Open Science*, 5(2).

Baronchelli, A., Dall'Asta, L., Barrat, A., and Loreto, V. (2005). Strategies for fast convergence in semiotic dynamics. In Rocha, L. M. and et al., editors, *Artificial Life X*, pages 480–485. MIT Press.

Baronchelli, A., Dall'Asta, L., Barrat, A., and Loreto, V. (2007). Nonequilibrium phase transition in negotiation dynamics. *Phys. Rev. E*, 76(5):051102.

Baronchelli, A., Felici, M., Caglioti, E., Loreto, V., and Steels, L. (2006). Sharp transition towards shared vocabularies in multi-agent systems. *J. Stat. Mech.*, (P06014).

Baronchelli, Andrea and Dall'Asta, Luca and Barrat, Alain and Loreto, Vittorio (2006). Topology-induced coarsening in language games. *Phys. Rev. E*, 73:015102.

Bender, E. A. and Canfield, E. (1978). The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, 24(3):296 – 307.

Biancani, S. (2014). Measuring the Quality of Edits to Wikipedia. In *Proceedings of The International Symposium on Open Collaboration*, OpenSym '14, New York, NY, USA. ACM.

Blondel, V. D., Hendrickx, J. M., and Tsitsiklis, J. N. (2010). Opinion dynamics for agents with opinion-dependent connections. In *49th IEEE Conference on Decision and Control (CDC)*, pages 6626–6632. IEEE.

Bochman, O. (2012). The consensus game: Modeling peer decision protocols. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, WikiSym '12, pages 37:1–37:7, New York, NY, USA. ACM.

Bonabeau, E., Theraulaz, G., and Deneubourg, J.-L. (1995). Phase diagram of a model of self-organizing hierarchies. *Physica A: Statistical Mechanics and its Applications*, 217(3):373 – 392.

Brandes, U., Kenis, P., Lerner, J., and van Raaij, D. (2009). Network analysis of collaboration structure in wikipedia. In *Proceedings of the*

18th International Conference on World Wide Web, WWW '09, pages 731–740, New York, NY, USA. ACM.

Brigatti, E. (2008). Consequence of reputation in an open-ended naming game. *Phys. Rev. E*, 78(4):046108.

Burghardt, K., Alsina, E. F., Girvan, M., Rand, W., and Lerman, K. (2017). The myopia of crowds: Cognitive load and collective evaluation of answers on stack exchange. *PLOS ONE*, 12(3):1–19.

Castellano, C. (2005). Effect of network topology on the ordering dynamics of voter models. In *AIP Conf. Proc. 779: Modeling Cooperative Behavior in the Social Sciences*, pages 114–120.

Castellano, C., Fortunato, S., and Loreto, V. (2009). Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81:591–646.

Castellano, C., Vilone, D., and Vespignani, A. (2003). Incomplete ordering of the voter model on small-world networks. *Europhys. Lett.*, 63(1):153–158.

Chen, P. and Redner, S. (2005a). Consensus formation in multi-state majority and plurality models. *Journal of Physics A: Mathematical and General*, 38(33):7239.

Chen, P. and Redner, S. (2005b). Majority rule dynamics in finite dimensions. *Phys. Rev. E*, 71:036101.

Clifford, P. and Sudbury, A. (1973). A model for spatial conflict. *Biometrika*, 60(3):581–588.

Daley, D. J. and Kendall, D. G. (1964). Epidemics and rumours. *Nature*, 204(1118).

Dall'Asta, L., Baronchelli, A., Barrat, A., and Loreto, V. (2006a). Agreement dynamics on small-world networks. *EPL (Europhysics Letters)*, 73(6):969.

Dall'Asta, L., Baronchelli, A., Barrat, A., and Loreto, V. (2006b). Non-equilibrium dynamics of language games on complex networks. *Phys. Rev. E*, 74:036105.

d'Aquin, M., Adamou, A., Dietze, S., Fetahu, B., Gadiraju, U., Hasani-Mavriqi, I., Holtz, P., Kimmerle, J., Kowald, D., Lex, E., López-Sola, S., Maturana, R. A., Sabol, V., Troullinou, P., and Veas, E. E. (2017). AFEL: towards measuring online activities contributions to self-directed learning. In *Proceedings of the 7th Workshop on Awareness and Reflection in Technology Enhanced Learning co-located with the 12th European Conference on Technology Enhanced Learning (EC-TEL 2017), Tallinn, Estonia, September 12, 2017.*

David, E. and Jon, K. (2010). *Networks, Crowds, and Markets: Reasoning About a Highly Connected World.* Cambridge University Press, New York, NY, USA.

de Oliveira, V. M., Gomes, M., and Tsang, I. (2006). Theoretical model for the evolution of the linguistic diversity. *Physica A: Statistical Mechanics and its Applications*, 361(1):361 – 370.

Debole, F. and Sebastiani, F. (2003). Supervised Term Weighting for Automated Text Categorization. In *Proceedings of the 2003 ACM Symposium on Applied Computing*, SAC '03, New York, NY, USA. ACM.

DeDeo, S. (2016). Conflict and computation on wikipedia: A finite-state machine analysis of editor interactions. *Future Internet*, 8(3).

Deffuant, G., Neau, D., Amblard, F., and Weisbuch, G. (2000). Mixing beliefs among interacting agents. *Advances in Complex Systems*, 03(01n04):87–98.

Ehmann, K., Large, A., and Beheshti, J. (2008). Collaboration in context: Comparing article evolution among subject disciplines in wikipedia. *First Monday*, 13(10).

Erdös, P. and Rényi, A. (1959). On random graphs i. *Publicationes Mathematicae Debrecen*, 6:290.

Fernández-Gracia, J., Suchecki, K., Ramasco, J. J., San Miguel, M., and Eguíluz, V. M. (2014). Is the voter model a model for voters? *Phys. Rev. Lett.*, 112:158701.

Flöck, F., Erdogan, K., and Acosta, M. (2017). Toktrack: A complete token provenance and change tracking dataset for the english wikipedia. In *Proceedings of the Eleventh International Conference on Web and Social Media, ICWSM 2017, Montréal, Québec, Canada, May 15-18, 2017.*, pages 408–417.

Friedkin, N. and Johnsen, E. (1990). Social influence and opinions. 15:193–206.

Friedkin, N. and Johnsen, E. (1999). Social influence networks and opinion change. 16.

Friedkin, N. E. (1998). *A Structural Theory of Social Influence*. Structural Analysis in the Social Sciences. Cambridge University Press.

Friedkin, N. E., Proskurnikov, A. V., Tempo, R., and Parsegov, S. E. (2016). Network science on belief system dynamics under logic constraints. *Science*, 354(6310):321–326.

Galam, S. (2000). Real space renormalization group and totalitarian paradox of majority rule voting. *Physica A Statistical Mechanics and its Applications*, 285:66–76.

Galam, S. (2002). Minority opinion spreading in random geometry. *Eur. Phys. J. B*, 25(4):403–406.

Gandica, Y., Carvalho, J., and Sampaio dos Aidos, F. (2015). Wikipedia editing dynamics. *Phys. Rev. E*, 91:012824.

Gandica, Y., dos Aidos, F. S., and Carvalho, J. (2014). The dynamic nature of conflict in wikipedia. *EPL (Europhysics Letters)*, 108(1):18003.

Gao, Y., Chen, G., and Chan, R. H. M. (2013). Naming game on networks: Let everyone be both speaker and hearer. *CoRR*, abs/1310.4169.

Garcia, D., Mavrodiev, P., Casati, D., and Schweitzer, F. (2017). Understanding popularity, reputation, and social influence in the twitter society. *Policy and Internet*, 9(3):343–364.

Goldhamer, H. and Shils, E. A. (1939). Types of power and status. *American Journal of Sociology*, 45(2):171–182.

Görögh, E., Vignoli, M., Gauch, S., Blümel, C., Kraker, P., Hasani-Mavriqi, I., Luzi, D., Walker, M., Toli, E., and Sifacaki, E. (2017). Opening up new channels for scholarly review, dissemination, and assessment. In *Proceedings of the 13th International Symposium on Open Collaboration*, OpenSym '17, pages 6:1–6:11, New York, NY, USA. ACM.

Guha, R., Kumar, R., Raghavan, P., and Tomkins, A. (2004). Propagation of trust and distrust. In *Proceedings of the 13th International Conference on World Wide Web*, WWW '04, pages 403–412, New York, NY, USA. ACM.

Hajian, B. and White, T. (2011). Measuring Semantic Similarity using a Multi-tree Model. In *Proceedings of the 9th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems*, ITWP 2011. CEUR Workshop Proceedings.

Halavais, A., Kwon, K., Havener, S., and Striker, J. (2014). Badges of friendship: Social influence and badge acquisition on stack overflow. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, pages 1607–1615.

Hasani-Mavriqi, I. (2011). Supporting creation of networked knowledge by automatically generated links. In *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*, i-KNOW '11, pages 14:1–14:8, New York, NY, USA. ACM.

Hasani-Mavriqi, I., Flöck, F., Maribel, A., Lex, E., and Helic, D. (2018a). Empirical analysis of editing dynamics in english wikipedia. *Work in Progress.*

Hasani-Mavriqi, I., Geigl, F., Pujari, S. C., Lex, E., and Helic, D. (2015). The influence of social status on consensus building in collaboration networks. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, ASONAM '15, pages 162–169, New York, NY, USA. ACM.

Hasani-Mavriqi, I., Geigl, F., Pujari, S. C., Lex, E., and Helic, D. (2016). The influence of social status and network structure on consensus building in collaboration networks. *Social Network Analysis and Mining*, 6(1):1–17.

Hasani-Mavriqi, I., Kowald, D., Helic, D., and Lex, E. (2018b). Consensus dynamics in online collaboration systems. *Computational Social Networks*, 5(1):24.

Hasani-Mavriqi, I., Leitner, H., Helic, D., and Maurer, H. (2011). Implementation of a wiki-based information and communication system for academia europapaea. In *Proceedings of the ITI 2011, 33rd International Conference on Information Technology Interfaces*, pages 193–198.

Hegselmann, R., Krause, U., et al. (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3).

Helic, D., Hasani-Mavriqi, I., Wilhelm, S., and Strohmaier, M. (2011). The effects of navigation tools on the navigability of web-based information systems. In *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*, i-KNOW '11, pages 16:1–16:8, New York, NY, USA. ACM.

Holley, R. A. and Liggett, T. M. (1975). Ergodic theorems for weakly interacting infinite systems and the voter model. *The Annals of Probability*, 3(4):643–663.

Immorlica, N., Stoddard, G., and Syrgkanis, V. (2015). Social status and badge design. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pages 473–483, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Iniguez, G., Török, J., Yasseri, T., Kaski, K., and Kertesz, J. (2014). Modeling social dynamics in a collaborative environment. *EPJ Data Science*.

Jackson, J. M. (1987). *Social Impact Theory: A Social Forces Model of Influence*, pages 111–124. Springer New York, New York, NY.

Jurgens, D. and Lu, T.-C. (2012). Temporal motifs reveal the dynamics of editor interactions in wikipedia. In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media (ICWSM)*, pages 162–169.

Kalyanasundaram, A., Wei, W., Carley, K. M., and Herbsleb, J. D. (2015). An Agent-based Model of Edit Wars in Wikipedia: How and when is Consensus Reached. In *Proceedings of the 2015 Winter Simulation Conference*, WSC '15, Piscataway, NJ, USA. IEEE Press.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Katz, E. and Lazarsfeld, P. F. (1955). *Personal influence: the part played by people in the flow of mass communications.* Free Press, New York, NY, USA.

Ke, J., Minett, J. W., Au, C.-P., and Wang, W. S.-Y. (2002). Self-organization and selection in the emergence of vocabulary. *Complexity*, 7(3):41–54.

Keegan, B. C., Lev, S., and Arazy, O. (2016). Analyzing organizational routines in online knowledge collaborations: A case for sequence analysis in cscw. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 1065–1079, New York, NY, USA. ACM.

Klemm, K., Eguíluz, V. M., Toral, R., and Miguel, M. S. (2003). Global culture: A noise-induced transition in finite systems. *Phys. Rev. E*, 67:045101.

Kopeinik, S., Kowald, D., Hasani-Mavriqi, I., and Lex, E. (2017). Improving collaborative filtering using a cognitive model of human category learning. *The Journal of Web Science*, 2(4):45–61.

Krause, U. (2011). Opinion dynamics–local and global. In *Proceedings of the International Workshop Future Diretions in Difference Equations*, pages 113–119.

Kuperman, M. N. (2006). Cultural propagation on social networks. *Phys. Rev. E*, 73:046139.

L. Dall'Asta, A. Baronchelli, A. Barrat, and V. Loreto (2006). Nonequilibrium dynamics of language games on complex networks. *Phys. Rev. E*, 74:036105.

Lacic, E., Kowald, D., and Lex, E. (2017). Tailoring recommendations for a multi-domain environment. In *RecSysKTL*, volume 1887 of *CEUR Workshop Proceedings*, pages 42–45. CEUR-WS.org.

Lacic, E., Kowald, D., Parra, D., Kahr, M., and Trattner, C. (2014). Towards a scalable social recommender engine for online marketplaces: The case of apache solr. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14 Companion, pages 817–822, New York, NY, USA. ACM.

Lambiotte, R. (2007). How does degree heterogeneity affect an order-disorder transition? *EPL (Europhysics Letters)*, 78(6):68002.

Lambiotte, R. and Redner, S. (2007). Dynamics of vacillating voters. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(10):L10001.

Latané, B. (1981). The psychology of social impact. *American Psychologist*, 36:343–365.

Latané, B. (1996). Dynamic social impact: The creation of culture by communication. *Journal of Communication*, 46(4):13–25.

Leskovec, J., Adamic, L. A., and Huberman, B. A. (2007). The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)*, 1(1):5.

Leskovec, J., Huttenlocher, D., and Kleinberg, J. (2010). Governance in social media: A case study of the Wikipedia promotion process. In *ICWSM 2010*.

Li, B., Chen, G., and Chow, T. W. (2013). Naming game with multiple hearers. *Communications in Nonlinear Science and Numerical Simulation*, 18(5):1214 – 1228.

Li, D., Han, D., Ma, J., Sun, M., Tian, L., Khouw, T., and Stanley, H. E. (2017). Opinion dynamics in activity-driven networks. *EPL (Europhysics Letters)*, 120(2):28002.

Liu, J. and Ram, S. (2011). Who does what: Collaboration patterns in the wikipedia and their impact on article quality. *ACM Trans. Manage. Inf. Syst.*, 2(2):11:1–11:23.

Liu, R.-R., Wang, W.-X., Lai, Y.-C., Chen, G., and Wang, B.-H. (2011). Optimal convergence in naming game with geography-based negotiation on small-world networks. *Physics Letters A*, 375(3):363 – 367.

Lorenz, J. (2007). Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C*, 18(12):1819–1838.

Lu, Q., Korniss, G., and Szymanski, B. (2009). The naming game in social networks: community formation and consensus engineering. *Journal of Economic Interaction and Coordination*, 4(2):221–235.

Maity, S. K., Porwal, A., and Mukherjee, A. (2013). Understanding how learning affects agreement process in social networks. In *Social Computing (SocialCom), 2013 International Conference on*, pages 228–235.

Markovsky, B., Skvoretz, J., Willer, D., Lovaglia, M. J., and Erger, J. (1993). The seeds of weak power: An extension of network exchange theory. *American Sociological Review*, 58(2):197–209.

Martin, T., Ball, B., and Newman, M. E. J. (2016). Structural inference for uncertain networks. *Phys. Rev. E*, 93:012306.

Metropolis, N. and Ulam, S. (1949). The monte carlo method. *Journal of the American Statistical Association*, 44(247):335–341.

Mobilia, M. and Georgiev, I. T. (2005). Voting and catalytic processes with inhomogeneities. *Phys. Rev. E*, 71:046102.

Molloy, M. and Reed, B. (1995). A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6(2-3):161–180.

Movshovitz-Attias, D., Movshovitz-Attias, Y., Steenkiste, P., and Faloutsos, C. (2013). Analysis of the reputation system and user contributions on a question answering website: Stackoverflow. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '13, pages 886–893, New York, NY, USA. ACM.

Muller, P. (2006). Reputation, trust and the dynamics of leadership in communities of practice. *Journal of Management & Governance*, 10(4):381–400.

Müller-Birn, C., Dobusch, L., and Herbsleb, J. D. (2013). Work-to-rule: The Emergence of Algorithmic Governance in Wikipedia. In *Proceedings of the 6th International Conference on Communities and Technologies*, C&T '13, New York, NY, USA. ACM.

Newman, M. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582.

Newman, M. (2010). *Networks: An Introduction.* Oxford University Press, Inc., New York, NY, USA.

Newman, M. E. J. (2003). Mixing patterns in networks. *Phys. Rev. E*, 67:026126.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM REVIEW*, 45:167–256.

Noldus, R. and Van Mieghem, P. (2015). Assortativity in complex networks. *Journal of Complex Networks*.

Nowak, A., Szamrej, J., and Latané, B. (1990). From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review*, 97(3):362–376.

Nowak, M. A. and Krakauer, D. C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences*, 96(14):8028–8033.

Nowak, M. A., Plotkin, J. B., and Krakauer, D. C. (1999). The evolutionary language game. *Journal of Theoretical Biology*, 200(2):147 – 162.

Osman, K. (2013). The Role of Conflict in Determining Consensus on Quality in Wikipedia Articles. In *Proceedings of the 9th International Symposium on Open Collaboration*, WikiSym '13, New York, NY, USA. ACM.

Palla, G., Barabasi, A.-L., and Vicsek, T. (2007). Quantifying social group evolution. *Nature*, 446(7136):664–667.

Papadopoulos, F., Kitsak, M., Serrano, M., Boguñá, M., and Krioukov, D. (2012). Popularity versus Similarity in Growing Networks. *Nature*, 489:537–540.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175.

Pedone, R. and Conte, R. (2001). *The Simmel Effect: Imitation and Avoidance in Social Hierarchies*, pages 149–156. Springer Berlin Heidelberg, Berlin, Heidelberg.

Pettijohn, T. F. (1998). *Psychology: A Connectext*. McGraw-Hill Higher Education, Pennsylvania Plaza, New York City.

Pfeil, U., Zaphiris, P., and Ang, C. S. (2006). Cultural differences in collaborative authoring of wikipedia. *Journal of Computer-Mediated Communication*, 12(1):88–113.

Rosenberg, D., Solan, E., and Vieille, N. (2009). Informational externalities and emergence of consensus. *Games and Economic Behavior*, 66(2):979 – 994.

Rudas, C., Surányi, O., Yasseri, T., and Török, J. (2017). Understanding and coping with extremism in an online collaborative environment: A data-driven modeling. *PLOS ONE*, 12(3):1–16.

Sahlgren, M. (2005). An Introduction to Random Indexing. In *Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering*, TKE '05, Copenhagen, Denmark.

Salton, G. and Buckley, C. (1988). Term-Weighting Approaches in Automatic Text Retrieval. *Information Processing and Management: an International Journal*, 24(5).

Sayama, H. and Sinatra, R. (2015). Social diffusion and global drift on networks. *Phys. Rev. E*, 91:032809.

Scholz, M. (2010). Node similarity is the basic principle behind connectivity in complex networks. *CoRR*, abs/1010.0803.

Schooler, C. (2013). Social class and social status. *Psychology*.

Schulze, C. and Stauffer, D. (2005). Monte carlo simulation of the rise and the fall of languages. *International Journal of Modern Physics C*, 16(05):781–787.

Seitlinger, P., Kowald, D., Kopeinik, S., Hasani-Mavriqi, I., Ley, T., and Lex, E. (2015). Attention please! a hybrid resource recommender mimicking attention-interpretation dynamics. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 339–345, New York, NY, USA. ACM.

Seitlinger, P., Ley, T., Kowald, D., Theiler, D., Hasani-Mavriqi, I., Dennerlein, S., Lex, E., and Albert, D. (2017). Balancing the fluency-consistency tradeoff in collaborative information search with a recommender approach. *International Journal of Human–Computer Interaction*, 34(6):557–575.

Sekara, V., Stopczynski, A., and Lehmann, S. (2016). Fundamental structures of dynamic social networks. *Proceedings of the National Academy of Sciences*, 113(36):9977–9982.

Sepehri-Rad, H. and Barbosa, D. (2015). Identifying controversial wikipedia articles using editor collaboration networks. *ACM Trans. Intell. Syst. Technol.*, 6(1):5:1–5:24.

Shirakawa, M., Nakayama, K., Hara, T., and Nishio, S. (2013). Probabilistic semantic similarity measurements for noisy short texts using Wikipedia entities. CIKM '13, New York, NY, USA. ACM.

Simmel, G. (1957). Fashion. *The American Journal of Sociology*, 62(6):541–558.

Singer, P., Helic, D., Hotho, A., and Strohmaier, M. (2015). Hyptrails: A bayesian approach for comparing hypotheses about human trails on the web. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pages 1003–1013, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Singer, P., Helic, D., Taraghi, B., and Strohmaier, M. (2014). Detecting memory and structure in human navigation patterns using markov chain models of varying order. *PLOS ONE*, 9.

Sire, C. and Majumdar, S. N. (1995). Coarsening in the q-state potts model and the ising model with globally conserved magnetization. *Phys. Rev. E.*, 52:244–254.

Sood, V. and Redner, S. (2005). Voter model on heterogeneous graphs. *Phys. Rev. Lett.*, 94:178701.

Stanisavljevic, D., Hasani-Mavriqi, I., Lex, E., Strohmaier, M., and Helic, D. (2016). Semantic stability in wikipedia. In *Complex Networks & Their Applications V*, pages 379–390, Cham. Springer International Publishing.

Stauffer, D. (2002). Sociophysics: the sznajd model and its applications. *Computer Physics Communications*, 146(1):93 – 98. Proceedings of the STATPHYS Satellite Conference: Challenges in Computational Statistical Physics in teh 21st Century.

Stauffer, D. (2003). Phase transition in hierarchy model of bonabeau. *International Journal of Modern Physics C*, 14(02):237–239.

Stauffer, D., Castelló, X., Eguíluz, V. M., and Miguel, M. S. (2007). Microscopic abrams–strogatz model of language competition. *Physica A: Statistical Mechanics and its Applications*, 374(2):835 – 842.

Stauffer, D., Sousa, A., and Schulze, C. (2004). Discretized opinion dynamics of the deffuant model on scale-free networks. *J. Artificial Societies and Social Simulation*, 7.

Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial Life*, 2:319–332.

Stefanescu, D., Rus, V., Niraula, N. B., and Banjade, R. (2014). Combining Knowledge and Corpus-based Measures for Word-to-Word Similarity. In *Proceedings of the Twenty-Seventh International Florida Artificial Intelligence Research Society Conference*, FLAIRS '14, Palo Alto, California. AAAI Press.

Sumi, R., Yasseri, T., Rung, A., Kornai, A., and Kertesz, J. (2011). Edit wars in wikipedia. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 724–727.

Sznajd-Weron, K. (2005). Sznajd model and its applications. HSC Research Reports HSC/05/04, Hugo Steinhaus Center, Wroclaw University of Technology.

Sznajd-Weron, K. and Sznajd, J. (2000). Opinion evolution in closed community. *International Journal of Modern Physics C*, 11(06):1157–1165.

Takale, S. A. and Nandgaonkar, S. S. (2010). Measuring Semantic Similarity between Words Using Web Documents. *International Journal of Advanced Computer Science and Applications, IJACSA*, 1(4).

Tang, J., Lou, T., and Kleinberg, J. (2012). Inferring social ties across heterogenous networks. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, pages 743–752, New York, NY, USA. ACM.

Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., and Su, Z. (2008). Arnetminer: Extraction and mining of academic social networks. In *KDD'08*, pages 990–998.

Tausczik, Y. R., Kittur, A., and Kraut, R. E. (2014). Collaborative problem solving: A study of mathoverflow. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, pages 355–367, New York, NY, USA. ACM.

Tischler, H. (2010). *Introduction to Sociology.* Wadsworth Publishing Company.

Török, J., Iñiguez, G., Yasseri, T., San Miguel, M., Kaski, K., and Kertész, J. (2013). Opinions, conflicts, and consensus: Modeling social dynamics in a collaborative environment. *Phys. Rev. Lett.*, 110.

Trattner, C., Hasani-Mavriqi, I., Helic, D., and Leitner, H. (2010). The austrian way of wiki(pedia)!: Development of a structured wiki-based encyclopedia within a local austrian context. In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*, WikiSym '10, pages 9:1–9:10, New York, NY, USA. ACM.

Tsvetkova, M., García-Gavilanes, R., Floridi, L., and Yasseri, T. (2017). Even good bots fight: The case of wikipedia. *PLOS ONE*, 12(2):1–13.

Tsvetkova, M., García-Gavilanes, R., and Yasseri, T. (2016). Dynamics of disagreement: Large-scale temporal network analysis reveals negative interactions in online collaboration. *CoRR*, abs/1602.01652.

Turney, P. D. and Pantel, P. (2010). From Frequency to Meaning Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, 37(1).

Urdan, T. C. (2010). *Statistics in Plain English.* Routledge, New York, NY, USA.

Vazquez, F., Krapivsky, P. L., and Redner, S. (2003). Constrained opinion dynamics: freezing and slow evolution. *Journal of Physics A: Mathematical and General*, 36(3):L61.

Waagen, A., Verma, G., Chan, K., Swami, A., and D'Souza, R. (2015). Effect of zealotry in high-dimensional opinion dynamics models. *Phys. Rev. E*, 91:022811.

Wagner, C., Singer, P., Strohmaier, M., and Huberman, B. A. (2014). Semantic Stability in Social Tagging Streams. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, New York, NY, USA. ACM.

Walker, H. A., Thye, S. R., Simpson, B., Lovaglia, M. J., Willer, D., and Markovsky, B. (2000). Network exchange theory: Recent developments and new directions. *Social Psychology Quarterly*, 63(4):324–337.

Wang, W. X., Lin, B. Y., Tang, C. L., and Chen, G. R. (2007). Agreement dynamics of finite-memory language games on networks. *The European Physical Journal B*, 60(4):529–536.

Watts, D. J. (1999). Networks, dynamics, and the small world phenomenon. *American Journal of Sociology*, 105(2):493–527.

Watts, D. J. (2004). *Six Degrees the Science of a Connected Age.* W. W. Norton and Company, 500 Fifth Avenue, New York.

Webber, W., Moffat, A., and Zobel, J. (2010). A similarity Measure for Indefinite Rankings. *ACM Transactions on Information Systems, TOIS*, 28(4).

Weber, M. (1964). *The theory of social and economic organization.* The Free Press.

Willer, D. (1999). *Network Exchange Theory.* Praeger, Westport, CT.

Xia, H., Wang, H., and Xuan, Z. (2011). Opinion dynamics: A multidisciplinary review and perspective on future research. *Int. J. Knowl. Syst. Sci.*, 2(4):72–91.

Yang, H.-X., Wang, W.-X., and Wang, B.-H. (2008). Asymmetric negotiation in structured language games. *Phys. Rev. E*, 77:027103.

Yasseri, T. and Kertész, J. (2013). Value production in a collaborative environment. *Journal of Statistical Physics*, 151(3).

Yasseri, T., Sumi, R., Rung, A., Kornai, A., and Kertész, J. (2012). Dynamics of conflicts in wikipedia. *PLOS ONE*, 7(6):1–12.

Zaman, A. (2012). Stop Word Lists in Document Retrieval Using Latent Semantic Indexing: an Evaluation. *Journal of E-Technology*, 3(1).

Zhang, W., Lim, C. C., Korniss, G., and Szymanski, B. K. (2014). Opinion dynamics and influencing on random geometric graphs. *Sci. Rep.*, 4.