# if we interact:

# we change

Sasha Piccione

PhD Student in Management

XXXVI Cycle

Supervisor: Marco Tolotti

Venice School of Management

Ca' Foscari University of Venice

# Contents

**1 Acknowledgements** **4**

**2 Introduction** **5**

2.1 Data usage and ethical issues . . . . . 7

2.2 Terminology . . . . . 10

**3 The role of Simmelian ties and sporadic communication in opinion dynamics with bounded confidence** **14**

3.1 Introduction . . . . . 14

3.2 Literature Review . . . . . 17

3.2.1 The role of network structures: where it happens . . . . . 18

3.2.2 Opinion dynamics: how and when it happens . . . . . 20

3.3 The model . . . . . 22

3.4 Simulations and results . . . . . 27

3.4.1 Types of networks . . . . . 27

3.4.2 Scenarios related to communication . . . . . 28

3.4.3 Simulation procedure . . . . . 30

3.4.4 Numerical results - Descriptive analysis . . . . . 31

3.4.5 Numerical results - Inference analysis . . . . . 34

3.5 Conclusion . . . . . 40

**4 Liking those you met the most** **43**

4.1 Introduction . . . . . 44

4.2 Literature Review . . . . . 46

4.2.1 Online and offline networks . . . . . 48

4.3 Methodology . . . . . 50

4.3.1 Data description and Network construction . . . . . 51

4.3.2 Sentiment Analysis . . . . . 56

4.4 Results and Discussion . . . . . 60

**5 I'm Politer than you just because I think you are** **63**

5.1 Introduction . . . . . 64

5.2 Literature review . . . . . 67

5.2.1 The role of emotions in communication . . . . . 68

5.2.2 Communication and coordination in Online Open Collaborative Communities . . . . . 69

5.2.3 Wikipedia: an exceptional case of anonymous cooperation . . . . . 71

5.2.4 Hypotheses on emotion expression . . . . . 73

5.3 Data and Methods . . . . . 74

5.3.1 Data Collection . . . . . 75

5.3.2 Measuring emotions . . . . . 77

5.3.3 Word Frequency Averaging . . . . . 77

5.3.4 Word embedding . . . . . 78

5.3.5 Variables . . . . . 79

5.3.6 The dependent variables: the comment tone . . . . . 80

5.3.7 Independent variables: the context . . . . . 80

5.4 Results . . . . . 82

5.4.1 Descriptive statistics . . . . . 82

5.4.2 Regression analyses and hypotheses checking . . . . . 82

5.5 Discussion and Conclusion . . . . . 87

**6 Final Discussion** **92**

**7 Appendix** **112**

7.1 Appendix to Chapter I . . . . . 112

7.2 Appendix to Chapter II . . . . . 116

7.3 Appendix to Chapter III . . . . . 117

# 1 Acknowledgements

You start asking yourself what it feels like getting to a point and you end up asking yourself how you got there. This PhD journey is indeed coming to an end and it's now time to look back and see what has been done. This journey allowed me to learn many things, research-related and not.

But, as the core of this thesis points, I wasn't alone during this journey.

A special thank goes to my supervisor, Marco Tolotti who followed me during this journey and who always pushed me to learn new things and meet new people as well as to value myself and my research.

A second mention goes to Nicolas Jullien who followed me for a considerable part of this journey and helped me know more about what a computer is and showed me where Britain is on the map.

I also want to thank Viviana Amati, Rachele Cavara, Caterina Cruciani, Nicolas Duminy, Anna Moretti, Francesco Rullani, Francesco Zirpoli for their inestimable help and input in my research.

I want to thank Marta Colombini, for her enduring support and assistance.

I want to thank my mother, Franca Ricchiuti, silent and solid oak of my life.

I want to thank Gaia Colangelo, Cecilia Ivardi Ganapini, Giulio Migotto, Michele Morabito and Costanza Sartoris for being my closest friends and bringing joy and laughs as well as wisdom to my days.

I want to thank all my friends and colleagues who showed to me their kindest side and tailored their life for allowing enough space for a cup of tea and a good banter.

I want to thank all those unknown individuals who produced massive knowledge and tutorials regarding Python language without which this thesis wouldn't be possible.

I finally want to thank myself, for having done it.

# 2 Introduction

We spend a significant part of our lives perceiving ourselves as autonomous entities, moving independently through the cosmic void and believing that our choices and opinions are the sole result of our own free will. This belief reinforces a comforting illusion of control, allowing us to attribute our successes and failures to internal rather than external factors.

In recent years, however, it has become increasingly evident that the context in which choices are made plays a crucial role in shaping those very decisions. As inherently social beings, the most influential context for us is undoubtedly the social context. Our capacity to use complex language and build intricate relationships has allowed our species not only to dominate others but also to develop mechanisms of self-regulation. We observe and learn from one another, internalizing norms, behaviours, and expectations. For example, a person might take up smoking not because of a deliberate decision but because someone they admire or perceive as "cool" smokes (see Mercken et al., 2010), subtly influencing their behaviour without explicit coercion.

Just as we have long believed ourselves to be independent actors, we have also assumed that only strong relationships—what Granovetter (1973) refers to as "strong ties"[1] - are capable of shaping our decisions and opinions. However, Granovetter's seminal work has demonstrated that weak ties can exert influence as well. This raises a fascinating question: how weak can a tie be before it ceases to have an impact on us? At what threshold does a relationship become strong enough to influence our choices and actions? Could even the most fleeting and seemingly insignificant interactions subtly alter our decisions and behaviours?

This overarching question—whether explicitly or implicitly—has been a driving force behind the research presented in this thesis. It reflects a broader interest in understanding how relationships, contexts, and interactions shape human behaviour, particularly in the digital age.

The central theoretical construct underpinning this thesis is the concept of network. A network is not merely a collection of entities but a structured system in which all elements are interconnected through various forms of interdependencies. These interdependencies can differ

[1] Strong ties are generally considered ties that, due to relational characteristics, entail a stronger relationship. For instance, in the author refers to the "combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie" (Granovetter, 1973, : p. 1361)

in nature depending on the type of network. In social networks, for instance, the focus is on social relationships and the ways in which individuals influence one another through direct and indirect connections. Networks, by their very nature, provide a powerful lens through which to study the dynamics of influence, cooperation, and decision-making.

The first chapter of this thesis investigates interpersonal relationships and opinion dynamics, with a particular focus on bounded confidence models. These models examine how individuals' opinions evolve based on their interactions within a network and the thresholds of tolerance they exhibit toward differing viewpoints. Specifically, we explored whether sporadic communication—introduced to model interaction preferences—affects the outcome of the opinion formation process within a network. By leveraging computer simulations, we were able to test exploratory hypotheses and observe how interaction frequency and structure influence collective opinion dynamics.

The second chapter shifts focus to LinkedIn relationships among individuals involved in the development of the European EURO 7 regulation. This chapter seeks to determine whether informal interactions on LinkedIn are connected to the formal network of relationships derived from official meetings related to EURO 7. Our findings reveal not only a tangible link between these network layers but also a series of structural characteristics that highlight specific social behaviours among the participants. For example, we demonstrate how certain lobbyists, by virtue of their strategic positions within the relational structure, were able to influence key aspects of the regulation that was ultimately adopted in March 2024. This chapter underscores the importance of informal networks in shaping formal outcomes, particularly in high-stakes regulatory environments.

The final chapter examines online cooperative communities and the ways in which interactions between participants shape individual behaviours. Using Wikipedia as a case study, we analyzed 1.4 million interactions between contributors to understand how the tone—evaluated in terms of arousal and valence—expressed by one participant influences the language and behaviour of others. While this research draws on well-established psychological theories such as mimicry and appraisal theory, its novelty lies in the unique context in which these theories are tested. Unlike typical social interactions, Wikipedia contributors lack access to nonverbal cues

or other social signals that are often implicitly used to modulate language and behaviour. Instead, the only available medium of communication is the text itself. This text not only provides instructions on what needs to be done (a concept known as stigmergy) but also conveys subtle social cues that influence how contributors interact with one another. Our findings demonstrate that even in asynchronous, non-reciprocal interactions, the textual environment significantly impacts individual behaviour, offering new insights into the dynamics of collaboration in virtual settings.

The findings presented in this thesis highlight how relationships and interactions between individuals are deeply influenced by their context. Context, in this sense, is not limited to what individuals can directly observe but also includes the topological characteristics of the network, which may not be immediately accessible to individual actors. At the same time, we show how structural and topological features of networks can reveal the presence of complex and nuanced dynamics that go far beyond the surface-level interactions visible to participants.

Each of the three chapters has benefited significantly from the use of the Python programming language. Nearly all of the data collected for these studies was obtained using custom-developed software, tailored to the specific needs of each research project. Since the data analyzed was public and online, the primary challenge lay not in collecting the data but in reorganizing it into a usable format. In practical terms, collecting this data simply involved copying magnetic alterations from a remote server's hard disk to our own. Because the data was publicly accessible, we have committed to fully embracing the principles of open science by making all collected data and the associated Python scripts publicly available. To protect the anonymity of the individuals involved, all published data has been pseudonymized using unique identifiers, and any elements that could potentially reveal personal identities—such as the text produced by contributors—have been removed. This approach ensures that the research adheres to the highest ethical standards while fostering transparency and reproducibility.

## 2.1 Data usage and ethical issues

In the digital age, the widespread availability of data has revolutionized research and business methodologies. A key element of this research is the extensive use of web scraping techniques

to gather data. Web scraping refers to the automated extraction of data from publicly accessible websites. Specifically, Chapter 5, focusing on focusing on Wikipedia, the data collection process was almost entirely automated. Conversely, Chapter 4, devoted on lobbying activities related to the EURO 7 directive, only a portion of the data was collected through web scraping, always in compliance with the platform's Terms of Service (ToS).

Although these techniques have significantly expanded the scope and volume of data that can be collected, they have also raised important ethical concerns, particularly regarding the distinction between public, personal, and private data. These ethical challenges have been accompanied by legal and regulatory interventions, progressively establishing clearer ethical and deontological boundaries. Understanding these boundaries is essential for responsible data collection and management.

Public data refers to information that is openly accessible without restrictions, such as content on public websites, government publications, and data released under open licenses. Ethically, scraping public data is generally acceptable, provided that it does not violate the ToS of the website or places an unnecessary burden on the server. Adhering to guidelines contained in files like "robots.txt", which specify permissible scraping activities, is a standard ethical practice. Moreover, it is critical to ensure that the data is used in alignment with its intended purpose and does not harm the individuals to whom the data pertains.

Social media platforms present a complex and evolving landscape. Users often share personal information voluntarily, sometimes without fully understanding the implications. This dynamic creates a blurred line between public and private data, as accessibility does not negate the sensitive nature of certain information. While the availability of this data has facilitated advancements in fields such as social network analysis, sentiment analysis, and behavioural studies, it also raises significant ethical and legal concerns regarding privacy, consent, and data protection.

The ethical considerations around personal data are grounded in the fundamental principle that individuals have a right to privacy, regardless of whether the data is publicly accessible. Consent is central to this debate: merely sharing personal information on a public platform does not imply consent for processing, analyzing, or repurposing it for unintended purposes.

This tension between accessibility and ethical responsibility becomes particularly acute in cases where personal data is used for commercial gain, surveillance, or profiling without the explicit knowledge or consent of the individual.

The European Union's General Data Protection Regulation (GDPR), effective since 2018, provides a comprehensive framework for addressing these issues. It emphasizes the necessity of explicit and informed consent for processing personal data, even if such data is publicly available. Article 6 of the GDPR outlines the conditions under which data processing is lawful, imposing strict compliance requirements on entities handling personal information, including social media platforms and third-party aggregators.

Before the GDPR's adoption, courts played a crucial role in defining users' rights. For instance, in Google Spain SL, Google Inc. v. Agencia Española de Protección de Datos, Mario Costeja González (2014), the Court of Justice of the European Union (CJEU) introduced the "right to be forgotten", allowing individuals to request the removal of links to personal information that is "inadequate, irrelevant, no longer relevant, or excessive". This ruling underscored the importance of contextual use of data and the potential harm it may cause, balancing the right to privacy with the public's right to access information.

Other cases, such as Bodil Lindqvist (C-101/01, 2003), highlighted that publishing identifiable personal information on a website constitutes data processing under EU law. Similarly, in Patrick Breyer v. Bundesrepublik Deutschland (C-582/14, 2016), the CJEU ruled that data cannot be considered anonymized if it can be combined with other legally obtained information to identify an individual. This distinction reinforces the importance of truly anonymizing data to protect privacy.

Private data, such as emails, private messages, and confidential records, is not publicly accessible. Collecting such data without explicit consent is both unethical and illegal. Unauthorized access violates privacy rights and can result in legal consequences under regulations such as the GDPR and the U.S. Computer Fraud and Abuse Act (CFAA). For instance, in hiQ Labs, Inc. v. LinkedIn Corp., the Ninth Circuit ruled that scraping publicly available data did not violate the CFAA but emphasized the importance of respecting the website's ToS and users' privacy. This ruling illustrates the dual interests at stake in social media scraping: the privacy

of the individuals whose data is collected and the platform's right to enforce its ToS.

Ethical and legal considerations extend beyond data collection to its subsequent use, management, and sharing. For this reason, our research adhered to a strict protocol. For the data collection from Wikipedia, the platform does not impose explicit restrictions on the use of its publicly available data or automated tools. Therefore, we used web crawlers implemented with Selenium to collect the data. After processing, the data was anonymized by removing textual content irrelevant to our analysis. In the case of LinkedIn, we complied with the platform's policy regarding the use of bots or automation tools. Data collection was conducted semi-manually: pages were loaded manually, and the information was subsequently organized and processed.

By following these measures, we ensured that our data collection and processing adhered to ethical standards and legal requirements, balancing research objectives with respect for privacy and platform policies.

At the same time, we fully embraced an open science policy in this research. The data and Python scripts used for data processing are publicly available in an open repository. To ensure the anonymity of the individuals whose data was collected, the published data is pseudonymized using unique identification codes, and any elements that could potentially reveal their identity—including any texts they may have produced—have been removed.

## 2.2 Terminology

This thesis focuses on three main themes: social network analysis, opinion dynamics, and interpersonal interactions. Before delving into the research, it is essential to provide a brief overview and definitions of key terminology to ensure a robust understanding of the concepts used throughout the document.

To facilitate navigation of the terminology employed within this thesis, we have chosen to organize the terms according to their respective chapters, with particular attention to highlighting any nuanced differences in meaning between analogous terms.

One of the most widely used terms throughout this research is "tie". A *tie* refers to a connection or link between two individuals. The specific definition of a *tie* varies depending on the dataset and context under consideration. Unless otherwise specified, *tie* in this thesis refers

to a relationship of friendship or acquaintance. The flexibility of this term allows it to adapt to different types of relationships as dictated by the focus of each study.

In the first chapter, the focus is on concepts such as *opinion*, *trust*, and *bounded confidence*. For the purpose of this chapter, the *opinion* refers to the position an individual holds on a given topic, expressed as a continuum between support and opposition. For example, the opinion of an individual $i$, represented as a value within the continuous interval $[0, 1]$, may reflect their level of support for or opposition to a specific policy. Adopting a single-dimensional scale simplifies the analysis, aligning with the original model by DeGroot (1974). While more complex approaches allow for the simultaneous consideration of multiple opinions or dimensions (Parsegov et al., 2016), the choice of a single dimension in this research is designed to streamline the modeling process and focus on core opinion dynamics.

*Trust* is another crucial concept, although its definition often varies depending on the context and nature of the research. In this thesis, *trust* refers to the perceived reliability or credibility an individual assigns to their neighbours, *i.e.*, the set of people with whom they share a trusted relationship. More specifically, the *trust* that an individual $i$ has in another individual $j$ corresponds to the weight $i$ assigns to the opinion of $j$. This conceptualization allows us to quantify *trust* as a tangible influence factor within opinion dynamics.

The concept of *bounded confidence* describes the conditions under which opinions are constrained and influenced. Broadly speaking, bounded confidence models stipulate that opinion formation is dependent on specific criteria, such as the distance between two individuals' opinions. In this research, *bounded confidence* is modeled as a function of this distance, which determines the extent to which two individuals influence one another.

The second chapter addresses lobbying activities in the context of preparatory work for adopting the European EURO 7 directive. In this chapter, we employ a multiplex network approach to analyse social interactions. Multiplex networks are a specific subset of multilayer networks, which include layers in addition to nodes and edges. These structures enable the representation of systems characterized by multiple levels, types of edges, or other complex features. In the most general framework, each node can belong to any subset of layers, and edges can encompass pairwise connections across all possible combinations of nodes and lay-

ers. This framework is further extensible to include hyperedges, which represent connections among more than two nodes. Additionally, multilayer networks allow for multidimensional layer structures, where each layer aspect—such as edge type or temporal context—introduces an additional dimension to the representation. This versatility makes multilayer networks a powerful tool for modelling complex systems (Kivelä et al., 2014).

In particular, this research focuses on *multiplex networks*, a more constrained subset of multilayer networks where the set of nodes remains identical across all layers, but each layer represents a distinct type of tie (De Domenico et al., 2013). Multiplex networks are commonly used in studying multi relational networks, as they allow for an in-depth analysis of how different layers (or types of relationships) influence one another. In this thesis, multiplex networks provide a detailed and nuanced representation of the social structure of the observed sample. This approach is particularly suited to questions that aim to identify the impact one layer or network has on another.

In the third chapter, the focus shifts to a significantly different setting, examining emotions and the influence of context on them. Emotions are generally defined as complex psychological states that involve subjective experiences, physiological responses, and behavioural expressions. Building on established literature in emotional psychology, this research conceptualizes emotions expressed in written text along two primary dimensions: *arousal* and *valence*. *Arousal* represents the intensity or activation level of the emotion, while *valence* captures the positivity or negativity of the expressed emotion. This dual-dimensional approach provides a comprehensive framework for analysing emotional tone in text-based communication.

Understanding the concept of context is crucial in this chapter, particularly given the unique characteristics of Wikipedia as a platform. Specifically, the *Talk Pages*—discussion pages associated with each Wikipedia article—serve as the primary context for this analysis. The context of a *Talk Page* is further subdivided into distinct components. First, the *text corpus* includes the entire collection of comments posted on the page. Second, the *thread* refers to a topic initiated by one individual and all subsequent comments within that topic. Finally, the *dyadic relationship* encompasses all comments exchanged directly between two individuals in response to one another. Whenever we refer to the context of a comment, we specifically mean the content

present within the *Talk Page*—organized as described above—at the time the comment was posted. Defining context in this granular manner required significant methodological effort and represents one of the primary contributions of this research.

By carefully delineating these terms and approaches, this thesis establishes a strong foundation for the studies presented in subsequent chapters. This rigorous framework not only enhances clarity but also provides valuable insights into the intricate interplay between opinion dynamics, social relationships, and emotional expression.

# 3 The role of Simmelian ties and sporadic communication in opinion dynamics with bounded confidence

**Sasha Piccione and Marco Tolotti**

**Abstract**

In this article we study how the network structure and the timing of communication influence the opinion dynamic process when trust between subjects is bounded by personal traits and beliefs. Inspired by the classical DeGroot model of consensus, we propose a novel analytical model capable to describe *sporadic* pairwise interactions, influenced by personal traits, strength of ties, trust, and other social/network factors. Our findings indicate that sporadic communication affects more the most influential actors. In this respect, continuity in communication is essential for leaders. Additionally, we show that Simmelian ties shape influential actors, particularly when communication is fluent.

**Keywords**: Bounded confidence; network structure; opinion dynamics; Simmelian ties; sporadic communication

## 3.1 Introduction

Understanding how opinion spreads within organizations or groups appears of crucial relevance in different fields. The literature regarding advocacy coalitions (Schlager, 1995) has highlighted the relevance of the role of interpersonal relations between subjects in the formation of sets of beliefs (Larsen et al., 2006) and the consequent policy formation process (Howlett, 2002). Starting from the traditional DeGroot model (1974), researchers have tried to develop mathematical models to represent the interaction among agents and opinion formation based on such interactions by either leaning toward non deterministic models (Choi et al., 2021) or by allowing agents' behaviour to change (Zhou and Wu, 2022). A second stream of literature, closer to political studies, proposed opinion dynamics models characterized by bounded confidence (Hegselmann and Krause, 2002). Here, two subjects interact only if the distance between their opinions is below a certain threshold. Those studies investigate *how* opinions and beliefs are formed. However, in this literature, less emphasis has been placed on other important aspects of

the opinion dynamics process: *where* and *when*—specifically, which structural characteristics significantly influence opinion formation and the pace at which communication occurs.

Borrowing notions and theories coming from the literature on information flows in organisational networks (Haas, 2015), we identify a set of major topological characteristics to be considered when studying opinion formation. Specifically, we devote our attention to a specific segment of such vast literature trying to complement the classical theories of information transfer with the studies of networks and their properties (Hansen, 1999, Tortoriello and Krackhardt, 2010)). Our purpose is to take inspiration from information transfer theories to specify the role that structural characteristics of a network have on the opinion formation process.

Among the structural characteristics that, according to information transfer literature, foster the transfer of information, we decided to focus on strong relationships (Granovetter, 1973). In particular, due to the difficulty of assessing and measuring the strength of a tie, we rely on the concept of *Simmelian ties* (Simmel, 1950, Krackhardt, 1998), *i.e.*, ties within a triad, which have often been used as signaling the presence of a relationship that is stronger if compared to generic dyads.

Starting from the assumption that each Simmelian tie is a strong tie allows us to bypass the issue of measuring and calculating tie strength. However, it remains unclear whether this strength arises from the social dynamics that such a structural configuration implies or if it is somehow linked to topological characteristics.

We also consider the fact that individuals may convey a message with different frequency based on their personal preferences or position in the network and the traits of the recipient of the message. To this aim, we introduce in the process a random component modelling intermittent (sporadic) communication, whose properties explicitly depend on the personal traits of the agents involved in the communication. Communication is not a deterministic process, rather, interactions may happen (or not) at different time due to internal or external factors. Some researches have explored sporadic or fuzzy interactions (Choi et al., 2021) or fuzzy linguistic (Giráldez-Cru et al., 2021). However, in these papers, authors do not discuss how personal (or social) traits interfere with preferences in interaction (Xia et al., 2020). In this research, we divide the *communication episode* (Yale and Gilly, 1995) into two parts, recognizing the dis-

tinction between the preferences of the individual sharing their opinion and those of the one receiving it. Specifically, we represent an individual's willingness to share their opinion with a neighbor as a sharing probability. Additionally, drawing from recent advances in the literature on bounded confidence models (Hegselmann and Krause, 2002), we consider the probability of interaction between two subjects to depend on the distance between their opinions at a given moment. On the other hand, the agency and preferences of the opinion receiver are represented, as in the classic DeGroot model, by the level of trust they place in the received opinion. Summing up, we propose a new formal model capable to account for agent's preferences and the social context (differences in terms of hierachical levels within an organization), personal traits (bounded confidence), relational structure (Simmelian ties) and specific traits of the communication process itself (sporadic communication). The primary objective of this study is to explore, by means of a novel analytical model, opinion dynamics with a focus on the influence of the network structures (*where* it occurs) and the mechanisms governing the consensus formation process (*how* and *when* it occurs). From a methodological viewpoint, we build an opinion dynamics model, inspired by the pioneering DeGroot (1974) model and by some more recent developments in the field of diffusion dynamics (see Assenova, 2018). Our model allows for sporadic (random) pairwise interactions that may depend on several antecedents such as the distance of opinions between subjects and other structural characteristics mimicking complex social structures (Choi et al., 2021). At the same time, our analytical model is capable to contribute to the theoretical framework studying the role of Simmelian ties in information transfer from which we take inspiration (Tortoriello and Krackhardt, 2010, Gómez-Solórzano et al., 2019). The goal is to understand whether, in the case of opinion formation *à la* DeGroot, Simmelian ties play a role also by virtue of their topological characteristics and not only by social traits, as suggested by Gómez-Solórzano et al. (2019). We further enhance the model by adding a personal preference aspect that influences the weight an individual assigns to the opinion distance with their neighbors. This introduces an idiosyncratic preference that is strictly subject-specific. In this research, we test this preference element by using hierarchical "classes" as a simplifying device; that is, the weight an individual assigns to opinion distance with another depends on both their own class and that of the other individual. This approach allows us

to envision our model applied to organizational contexts with a hierarchical structure, where the goal is to investigate the role that differences in hierarchical level may play in opinion formation processes.

Specifically far as results are concerned, we show that sporadic communication may negatively impact influential actors, especially in the presence of interaction imbalances among subjects. Indeed, some individuals that have a large social influence in the deterministic opinion dynamics *à la* DeGroot, result to suffer from intermittent communication. Moreover, they also suffer a larger variance in the possible outcome of their relative influence. Therefore, especially for leaders role, it becomes crucial to ensure continuity in communication flows to avoid such unpleasant outcome. As far as the network structure is concerned, we show that Simmelianity plays a role in shaping influential actors in the network, especially when communication is sufficiently fluent. In particular we highlight how also the characteristics of the neighborhood of the analysed subject is relevant to their influence.

The remainder of the paper is structured as follows. In the next section, we briefly review the literature in the context of information transfer and network structural properties, with a specific emphasis on Simmelian ties; we also discuss the classical and more recent literature related to opinion dynamics. In Section 3, we introduce our model, discussing its main ingredients, and their role in the opinion formation process. In Section 4, we analyze our model by means of numerical simulations with the scope of understanding what is the impact of the introduction of stochastic interaction and, in particular, what is the role that is played by Simmelian ties by virtue of their topological peculiarities. Section 5 concludes with some final considerations.

## 3.2 Literature Review

As summarised in the Introduction, the nature of this article is twofold; on the one hand, we are aiming at understanding how *Simmelian ties* might impact the influence a subject has in the opinion formation process within an organization. On the other hand, we want to further the discussion regarding the micro-motives giving rise to the spread of a specific opinion, especially when communication is sporadic.

To facilitate the reading of the literature review and the rest of the paper, we have included

all necessary notation used in this chapter in Table 9.

### 3.2.1 The role of network structures: where it happens

In discussing the role of network structure in the diffusion of opinion, there are specific aspects regarding the characteristics of the overall network that are considered to either facilitate or hinder the diffusion of an opinion. In this section, we briefly explore the different levels of analysis that researchers have taken into account in their studies regarding the transfer of information.

Information transfer and opinion dynamics are two topics that have often been addressed separately by different streams of literature that sought to deepen their understanding. However, there have been cases where this divide was bridged by combining the two bodies of literature, yielding significant results. For example, Assenova (2018), building on the classic DeGroot model (DeGroot, 1974), developed a technology diffusion model using a threshold. The main assumption was that if individual $i$ was sufficiently convinced of the relevance of a particular piece of information or good, then they would share their information. Similarly, a large body of literature has examined the role of information transfer in opinion formation processes, as information is an integral part of the opinion being shared (Price et al., 2006). Furthermore, the literature on information transfer allows us to develop a broad and robust understanding of what structural characteristics facilitate the transfer of opinions within a group and can thus better provide valid insights into studies regarding opinion dynamics.

In line with previous studies, we distinguish among different levels of analysis pointing attention to the overall network, to the ego-level network (Burt and Celotto, 1992, Cohen et al., 1990), on the characteristics of the dyad (Granovetter, 1973) and, finally, on more complex relational conformations (Krackhardt, 1998; (Gómez-Solórzano et al., 2019). This part of the literature review has the scope of building a solid understanding of *where* information moves within a network, *i.e.* what structural characteristics facilitate its movement.

Subjects in sparse networks are considered to be reaching areas possessing diverse information (Rowley et al., 2000) and to be facing less constraints caused by the "prevailing cognitive scheme" (Wang et al., 2014 p. 486). On the contrary, dense networks promote trust and cooperation among its members (Coleman, 1988) due to reputation effects (Kreps et al., 1990).

The studies that focus on the ego subjects, either analyze their personal qualities or their position within the network. Haas (2015) has pointed out the presence of three "types" of individuals that appear to be particularly relevant in the matter of information transfer: brokers, gate-keepers and boundary spanners (Cohen et al., 1990, Tushman, 1977). Brokers are considered to be subjects that find themselves to be bridging to otherwise disconnected groups of people and that can, thus, exploit the advantage caused by such condition (Burt and Celotto, 1992, Burt, 2004). On the other side, the advantage that gate-keepers enjoy is in relation with the company or the organization they operate for (Haas, 2015). Finally, boundary spanners (Tushman, 1977) are those subjects that are "capable of translating contrasting coding schemes [...] between the work unit and external information areas" (p. 591). (Friedman and Podolny, 1992), provide an even broader definition, talking about a "central role in intergroup relations" (p. 28).

When it comes to ties' characteristics and information transfer, tie strength is one of the characteristics that is generally deemed as relevant (Gargiulo et al., 2009, Friedkin, 1982, Hansen, 1999). In the literature, ties' strength is considered as the "combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie" (Granovetter, 1973, : p. 1361).

Some more recent researches have highlighted how the strength of a tie can be inferred from the structural characteristics surrounding the edge itself. More precisely ties composing cliques, *i.e.* a subgroup of a network that are composed by members that are all connected to each other, are generally deemed to be strong as such a tightly knitted mesh of relationships involves several social dynamics (Gómez-Solórzano et al., 2019). There's a relevant branch of the research that focuses on cliques composed by three subjects, *i.e.*, triads (Krackhardt, 1998, 1999, Kilduff and Lee, 2020). Triads are particularly interesting due to the psycho-social dynamics that come into play and that affect each relationship. As a matter of fact, such type of relationship are called "Simmelian ties" (Simmel, 1950, Krackhardt, 1998). A Simmelian tie is a tie between two subjects who have, at least, a third relationship in common, *i.e.*, a Simmelian tie is a tie within a triad (Krackhardt, 1998). More formally, nodes $i$ and $j$ have a Simmelian tie if such there exist a common neighbor $k$ such that $i$ and $j$ are part of a closed triad:

$$(i,j) \in E \text{ is Simmelian } \implies \exists k : (i,k) \in E \ \ \& \ \ (j,k) \in E,$$

where $E$ represents the set of ties in the network. There are three main aspects that differentiate Simmelian ties from other dyads (Gómez-Solórzano et al., 2019). The first group of elements regards the reciprocity, the symmetry and the transparency that characterize the triad (Simmel, 1950); all subjects in the triad know each other and everyone is aware of it (Krackhardt, 1999). Secondly, since everyone knows each other, when two subjects are interacting, the third party behaves as an indirect controlling dispositive (Krackhardt, 1998). In this respect, the third party acts as a counterbalance to potential power imbalances (Gómez-Solórzano et al., 2019, Krackhardt, 1999). Gómez-Solórzano et al. (2019) argued that triads best express the necessary social normative power that prevents one of the subjects to behave in an "unregulated" way. Finally, the presence of a third subject permits the triad to be less susceptible to external influence (Gómez-Solórzano et al., 2019). (Tortoriello and Krackhardt, 2010). By increasing the stability of the relationships within the triads, favor the development of a "common language and shared understanding among the parties involved" (Tortoriello and Krackhardt, 2010). Having highlighted the characteristics of Simmelian ties, in this research we want to explore whether the impact of Simmelian ties in opinion formation processes and, in particular, in the influence one can have on such process depends on the social implications here highlighted or also by virtue of their topological characteristics.

### 3.2.2 Opinion dynamics: how and when it happens

There are several analytical approaches to the study of opinion dynamics, differing for their baseline assumptions, the complexity of the implemented models, and their applicability to real situations. One of the most inspiring ones is undoubtedly the (DeGroot, 1974) model, developed with the aim of studying how the opinion of subjects changes depending on their starting opinion, the opinion of the others and the relationship they have among each others. The model is based on a graph structure $G(\mathcal{N}, \mathcal{E})$ composed by $N$ nodes (individuals) and $E$ ties. The opinion of an individual ($p_i(t)$) regarding a specific topic/idea vary in time due to the influence caused by the interaction with individuals to which $i$ is connected. Opinions at

any generic time $t$ are represented by $\mathbf{p}(t) = (p_1(t), ...p_n(t), ...p_N(t))$, where $p_n(t) \in [0, 1]$; similarly, $\mathbf{p}(0)$ is the vector that represents the set of initial opinions. In line with opinion dynamics literature, actor *j* may influence actor *i* depending on the trust that *i* places on the opinion of *j* ($T_{ij}$). The matrix **T**, represents the overall set of "trust", ties within the entire population of actors. Note that, **T**'s rows sum to 1, and is constant in time. The opinion of the sample are updated at each time period so that:

$$\mathbf{p}(t) = \mathbf{T}\mathbf{p}(t-1), \tag{1}$$

meaning that, for Agent $i$, we have

$$p_i(t) = \sum_{j=1}^{N} T_{i,j} p_j(t-1). \tag{2}$$

This classical approach has been developed in many directions. Recently, among the others, Ding et al. (2019) explores the impact of centrality in reaching consensus; Shen et al. (2023) studies the impact intermittent interaction by specific actors; Assenova (2018) proposes several settings to describe different social dynamics such as innovation diffusion; finally, Risco (2025) applies the model to information transfer.

One basic property of the DeGroot model is trust as, according the the model as in (1), it directly determines the magnitude of influence other can have on one's opinion. However, trust is not only relevant for opinion dynamics but also for information transfer (Tsai and Ghoshal, 1998, Rutten et al., 2016). Some studies have highlighted the importance of trust and perceived leadership in the active search for information (Yale and Gilly, 1995, Khvatova et al., 2016). Trust of actor $i$ in shaping the opinion of actor $j$ can be defined as the weight actor $j$ puts on the opinion of $i$ when forming her own opinion. Therefore, mathematically, putting together all values of trust, we form a matrix, the so called trust matrix **T**; this is of paramount importance to shape the final opinion reached by the society. Under mild assumptions on the (deterministic and static) trust matrix, the final opinion reached is determined by the relative importance of the individuals in the network.

Despite its importance, the DeGroot model has some limitations, already acknowledged

by the literature. One aspect, not often considered so far, is the focus on *when* individuals communicate. Classical models assume that each person interacts with all her neighbour at any generic iteration of the process. Some recent studies have proposed random communication channels and/or non constant communication flows (Choi et al., 2021).

Another relevant stream of research regards bounded confidence models. In the original DeGroot model, agents spread their opinion at any iteration and no matter of their confidence level in the opinion. This assumption has been relaxed in Hegselmann and Krause (2002) where users will share their opinion if the difference between their opinions is within a certain threshold. The model has been widely used in understanding modern days phenomena such as echo chambers (Baumann et al., 2020) and also in decision making processes (Dong et al., 2018). The model has also been adapted to different specifications such as time delayed interactions (Choi et al., 2021), heterogenous confidence (Liang et al., 2013) diverse generalisations that tried to include a more complex agent's reasoning and behaviour to others' opinions (Burfurd and Wilkening, 2022).

In the next section we propose a model where some of the aforementioned aspects are considered in order to analyze the relative social influence of individuals depending on their personal traits, their confidence, their willingness to communicate, their position in the network, and the characteristics of the whole network itself.

## 3.3 The model

We start this section by refreshing some basic ideas behind the original DeGroot model as defined in (1). In line with opinion dynamics literature, actor *j* may influence actor *i* depending in the trust that *i* places on the opinion of *j*. The matrix $\mathbf{T}$, represents the overall set of "trust", ties within the entire population of actors. Eventually, if the network that is analyzed is strongly connected and aperiodic (DeGroot, 1974, Anderson and Ye, 2019), a consensus is reached. In order to precisely understand the impact that each subject plays during the opinion formation process, we focus on those cases in which the consensus is reached.

This means that, after a sufficient number of iterations, $\mathbf{p}(\mathbf{t}) \rightarrow \mathbf{p}^*$, where $\mathbf{p}^*$ is a constant vector, namely, each subject has the same opinion. In this case, it turns out that $\mathbf{p}^*$ is the only

solution to the fixed point problem

$$\mathbf{p} = \mathbf{T}^{\infty}\mathbf{p}, \tag{3}$$

where $\mathbf{T}^{\infty} = \mathbf{lim_t}\mathbf{T^t}$ is well-defined (see DeGroot, 1974). In case of consensus, it can be proved that all rows of $\mathbf{T}^{\infty}$ are equal and represent the *social influence vector* $\mathbf{s}$, such that the value of consensus, $\mathbf{p}^*$, can be obtained as $p^* = \mathbf{s} \cdot \mathbf{p}(\mathbf{0})$ (Golub and Jackson, 2010). In this respect, the $\mathbf{s}$ vector represents the overall impact that each subject had on the achieving the consensus and is the key ingredient to answer the question *'who is the most influential?'* as the larger is $s_i$, the *more influential* is actor $i$ in the process.

After having briefly collected some of the main characteristics of the DeGroot model, we discuss how we generalize it. In this research, we aim to treat each interaction between two individuals as a *communication episode* (Yale and Gilly, 1995), a moment in which interaction entails communication and the exchange of information or opinions. The authors use this term to highlight that it is not possible to fully understand the motivations that *stand behind* the communication episode without considering the perceptions and preferences of both parties involved.

Similarly, we aim to divide this interaction into two parts: on one side, the perception and preferences of the opinion receiver are already considered in the classical DeGroot model and many of its generalizations, represented through the trust matrix $\mathbf{T}$, which indicates the weight that subject $j$ assigns to the opinion of subject $i$. The innovation we introduce in these models is the inclusion of subject $i$'s willingness to interact with $j$. Specifically, we represent this willingness as a probability, $\xi_{i,j}(t)$.

The overall interaction occurring at time step *t* is summarized by the matrix $\mathbf{B}(t) = (B_{i,j}(t))$, where $B_{i,j}(t)$ expresses whether subjects *i* and *j* have interacted at time *t*. More formally, the matrix $\mathbf{B}(t)$ is specified as follows for each $(i, j) \in E$ and $t > 0$:

$$B_{i,j}(t) = \begin{cases} 1 & \text{with probability } \xi_{i,j}(t) \\ 0 & \text{with probability } 1 - \xi_{i,j}(t) \end{cases} \tag{4}$$

The interaction probability may depend on individual factors related to $i$, dyadic factors concerning the relationship $ij$, or even exogenous factors. Furthermore, it is important to note

that $\xi_{ij}(t)$ may differ from $\xi_{ji}(t)$.

The possibility of having two different probabilities of opinion sharing between two individuals allows us to represent imbalances in terms of interaction preferences. Specifically, the ability to make the probability of opinion sharing from $i$ to $j$ different from that of $j$ to $i$ enables us to represent interaction preferences that may be due to exogenous factors, as in the example we wish to illustrate here of opinion dynamics between individuals belonging to two different hierarchical classes.

In our specific case, drawing from the literature on bounded confidence models, we aim to make this probability dependent on the distance between the opinions of the two subjects at a given moment. Formally:

$$\xi_{i,j}(t) = \xi^0_{i,j} \cdot \big(1 - |p_i(t-1) - p_j(t-1)| \cdot \gamma_{i,j}\big). \tag{5}$$

Some clarifications on (5) are warranted. First, note that the probability of interaction has two components: an ex-ante interaction probability $\xi^0_{i,j}$ and a time-dependent factor. The latter depends on $|p_i(t-1) - p_j(t-1)|$, which represents the opinion distance between $i$ and $j$ at the previous step. Thus, the greater the distance in opinions between $i$ and $j$, the more sporadic their interaction. This approach allows us to disentangle two distinct motives for sporadic communication: an ex-ante eagerness to communicate, potentially dependent on characteristics of actors $i$ and $j$, and an ex-post interaction probability influenced by the process itself. We moderate the relevance of this time-dependent (endogenous) factor by virtue of a parameter $\gamma_{i,j}$, which takes value within $[0, 1]$ and represents the weight that subject $i$ gives to the distance of opinion between subject $j$. Suppose that Subject $i$ does not put excessive importance on the distance of opinion with a peer $j$; then, $\gamma_{i,j}$ will be small. Note that if $\xi_{ij} = 1 \; \forall \; ij$ and $\gamma_{ij} = 0 \; \forall \; ij$ it becomes deterministic model à la DeGroot. The decision to introduce $\gamma_{i,j}$ lies on the assumption that the propensity of a subject to share their opinion with someone is not equally distributed among the neighbours, but depends on the specific characteristics of the relationship that the subjects share. At the same time, making the interaction probability to be linearly dependent on the distance of opinion would be simplistic.

By including $B_{i,j}$, we are taking into consideration the fact that $i$ and $j$ are connected it

does not necessarily mean that they are always communicating and that, in particular, the first will share the opinion to the second. The introduction of a stochastic element in the dynamic process, forces us to apply some adjustments to the classical trust matrix as defined above. Specifically, Equation (1) has to be generalised to encompass sporadic communication, namely, the possibility that, at a specific time, agents $i$ and $j$ do not communicate. In this case, when updating the opinion vector, the trust $i$ poses on *j*'s opinion has to be redistributed to the other neighbours, included *i*. Such arrangement is necessary so that the rows of resulting trust matrix sum to 1 regardless the time step. The natural generalisation to sporadic communication of the original model is represented by a new trust matrix $\tilde{\mathbf{T}}(\mathbf{t})$ (the *sporadic trust* matrix), representing the redistributed weights that the subjects pose on the persons they talked with at time *t*. Specifically, $\tilde{T}_{i,j}(t)$ is defined as follows:

$$\tilde{T}_{i,j}(t) = \frac{T_{i,j}B_{j,i}(t)}{\sum_{j=1}^{N} T_{i,j}B_{j,i}(t)}. \tag{6}$$

Note that the new sporadic trust matrix is stochastic and time-varying, where the numerator can either be equal to 0 or to $T_{i,j}$ while the denominator acts as a normalization factor that takes into consideration all the interactions that subject *i* had at time *t*. For instance, if *i* does not interact with any of her neighbour at time *t*, the trust she poses on her own opinion ($T_{i,i}$) will be equal to 1. On the contrary, if *i* interacts with all her neighbours at time *t*, $\tilde{T}_{i,j}(t) = T_{i,j}$. Moreover, $\tilde{\mathbf{T}}(t)$ is still a trust matrix, in fact:

$$\sum_{i,j}^{N} \tilde{T}_{i,j}(t) = 1, \tag{7}$$

for all $i, j$ and $t$, and

$$\tilde{T}_{i,i} = \frac{T_{i,i}}{\sum_{j=1}^{N} T_{i,j}B_{j,i}}. \tag{8}$$

In particular, if all $B_{i,j} = 0$, $\tilde{T}_{i,i} = 1$. Generally speaking, under the new sporadic trust model, each subject will update its opinion as follows:

$$p_i(t) = \sum_{j=1}^{N} \frac{T_{i,j}B_{j,i}(t)p_j(t-1)}{\sum_{j=1}^{n} T_{i,j}B_{j,i}(t)}, \tag{9}$$

which can be simplified as:

$$p_i(t) = \sum_{j=1}^{N} \tilde{T}_{i,j}(t) p_j(t-1). \tag{10}$$

It is important to stress that, due to the way in which we designed our model, Subject $i$ can be informed by her $j$th neighbour if and only if the condition expressed in (4) is true; eventually, Subject $i$ updates her opinion taking into consideration those neighbours that are likely to interact to her. Such likelihood, again, depends on the distance of opinion and on the weight that is posed on such distance. The opinions of the sample are updated at each time period so that:

$$\mathbf{p}(t) = \tilde{\mathbf{T}}(t)\mathbf{p}(t-1). \tag{11}$$

Differently from the classical DeGroot model, it is not possible to find $\mathbf{p}^*$ as the solution of the fixed point problem $\mathbf{p} = \tilde{\mathbf{T}}\mathbf{p}$, since $\tilde{\mathbf{T}}$ is not constant in time. Nevertheless, Chatterjee and Eliashberg (1990) suggested an approximation to the actual consensus, based on the following procedure. By definition, for all $t$,

$$\mathbf{p}(t) = \tilde{\mathbf{T}}(t) \cdot \tilde{\mathbf{T}}(t-1) \cdot ... \cdot \tilde{\mathbf{T}}(1) \cdot \mathbf{p}(0). \tag{12}$$

Furthermore, in case of convergence to a consensus, we can estimate the value of the consensus, $p^*$, with an approximation, based on a sufficient number of iterations:

$$\mathbf{p}^*_\epsilon = \tilde{\mathbf{T}}(t^*) \cdot \tilde{\mathbf{T}}(t^*-1) \cdot ... \cdot \tilde{\mathbf{T}}(1) \cdot \mathbf{p}(0), \tag{13}$$

2here $t^*$ is the smallest $t$ such that

$$\|\mathbf{p}(t^*-1) - \mathbf{p}(t^*)\| < \varepsilon,$$

with $\varepsilon$ representing a small tolerance value. This stopping criterion identifies the point where the vector $\mathbf{p}(t)$ converges sufficiently close to a stable state.

Based on this, we define the **(approximated) equilibrium sporadic social influence matrix** as

$$\tilde{\mathbf{T}}^\infty := \tilde{\mathbf{T}}(t^*),$$

which approximates the long-term behaviour of the system. Using this equilibrium matrix, we derive an **approximation of the social influence vector** $\tilde{\mathbf{s}}$ by simply selecting any of its rows.[2]

## 3.4 Simulations and results

In this section we propose a set of numerical experiments based on simulations with the aim of analysing the peculiarities of our model and the role that each component has on the final equilibrium expressed in terms of consensus and the weight that each subject had in reaching such consensus. We perform several experiments based on different specifications of the structure of the network, of the characteristics of the subjects and of the model itself. The final aim is to isolate different aspects that we deem to have an impact on the overall process.

### 3.4.1 Types of networks

To explore the impact that the introduction of sporadic interaction and the preference for interaction between subjects due to personal characteristics, as well as the opinion distance between them, have on the social influence $\tilde{s}_i$ that each Subject $i$ has on reaching consensus, we will conduct ad hoc experimental simulations.

Specifically, we perform various simulations using two ad hoc network structures: The so-called "*Kite*", network as in Krackhardt (1990) (see also Brandes and Hildenbrand, 2014), and a "*Torus*", network by 36 nodes as depicted in Figure 1.

The first graph was chosen as it was designed to highlight differences in terms of centrality measures, given that the most central individual varies depending on the centrality measure considered. Essentially, this graph serves as a quick way to visualize whether certain centrality measures are more likely to correlate with a higher $\tilde{s}_i$. The Torus graph, in contrast, features nodes with the same centrality regardless of the measure used. This allows us to better examine the role that potential imbalances in preferences, due to exogenous factors, may play in determining social influence $\tilde{s}_i$.

[2]This step is justified by the classical DeGroot model, which ensures that, under consensus, all rows of $\tilde{\mathbf{T}}^{\infty}$ are identical and correspond to the social influence vector $\tilde{\mathbf{s}}$. For more details, see Golub and Jackson (2010).

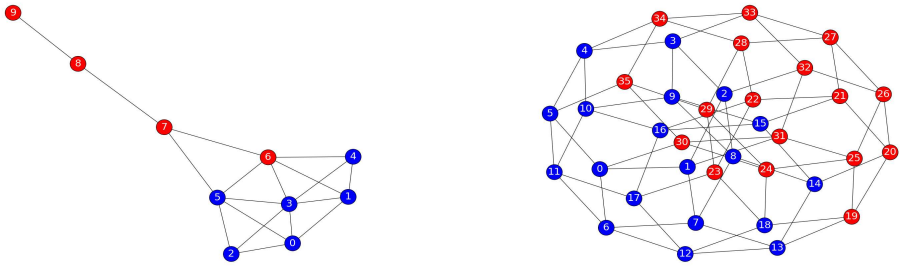


Figure 1: Kite graph (left) and Torus graph (right)

The aim is to visually determine whether and what effects the introduction of sporadic interaction has on the influence that an individual has in reaching a consensus.

Lastly, to isolate and understand the effect that Simmelian ties can have on $\tilde{s}_i$, we repeated the simulations on a Torus network to which we added 4 ad hoc ties and performed ANOVA analysis. The ultimate goal was to understand if and how having a Simmelian tie can have a positive effect on $\tilde{s}_i$ at different probability levels.

### 3.4.2 Scenarios related to communication

We now specify how we introduce sporadic communication as defined in (4). We identify a number of different scenarios based on specific values of the parameters $\xi^0_{i,j}$ and $\gamma_{i,j}$ as in (5). We first consider a benchmark situation, scenario (o), where $\xi^0_{i,j} = 1$ and $\gamma_{i,j} = 0$ for all actors. Under this assumption, considering (5), we can see that if $\gamma_{i,j}$ is equal to 0, then $\xi^0_{i,j}$ is multiplied by 1. Being $\xi^0_{i,j}$ set to one in this baseline scenario, $\xi_{i,j}(t) = 0$ at every time step. Such scenario, in which all individuals interact to with each neighbour at each time step is equivalent to the DeGroot one.

We then introduce six scenarios characterized by different values of the two aforementioned parameters.

In the creation and planning of scenarios, we aimed to consider the potential applicability of the model. We chose to account for two hierarchical classes within an organization or company with the intent of hypothesizing two specific behaviours. Specifically, Table 1 outlines such scenarios based on the model used and the criteria for determining the interaction probability.

Scenarios (i)-(iii) are related to cases where communication is sporadic but not dependent on the opinion of the two individuals, namely, $\gamma_{i,j} = 0$. By setting $\gamma_{i,j}$ equal to zero, the difference of opinion between $i$ and $j$ has no weight in determining the probability of interaction. To test different modelling assumptions, under the three scenarios, $\xi^0_{i,j}$ is characterized in different ways. In Scenario (i), the value of $\xi^0_{i,j}$ is fixed at an intermediate value for all actors. In Scenario (ii), it depends on the dyad under consideration; to this aim, we classify the subjects into two groups: "leaders", and "followers". Leaders are more likely to share their opinions with followers (resulting in a higher probability), while followers are less inclined to share their opinions with leaders (resulting in a lower probability). Members within the same group are indifferent to each other, resulting in equal probabilities for interaction and non-interaction. In Figure 1, we can see in blue the followers and in red the leaders. Finally, in scenario (iii), the communication probability is randomly assigned to each pair of actors, according to a beta distribution with parameters $\alpha = 1$ and $\beta = 1$.[3] As for Scenarios (iv)-(vi), we use the same rationale but letting $\gamma_{i,j}$ vary, and fixing $\xi^0_{i,j} = 1$. In this way, we focus on a communication probability that is ex-ante deterministic but then varies depending on the opinion's vector evolution and, more precisely, on the distance of opinions among subjects.

Table 1: Specification of interaction probabilities for each scenario

| Homogeneous | | Heterogeneous | Random |
|---|---|---|---|
| (i) | | (ii) | (iii) |
| $\gamma_{i,j} = 0$ | $\xi^0_{ij} = 0.5$ | $\xi^0_{ij} = \begin{cases} 0.2 & \text{if } i \text{ in "followers" and } j \text{ in "leaders"} \\ 0.5 & \text{if } i,j \text{ both in "followers" or in "leaders"} \\ 0.8 & \text{if } i \text{ in "leaders" and } j \text{ in "followers"} \end{cases}$ | $\xi^0_{ij} \sim \text{Beta}(1,1)$ |
| (iv) | | (v) | (vi) |
| $\xi^0_{i,j} = 1$ | $\gamma_{ij} = 0.5$ | $\gamma_{ij} = \begin{cases} 0.2 & \text{if } i \text{ in "followers" and } j \text{ in "leaders"} \\ 0.5 & \text{if } i,j \text{ both in "followers" or in "leaders"} \\ 0.8 & \text{if } i \text{ in "leaders" and } j \text{ in "followers"} \end{cases}$ | $\gamma_{ij} \sim \text{Beta}(1,1)$ |

[3] We have chosen a beta distribution as it allows a simple and not skewed opinion distribution within 0 and 1. For instance, even though the expected value of this beta distribution is 0.5, differently from scenario (i), each *ij* has a different probability allowing, thus, $\xi^0_{ij}$ and $\xi^0_{ji}$ to be different.

### 3.4.3 Simulation procedure

The simulations are carried out as follows. First, the type of network (Kite or Torus) and a scenario as in Table 1 are chosen. In this way, the graph $G$ and the matrices $\xi$ and $\gamma$ are identified. Second, the opinion vector is initialized according to a beta distribution with parameters $\alpha = 1$ and $\beta = 1$. Two technical parameters controlling for the convergence to consensus, as defined in (13), are introduced, namely, the maximum number of iterations allowed (*max_steps*) and $\varepsilon$. The algorithm is designed to stop either when the variance of the values of the opinion vector $\mathbf{p}(t)$ is smaller than $\varepsilon$, or when the number of iterations exceeds the predetermined upper bound. If the number of iterations exceeds the upper bound, the simulation is discarded and a new one is started.

The simulation procedure is described by Algorithm 1. The specific dynamic in which the opinion updating mechanism is carried out is specified in Algorithm 2. A simplified version of the Python code is given in Appendix 7.1.

The output of the procedure is a stationary (random) $\tilde{\mathbf{T}}^{\infty}$ matrix and the related stationary (random) social influence vector $\tilde{\mathbf{s}}$. We carried out 100,000 simulations for each scenario as in Table 1 and for each type of network (Kite and Torus). For each simulation, several metrics are then collected and eventually linked to each Subject $i$ through the value of $\tilde{s}_i$. Finally, we statistically analyse the properties of the social influence vector in terms of all the antecedents of the model described above.

**Algorithm 1** Simulation procedure

**Input:** $G$, $influence_matrix$, $probability_matrix$, $initial_opinions$, $max_steps$, $\epsilon$
**Output:** $opinions$
**Function** `simulations`($G$, $influence_matrix$, $probability_matrix$, $initial_opinions$, $max_steps$, $\epsilon$) **:**
  $step \leftarrow 0$ $opinions \leftarrow initial_opinions$
  **while** $step < max_steps$ ***and*** $\Delta(opinions) > \epsilon$ **do**
    $opinions \leftarrow$ `event`($G$, $influence_matrix$, $probability_matrix$, $opinions$) $step \leftarrow step + 1$
  **end**
  **return** $opinions$
**end**

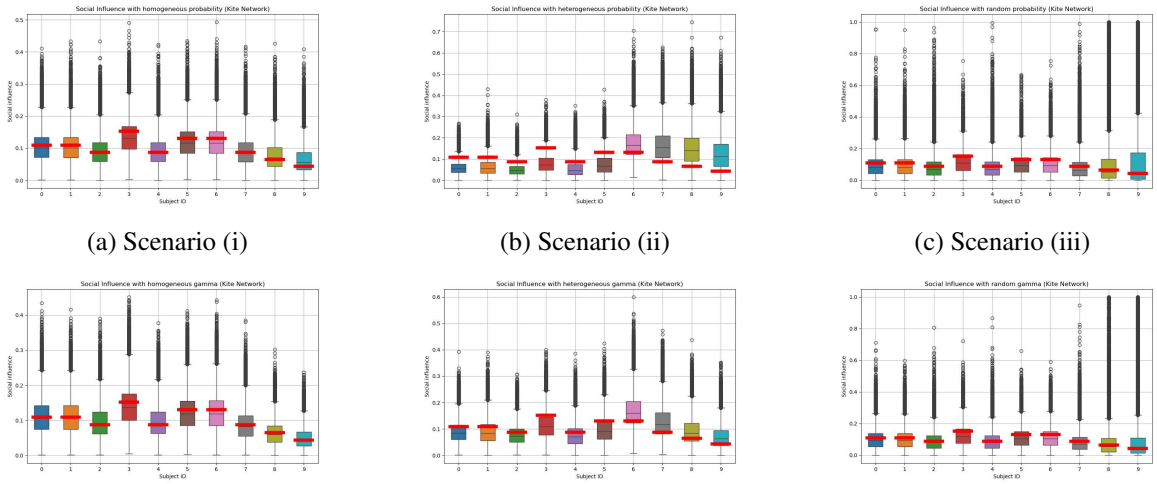
**Algorithm 2** Detailed update algorithm

**Input:** $G$, $influence_matrix$, $probability_matrix$, $opinions$
**Output:** $updated_opinions$
**Function** `event` ($G$, $influence_matrix$, $probability_matrix$, $opinions$) **:**
    $event_influence_matrix \leftarrow$ zeros matrix of size $num \times num$
    **for** *each node* $i$ *in* $G.nodes$ **do**
        **for** *each neighbor* $j$ *of* $i$ *in* $G$ **do**
            **if** *rand()* $> 1 - probability_matrix[i][j]$ **then**
                $event_influence_matrix[i][j] \leftarrow 0$
            **end**
            **else**
                $event_influence_matrix[i][j] \leftarrow influence_matrix[i][j]$
            **end**
        **end**
    **end**
    **for** *each row* $i$ *in* $event_influence_matrix$ **do**
        Normalize the values in row $i$ of $event_influence_matrix$
    **end**
    **for** *each node* $i$ *in* $G.nodes$ **do**
        $updated_opinions[i] \leftarrow \sum_j opinions[j] \times event_influence_matrix[i][j]$
    **end**
    **return** $updated_opinions$
**end**

### 3.4.4 Numerical results - Descriptive analysis

In Figure 2, we represent the distribution of social influence variable $\tilde{s}_i$ for each subject in Scenarios (i)-(vi). In each plot, it is also provided (in red) the value of the social influence in case of the baseline deterministic (DeGroot) model. All plots refer to the simulations carried out on the Kite network.

Initially, we notice that the social influence that subjects acquire in the baseline (DeGroot) model does not match the averages of the models where sporadic interaction is introduced. Interestingly, subjects with high social influence in the baseline model tend to lose relevance. Conversely, those who appear to have low social influence in the baseline model gain relevance in the model where sporadic interaction is introduced. As expected, the introduction of a preference in communication (Scenario (ii)) produces a visible impact on the distribution of social influence. Comparing Scenarios (i) and (ii) we can easily notice how being in the "leaders" group (from Subject 6 to Subject 9) represents an advantage for subjects despite the structural characteristics in which they are set. Finally, looking at Scenario (iii), it is noteworthy that the




(a) Scenario (i) (b) Scenario (ii) (c) Scenario (iii)

(d) Scenario (iv) (e) Scenario (v) (f) Scenario (vi)

Figure 2: Distribution of $\tilde{s}$ in Kite Network

subject who benefits at most from the introduction of random preference is Subject 9, *i.e.*, the most isolated one (see Figure 1). This result may be attributed to the quasi-isolated position of the subject, which may facilitate the emergence of a stubborn agent, *i.e.* an agent that influences but is not influenced by others. In particular, if Subject 8 does not frequently share its opinion with Subject 9, the latter will be able to disseminate its opinion through Subject 8 without being influenced by it. Referring to Figure 2, we can observe a greater presence of outliers and, simultaneously, a minor impact of the introduction of sporadic interaction. Specifically, the averages of the distributions do not deviate significantly from the social influence value of the deterministic model. While the differences are still noticeable, they are dampened by the fact that the interaction probability also depends on the opinion that the subjects within the network have at a given time $t$. These findings seems to suggest that it is not relevant whether or not the communication probability varies according to the vicinity of opinions at a given time, rather, it depends on the intrinsic characteristic of the dyads.

Lastly, Figure 13 in Appendix 7.1 13 represents the distribution of social influence in all the scenarios in the case of the Torus network. A review of Scenarios (ii) and (iv) reveals that being labelled as "leader", is not the sole determining factor in the ability to influence. The characteristics of one's own neighbourhood also play a role. For instance, Subject $i$'s ability to influence is contingent upon not only the likelihood of sharing its views with its neighbours, but

also the likelihood of its neighbours sharing theirs.

Consequently, in the upcoming stages, we seek to explore the role that potential differences in terms of the interaction probability between Subject $i$ and $i$'s neighbors play in determining a higher or lower $\tilde{s}_i$.

Another aspect that we want to explore in this phase is the role of simmelianity on shaping social influence. More specifically, we wonder whether, *ceteris paribus*, having a Simmelian tie could have an impact on the influence that a subject can have on the opinion dynamics. To this aim, we carry out a different simulation experiment. The Torus network (see Figure 1) has been augmented with additional ties, with the objective of enabling subjects with a slightly higher degree (namely, degree 5). Consequently, some of these newly established ties contribute to the formation of Simmelian ties. To check for the effect of simmelianity, differently from the previous analysis, we fix the interaction probability to values 0.1, 0.3, 0.5, 0.7, and 0.9 for each tie $ij$ in different runs of the experiment. Then, we group the subjects according to their degree and depending on whether they have *at least* one Simmelian tie. We collect in Appendix 7.1 all the descriptive statistics of the results related to this experiment. In particular, we carry out an ANOVA analysis with the aim of understanding whether the values in terms of influence were systematically different between the identified groups. From the results of the Tukey test carried out for different interaction probabilities (Table 10) we can see that the differences between groups are not significant for lower probabilities while, with the increase of communication probability, such differences become significant. From such results we can infer that, on the one hand, in case of too sporadic communication, the role of topological characteristics tends to decrease. On the other hand, there are intergroup differences that can be noticed. For instance, the role of the degree is undeniable and justifiable. However, it is relevant to notice that there are differences between subjects with the same degree having a Simmelian tie and subjects not having a Simmelian tie. Surprisingly, however, having a Simmelian tie would advantage subjects with a higher degree and results to be detrimental for subjects with a lower degree. Such results might be related to the way in which the network was designed. For instance, subjects with degree 4 and that had *at least* one Simmelian tie were bound to be next to two subjects that have a Simmelian tie and with a degree of 5 with a consequent imbalance

of degree. Nevertheless, there is a discernible yet statistically significant impact of Simmelian ties, as a result of their topological characteristics, on the influence that subjects exert in the opinion formation process.

We can conclude this section saying that the presence of sporadic communication, namely, introducing randomness into a traditional opinion dynamics model, impacts differently on the ex-ante social influence hierarchy. More specifically, leaders who are advantaged by their structural position are likely to see this advantage eroded by the introduction of sporadic interaction. similarly, such sporadic interaction may lead to the emergence of almost stubborn agents. What's more, the introduction of sporadic interaction seems to make the ability of $i$'s neighbour relevant in determining $i$'s social influence. As a matter of fact, being able to influence those who are influential appears to be relevant. Furthermore, we highlighted the role of simmelianity on the social influence of agents. Highlighting how their impact depends on other structural characteristics. Eventually, we can conclude that sporadic communication moderates the effect that structural characteristics, including Simmelian ties, have in shaping the social influence at the equilibrium.

### 3.4.5 Numerical results - Inference analysis

Having isolated and highlighted the impact of Simmelian ties on the social influence of subjects, we now want to see what is the role of other structural characteristics of the network and if there is an interaction between some of them and the simmelianity of a node.

To further understand which structural characteristics might impact $\tilde{s}_i$, we adopted a statistical inference approach. Specifically, we repeated the simulations as outlined above for the *Kite* and the *Torus* network and changed the network structure for each simulation by utilizing an Erdős-Rényi random network generator model (see Gómez-Gardeñes and Moreno, 2006). We chose this model over others commonly used, such as Barabasi-Albert or Watts-Strogatz, because it allows for a normal distribution of most of the variables we intend to take into consideration (see Harrison et al., 2016). As before, the number of collected simulations is 100,000.

The metrics collected and used as independent variables concern the node's position within the network, microstructural characteristics around the observed node (*e.g.*, the presence of

Simmelian ties), and differences in terms of interaction preference between the observed subject and its neighbours.

The first aspect we observe is the aggregate capacity of a subject to communicate and, in particular, whether it is more likely that it share its opinion to others or to receive the opinion of others. We look at the $out - degree$ and the $in - degree$ of the focus subject as they represent, respectively, the possible occasions to share the opinion to its neighbours and to receive the opinion of the neighbours. However, considering the mere $in - degree$ and $out - degree$ overlooks the role that the weight of ties (in our case, the probability to interact). At the same time, taking the simple mean of interaction might lose the relevance of having many ties. Trying to understand whether having more ties or having "heavier"ties is more relevant, Opsahl et al. (2010) propose the following metrics:

$$C^{\alpha}_{D-\text{out}}(i) = k_i^{\text{out}} \times \left( \frac{w_i^{\text{out}}}{k_i^{\text{out}}} \right)^{\alpha}, \tag{14}$$

$$C^{\alpha}_{D-\text{in}}(i) = k_i^{\text{in}} \times \left( \frac{w_i^{\text{in}}}{k_i^{\text{in}}} \right)^{\alpha}. \tag{15}$$

Here, $C^{\alpha}_{D-\text{out}}(i)$ represents the weighted out-degree measure of node $i$, while $C^{\alpha}_{D-\text{in}}(i)$ represents the weighted in-degree of node $i$. The term $k_i^{\text{out}}$ is the out-degree of node $i$, indicating the number of outgoing ties, and $k_i^{\text{in}}$ is the in-degree of node $i$, indicating the number of incoming ties. Similarly, $w_i^{\text{out}}$ denotes the total weight of outgoing ties of node $i$, and $w_i^{\text{in}}$ denotes the total weight of incoming ties of node $i$. The parameter $\alpha$ adjusts the influence of the tie weights: if $\alpha < 1$, nodes with lower $k_i^{\text{out}}$ or $k_i^{\text{in}}$ are favoured, whereas if $\alpha > 1$, nodes with higher $k_i^{\text{out}}$ or $k_i^{\text{in}}$ are favoured.

In our case, $w_i^{\text{in}}$ is calculated as the sum of the probabilities that the neighbours of a given node $i$ communicate with $i$ and $w_i^{\text{out}}$ as the sum of probabilities that $i$ communicates with its neighbours. We therefore name (14) as the Outward Probability and (15) as the Inward Probability.

To better stress the potential relevance that imbalances between $C^{\alpha}_{D-\text{out}}(i)$ and $C^{\alpha}_{D-\text{in}}(i)$, we

propose to consider a ratio between the two metrics. We, define

$$ProbRatio_{\alpha}(i) = C^{\alpha}_{D-\text{out}}(i)/C^{\alpha}_{D-\text{in}}(i)$$

as the ratio between $i$'s probability to share the opinion to others and $i$'s probability to receive the opinion from others. We set $\alpha$ to different values so that, similarly to Opsahl et al. (2010), we can see whether having more ties or having higher probabilities have a higher impact on the influence the subject has on the opinion formation process. A similar approach is proposed for what regards the weight $\gamma_{ij}$ as defined in (5).

Wanting to verify the potential role that Simmelian ties can have in determining the social influence exercised by the focal subject we consider to number of Simmelian ties a subject has as well as the number of Simmelian ties that the neighbours of the subject itself have. The *Neighbour Simmelian* metric evaluates the average number of Simmelian ties within the neighbourhood of a node $i$, denoted as

$$\frac{1}{|N(i)|}\sum_{j\in N(i)} \text{Number of Simmelian ties of } j.$$

This metric highlights the structural characteristics of the local network around a node, providing insights into the interconnectedness and the nature of the ties within its neighbourhood.

Eventually, considering the way in which interaction probabilities are defined in (5), we suggest to take into consideration the distance from the midpoint of the position of the focal subject at $t = 0$,

$$dist(p_i) = |p_i(0) - \overline{p}(0)|$$

where $p_i(0)$ is the initial opinion of subject $i$ at $t = 0$ and $\overline{p}(0)$ is the average of opinions of all the subjects at $t = 0$.

These metrics collectively offer a comprehensive understanding of the structural and interactional dynamics within the network, providing a robust framework for our analysis. In addition to these metrics, we wanted to take into account more traditional node importance measurement such as Betweenness Centrality, Eigenvector Centrality and Page-Rank Central-

ity (see Ding et al., 2019, Harrison et al., 2016, Freeman, 1978). An overall description of the metrics can be found in Table 2.

Table 2: Descriptive Statistics

| **Variable** | **Min** | **1Q** | **Mean** | **Median** | **3Q** | **Max** | **Var** |
|---|---|---|---|---|---|---|---|
| Influence | 0.000 | 0.029 | 0.050 | 0.046 | 0.065 | 1.000 | 0.001 |
| Weighted Indegree$_{0.5}$ | 1.000 | 3.981 | 5.026 | 4.950 | 5.992 | 12.687 | 2.177 |
| Weighted Indegree$_{1.0}$ | 1.000 | 2.960 | 3.828 | 3.740 | 4.600 | 10.060 | 1.442 |
| Weighted Indegree$_{1.0}$ | 0.552 | 2.186 | 2.939 | 2.847 | 3.596 | 8.816 | 1.077 |
| Weighted Outdegree$_{0.5}$ | 1.000 | 3.980 | 5.026 | 4.954 | 5.993 | 12.432 | 2.185 |
| Weighted Outdegree$_{1.0}$ | 1.000 | 2.960 | 3.828 | 3.740 | 4.600 | 10.180 | 1.450 |
| Weighted Outdegree$_{1.5}$ | 0.544 | 2.184 | 2.939 | 2.847 | 3.591 | 9.008 | 1.083 |
| Probability Ratio$_{0.5}$ | 0.400 | 0.913 | 1.010 | 1.000 | 1.095 | 2.149 | 0.021 |
| Probability Ratio$_{0.5}$ | 0.319 | 0.840 | 1.037 | 1.000 | 1.192 | 3.260 | 0.080 |
| Probability Ratio$_{0.5}$ | 0.184 | 0.771 | 1.082 | 1.000 | 1.299 | 5.879 | 0.204 |
| NProbRatio | 0.474 | 0.963 | 1.094 | 1.087 | 1.217 | 2.200 | 0.040 |
| Simmelian | 0.000 | 3.000 | 4.661 | 5.000 | 6.000 | 15.000 | 6.368 |
| NSimmelians | 0.000 | 4.500 | 5.361 | 5.429 | 6.300 | 12.000 | 1.941 |
| BetCent | 0.000 | 0.018 | 0.046 | 0.037 | 0.064 | 0.399 | 0.001 |
| EigCent | 0.003 | 0.160 | 0.211 | 0.212 | 0.263 | 0.482 | 0.005 |
| DegCent | 0.053 | 0.211 | 0.301 | 0.316 | 0.368 | 0.789 | 0.011 |
| InitPosition | 0.000 | 0.238 | 0.475 | 0.475 | 0.712 | 0.950 | 0.083 |
| Pagerank | 0.013 | 0.040 | 0.050 | 0.049 | 0.059 | 0.126 | 0.000 |

With the aim of better understanding what characteristics might affect the social influence we proceeded with a linear regression analysis. More precisely, we here opted for a Ridge regression. Ridge Regression is particularly advantageous over Ordinary Least Squares (OLS) linear regression when multicollinearity is present among the independent variables. In the presence of multicollinearity, OLS estimates become unstable, leading to large variances in the coefficient estimates, which means that even small changes in the data can lead to significant changes in the estimated coefficients. Ridge Regression addresses this issue by adding a regularization term to the cost function, which shrinks the coefficient estimates towards zero. This shrinkage reduces the variance of the coefficients, leading to more stable and reliable estimates. James et al. (2023) explain that ridge regression finds coefficient estimates, represented by $\beta$, by determining the values that minimize the expression shown in the following equation:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p}\beta_j^2, \tag{16}$$

where $\lambda$ is a regularization term that controls the intensity of penalization applied to the regression coefficients. To determine the optimal $\lambda$ value, we conducted a grid search paired with 5-fold cross-validation. Once the optimal $\lambda$ was identified, we retrained the model using this selected value. To evaluate the statistical significance of the identified regression coefficients, we utilized non-parametric bootstrap confidence intervals (Efron and Tibshirani, 1994). Specifically, if zero does not lie within a $(1-\alpha)\%$ confidence interval for a regression coefficient, we reject the null hypothesis that the coefficient is equal to zero at the $\alpha\%$ significance level. Previous researches have employed bootstrap methods in the context of ridge regression (Sabbaghi, 2019, Crivelli et al., 1995).

Before proceeding we the analyses, we normalized the independent variables to ensure that each variable has a mean of zero and a standard deviation of one. Normalizing the variables in this way helps the Ridge Regression algorithm apply penalization consistently across all features, preventing features with larger values from dominating the model.

This comprehensive fine tunes the model making it more interpretable, and robust, minimizing over-fitting while maintaining an optimal level of bias and variance.

Table 3 presents the OLS linear regression results examining the relationship between the social influence of each subject and the various independent variables above described in Scenarios (i), (iii), (iv) and (vi).

The results indicate that the intercept is consistently significant across all models, which implies that even in the absence of other predictors, the baseline level of the dependent variable is substantial. The *ProbRatio* variables show significant positive coefficients in various models and at different values of $\alpha$. As it can be seen, changing the value of $\alpha$ does not affect the impact that the ratio of the probabilities as defined in (14) and (15) on the Influence of the subject. Such lack of difference can be attributed to the normalization of the variables. Such result suggests that higher probability ratios are associated with increases in the social influence. Similarly, the GammaRatio variables, including GammaRatio0.5, GammaRatio1.0, and GammaRatio1.5, are significant in Models $(\text{vi})_a$, $(\text{vi})_b$, and $(\text{vi})_c$ respectively, indicating a positive association between higher gamma ratios and the Social Influence. It can be noticed that GammaRatio variables have a lower impact if compared with ProbRatio ones, and this can be explained

Table 3: OLS Linear Regression Results

| Variable | Model (i) | Model (iii)$_a$ | Model (iii)$_b$ | Model (iii)$_c$ | Model (iv) | Model (vi)$_a$ | Model (vi)$_b$ | Model (vi)$_c$ |
|---|---|---|---|---|---|---|---|---|
| Const. | 0.0500*** | 0.0500*** | 0.0500*** | 0.0500*** | 0.0500*** | 0.0500*** | 0.0500*** | 0.0500*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| ProbRatio$_{0.5}$ | | 0.0148*** | | | | | | |
| | | (0.000) | | | | | | |
| ProbRatio$_{1.0}$ | | | 0.0148*** | | | | | |
| | | | (0.000) | | | | | |
| ProbRatio$_{1.5}$ | | | | 0.0147*** | | | | |
| | | | | (0.000) | | | | |
| NeighProbRatio | | 0.0002 | 0.0000 | 0.0001 | | | | |
| | | (0.000) | (0.007) | (0.000) | | | | |
| GammaRatio$_{0.5}$ | | | | | | 0.0105*** | | |
| | | | | | | (0.000) | | |
| GammaRatio$_{1.0}$ | | | | | | | 0.0102*** | |
| | | | | | | | (0.000) | |
| GammaRatio$_{1.5}$ | | | | | | | | 0.0099*** |
| | | | | | | | | (0.000) |
| NeighGammaRatio | | | | | | -0.0000 | -0.0000 | -0.0000 |
| | | | | | | (0.000) | (0.0000) | (0.000) |
| Simmelian | 0.0078*** | 0.0060*** | 0.0061*** | 0.0062*** | 0.0080*** | 0.0069*** | 0.0071*** | 0.0070*** |
| | (0.000) | (0.000) | (0.0000) | (0.001) | (0.000) | (0.000) | (0.000) | (0.000) |
| NeighSimmelian | -0.0021*** | -0.0021*** | -0.0021*** | -0.0021*** | -0.0021*** | -0.0020*** | -0.0019*** | -0.0019*** |
| | (0.001) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| BetCentrality | 0.0046*** | 0.0036*** | 0.0037*** | 0.0037*** | 0.0043*** | 0.0037*** | 0.0037*** | 0.0037*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| DistanceOpinion | -0.0001 | 0.0003 | 0.0000 | 0.0000 | -0.0038*** | -0.0031*** | -0.0031*** | -0.0031*** |
| | (0.001) | (0.000) | (0.000) | (0.000) | (0.000) | (0.008) | (0.000) | (0.000) |
| $R^2$ | 0.2080 | 0.2766 | 0.2789 | 0.2741 | 0.2449 | 0.2465 | 0.2399 | 0.2323 |
| MSE | 0.0001 | 0.0007 | 0.0004 | 0.0007 | 0.0003 | 0.0006 | 0.0006 | 0.0006 |
| RMSE | 0.0204 | 0.0273 | 0.0272 | 0.0273 | 0.0199 | 0.0248 | 0.0241 | 0.0250 |
| MAE | 0.0160 | 0.0181 | 0.0180 | 0.0181 | 0.0157 | 0.0177 | 0.0177 | 0.0178 |
| $\lambda$ | 35.56 | 51.79 | 51.79 | 35.56 | 35.56 | 75.43 | 75.43 | 75.43 |

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

considering that the former regards a parameter that influences the probability of interaction while the latter regards the probability of interaction itself.

On the other hand, the neighborhood probability ratio (NeighProbRatio) and neighborhood gamma ratio (NeighGammaRatio) do not show consistent significance across the models, indicating a less clear relationship with the social influence. The Simmelian variable is consistently significant and positive across all models where it is included, suggesting a strong positive association with the social influence. Conversely, NeighSimmelian shows a significant negative association in all models, indicating that a higher presence of Simmelian ties within ego subject's neighborhood is associated with lower levels of the social influence.

Furthermore, Betweenness Centrality is significant in all models, suggesting a positive association with the dependent variable. The DistanceOpinion variable a negative and significant in Models (iv),(vi)$_a$,(vi)$_b$,(vi)$_c$, suggesting that greater distance in opinion from the midpoint of the considered range is associated with a lower social influence.

In conclusion, the results indicate that various node centrality measures, ProbRatio and Betweenness Centrality, are significant predictors of the social influence. What appears to be interesting, considering the scope of this research, is represented by the other measurements.

Looking at the role that Simmelian ties and the presence of Simmelian ties within the ego subject's neighborhood, we can confirm what we observed in the descriptive analysis, namely, having Simmelian ties is favourable but being connected with subjects who have Simmelian ties is detrimental. These findings highlight the importance of both direct and neighborhood-based network characteristics in understanding the underlying processes captured by the social influence. The use of a log transformation has effectively normalized the residuals, allowing for more robust and reliable interpretations of the regression coefficients.

## 3.5 Conclusion

In this research, we proposed a bounded confidence opinion dynamics model that accounts for both sporadic interaction between subjects and their interaction preferences. We explored the impact of introducing sporadic interactions and potential asymmetries in interaction preferences on social influence. We investigated how such traits interact with various structural character-

istics of the network. Specifically, we aimed to examine whether Simmelian ties and different centrality measures influence the role played by each subject.

By comparing the deterministic DeGroot model with several variations introduced in this study, we found that sporadic communication significantly reduces the impact of structural characteristics on determining the social influence of subjects. In particular, it appears that those with a higher influence in the deterministic model risk to lose more influence in the moment in which sporadic communication is introduced. Overall we highlighted that the introduction of sporadic interaction tends to decrease difference of expected influence. However, when the possibility of asymmetric preferences between subjects is introduced, different phenomena can be observed. Such asymmetry, in fact, allows for the convergence to diverse values of consensus, different from the deterministic model. In the simulations we carried out we were able to notice how, given specific circumstances, some peripheral subjects were able to acquire a major impact on the convergence achieved due to the fact that they would influence more than they are influenced. The introduction of sporadic interaction does not negate the structural characteristics of the node; rather, it allows for the configuration of cases that, although extreme, can occur in reality.

Similarly, our analysis demonstrates that the structural characteristics of the subjects and their neighbors retain their relevance, although this relevance diminishes as the probability of interaction decreases. Specifically, by using the hierarchical class stratagem, we aimed to highlight how belonging to a hierarchical class can impact the influence that a subject has during the consensus-building process. Starting from the assumption that certain categories of individuals are characterized by a greater or lesser willingness to share their opinion with other categories of individuals, it is possible to identify which subjects can more easily gain influence due to their role within an organization or company and their position within the network. It is important to note that it is not just the position of the analyzed subject that has an impact, but also the structural characteristics of their neighborhood. For instance, while having Simmelian ties offers advantages due to the topological characteristics they bring, it is equally true that having neighbors with a high number of Simmelian ties can be detrimental.

This research introduced sporadic interaction into opinion dynamics models to emphasize

the role that interaction frequency can play at the expense of the intensity of trust between two subjects. The main aim of this approach was to represent a two-sided event: on the one hand, the decision of subject to share their opinion with subject , and on the other hand, the decision of the latter regarding the weight to give to the former's opinion. Our analysis highlighted that the frequency of interactions and their distribution within the network can lead to complex outcomes that significantly differ from those predicted by the deterministic DeGroot model.

It is important to note that this research is not without limitations. Our findings are mostly based on computer simulations, with parameters set without proper calibration to real data. However, the analysis conducted can still inform future research, possibly based on empirical data, that aims to understand phenomena regarding followers, leaders, or the role of formal structures in information transfer within an organization. Future research would benefit from elucidating potential relationships between different characterizations of subjects, particularly focusing on exogenous factors.

# 4 Liking those you met the most

**Sasha Piccione**

**Abstract**

This research investigates the interactions among individuals who participated in the European legislative process for the adoption of EURO 7 Regulation, with the objective of identifying which actors played influential roles, based on the central positions they occupy within the interaction network. To achieve this goal, the study employs techniques and methodologies typical of social network analysis, with a particular focus on the presence of an informal network. This network is characterized by public interactions between participants on the social media platform LinkedIn, allowing for a deeper understanding of the underlying social dynamics. By utilizing LDA and sentiment analysis, we characterize the role and the expressed emosiont by the involved individuals, taking into account the linguistic features they expressed in their online activity. Using Exponential Random Graph Models (ERGM), the research ultimately demonstrates a causal connection between the formal and informal networks. This, in turn, helps to answer the research question regarding the relevance and centrality of specific stakeholders during the legislative process.

***Keywords*—** Lobbying, Social Network Analysis, Social Media Platform, Multilayer Networks.

## 4.1 Introduction

The study of interest groups and their influence on public policy has long been a cornerstone of political science. Early interest group theory, developed within the pluralist framework, posited that decentralization exists to accommodate the policy demands of a society fundamentally composed of numerous social and economic groupings of individuals and competing organizations, such as business associations, labor unions, and environmental groups (Dahl, 1961, Truman, 1951). These groups were regarded as essential for a functioning democracy, providing citizens with a voice in government decisions and ensuring that diverse interests could be represented within the political process.

However, pluralism has been criticized for overlooking significant power imbalances between groups, leading to the development of more nuanced models, including elite pluralism and regulatory capture theories (Lowi, 1969), neopluralism (McFarland, 2007), and structuralism (see Lindblom, 1982). Structuralism, in particular, shifted attention from the attributes of individual groups to a broader structural perspective, emphasizing the patterns of relationships among actors, including other interest groups, policymakers, and stakeholders (Knoke, 1990). This structural turn fundamentally reshaped the field, paving the way for methodologies like social network analysis (SNA) to map and analyze political networks with unprecedented precision.

SNA introduced a new dimension to the study of lobbying dynamics and interest group influence by highlighting the strategic positioning of actors within broader networks (Scott, 1991). Research by Baumgartner and Leech (1998) and Knoke et al. (1996) demonstrated that effective lobbying often depends less on direct financial contributions and more on an actor's position within a network. Strategic network positions, such as centrality and brokerage, allow groups to amplify their influence by controlling or facilitating the flow of information and resources (see also Mahoney and Baumgartner, 2015).

The application of SNA has proven instrumental in understanding how interest groups organize and mobilize to exert influence over policymakers. Studies such as those by Mahoney and Baumgartner (2015) emphasize that the success of lobbying efforts is not merely a function of financial or material resources but is significantly shaped by the structural positioning of groups

within networks of policymakers and stakeholders. Similarly, Knoke's comparative research on labor politics in the U.S., Germany, and Japan 1996 illustrates how differences in network structures across countries result in varying levels of influence exerted by labor unions on national policies. These studies highlight the role of structural and systemic factors in shaping political outcomes, offering a more nuanced understanding of the mechanisms that underpin interest group influence.

By leveraging SNA, researchers have uncovered the hidden structures that govern political influence, moving beyond traditional metrics of group power such as membership size or financial resources. Instead, the focus is on how interest groups and individuals navigate complex networks, form alliances, and leverage their positions to maximize their impact on policymaking (Zeng and Battiston, 2016). This network-based approach sheds light on the mechanisms of power that operate beneath the surface of pluralist democratic systems, offering a more sophisticated understanding of how influence is exerted and maintained.

Given the significant role of lobbying activities in shaping public interest, numerous national and supranational legislative efforts have sought to promote greater transparency in lobbying practices. For instance, initiatives such as the Transparency register [4] have aimed to regulate and disclose lobbying activities more effectively. However, systematically profiling the opinions and interests of lobbyists remains a complex challenge, often requiring innovative approaches and methodologies (Sluban et al., 2018).

This challenge has been amplified by the rise of social networking platforms, which have transformed how individuals express their positions and target specific groups or individuals. Social media platforms like LinkedIn have become powerful tools for professional networking and public engagement, enabling stakeholders to amplify their influence on public opinion and policy debates. Concurrently, these platforms have facilitated the profiling of interest group representatives, particularly lobbyists. Studies such as (Yilmaz et al., 2024) illustrate how digital engagement and online network positioning not only signal an individual's influence but also significantly shape it. In this context, LinkedIn emerges as a modern public *agora*, where lobbying efforts can intersect with broader social and professional dynamics.

[4] Available at https://transparency-register.europa.eu/index_en

Considering that lobbyists do not operate in isolation but rather "work in packs" (Mahoney and Baumgartner, 2015), understanding how individuals navigate between their digital presence and offline roles is critical for clarifying their broader impact on legislative processes. This study aims to bridge the gap between the in-person interaction network associated with the procedural work of the legislative process for the adoption of the EURO 7 regulation and the digital interaction network established by the same individuals on LinkedIn. Our analysis focuses on identifying key behavioural patterns to uncover potential connections between these in-person and online networks.

The goal of this research is to explore how online and offline interactions intersect, revealing whether structural correlations exist between these two spheres of activity. By doing so, we aim to understand how, in cases where offline lobbying activity information is missing or incomplete, analyzing online activity could help fill these gaps and provide complementary insights.

The remainder of this paper is organized as follows. In the next section, we review the most relevant research addressing lobbying activity through the lens of Social Network Analysis. We then present and describe our research setting and the data collection process. Following this, we introduce our methodology and the analyses performed, alongside the final results. The paper concludes with a discussion of the findings and their implications for the study of lobbying dynamics and interest group influence.

## 4.2 Literature Review

The application of Social Network Analysis (SNA) in lobbying research has gained significant traction in recent years, focusing on understanding the interactions and influence strategies among interest groups and different actors. One key element in this area of research is mapping the complex relationships between lobbyists, policymakers, and stakeholders within the political process. SNA enables researchers and practitioners to quantify and visualize not only the single interactions but also the overall structure of these interactions, which can significantly impact lobbying outcomes. Studies like Baumgartner and Leech (1998) have laid the foundation for network analysis by illustrating how power within the policy domain is concentrated

among key actors and how these relationships affect the policy process.

Bipartite networks (networks in which there are two classes of nodes and a link can connect only nodes belonging to two different classes) have been used to profile lobbyists interests as well as similarity among them(Scott, 2013), but also to connect political actors (Fischer et al., 2019).

More recent studies have built upon this foundation by integrating SNA into specific sectors, such as financial regulation, where networks have been shown to shift post-crisis. For instance, Redert (2022) explored the evolution of lobbying networks in the European financial regulatory environment following the 2007–2008 financial crisis, demonstrating that crises often lead to a diversification of lobbying actors and a breakdown of previously cohesive lobbying networks. This shift has broader implications for understanding how lobbying efforts adapt in times of regulatory change and crisis.

Furthermore, studies like the one of Kim and Kunisky (2020) proposed to apply advanced network models to reveal how special interest groups form coalitions and influence legislative decisions. In their research, they propose the integration of bipartite and community detection models in these studies has advanced the theoretical understanding of how interest groups operate within different political structures, such as the U.S. Congress.

The structural analysis of lobbying networks has also expanded beyond traditional political settings. Hofman and Aalbers (2017) discuss how spaces of lobbying are shaped by both physical and digital environments, offering a more spatial perspective on the flow of information and resources within lobbying networks. Their research adds an important dimension to the study of lobbying by examining the geographical and digital proximity of actors and their access to key decision-makers, opening the discussion regarding the definition of proximity with the advent of new communication technologies such as “the internet, phone, virtual meetings, and social media” (Hofman and Aalbers, 2017).

More recent studies have pushed the boundaries of Social Network Analysis (SNA), applying it to concepts and elements of discourse.

Fisher et al. (2013), for example, introduced the concept of Discourse Network Analysis (DNA), which examines the network of concepts linked together like a co-occurrence network,

representing the relationships between ideas in public discourse as well as in lobbying Leifeld (2017). DNA has gained traction in the last few years as an effective tool for studying the lobbying capacity of various actors, as demonstrated by Sick (2023), who applied this method to analyse the rhetoric used by lobbying groups.

One of the key challenges in accurately representing lobbying activities through Social Network Analysis (SNA) lies in profiling the actors, particularly in capturing their opinions and interests regarding the issues they aim to influence. Profiling stakeholders based on their roles, interests, and social positions within a network is an inherently complex task. This complexity is further exacerbated by the tactics and strategies adopted by lobbyists and policymakers to obscure or deliberately confuse the nature of their activities, despite efforts by national and supranational authorities to promote transparency, such as through initiatives like the Transparency Register (see Naurin, 2007, Veksler, 2016).

The aim of this research is to propose an innovative approach to addressing these challenges, particularly in filling the gaps surrounding the activities and profiling of individuals engaged in lobbying. In our specific case, this approach is applied to the European Union, where the unique institutional and legislative environment offers both opportunities and complexities for understanding lobbying dynamics. By integrating advanced SNA techniques with insights from digital engagement and network positioning, this research seeks to enhance the transparency and understanding of lobbying practices within the EU context. In the following chapter we outline the main approaches adopted with the scope of better profiling lobbyists and, eventually, propose our method of using a multiplex network.

### 4.2.1 Online and offline networks

Social Network Analysis (SNA) has been widely employed to understand and measure lobbyists' power (Jäger, 2013). However, as highlighted in the previous section, lobbyists' intent to obscure their activities can hinder researchers' ability to accurately portray the structure of lobbying networks. This limitation presents a significant challenge to the transparency and analysis of lobbying practices.

The emergence of social media platforms, however, has created and popularized new spaces

for interaction among individuals. While the nature and characteristics of individual interactions on these platforms differ from those in offline contexts, multiple studies have demonstrated behavioral similarities between the two. For instance, Dunbar et al. (2015), examining Twitter and Facebook, showed that structural characteristics found in online networks closely resemble those of offline networks. This suggests that despite the limited range of actions afforded by these platforms, online interactions can mirror offline relational patterns.

The purpose of social media platforms has evolved far beyond merely sharing moments and photos with friends. These platforms have rapidly assumed the role of a public *agora* (Colleoni et al., 2014, Kruse et al., 2018), where individuals share opinions, organize political actions, and engage in public discourse (Urman and Katz, 2022). While the rise of social media has raised concerns about its impact on political processes and public interest (see Boutyline and Willer, 2017), these platforms have also become arenas for institutionalized political actors (Sørensen, 2016). Specifically, Sørensen (2016) demonstrated how the active participation of Danish Members of Parliament on social media has recently increased direct engagement with citizens, driven primarily by the opening of direct channels of communication.

In this study, we ask whether similar behavioural patterns can be observed within lobbying dynamics. Specifically, we aim to explore the potential correlation between lobbyists' interactions on social media platforms and their targeting of specific individuals. By adopting a multiplex approach (De Domenico et al., 2013, Kivelä et al., 2014), we propose juxtaposing the formal network of meetings held during the legislative process for EURO 7 with the informal interaction network of the same actors on LinkedIn. While the formal meeting network is shaped by technical and procedural necessities (see Chalmers, 2013), the informal LinkedIn network reflects the free choices and initiatives of individual actors.

We hypothesize that the informal LinkedIn network will exhibit structural dependencies typical of social networks, such as reciprocity and popularity. Moreover, we anticipate a correlation between the formal meeting network and the informal interaction patterns observed on LinkedIn. Understanding these connections may offer valuable insights into how lobbyists navigate between formal institutional spaces and informal digital arenas, potentially shaping their influence in the legislative process. This approach not only sheds light on the interplay between

formal and informal lobbying practices but also highlights the increasing relevance of digital platforms in modern political dynamics.

To verify these hypotheses, we employ Exponential Random Graph Models (ERGMs), which are statistical tools used to analyse and represent the structure of networks by estimating the probability of connections between nodes based on specific patterns or attributes. ERGMs capture network dependencies such as mutual relationships, triadic closure, and node-specific characteristics, providing a realistic model of social interactions. These models are essential for understanding the social dynamics within networks, predicting future connections, and identifying influential actors or behaviours in complex systems.

## 4.3 Methodology

The data utilized in this research is multifaceted but can be preliminarily divided into two main categories. The first category pertains to the EURO 7 legislative process. Specifically, we collected and mapped the meetings that involved members of EU institutions, company representatives, and NGOs representatives. Gathering information on these meetings between EU institutions and stakeholders enables us to develop a comprehensive understanding of the social dynamics that unfolded throughout the legislative process. In particular, by analyzing the topological position of the participants within the process, we were able to identify the key players who played a significant role in influencing the opinions of both EU institutions and stakeholders.

Another crucial aspect tied to the EURO 7 legislative process is the ensemble of public consultations and feedback submitted by stakeholders to the European Commission via its institutional channel. At the time of data collection, a total of 177 documents were available at "have your say" dedicated page. These documents are critical to our analysis as the corpus of text they form allows us to identify the most important topics brought up when talking about the EURO 7 and, thus, which LinkedIn posts are relevant to our research.

The second category of data concerns LinkedIn activities carried out by individuals involved in the legislative process in various capacities. Specifically, we focused on the posts written by these individuals and the publicly traceable interactions among those in our dataset of meet-

ings. The aim of this dataset is twofold: first, to understand the emotions expressed in social media posts by individuals involved in the legislative process, and second, to trace and map the interactions among these individuals.

The data collection outlined here ultimately aims to establish a causal connection between the two groups of collected data. Specifically, this research seeks to produce a multilayer network, where the layers represent, respectively, the network of meetings held during the legislative procedure that led to the adoption of EURO 7 and the network of online interactions among the same individuals. The objective of this approach is to develop an understanding of social dynamics not only in isolation but in a way that takes into account potential correlations between the two layers.

### 4.3.1 Data description and Network construction

The dataset consists of 291 meetings (hereafter referred to as an *event*), with each *event* attended by, at least, 2 participants. Given the nature of the data we possess, it is necessary to develop a method to define our network. Specifically, the network is constructed as follows: for each dyad $i, j$ within the set of individuals who participated in an *event* , there is an edge $ij$.

$$\forall\,(i,j) \in V, \text{ if } i,j \in \textit{event} \Rightarrow \text{edge}(i,j) = 1$$

Thus, every *event* is characterized by a *clique*, as an edge is assigned to every possible pair of individuals participating in the same *event*. However, since there are overlaps between the groups of participants in various events, using the frequency of meetings between any two individuals $i$ and $j$ as the weight of that edge results in a network where individuals exhibit highly heterogeneous topological properties. In Figure 3, a graphical representation of this network is shown. The network consists of 123 subjects and 1779 edges.

In Figure 3, a graphical representation of the network in question is shown. The thickness of the ties connecting two nodes represents the weight of the connection, *i.e.*, the total number of meetings between the two connected individuals. The color of the nodes represents their affiliation, with EU institution members in blue and external entities in green. The size of the node reflects the betweenness centrality Freeman et al. (1991) of the node.

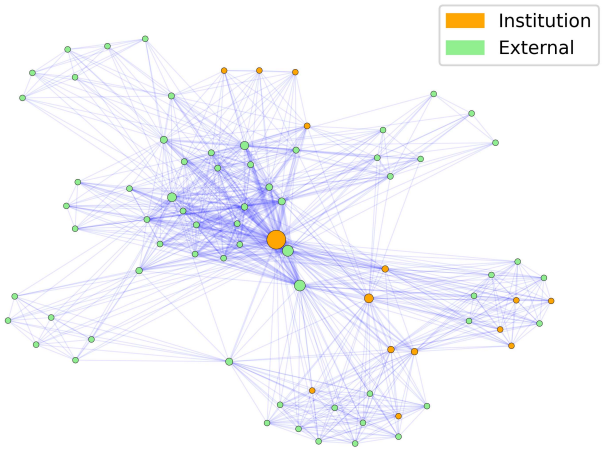
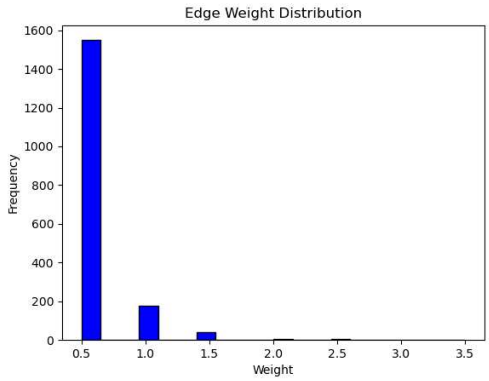



Figure 3: Meetings Network

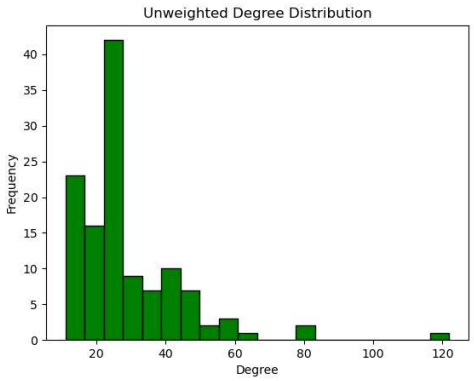
As seen in Figure 3, several cliques can be easily identified, particularly located in the peripheral areas of the graph. This suggests that these individuals participated in fewer meetings with other individuals. At the same time, several central actors can be identified at the core of the network. This central positioning is a result of their participation in multiple meetings with a diverse range of individuals, allowing them to establish a broad set of heterogeneous connections.

In Figure 4, we provide a summary of the characteristics of the network under analysis. As shown in Figure 4a, the majority of the edges have a weight of 0.5. Since this is an undirected network, the weight of each edge is halved. The maximum edge weight of 3.5 indicates that, at most, two individuals met multiple times. By observing the degree distributions (Figures 4b and 4c), we can see that most connections between subjects are relatively weak, especially when considering weighted edges.

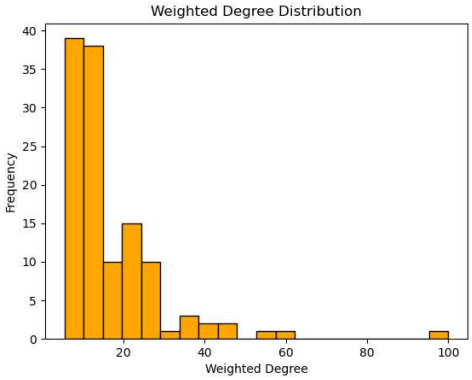
In Figure 4d, we can clearly observe that the vast majority of subjects have a betweenness centrality close to zero. This result stems from the fact that the network is composed of overlapping sets of cliques, meaning that very few individuals act as bridges between groups. Referring back to Figure 3, we can see that only a small number of subjects are bridging across different clusters. This feature, however, is not necessarily a drawback, as it highlights the significance



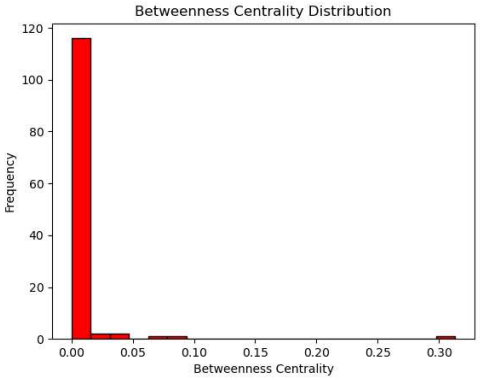

(a) Edge Weight Distribution




(b) Unweighted Degree Distribution




(c) Weighted Degree Distribution




(d) Betweenness Centrality Distribution

Figure 4: Network Analysis Plots

of an individual's role by virtue of their position within the network.

At this point, it is time to introduce the second dataset, which concerns LinkedIn activity.

With respect to LinkedIn data, we collected a total of 1,630 posts written by the subjects from either the moment their accounts were created or from the $1^{st}$ January of 2014. The posts included in this dataset were either written or reposted by individuals who participated in the meetings related to the EURO 7 legislative process and, as such, are part of the set of individuals forming the meeting network introduced earlier.

For each post, all available metadata was collected. Specifically, we gathered the text, hashtags, whether the post was a repost (and, if so, the original author), as well as information on individuals who liked, commented (including the text of the comment), or reposted the post. While these likes, comments, and reposts do not have a time-stamp and therefore cannot be precisely positioned in time, they provide valuable insight into the network of people interacting with these posts, and, in particular, the network of interactions among individuals involved in

the EURO 7 process. Simultaneously, we gathered information about the posts with which the individuals belonging to our meeting network interacted. This step is necessary to provide a detailed representation of the interaction network, both among the focus subjects and the potential neighbors they may have in common.

The construction of this network is much more straightforward: if a person liked, commented on, or reposted someone's post, a tie is created between the individual interacting with the post and the post's author. Unlike the meeting network, the LinkedIn interaction network is directed.

The resulting graph consists of 58357 nodes and 81991 ties. Focusing exclusively on individuals who participated in the EURO 7 discussions, the LinkedIn network is composed of 111 nodes and 722 ties. Specifically, the number of nodes in this latter graph is smaller than the number of nodes in the meeting network because many individuals either did not have a LinkedIn account or were not active on the platform. In Table 4, we can see a brief description of the networks and the specification regarding the condition for edge existence. In Figures 5a, 5b and 5c we can see the graphical representations of, respectively, the repost network, the comment network and the like network. The size of the nodes represent the Degree Centrality of each node while the color represents the emotions they expressed in posts regarding the EURO 7 (see *infra*). Finally, in Figures 6, 7 and 8 we can see the graphical description of the amount degree, weights and betweenness centrality distributions of the, respecctively, the repost, the comment and the like networks [5].

| Network Name | Condition for Link | N. Nodes | N. Edges | Density |
|---|---|---|---|---|
| Network Meetings | $i$ and $j$ participated in the same meeting | 168 | 1783 | 0.127 |
| **Likes Network** | $i$ **likes a post of** $j$ | **57002** | **71368** | **0.000** |
| Filtered Likes Network | $i$ and $j$ are present in Network Meetings and $i$ likes a post of $j$ | 105 | 538 | 0.051 |
| **Comments Network** | $i$ **comments on a post of** $j$ | **8069** | **8568** | **0.000** |
| Filtered Comments Network | $i$ and $j$ are present in Network Meetings and $i$ comments on a post of $j$ | 80 | 129 | 0.020 |
| **Repost Network** | $i$ **reposts content from** $j$ | **1865** | **2055** | **0.000** |
| Filtered Repost Network | $i$ and $j$ are present in Network Meetings and $i$ reposts $j$'s content | 46 | 55 | 0.026 |

Table 4: Networks Summary

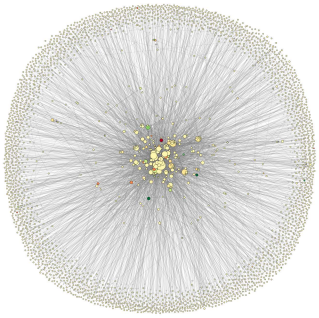
As it can be seen in Table 4, the networks relating to LinkedIn interactions are quite sparse and the degree distribution is particularly skewed. In Figures 5b and 5c we can see that the

[5] In the creation of these graphs, we reduced the set of nodes and edges by eliminating all the nodes that were not reciprocated. Such choice was taken due to the fact that, considering the way in which data were collected, most nodes had only one link.

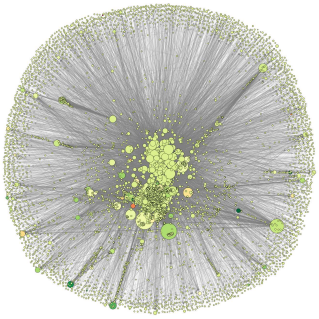
networks are characterized by a dense core composed by central agents. The ones that compose the cloudy periphery are the subjects that have interacted with the subjects that participated to the works for the EURO 7 but that are not strongly connected with the rest of the network. It appears that these subjects are unrelated individuals that are loosely interested in the posts produced by the focus subjects.



(a) Repost Network



(b) Comment Network



(c) Like Network

Figure 5: Overview of Interaction Networks: Repost, Comment, and Like Networks

In order to construct a multilayer network with the meeting network and the one regarding online interactions, it was necessary to link the two datasets. Since the data came from different sources, we had to perform a "names game" (Raffo and Lhuillery, 2009), *i.e.*, finding a way to ensure a safe matching process while avoiding false positives. After having pre-processed the names for an easier matching (see Raffo and Lhuillery, 2009), we opted for fuzzy matching, a

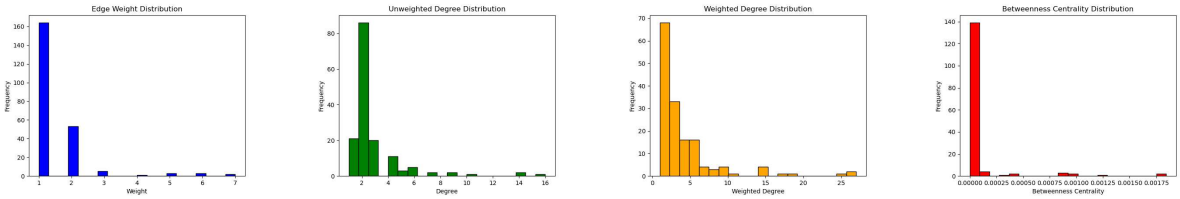



Figure 6: Network Metrics Distributions of Repost Network

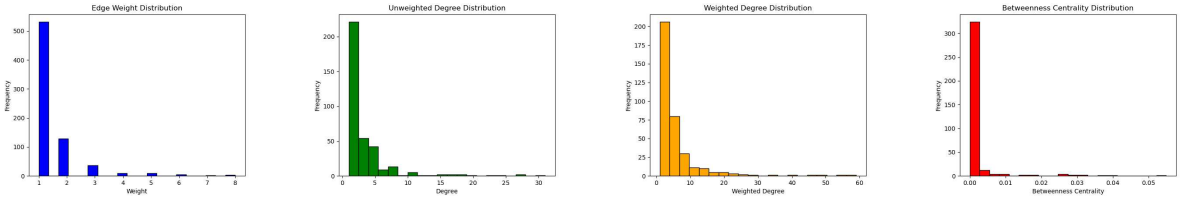



Figure 7: Network Metrics Distributions of Comment Network

technique used to compare strings by measuring their similarity rather than requiring an exact match. Specifically, in our case, we calculated the Levenshtein distance (the difference in terms of insertions, deletions, and substitutions) (Levenshtein, 1966, Cohen et al., 2003), obtaining a similarity score. If this score exceeded a threshold of 0.95, the two names were considered identical. Finally, we performed a manual check for matches with a similarity score above 0.8 to ensure that important observations were not missed.

### 4.3.2 Sentiment Analysis

Before proceeding with the network analysis, we sought to characterize the individuals based on the emotions expressed in their posts regarding EURO 7. To this aim, we filtered the posts for relevance and then analyzed them based on the emotions expressed within the content.

Applying topic modeling to social media posts can uncover the most prominent topics, while sentiment analysis determines whether the public discourse on these topics is generally positive, negative, or mixed. This approach has been widely used in public policy and social

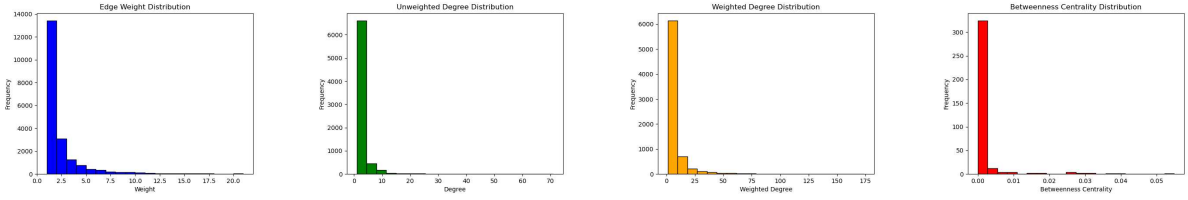



Figure 8: Network Metrics Distributions of Like Network

media studies, offering researchers valuable insights into both the subject matter and the public's emotional response (Blei et al., 2003, Pang and Lee, 2008).

*Latent Dirichlet Allocation* (LDA), introduced by Blei et al. (2003), is a popular method for topic modeling, particularly for large text datasets. LDA assumes that documents are mixtures of topics, where each topic is a distribution over words. Mathematically, the model is described as:

$$P(w|d) = \sum_{k=1}^{K} P(w|z = k) \cdot P(z = k|d)$$

where $P(w|d)$ is the probability of word $w$ in document $d$, $P(w|z = k)$ is the probability of word $w$ in topic $k$, and $P(z = k|d)$ is the probability of topic $k$ in document $d$.

In our study, the textual corpus used to train the LDA model consisted of 177 documents coming from public consultations and feedback submitted by stakeholders to the European Commission. This choice of corpus was driven by the desire to obtain texts from a diverse group of stakeholders, all converging on the central theme—EURO 7. This allowed us to bypass the need for additional classification steps regarding document relevance.

The LDA model was subsequently applied to the LinkedIn dataset, with posts considered relevant to a topic if the proportion of that topic exceeded 0.45 . This enabled us to analyze posts most closely related to the identified topics. The number of posts selected for further analysis was 1331.

Sentiment analysis was then performed on these posts using TextBlob, which assigns a polarity score ranging from -1 (negative) to 1 (positive). The assignation of a polarity score is single-word-wise and is based on a previously evaluated corpora of words. The sentiment scores were categorized into five groups: *extremely negative*, *mildly negative*, *neutral*, *mildly positive*, and *extremely positive*, based on the quantiles of the score distribution. This approach provided a nuanced view of the emotional tone across the LinkedIn posts (Pang and Lee, 2008, Shirky, 2011).

In Figure 9, we can see a stacked volume plot of the posts published on topics related to EURO 7, categorized by the emotion expressed in the text of the post. Specifically, some notable observations can be made. There is a noticeable increase in the volume of posts starting

from 2022. Additionally, it is relevant to highlight that there were distinguishable spikes in activity around the highlighted dates ($10^{th}$ November 2022, $9^{th}$ November 2023 and $24^{th}$ April 2024. The presence of these activity spikes suggests a connection between the progress of the work on EURO 7 and the activities (of the workers themselves) on the social media platform LinkedIn. In the next section, we will attempt to analytically explore the networks of meetings and LinkedIn interactions, with the aim of identifying a correlation between these two relational structures.

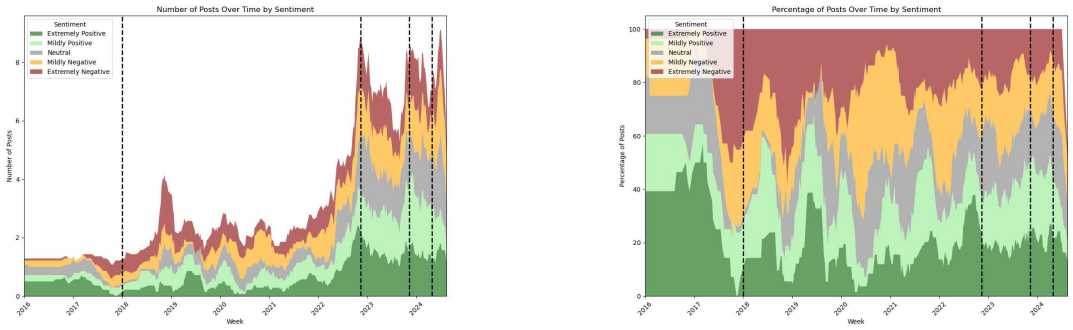



Figure 9: Stacked volume of posts divided by emotion (Percentage on the right)

In order to better understand the structural characteristics of the networks we are dealing with, we decided to utilize Exponential Random Graph Models (ERGM). The use of ERGMs in this study aims to explore the relationship between two networks: the in-person meeting network and the online LinkedIn interaction network. By applying a multilayer approach, we seek to understand how face-to-face interactions might influence online behaviours among the same individuals, and whether structural correlations exist between these networks.

ERGMs are essential for capturing the complexity of interactions by modeling dependencies between ties, offering insights into how one network (*e.g.*, meetings) might drive the dynamics of another (LinkedIn) (Robins et al., 2007). For instance, face-to-face interactions might facilitate online connections, or online exchanges might influence in-person negotiations (Lusher et al., 2013).

This multilayer framework considers both the formal nature of the meeting network and the informal dynamics of LinkedIn, where interactions extend beyond formal contexts (Snijders et al., 2006). Linking these layers allows us to assess whether frequent or intense in-person

meetings lead to a higher likelihood of online engagement, and whether the same actors hold central positions in both networks. These key actors may serve as bridges between formal policy processes and informal public discussions, amplifying influence across platforms (Yilmaz et al., 2024).

By examining how the meeting network impacts online interactions, this approach provides a deeper understanding of how social dynamics in policy making unfold across both structured and informal environments.

Table 5: ERGMs Coefficient Table for *like-network*

| | Model 1 | Model 3 | Model 5 | Model 7 | Model 9 | Model 11 | Model 13 | Model 15 | Model 17 | Model 19 |
|---|---|---|---|---|---|---|---|---|---|---|
| edges | -4.295*** | -4.416*** | -4.478*** | -4.365*** | -4.368*** | -4.594*** | -6.411*** | -3.513*** | -4.056*** | -6.781*** |
| | (0.059) | (0.065) | (0.064) | (0.062) | (0.062) | (0.072) | (0.124) | (0.155) | (0.177) | (0.264) |
| mutual | 3.558*** | 3.417*** | 3.382*** | 3.530*** | 3.549*** | 3.359*** | 1.606*** | 3.570*** | 3.357*** | 1.514*** |
| | (0.157) | (0.167) | (0.164) | (0.157) | (0.161) | (0.164) | (0.202) | (0.154) | (0.163) | (0.204) |
| comment | 5.002*** | | 4.768*** | | | 4.776*** | | | 4.797*** | 3.557*** |
| | (0.319) | | (0.313) | | | (0.335) | | | (0.324) | (0.382) |
| meeting | | 0.935*** | 0.857*** | | | 0.839*** | | | 0.814*** | 0.530*** |
| | | (0.082) | (0.088) | | | (0.088) | | | (0.088) | (0.113) |
| betweenness_like (in) | | | | | | | -0.002*** | | | -0.003*** |
| | | | | | | | (0.0002) | | | (0.0003) |
| betweenness_like (out) | | | | | | | -0.003*** | | | -0.003*** |
| | | | | | | | (0.0002) | | | (0.0003) |
| degree_like (in) | | | | | | | 0.131*** | | | 0.140*** |
| | | | | | | | (0.007) | | | (0.008) |
| degree_like (out) | | | | | | | 0.124*** | | | 0.127*** |
| | | | | | | | (0.007) | | | (0.008) |
| emotion (in) | | | | | | | | -0.728 | -0.012 | 1.850* |
| | | | | | | | | (0.696) | (0.745) | (0.810) |
| emotion (out) | | | | | | | | -3.147*** | -2.944*** | -1.377. |
| | | | | | | | | (0.714) | (0.743) | (0.826) |
| inst→ext | | | | | 0.199 | 0.072 | | | 0.023 | 0.154 |
| | | | | | (0.135) | (0.146) | | | (0.143) | (0.161) |
| ext→inst | | | | | 0.663*** | 0.613*** | | | 0.608*** | 0.394* |
| | | | | | (0.119) | (0.124) | | | (0.125) | (0.156) |
| inst→inst | | | | | 0.841*** | 0.398. | | | 0.339. | 0.430. |
| | | | | | (0.164) | (0.208) | | | (0.205) | (0.258) |
| AIC | 4047 | 4225 | 3960 | 4297 | 4297 | 3942 | 3075 | 4314 | 3929 | 2948 |
| BIC | 4071 | 4249 | 3992 | 4321 | 4337 | 3998 | 3123 | 4346 | 4001 | 3052 |

. $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: ERGMs Coefficient Table for *comment-network*

| | Model 2 | Model 4 | Model 6 | Model 8 | Model 10 | Model 12 | Model 14 | Model 16 | Model 18 | Model 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| edges | -7.347*** | -6.474*** | -7.526*** | -6.553*** | -6.579*** | -7.869*** | -7.848*** | -5.251*** | -7.958*** | -8.378*** |
| | (0.267) | (0.180) | (0.279) | (0.189) | (0.200) | (0.316) | (0.282) | (0.443) | (0.619) | (0.812) |
| mutual | 1.825** | 3.552*** | 1.522* | 3.541*** | 3.552*** | 1.347* | 0.716 | 4.020*** | 1.286* | 0.291 |
| | (0.614) | (0.595) | (0.623) | (0.608) | (0.608) | (0.629) | (0.686) | (0.551) | (0.587) | (0.696) |
| like | 4.990*** | | 4.817*** | | | 4.742*** | | | 4.753*** | 3.742*** |
| | (0.310) | | (0.320) | | | (0.326) | | | (0.319) | (0.384) |
| meeting | | 1.594*** | 0.772** | | | 0.691** | | | 0.696** | 0.035 |
| | | (0.247) | (0.279) | | | (0.280) | | | (0.271) | (0.341) |
| betweenness_comment (in) | | | | | | | -0.035*** | | | -0.018** |
| | | | | | | | (0.005) | | | (0.006) |
| betweenness_comment (out) | | | | | | | -0.028*** | | | -0.020** |
| | | | | | | | (0.006) | | | (0.007) |
| degree_comment (in) | | | | | | | 0.597*** | | | 0.331*** |
| | | | | | | | (0.054) | | | (0.066) |
| degree_comment (out) | | | | | | | 0.495*** | | | 0.302*** |
| | | | | | | | (0.065) | | | (0.081) |
| emotion (in) | | | | | | | | -3.220. | -2.125 | -3.166 |
| | | | | | | | | (1.890) | (2.145) | (2.799) |
| emotion (out) | | | | | | | | -0.902 | 2.535 | 3.929 |
| | | | | | | | | (1.844) | (2.047) | (2.559) |
| inst→ext | | | | | 1.382*** | 1.128*** | | | 1.160*** | 1.055** |
| | | | | | (0.316) | (0.334) | | | (0.333) | (0.403) |
| ext→inst | | | | | 1.010** | 0.436 | | | 0.438 | -0.445 |
| | | | | | (0.350) | (0.437) | | | (0.437) | (0.445) |
| inst→inst | | | | | 2.241*** | 1.579*** | | | 1.591*** | 0.916. |
| | | | | | (0.367) | (0.425) | | | (0.425) | (0.557) |
| AIC | 552.2 | 806.4 | 545.4 | 811.5 | 813 | 533.5 | 608.7 | 845.6 | 535.4 | 493.7 |
| BIC | 576.2 | 830.4 | 577.4 | 835.5 | 853 | 589.5 | 656.7 | 877.6 | 607.5 | 597.7 |

. $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

## 4.4 Results and Discussion

The two tables present the results of several Exponential Random Graph Models (ERGMs) applied to both the *like-network* and the *comment-network*, which provide insights into the structural features that influence online interactions related to EURO 7. In these models, the networks are weighted, where the weight of an edge indicates the frequency of in-person meetings between the connected individuals. The ERGMs explore multiple factors that contribute to the formation and persistence of ties within these networks, including mutual interactions, online engagement (likes and comments), and the influence of in-person meetings.

In Table 5, which focuses on the *like-network*, the coefficient for edges is consistently negative and highly significant across all models suggesting network sparseness. The positive and significant coefficient for mutual interactions highlights that reciprocity is a dominant feature in the network. Users who like each other's posts are likely to reciprocate, forming stronger mutual ties over time.

Models that include the comment variable show a strong positive relationship between likes and comments. This indicates that when individuals comment on a post, they are more likely to also like that post. This is an expected behaviour, as deeper engagements often involve multiple forms of interaction.

What is particularly relevant to this study is the role of in-person meetings. The variable representing meetings is significant across most models, demonstrating that individuals who meet face-to-face are more likely to interact online through likes. This finding suggests that real-world interactions foster stronger virtual engagements, reinforcing the idea that personal meetings enhance online network activity.

Additionally, we see that certain attributes of the participants also play an important role in shaping online engagement. The directed ties from institutional actors to external stakeholders show positive and significant coefficients in several models, suggesting that institutional actors are central to promoting online engagement. This highlights the role of institutions in creating and fostering online interactions, especially through likes. The AIC values suggest that Model 11, which includes mutual interactions, comments, meetings, and directed ties, provides the best fit for the structure of the *like-network*.

Turning to Table 6, which focuses on the *comment-network*, similar trends are observed. The edge coefficient remains negative and highly significant, indicating that not all potential ties between nodes are realized, which again suggests network sparseness. However, the mutual interaction coefficient remains positive and significant, indicating that users tend to engage in bidirectional communication, reciprocating by commenting on each other's posts. This reciprocity mirrors the findings from the *like-network*, reinforcing the idea that social exchanges online are often reciprocal.

The inclusion of the variable for likes shows a strong positive correlation, meaning that individuals who like posts are also likely to comment on them. This finding indicates that liking is often a precursor to deeper engagement in the form of comments. The meeting variable, once again, is significant, indicating that offline interactions lead to increased online interactions, this time through comments. This supports the hypothesis that real-world connections amplify digital interactions across both likes and comments.

Directed ties between institutional actors and external stakeholders are significant in several models, particularly in Model 12. The positive coefficients for these ties suggest that institutional actors are instrumental in driving online discussions, actively engaging with external stakeholders and reinforcing the ties between these two groups.

Across both tables, several key themes emerge. First,in line with Lewis (2015), reciprocity is a dominant force in both the *like-network* and the *comment-network*. The likelihood of mutual interactions is high, reinforcing the bidirectional nature of online engagement. Second, we can notice how degree centrality, have a positive impact on the likelihood of giving and receiving a like or comment, reflecting a reinforcement mechanism, where higher centrality leads to continued activity. The correlation between offline meetings and online interactions is evident. Individuals who meet in person are more likely to engage with one another online, whether through likes or comments. This reinforces the idea that offline relationships are vital for fostering online networks, suggesting a strong link between face-to-face meetings and subsequent digital engagement. Finally, institutional actors play a central role in shaping both networks, as they not only influence the formation of ties but also drive the conversations and interactions that occur within these networks. At the same time we found mixed or non significant results

when it comes to the impact that the expressed emotions had on the relations among subjects. Future research could consider different instruments to assess the positivity and negativity of the content in a post. Furthermore, another venue for research would regard the opinion regarding the EURO 7, rather than the emotion, expressed in a post. This type of approach could also focus on different structural conformations such as echo chambers and better help with the profiling. As a matter of fact, for some subjects it is not possible to assess their opinion regarding the EURO 7 regulation and the linked topics. However, it could be possible to estimate the opinion of one subject by taking into consideration the users they're interacting with.

This paper wanted to present the characteristics of the networks representing, on the one hand, the meetings that incurred among subjects within the EURO 7 works and, on the other hand, the interactions occurred among the same subjects on the social media platform LinkedIn. The scope of this approach is to propose a multilayer network composed by a formal and a more informal network and to, by doing so, develop a better comprehension of the relational dynamics that are undergoing and of the prominence of specific actors by virtue of their position within the position they occupy within one network. With the help of LDA and sentiment analysis we also tried to characterize the involved subjects according to the opinion they expressed in the post they shared on the platform with the eventual aim of understanding whether such emotion can influence the structure of the network. We found some correlations with the two networks and also highlighted the role that institutional subjects have in driving the interactions on the online platform but we didn't find sufficient results for the emotional aspect of our research. Nevertheless, the preliminary results here highlighted prove to be sufficiently strong and significant to suggest further research.

# 5 I'm Politer than you just because I think you are

**Sasha Piccione and Nicolas Jullien**[6]

**Abstract**

This study examines how participants in Online Open Collaborative Communities, specifically Wikipedia, express emotions while cooperating considering the limitations of Computer Mediated Communication (CMC). Using sentiment analysis, we analyse over 312,000 Wikipedia Talk Pages to explore the influence of context on emotional expression. We assess the arousal (emotional charge) and valence (pleasantness) of comments, correlating these with the tone of surrounding discussions and the emotional patterns of users. Our findings show that context, especially at the thread level, significantly influences the emotions expressed, with mimicry and emotional appraisal playing key roles. This research contributes to understanding the complexities of online collaborative work and offers a valuable dataset for future studies. The insights could also inform the development of tools to manage emotions and prevent toxic behavior in online communities.

***Keywords—*** Online Open Collaborative Communities, Wikipedia, Sentiment Analysis.

[6]IMT Atlantique, LEGO-M@rsouin. Technopole de Brest Iroise, 29238 Brest, France. Nicolas.Jullien@imt-atlantique.fr

## 5.1 Introduction

The COVID-19 pandemic has fostered the diffusion of virtual teamwork, *i.e.* team members working independently and communicating through digital infrastructures. This tendency is connected to the rise of Online Open Collaborative Communities (OOCCs), *i.e.* communities in which any users can intervene and participate to the creation process that we witnessed in the last few decades. Such organizational shift entails several challenges (Marlow et al., 2017, Morrison-Smith and Ruiz, 2020) as the technical characteristics of such infrastructures constrain what actions are possible and how the collective work is organized (Ronzhyn et al., 2023),

and more specifically in this research, the ability of group members to communicate and coordinate effectively.

Indeed, when looking at the main theories of group coordination, we can discern two branches. The first group pertains to explicit coordination, encompassing situations where individuals directly communicate their actions or intentions to identify and manage tasks effectively. While theories of implicit coordination (e.g. Rico et al., 2008) suggest that, by sharing mental models, team members can predict and adjust behaviours without communication. However, the reliance on shared mental models poses limits on the viability of implicit coordination. For example, in OOCCs there is no formal authority to impose a particular way of working or solving disputes among participants.

A third alternative to the traditional explicit-implicit coordination dichotomy is coordination through the work product itself, a particularly effective approach in text-based online organizations (Bolici et al., 2016, Arazy et al., 2016). This mechanism, akin to the biological process of "stigmergy" (Grassé, 1959, Marsh and Onof, 2008), is defined by Heylighen (2007) as a process where "the work [...] done by one agent provides a stimulus ('stigma') that entices other agents to continue the job."

In this context, the artifact serves as "the sign (or contribution), the interface, the location, and the processing defining the semantics of stigma in stigmergy" (Ricci et al., 2007). While stigmergy effectively explains why users converge on whether a task should be addressed, it falls short of explaining how they agree on the approach to completing it. Coordination requires not

only consensus on what should be done but also on how to build upon others' contributions. These complex actions often necessitate explicit communication Dipple et al. (2014), even in online environments (see Weltevrede and Borra, 2016, Bolici et al., 2016).

Considering the importance of asynchronous communication in these contexts, the quality of communication remains a critical factor in determining a virtual group's performance. High-quality communication fosters trust (Jarvenpaa and Leidner, 1999, Jarvenpaa and Keating, 2021, Bleakley et al., 2022) and minimizes the risk of misunderstandings (Hwang, 2013). However, the anonymity characteristic of Computer Mediated Communication (CMC) poses significant challenges, stripping interactions of social cues and reducing the emotional depth of exchanges (Greineder and Blohm, 2020).

In such settings, the only emotions contributors in OOCCs can perceive are those conveyed through the text used for coordination. This corpus of text serves a dual role: it provides information about completed and pending tasks while also reflecting the emotional and relational dynamics at play during earlier stages of the creative process. This interplay highlights the importance of understanding how text-based communication impacts group dynamics and individual emotional expression.

While previous research has focused on how demographic factors influence contributors' emotions (Gallus and Bhatia, 2020), less attention has been paid to how the collaborative context itself shapes these expressions. More precisely this paper aims to investigate how the emotions embedded in the text corpus used for coordination influence the emotions expressed by individuals in subsequent interactions, offering new insights into the emotional dynamics of virtual collaboration.

To do so, we draw on well-established psychological theories, such as emotional mimicry (Hess and Fischer, 2022) and Communication Accommodation Theory (Bernhold and Giles, 2022). Both theories emphasize the tendency to adapt or imitate the language and emotional tone of one's interlocutors. We aim to investigate how the emotions embedded in the text corpus used for coordination influence the emotions expressed by individuals in subsequent interactions, offering new insights into the emotional dynamics of virtual collaboration.

The dataset used in this research consists of collected and processed data from discussions

taking place in *Talk Pages* associated with 312,088 Wikipedia *Articles*. As one of the most prominent computer-mediated cooperative platforms, Wikipedia serves as an ideal setting for this study. The platform lacks a hierarchical conflict resolution mechanism (Lerner and Lomi, 2020), relying instead on the autonomous management of conflicts and disputes (Weltevrede and Borra, 2016).

This research leverages Wikipedia's fully traceable interaction history, where all comments and modifications are publicly accessible, to build an extensive and detailed dataset. These unique characteristics make Wikipedia an exceptional case study for exploring the complexities and effectiveness of communication mechanisms in virtual teamwork.

Guided by key theories in emotion psychology (Scherer, 2005, Gallois et al., 2005), we investigated the factors influencing the tone and emotions expressed by users on Wikipedia's *Talk Pages*. These pages serve as the primary space for coordination and planning of knowledge contributions on Wikipedia (Viégas et al., 2007a).

In this study, emotions are evaluated along two dimensions: *arousal* and *valence*. *Arousal* reflects the degree of excitement and energy associated with a word or text, while *valence* refers to its pleasantness (Warriner et al., 2013). Our analysis spans the *Talk Pages* of 312,088 articles, assessing the contextual factors that shape the emotional tone of user comments, such as the general tone of the page, the specific tone of the comment being replied to, and the usual emotional behaviour of the comment's author.

Unlike previous studies on Wikipedia's *Talk Pages* (Gallus and Bhatia, 2020, Iosub et al., 2014), this research builds its variables solely on the information visible to participants—specifically, the text of the comments. By focusing exclusively on observable data, such as user interactions within *Talk Pages*, we aim to provide actionable insights and recommendations for practitioners. For this reason, we excluded unobservable elements, such as demographic information, which are inaccessible to Wikipedia contributors and hence irrelevant for the scope of our study.

The results show that the emotions expressed in the context are positively correlated with the emotions expressed by users in subsequent comments. Furthermore, in line with Communication Accommodation Theory (CAT), we demonstrate that the subjective evaluation of emotions within the context influences deviations from the author's emotional baseline. This mechanism

also emerges in cases of extreme behaviour, where the context significantly impacts the likelihood of high-*valence* or high-*arousal* comments. Interestingly, the magnitude of contextual influence varies across levels, with the thread level consistently emerging as the most impactful.

This research contributes in two primary ways. First, it provides insights into how the emotions expressed by users in collaborative settings are shaped by the ongoing discussion and context. Second, it proposes a tailored methodology to test well-established theories, such as mimicry and CAT, in a setting characterized by a substantial number of observations and near-complete transparency of the information accessible to participants. This transparency minimizes the number of extraneous variables, enabling a more focused and robust analysis. The chosen research setting not only facilitated the production of a large dataset—an important contribution in itself—but also allowed us to construct it with a comprehensive understanding of the data available to the observed subjects.

In sum, this study not only advances our understanding of the role of emotions in collaborative online environments but also offers a replicable methodological framework for exploring the interplay between emotion, context, and cooperation in other settings.

The article is structured as follows. In Section 5.2 we provide provide an overview of the characteristics of the relationship between coordination with a particular focus on Computer Mediated Communication environments. Taking inspiration from main psychology theories regarding emotion expression, we provide insights regarding the role of the context in defining one's emotional expression. In Section 5.3 we present our case study and the method we used to construct the dataset. Section 5.4 present our results, before discussing in Section 5.5 the relevance of our first results, but also their limitations and how they can be addressed with the aim of furthering the research area in which this paper is set.

## 5.2 Literature review

In this research we aimed at studying the contextual factors that influence the way in which people express their emotions in Wikipedia *Talk Pages*. In order to do so, in this section we will present the role that emotions play in cooperative process and how they are experienced and expressed by individuals. We, then, explore the peculiarities of Computer Mediated Commu-

nication (CMC) with particular attention to the role of communication and its different forms. We, then introduce Wikipedia, our case study that exemplifies CMC environments. We present its peculiarities and recent findings in studies relevant to our research. This will allow us to propose a specification of the emotions we intend to analyse and to define our hypotheses.

### 5.2.1 The role of emotions in communication

Communication does not regard only the transfer of information but also contextual aspects. Affective communication is used to communicate information on the goals, beliefs and intents of an individual towards social exchange (Van Kleef, 2009). This "information" helps group members to coordinate their efforts. For instance, (Tsai et al., 2012) found that higher positivity in the tone expressed by group members can compensate for a lack of trust within the group, positively impacting creativity in low-trust contexts. Furthermore, according to the literature on psychological safety and voice behaviour, when team members feel safe and comfortable in their environment, they are more likely to express their thoughts and encourage others to participate in team tasks (Edmondson, 1999, LePine and Van Dyne, 1998). In high-trust groups, the communication of anger and disappointment can foster cooperation by preventing retaliation dynamics (Wubben et al., 2009). Additionally, the communication of emotions plays a crucial role in building trust, as it helps determine and enforce social roles (Vivian et al., 2016).

Several studies have proved the idiosyncratic aspects that influence emotion expression, such as gender, social status (Gallus and Bhatia, 2020, Hargrave and Langengen, 2021), as well as the language and the vocabulary one can use to communicate with others (Panayiotou, 2004). Positions of power also affect how people express themselves (Berdahl and Martorana, 2006), and this can be leveraged by the recipients of the communication. This last aspect seems to indicate that emotions are also context-dependent: one feels and expresses an emotion according to one's perception and evaluation of the context. Indeed, according to general appraisal theory (Scherer, 2005), the emotions one experiences and expresses - through facial and body expressions, voice tone, and lexical choice - result from an individual's evaluation of a specific external stimuli. In various contexts, individuals use personal dispositions and social principles to appraise and react to external stimuli (Scherer, 2021). In teamwork contexts, emotions can

be triggered by internal or external stimuli if they are evaluated as goal-relevant (Watzek et al., 2019).

Shifting to an approach that adopts a more interactive view, Communication Accommodation Theory (CAT) (Bernhold and Giles, 2022) and theories regarding emotional mimicry (Hess and Fischer, 2022) focus on the interaction between subjects and the perception-behaviour link (Chartrand and Bargh, 1999). They suggest that people involved in an exchange tend to converge their verbal and nonverbal communication (Swaab et al., 2011). Such convergence involves enacting behaviours and emotional expressions similar to those of the interlocutor, as speakers use both conscious and unconscious mimicry mechanisms to signal the sharing of appraisals and intentions (Hess and Fischer, 2022).

While these theories primarily address contexts rich in social and contextual cues, such as those providing information about the identity or emotions of the interlocutor, the dynamics can be markedly different in online open collaborative communities. In these environments, the reconstruction of social identity and the identification of interlocutors' emotions must be achieved with far less contextual information, relying almost entirely on the textual content of interactions. This shift from rich, face-to-face interactions to sparse, text-based communication introduces unique challenges and adaptations that warrant closer examination. These nuances, specific to online contexts, will be explored in greater detail in the following subsection.

### 5.2.2 Communication and coordination in Online Open Collaborative Communities

Online Open Collaborative Communities (OOCCs) are organizations formed by geographically dispersed users. In these communities, the identities of participants are not necessarily known and, in some cases, may even be indiscernible (Tanis and Postmes, 2007, Faraj et al., 2011). These users collaborate to achieve a shared goal, distinguishing OOCCs from general Online Communities (OCs) such as Facebook and Instagram.

The fluid nature of OOCCs enables the definition of goals of interest through stigmergy, as noted by Elliott (2016), Crowston and Rezgui (2020), Zheng et al. (2023). The concept of stigmergy was proposed by Grassé (1959) to describe how termites indicate to each other the position in which they are expected to leave building material by impregnating such build-

ing material with specific hormones. Such dynamic can be used to explain how ant can guide others to food sources or how architects and builders coordinate their tasks through artifacts such as drawings Christensen (2007, 2013, 2014, 2013). Stigmergic coordination has been emphasized in more complex processes where individuals must coordinate their actions using a particularly limited amount of information. Specifically, Bolici et al. (2016), in analysing interactions among individuals in Free/Libre Open Source Software (FLOSS) projects, highlighted how the artifact itself serves not only as a product but also as a coordination tool. Each individual coordinates their actions by observing what has already been done, ensuring that their contribution aligns with previous work and facilitates coordination for future contributors. Similarly, in OOCCs, contributors can coordinate by relying on indirect communication, either by creating artifacts (Susi and Ziemke, 2001) or by modifying them (Dorigo et al., 2000). In the case of OOCCs, when individuals collaborate on a shared project, their efforts and the resulting changes serve as encouragement for others to participate and contribute as well (see Dorigo et al., 2000).

Stigmergy, while effective in explaining how users converge on the decision of which topics to address, falls short of explaining how they converge on how these topics should be addressed. As a result, horizontal organizations often require dedicated spaces for explicit communication, enabling participants to coordinate not only on what needs to be done but also on how to contribute to others' work.

Computer-mediated communication (CMC) in such contexts, however, is inherently challenging due to issues related to anonymity, deindividuation, and the fluidity of social identities. The fluid nature of these communities complicates the definition of contributors' social identities, normalizing anonymity and fostering ambiguity regarding the identities of participants (Faraj et al., 2011). Anonymity, in turn, has mixed effects on collective production processes (Tanis and Postmes, 2007). According to general deindividuation theory (Coe et al., 2014), anonymity can encourage aggressive behaviour. Conversely, Social Identity theories (see Hornsey, 2008, for a review) suggest that anonymity may enhance the salience of group identity, increasing the influence of social norms on individual behaviour.

In the absence of clearly defined social identities, contributors must rely on limited social

cues, which are often difficult to discern. The spaces where users interact to coordinate collective creation processes also serve as the primary source of social cues and, consequently, of social roles that guide users' expectations about others' behaviour (Walther, 1997). In particular, written messages left by contributors act as artifacts that, in a manner resembling stigmergic mechanisms, signal not only the content—the information to be transferred—but also the context, including emotions, appraisals, and social cues.

These textual artifacts are thus central to coordination. As noted by Arazy et al. (2016), the artifact encompasses not only the final output but also the entire history of changes, contributions, and communication among contributors. This corpus represents the foundation of asynchronous, computer-mediated communication that characterizes these platforms. Text provided by peers serves as normative guidance to which contributors tend to adhere (Rösner and Krämer, 2016). However, despite the importance of expressed emotions in collaborative processes, the role of contextual factors in regulating the emotions expressed by users in CMC requires further investigation. The next subsection explores why Wikipedia provides a rich case study for addressing these questions.

### 5.2.3 Wikipedia: an exceptional case of anonymous cooperation

Wikipedia is a social media platform that has increasingly attracted attention due to its peculiar affordances (social and technical characteristics that allow specific actions - and dynamics (Weltevrede and Borra, 2016) and the relevance of its main output available to subjects who are not contributors. The attention it has gathered ranges from different disciplines, from organisation theory (Forte et al., 2009) to gender studies (Gallus and Bhatia, 2020) and to political sciences (Rijshouwer et al., 2023). Even though Wikipedia exemplifies an emerging collaborative organization (Padgett and Powell, 2012), its collaboration process is neither linear nor peaceful, as conflicts and "edit wars" are common occurrences on the platform (Weltevrede and Borra, 2016).

From an editorial point of view, the core affordance of the system is to organise the collaborative creation of an encyclopedia, and to regulate the contributions as they all come together to form a single corpus. Such mechanism is summarised in the opening sentence on its welcome

page: “the free encyclopedia that anyone can edit.”

More concretely, the Information System is built to accept only those contributions that are both good (that add knowledge to the existing set) and correct, *i.e.*, that the content of the contribution can be agreed upon by users (Rijshouwer et al., 2023).

This organisational layout and these processes are embedded in the tools that allow contributors to participate, and allow for some forms of “stigmergic” dynamics and mechanisms (Zheng et al., 2023): contributors can coordinate by relying on indirect communication by creating artifacts (Susi and Ziemke, 2001) or leaving specific traces Dorigo et al. (2000). For instance, Zheng et al. (2023) suggest that, by contributing - *i.e.* adding content, sources or editing - users of the community signal the relevance of a given topic to others and, by doing so, elicit their intervention.

However, coordinating suppose a consensus on what is to be done, and how one can add to someone else’s work. Considering that contributors intervene when they believe that the reported information is wrong or not satisfying, it takes time to reach a collective agreement and coordination has phases of discussion, debate, even fights (Kittur et al., 2007).

A significant proportion of the planning and quality control processes associated with Wikipedia *articles* are conducted on the platform’s *Talk pages*. This observation is supported by the findings of Viégas et al. (2007a) and Gallus and Bhatia (2020). A *Talk page* is associated with each Wikipedia article. This page is intended for the submission, discussion and critique of contributions or changes made by other users. Similarly, the discussion process follows a thread format, whereby users can either initiate a new discussion or reply to a specific comment.

In light of the significance of a productive and accommodating discussion in terms of knowledge production (Jarvenpaa and Keating, 2021), it is crucial to identify the factors that contribute to a positive and constructive discussion in this environment. It is therefore necessary to ascertain whether the emotion expressed in written comments on Wikipedia’s *Talk Pages* is linked to the subject matter and to determine the impact of such comments, as the primary contextual elements, on the tone of a new comment. In the following subsection we expose the dimensions of emotions that we intend to explore and propose our hypotheses.

### 5.2.4 Hypotheses on emotion expression

In this study, similarly to (Gallus and Bhatia, 2020), we consider emotions in two dimensions, *i.e.*, *valence* and *arousal*. The former indicates the positivity of an emotion while the latter regards the intensity of that emotion. The reasoning behind this choice is twofold. On the one hand, these two emotions, together with *power*, have largely been considered the main emotional components of verbal expressions (Bradley and Lang, 1999). At the same time, the dataset provided by Warriner et al. (2013) comprises a particularly wide set of evaluated words which, as we will see, will facilitate our approach to estimate the emotions expressed within a corpus of text. In Section 5.3, we provide a thorough description of the measurement of these emotions. The choice of these two dimensions is driven by their frequent use in psychology research when evaluating the emotions expressed through words (Bradley and Lang, 1999, Osgood et al., 1957). We decided to consider only *valence* and *arousal*

As previously discussed, the theoretical explanations regarding the contextual factors that influence the emotional expression of a subject can be divided into two main groups.

The *mimicry* behaviour, *i.e.*, when someone's verbal and non-verbal communication style and content converge with that of their interlocutor. In the case of Wikipedia's Talk Pages, the context is limited to the text that other users have written. This theory aligns with Rösner and Krämer (2016), who suggest that the text embodies a normative set of guidelines to which users tend to adhere. The emotional intensity expressed by the corpus of text representing the context that the user faces should influence the new comment's emotional intensity. We, therefore, hypothesize:

*H1. The emotional valence (arousal) of a comment is positively associated with the valence (arousal) level of the contextual text in which it is embedded.*

The second group of theories that seek to understand how context influences the emotions experienced and expressed by an individual, which we have considered in this research, aims to highlight how such a connection must involve the subject's evaluation.

Emotional appraisal theory (Scherer, 2005) and Communication Accommodation Theory (Gallois et al., 2005), in fact, imply that an individual experiences and expresses an emotion based on their evaluation of an external stimulus.

In our specific case, the external stimulus is always the corpus of text visible to an individual within the *Talk Page*. In this research, we assume that appraisal occurs by using each individual's own language as the baseline. More specifically, subjects will perceive a comment as having a high level of arousal if its arousal exceeds what they typically express and consider to be the norm. According to the (CAT), if an individual appraises the emotion expressed in the context as having higher valence (arousal), they will adjust their language to align more closely with the perceived language of the context.

We, therefore, hypothesize that:

*H2. A user's deviation from their baseline emotional expression is positively associated with the difference between their idiosyncratic arousal (valence) and the arousal (valence) expressed in the context.*

In the next section we show the data we utilized and will discuss the method to verify our hypotheses.

## 5.3 Data and Methods

This section outlines the data collection and methodology used to analyse the emotions expressed in comments on Wikipedia Talk pages. By linking each comment to its contextual elements and the author's previous activity, we aim to understand the factors influencing the emotions conveyed.

The data collection and methodology were designed with a primary focus on the data to be analysed, specifically the emotions of a comment on a Wikipedia Talk page. To understand which factors might have led to a particular emotion, we aimed to gather all information regarding the contextual elements that the author of the analysed comment might have seen before posting the comment. To achieve this, we linked each observation (the comment) with all the preceding contextual aspects present on the *Talk Page*, as well as with the characteristics of the author's previous activity, including their prior contributions to any article's content page and *Talk Page*.

### 5.3.1 Data Collection

The scope of the data collection was limited to content created on English Wikipedia as the methodology we used is based on a sentiment measurement based on a survey of English speakers Warriner et al. (2013). The data collection process was carried out through a web scraping bot developed using a Python script[7].

The data collection protocol was designed in three stages. The first phase is represented by the selection of the *Articles* that would constitute the basis of our dataset. After such selection process, we would collect the information about the discussion present in the *Talk pages* linked with the selected articles. Finally, we would check the history of the activity of each subject that we identified in the previous phase and would collect all the information regarding direct contributions to *Articles* and all the comments on *Talk Pages*.

We first selected all the *Today's Featured Articles* highlighted on the English Wikipedia welcome page between December $1^{st}$ 2019 and December $31^{st}$ 2019[8].

The scope of such choice is twofold. On the one hand, we wanted to take into account articles that are generically considered to be high quality articles (Viégas et al., 2007b) and, thus, according to Viégas and Smith (2004), the related *Talk Pages* should be particularly dense. At the same time, however, differently from other scholars (Gallus and Bhatia, 2020, Prabhakaran and Rambow, 2016), we did not want to start from articles that were already considered as controversial but rather have a representation of the diversity of the topics covered by Wikipedia, and thus of its contributors, as according to the subject the composition of the teams writing the article varies (as the level of subject expert participation in the text produced, see Yarovoy et al., 2020).

Having collected the first 31 articles, we looked for the hyperlinks present in each *Article* that lead to other *Articles* and added such *Articles* to our initial set, obtaining a temporary set of articles. The same process was repeated on these articles obtaining, thus, 312,088 articles. This second phase of the process was designed to guarantee a sufficient degree of both similarity and diversity among the articles considered. This degree of diversity allows to control for topic-

[7]The code will be available in a Git repository.

[8]A *Featured Article* is an article chosen by Wikipedia's editors as considered to be exemplary for its accuracy, neutrality and completeness. The choice is carried out through an election. Currently, there are 6424 *Featured Articles*, *i.e.* $0,09\%$ of all Wikipedia's Articles.

specific confounding factors.

When it comes to observe the context, we must be careful and consider the context that the user is actually seeing. For instance, we expect that the thread in which a comment is posted has a stronger impact on the emotion expressed than the overall *Talk Page*. We, therefore, propose to consider different layers of context, *i.e.*, the overall *Talk Page*, the *Thread* and comments that have been previously exchanged with the intended recipient (the "dyadic" level). The comments sent by users to other users in *Talk Pages* represent the core object of analysis in this research. For each comment, we needed to define the sender, the receiver, the thread in which it was posted, the state of the *Talk Page* at the moment in which it was posted. The latter aspect would define the contextual aspects that we consider together with the comment, *i.e.* what the author of the observed comment could see before posting the comment. Looking at Figure 10, we can see an hypothetical scenario. When collecting Comment (C), we would consider the identity of the receiver (the author of comment (B)), the comments posted in the thread ($\alpha$) before comment (C) was posted - thus, comments (A) and (B) - and the comments posted in the talk page before comment (C) was posted - thus, comments (A), (B) and (E). The combination of these elements provides a clear understanding of contextual aspects and distinguishes between events and activities before and after the comment. To evaluate the emotion within the thread, comments (A) and (B) are taken into consideration. To evaluate the emotion within the overall *Talk Page*, comments (A), (B) and (E) are considered. To make sure that we would cover all the comments posted in a *Talk page* we took into consideration also the archived discussions with the aim of having all the information regarding the *Talk Page* at the moment in which the observed comment was posted. During this phase we discarded all the comments posted anonymously as we could not track their history to study their baseline emotional expression. We, nevertheless, considered them when rebuilding the context. Eventually, we obtained 1.396.531, each of which represent an observation in our dataset.

We then collected the activities carried out by the subjects identified in the previous process. For what regards the contributions to articles, we collected the time-stamp of each subject's contribution, the byte value of the contribution, and the article in which the contribution was made. As for comments, we collected all comments of each subject regardless of the *Talk Page*

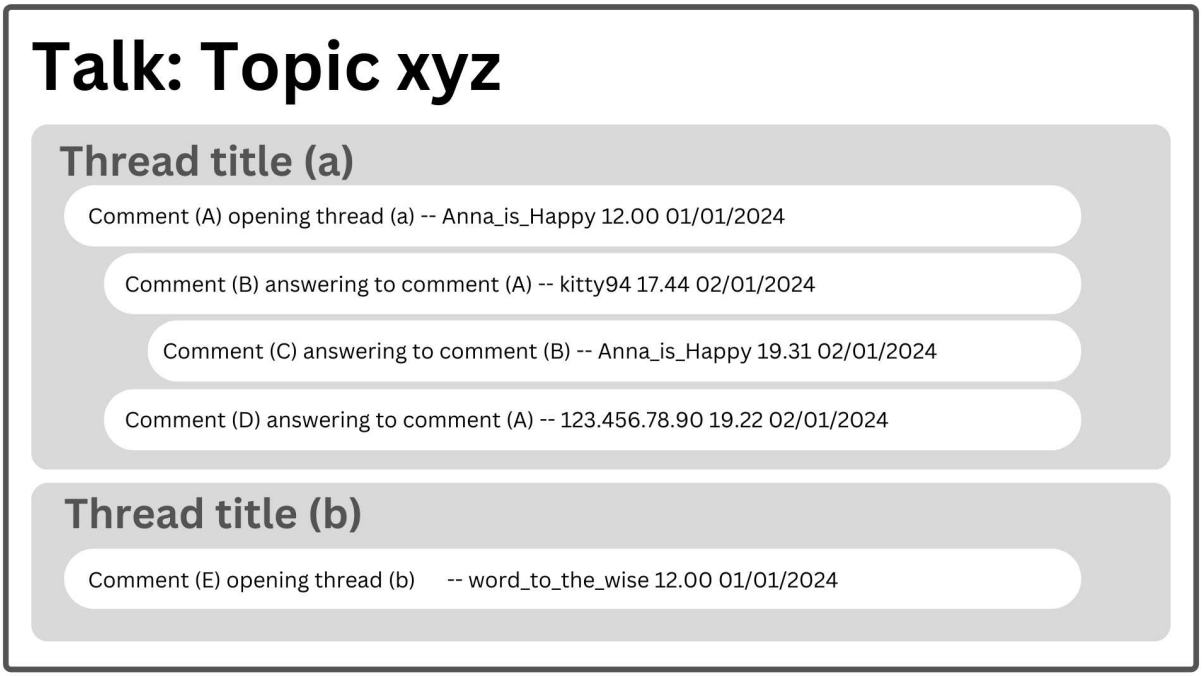



Figure 10: Stylized structure of Talk page

on which they were posted. For each comment, we collected information about the sender and receiver, the time-stamp, and the Talk page on which it was posted. It is important to note that the comments collected during this phase are not the object of analysis but rather serve to characterize the past activity of the subjects.

### 5.3.2 Measuring emotions

As mentioned earlier, in this study we take into consideration emotions in two dimensions, *i.e.* *Valence* and *Arousal*. In order to estimate and measure the emotion that a comment has we rely on the dataset of ratings created by Warriner et al. (2013). Figure 11a shows a sample of words evaluated within such dataset.

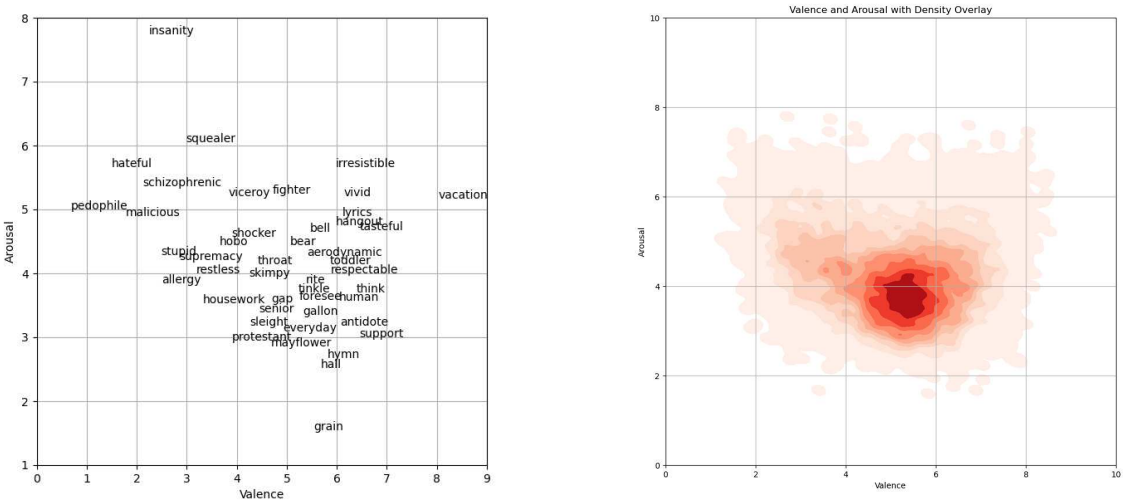
For what regards the measurement tools, we rely on two main methods, *i.e.* Word Frequency Averaging (WFA) and Embeddings (EMB) methods.

### 5.3.3 Word Frequency Averaging

Word Frequency Averaging, put simply, measures how often a word appears in a text (Humphreys and Wang, 2018). In order to do so, the comment is first tokenized by converting it to lowercase, removing all punctuation, and splitting it by white space. This transforms the comment into a

"bag-of-words" representation, consisting of a set of component words and their frequencies. Next, the WFA method queries the Warriner et al. (2013) lexicon for the valence and arousal ratings of each word. The valence or arousal ratings for all words present in the lexicon are then averaged to obtain an overall measure of the comment's valence or arousal.

For example, if a comment contains words that are all found in the Warriner et al. lexicon, the method will average the ratings for each of those words to produce a single *valence* or *arousal* score for the entire comment. This provides a straightforward way to quantify the emotional content of comments based on the known emotional ratings of the words they contain (Warriner et al., 2013).




(a) Sample words illustrating Valence and Arousal

(b) Density Plot of Valence and Arousal

Figure 11: Overview of Valence and Arousal in Warriner et al. 2013

### 5.3.4 Word embedding

One limitation of the WFA (Word Frequency Analysis) method is that not all comments contain words from the Warriner et al. (2013) lexicon, and many commonly used words are missing from this lexicon. To address this and improve our results, we use another approach: the word embeddings method.

Word embeddings are tools in computational linguistics that represent word meanings as high-dimensional vectors, based on word co-occurrence patterns in language. These vectors are

useful for various text analysis tasks. For more details, see (Gallus and Bhatia, 2020) or (Lenci, 2018).

In this study, we use word embeddings to extend the valence (emotional value) and arousal (emotional intensity) ratings from Warriner et al. to other words found in Wikipedia comments that are not in their lexicon. We use the Google News word2vec embeddings model, which provides 300-dimensional representations for over 3 million words and phrases (Mikolov et al., 2013).

We first map each word in the Warriner et al. (2013)lexicon to the word2vec embedding space and use their valence and arousal ratings to train an algorithm to predict these ratings for the 3 million words in the word2vec model. This allows us to measure the valence and arousal of many words in Wikipedia comments, even those not in the Warriner et al. (2013) dataset. For more details on this method, see Hollis et al. (2017).

Each word in the Warriner et al. (2013) lexicon is represented as a vector in the word2vec space. We use the valence and arousal ratings to develop a function that maps these vectors to their ratings. Assuming this function is linear, we train it using ridge regression from the sci-kit learn Python library.

With this trained function, we can predict the valence or arousal of any word or comment. For a Wikipedia comment, we tokenize it (convert to lowercase and split by spaces), average the word vectors for each word in the comment, and pass this comment vector through our function to get the predicted rating. This approach combines the vectors of all words in a comment into a single vector before predicting the valence or arousal rating, instead of predicting ratings for each word individually and then averaging them as in the WFA method. Since our function is linear, this method gives the same result.

### 5.3.5 Variables

In this research we focus on the emotions and the tone expressed by users when commenting in *Talk Pages* and the motivation that might drive such tone. When considering the dependent variables, which include all possible emotions expressed through comments, the choice and selection criteria are straightforward. However, when it comes to the independent variables, it is

important to keep in mind the premises formulated during hypothesis development. Considering the literature regarding social cues (see Tanis and Postmes, 2003, Rutter and Stephenson, 1979) and contextual characteristics, we focused on the elements that users see in the Talk Page before commenting. We analysed the impact of different levels of granularity on the tone of the user's comment.

At the same time, however, with the aim of better understanding the potential differences between current and past behaviours, we wanted to consider and collect different aspects of user activity history in terms of knowledge contributions on the *Article Pages* and discussion contributions on the *Talk Pages*.

### 5.3.6 The dependent variables: the comment tone

The main focus of this research is the emotion expressed by users when answering to someone else in Wikipedia's *Talk pages*. The *Valence* and the *Arousal* of each comment was evaluated using both WFA and EMB methods represent our main dependent variables .

To test hypothesis 2, we needed to assess the author's evaluation of the contextual emotion. We define the subjective estimate as the difference between the *valence* (*arousal*) previously expressed by the author and the *valence* (*arousal*) present in the context with which the author is confronted. Similarly, we defined the change in the comment, the dependent variable used to verify hypothesis H2, as the difference between the *valence* (*arousal*) expressed in the comment and the *valence* (*arousal*) previously expressed by the author.

In Table 7.3, we provide an overall description of these variables.

### 5.3.7 Independent variables: the context

The context is defined by the comments that can be seen in the *Talk Page* by the author of the comment that is taken into consideration. In our specific setting, the context can be split in three levels: the previous comment in the discussion (the one the author is responding to), the thread, and the page.

The *Dyadic Arousal* and *Dyadic Valence* respectively represent the *Arousal* and the *Valence* of the comments received by the sender from the author of the comment the sender is answering

to. In order to measure the tone, we joined together all the considered comments in a unique string. The string was then analysed using WFA and using Word Embeddedness for the words not contained in the dictionary provided by Warriner et al. (2013). Taking a broader perspective, we consider the *Thread* in which the author posted the observed comment, calculating, thus, the *Thread Arousal* and *Thread Valence*. The same logic has been applied to the *Page Arousal* and *Page Valence*.

Similarly to what we have done for the observed comment, we wanted to take into consideration the difference between the tone of the comments that the user would be facing before posting their own comment and the user-specific language before the comment *i.e.* the *Subjective Thread Arousal* (*Subjective Thread Valence*) .

For what concerns users' idiosyncratic tone, we defined the *User Arousal* and *User Valence* as the user-specific emotional language used before the observed comment. In order to calculate such variables, we collected all user's comment posted in *Talk Pages* before the comment under analysis and joined them in a unique string. The *Arousal* and the *Valence* of the obtained string would be estimated using both WFA and EMB methods.

We also planned to collect metadata regarding users' activity on Wikipedia with the aim of enriching our comprehension of the phenomenon at hand. For instance, we collected information about the contributing activity (edits to articles). We observed the total amount of contribution in *Articles* (*User Contribution Production*) and in *Talk Pages* (*User Comment Production*). At the same time we considered *Age* as the time since the first comment (in months). The overall idea behind the variables chosen for representing the users activity was to, on the one hand, obtain an idea regarding the experience of the user (*e.g.*, *User Contribution Production*) and, on the other hand, the characteristics of users' contribution activities (*e.g.*, User Comment Density).

A thorough description of the listed variables can be found in page 119.

## 5.4 Results

### 5.4.1 Descriptive statistics

In order to evaluate the emotions expressed in the focus comment and to connect them to the different contextual levels, it was necessary to ensure that certain requirements were met. For instance, a comment could be considered if it was responding to another comment. Consequently, we were unable to consider as an object of analysis the comments that were starting a thread. Similarly, it was necessary to ensure that the comments being analysed were on a page with at least two threads. If there was only one thread, the text at the thread and page levels would be identical. Finally, the focus comment had to be written by a subject who had previously logged in and contributed at least once. Due to the fact that many of the collected comments represented users' first comment's and, therefore, we were not able to evaluate the emotions expressed by the user in previously posted comments. For such variable, in fact, we have 1,148,677 observations. For the same reason, we have fewer observations regarding the tone expressed by the receiver of the comment analysed 1,160,575. When it comes to the contextual factors, we lack the pages in which there was no previous thread (obtaining, thus, a total of 1,296,621 observations) and lack the threads in which the comments preceding the observed comment were non-readable, absent or written by an IP (obtaining a total of 1,038,095 observations). The number of observations do not depend on the method utilized to evaluate the comments. In Table 7.3 we can see an overview of the variables we collected.

### 5.4.2 Regression analyses and hypotheses checking

In order to check and validate our hypotheses, we carried out a linear regression analysis. We opted for Ridge Regression as the primary method. Ridge Regression offers significant advantages over Ordinary Least Squares (OLS) regression, particularly in cases where multicollinearity exists among the independent variables. Multicollinearity causes instability in OLS estimates, leading to high variance in the coefficient estimates, which in turn makes them highly sensitive to small variations in the data. Ridge Regression addresses this issue by adding a regularization term to the cost function, effectively shrinking the coefficients towards zero. This shrinkage reduces variance, resulting in more stable and reliable estimates. As described by

James et al. (2023), Ridge Regression estimates the coefficients, $\beta$, by minimizing the following expression:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p}\beta_j^2, \tag{17}$$

where $\lambda$ is a regularization parameter controlling the extent of penalization applied to the regression coefficients. To identify the optimal value of $\lambda$, we employed a grid search combined with 5-fold cross-validation. Using the selected $\lambda$, the model was retrained for final analysis.

To evaluate the statistical significance of the regression coefficients, we utilized non-parametric bootstrap confidence intervals (Efron and Tibshirani, 1994). Specifically, a coefficient was deemed statistically significant at the $\alpha\%$ level if zero was not contained within the $(1-\alpha)\%$ confidence interval. Previous studies have demonstrated the applicability of bootstrap methods in Ridge Regression (Sabbaghi, 2019, Crivelli et al., 1995).

Before conducting the analysis, all independent variables were normalized to have a mean of zero and a standard deviation of one. This normalization ensures consistent penalization across features, preventing variables with larger magnitudes from disproportionately influencing the model.

This methodological framework enhances the interpretability and robustness of the model, effectively minimizing overfitting while maintaining an optimal balance between bias and variance.

Table 7 shows the results of the Ridge linear regression analysis regarding the contextual emotion evaluated with the WFA method while Table 8 considers the emotions evaluated with the Embedded method.

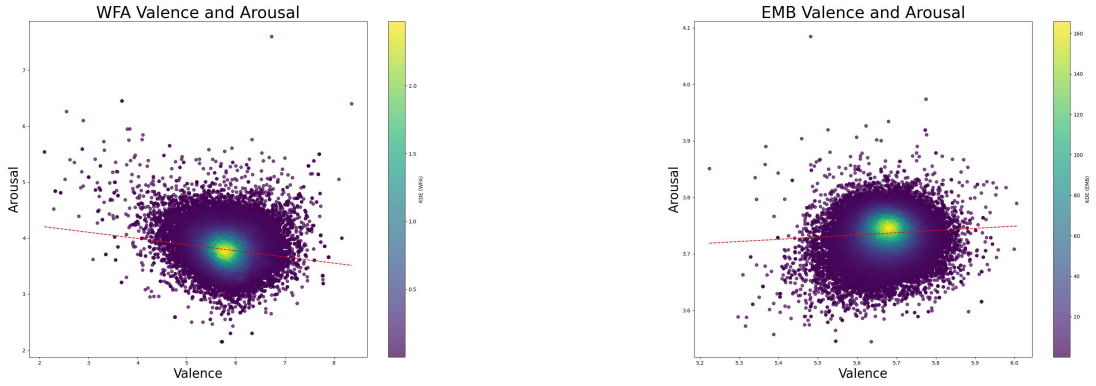
Comparing the models reveals several key insights. First, examining the relationship between the two emotional dimensions, we observe that they are positively correlated within individual comments. This finding may reflect a specific characteristic of language and the way individuals express themselves. However, as shown in Table 8, Models $1_b$ and $2_b$ exhibit opposite behaviour. More precisely, the correlation between *arousal* and *valence* depends on the method used to evaluate these dimensions within a text. We can see better the relationship between the two emotional dimension compared in Figure 11. This divergence in the relation-

Table 7: Ridge Regression with contextual emotions (WFA)

| **Model** | Model $1_b$ **Arousal** | Model $2_b$ **Valence** | Model $3_b$ **$\Delta$ Arousal** | Model $4_b$ **$\Delta$ Valence** |
|---|---|---|---|---|
| Const. | 0.0000 | 0.0000 | 0.0000 | -2.666e-17 |
| | (0.002) | (0.002) | (0.002) | (0.005) |
| Comment Valence | -0.1749*** | | | |
| | (0.002) | | | |
| Comment Arousal | | -0.1777*** | | |
| | | (0.004) | | |
| Page Valence | 0.0038*** | 0.0715*** | | |
| | (0.008) | (0.003) | | |
| Page Arousal | 0.0718*** | 0.0121*** | | |
| | (0.003) | (0.003) | | |
| Thread Arousal | 0.2293*** | 0.0269*** | | |
| | (0.002) | (0.003) | | |
| Thread Valence | 0.0130*** | 0.1910*** | | |
| | (0.003) | (0.003) | | |
| User Valence | 0.0047* | 0.0635*** | | |
| | (0.007) | (0.002) | | |
| User Arousal | 0.0678*** | 0.0152*** | | |
| | (0.002) | (0.002) | | |
| Receiver Arousal dyadic | 0.0135*** | 0.0038 | | |
| | (0.003) | (0.003) | | |
| Receiver Valence dyadic | -0.0050 | 0.0240*** | | |
| | (0.007) | (0.002) | | |
| Comment Valence (rel) | | | -0.1759*** | |
| | | | (0.002) | |
| Comment Arousal (rel) | | | | -0.1777*** |
| | | | | (0.002) |
| Valence Page (rel) | | | -0.0207*** | 0.0738*** |
| | | | (0.002) | (0.003) |
| Arousal Page (rel) | | | 0.0977*** | 0.0210*** |
| | | | (0.003) | (0.003) |
| Valence Thread (rel) | | | 0.0271*** | 0.1767*** |
| | | | (0.002) | (0.003) |
| Arousal Thread (rel) | | | 0.2029*** | 0.0216*** |
| | | | (0.003) | (0.003) |
| Valence Receiver Dyadic (rel) | | | -0.0034*** | 0.0613*** |
| | | | (0.004) | (0.004) |
| Arousal Receiver Dyadic (rel) | | | 0.0626*** | -0.0085*** |
| | | | (0.004) | (0.004) |

Table 8: Ridge Regression with contextual emotions (EMB)

| **Model** | Model $1_a$ **Arousal** | Model $2_a$ **Valence** | Model $3_a$ **Δ Arousal** | Model $4_a$ **Δ Valence** |
|---|---|---|---|---|
| const | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Comment Valence | 0.1037*** | | | |
| | (0.002) | | | |
| Comment Arousal | | 0.1008*** | | |
| | | (0.002) | | |
| Page Valence | -0.0063** | 0.0691*** | | |
| | (0.003) | (0.003) | | |
| Page Arousal | 0.0558*** | -0.0039 | | |
| | (0.003) | (0.002) | | |
| Thread Arousal | 0.2114*** | -0.0192*** | | |
| | (0.003) | (0.002) | | |
| Thread Valence | -0.0057 | 0.2269*** | | |
| | (0.003) | (0.002) | | |
| User Valence | -0.0111*** | 0.0935*** | | |
| | (0.002) | (0.003) | | |
| User Arousal | 0.0806*** | -0.0022 | | |
| | (0.002) | (0.003) | | |
| Receiver Arousal dyadic | 0.0444*** | 0.0073 | | |
| | (0.003) | (0.003) | | |
| Receiver Valence dyadic | -0.011 | 0.0348*** | | |
| | (0.003) | (0.003) | | |
| Comment Valence (rel) | | | 0.1020*** | |
| | | | (0.002) | |
| Comment Arousal (rel) | | | | 0.1013*** |
| | | | | (0.003) |
| Valence Page (rel) | | | -0.0114*** | 0.0738*** |
| | | | (0.002) | (0.002) |
| Arousal Page (rel) | | | 0.0648*** | -0.0110*** |
| | | | (0.003) | (0.003) |
| Valence Thread (rel) | | | -0.0044 | 0.2111*** |
| | | | (0.003) | (0.003) |
| Arousal Thread (rel) | | | 0.2018*** | -0.0176*** |
| | | | (0.002) | (0.002) |
| Valence Receiver Dyadic (rel) | | | 0.0024 | 0.0273*** |
| | | | (0.002) | (0.002) |
| Arousal Receiver Dyadic (rel) | | | 0.0366*** | 0.0085 |
| | | | (0.002) | (0.003) |
| $R^2$ | 0.117 | 0.141 | 0.088 | 0.0947 |
| MSE | 0.884 | 0.860 | 0.921 | 0.915 |
| RMSE | 0.940 | 0.927 | 0.960 | 0.956 |
| MAE | 0.703 | 0.691 | 0.717 | 0.713 |
| $\lambda$ | 212.095 | 212.095 | 212.095 | 212.095 |




(a) WFA Valence and Arousal

(b) EMB Valence and Arousal

Figure 12: Comparison of Valence and Arousal scatter plots for WFA and EMB datasets.

ship between *arousal* and *valence* is also evident across other contextual levels when comparing Tables 7 and 8. Nevertheless, if we focus on the relationship between the *arousal* (*valence*) expressed at different contextual levels, we find that their sign remains consistent across the two tables, indicating robustness to differences in the methods used for emotion measurement.

For instance, across all models, the variables representing the contextual corresponding emotion consistently exert a positive impact on the dependent variable, regardless of the specific emotion analyzed or the method used to evaluate emotional intensity. This consistency strongly supports our initial hypothesis, which posits a positive correlation between the *arousal* (*valence*) expressed in the context and that expressed in the focus comment. This relationship, driven by mimicry behaviour, highlights how users adapt their emotional tone to align with the emotional tone present in the context. Remarkably, this dynamic is evident across all models, irrespective of the method employed for emotion measurement, further reinforcing the robustness of these findings.

In contrast, the influence of one emotional dimension in the context on the other emotional dimension in the observed comment tends to be either nonsignificant or slightly negative. Specifically, in Model $1_a$, we observe that contextual *valence* generally exerts an insignificant influence on the *arousal* of the observed comment, and when such an effect is significant, its magnitude remains relatively small. Similarly, in Model $2_a$, contextual *arousal* has a negligible or slightly negative impact on the *valence* of the observed comment. Interestingly, as observed for the emotions expressed within individual comments, the relationship between the

two emotional dimensions at the contextual level and those expressed in the observed comment tend to exhibit opposing signs when comparing the two methods of emotion evaluation. This divergence suggests that differences may arise from methodological discrepancies rather than inherent emotional dynamics, highlighting the importance of standardizing measurement approaches or accounting for these differences in future research.

Turning to Models $3_a$ and $4_a$, we observe similar findings regarding the perceived emotional intensity, which refers to the impact of the emotion expressed in the context—as perceived by the author of the comment—on how the author adjusts their language to align with the contextual tone. These results provide strong evidence for the appraisal mechanism, demonstrating that the subjective evaluation of the context significantly influences deviations in the emotional intensity of the comment from the author's baseline. The consistency and significance of these results across models lend robust support to our hypothesis regarding this mechanism. Specifically, these models highlight the factors driving changes in emotional intensity relative to the user's benchmark, offering valuable insights into the dynamics of emotional adaptation in collaborative online settings.

Another important thing to stress is the differences in terms of magnitude across different contextual levels. For instance, it can be noticed that the thread level has an overall higher impact compared to the other levels (page and dyadic level).

## 5.5 Discussion and Conclusion

The aim of this research was to investigate whether the emotions expressed in the corpus of text on Wikipedia's *Talk Pages* influence the emotions expressed in subsequent comments. To achieve this, we adapted theories of emotional expression to the specific context of Online Open Collaboration Communities (OOCCs). The results highlighted the significant role that context plays in shaping the emotions expressed by users and, in alignment with Appraisal Theory, how these emotions are modulated by the subjective perception of the context.

Furthermore, the findings demonstrated that the magnitude of the contextual influence on a user's emotions depends on the specific contextual level being considered. Notably, our analysis consistently showed that the *Thread* level has a greater impact compared to other levels

of context. Interestingly, contrary to what might be expected, it is not the dimension closest to the user—such as the *dyadic* level, representing direct exchanges between individuals—that exerts the strongest influence. Instead, the *dyadic* level, while relevant, has a lower impact than the *Thread* level. This reduced relevance may stem from several factors. On the one hand, the scarcity of social cues and the anonymity intrinsic to such environments can lead, as highlighted in prior research (see Faraj et al., 2011, Greineder and Blohm, 2020), to dynamics of deindividuation where the interlocutor becomes less distinct. On the other hand, this finding underscores the importance of studying these theories while carefully considering the peculiarities of the specific setting. In this chapter, we emphasised the role of *affordances* in defining the potential behaviours of users.

Specifically, the *Talk Page* section of Wikipedia is designed to facilitate coordination among individuals during a collaborative process aimed at achieving a shared goal, *i.e.*, the creation of an *Article* on which all contributors converge. As previously noted by Watzek et al. (2019), the stimuli that trigger emotions in collaborative processes must be goal-relevant. Given that the *Thread* is the contextual level most useful for understanding the state of a specific topic, it is plausible that it exerts the strongest influence because it represents the most goal-relevant corpus of text. Building on the idea that the entire text corpus used for coordination serves a normative function (Arazy et al., 2016), we suggest that determining the relevance of context requires accounting for the *afforded* actions and ongoing dynamics of cooperation. For example, if a user seeks to modify a detail regarding a specific topic, they must coordinate within the relevant *Thread*. This coordination necessitates understanding the preceding dynamics of discussion and collaboration within that thread. The collective and procedural dimensions, amplified by anonymity due to the absence of social cues, play a fundamental role in shaping these interactions.

The novelty of this research lies not in the highlighted social behaviour itself but in the unique setting considered and the framing of the findings. Online Communities (OCs) and OOCCs have increasingly attracted attention, particularly due to the prevalence of toxic social behaviours in these environments, raising concerns across various disciplines. In this study, we proposed analysing specific social behaviours by taking into account the peculiarities of these

settings and adapting well-established theories from emotional psychology to fit this context. The results are particularly relevant as they illuminate not only how emotions are expressed in online cooperative processes but also how users in online communities collaborate. Moreover, the findings underscore how the process as a whole, rather than individual interactions, influences the emotions verbally expressed by users.

Practitioners can benefit from these results by gaining a deeper understanding of group dynamics in virtual, text-based collaborative teams. These insights could pave the way for the development of tools to detect early signs of toxic dynamics, potentially enabling intervention before such behaviours escalate. By focusing on the broader dynamics of emotional expression, this research provides actionable strategies for mitigating conflict and fostering healthier collaborative environments.

Another aspect of novelty in this research lies in the data collection methodology and its underlying criteria. The cornerstone of the data collection process was the individual comment, with the primary criterion being the perspective of the author of the focus comment. The dataset was structured to include all the elements visible to the author at the time of writing the comment. This methodological approach ensures a detailed and contextually accurate understanding of the emotional processes involved.

This research, however, is not without limitations, which simultaneously open avenues for future investigation. First, the study is based on a single, highly specific setting. Users may have developed a unique vocabulary and tone of interaction within this environment. Furthermore, as our results suggest, the platform itself signals and enforces norms that influence behaviour (Dang Nguyen et al., 2017). A comparative study across different platforms would be valuable in validating and extending the findings presented here.

While *Valence* and *Arousal* have long been considered critical properties of emotional experience (see Kuppens et al., 2013), they do not provide a comprehensive description of the full spectrum of emotions that can be expressed verbally. Future research could explore whether other emotional dimensions or combinations of emotions exhibit similar dynamics to those identified in this study or whether these patterns are unique to specific parts of the emotional spectrum (Mohammad, 2016).

Another promising direction for future work involves investigating new language paradigms (Fernández-Gavilanes et al., 2018). For instance, the increasing use of emoticons and emojis in Computer-Mediated Communication (CMC) offers an intriguing avenue for study, as these elements help reduce uncertainty about the emotions being conveyed in text (Robertson et al., 2021). Additionally, a cross-linguistic study could enrich the research by accounting for linguistic and cultural variations in emotional expression (Kruspe et al., 2020).

Further research could also connect the "emotional temperature" of a discussion with the actual state and evolution of the *Article Page*. As highlighted by Weltevrede and Borra (2016), understanding the value of disputes in such collaborative settings is crucial. For example, it would be interesting to explore whether more heated disputes positively impact the quality of the *Article* by fostering engagement or whether they discourage collaboration among contributors. Similarly, identifying specific moments or stages in the editing process—such as restructuring the page or expanding its coverage—that are more prone to disputes could provide valuable insights into managing conflict.

Moreover, future research could better fine-tune the magnitude of the different contextual levels as a function of time. Specifically, it would be interesting to investigate how long a text continues to impact future comments and whether time plays a role in diminishing its influence. Considering the cogency of the text within a thread when a stigmergic approach is insufficient, it is worth exploring whether a prior, no longer relevant conversation can still impact the emotions expressed in subsequent discussions.

The main contribution of this paper is to advocate for a paradigm shift in data collection and analysis, urging researchers to consider not only individual behaviours but also the context-specific factors that shape emotional processes. Recognising that emotions are influenced by both individual and contextual factors compels us to consider the affordances of the platform and the interactions mediated through them. Given the growing importance of online collaboration and the increasing concerns about toxic behaviours and polarisation, this shift in perspective is essential. It offers a framework for understanding and, where necessary, detecting mechanisms that lead to conflict or disengagement. Ultimately, this approach aims to mitigate toxic dynamics, foster constructive collaboration, and prevent the premature withdrawal of

contributors from platforms such as Wikipedia and other collaborative projects.

# 6 Final Discussion

In this research, we proposed three studies aimed at understanding the dynamics at play in interactions among individuals. Specifically, we sought to analyze this phenomenon from two different perspectives. The first employs a network approach, examining how interactions influence the processes of opinion formation across all individuals considered and how these interactions are shaped by exogenous dynamics, such as those observed on LinkedIn and their relationship with lobbying activities during the preparatory work for the adoption of the EURO 7 directive. The second perspective focuses on asynchronous interaction mediated by an artifact, as exemplified by Wikipedia and its *Talk Pages*.

Interaction with others and complex communication are defining characteristics of our species, as they have enabled both social and technological development. Through communication and the exchange of information and opinions, humans can reduce uncertainty and solve collective action problems (CAPs) (Smith, 2010). While network theory and social network analysis are not new concepts, the advent of new technologies has both deepened this approach and introduced new questions and challenges. On the one hand, digital technologies have vastly expanded communication opportunities and interaction modalities. On the other hand, they have also contributed to a form of communication impoverished by the lack of essential social cues, undermining its effectiveness and completeness. In some ways, this has reintroduced a degree of uncertainty and ambiguity in interpersonal communication (Faraj et al., 2011). At the same time, society has witnessed the emergence of complex collective projects in non-hierarchical contexts, such as Open Online Collaborative Communities (OOCCs) like Debian and, in this case, Wikipedia. These communities present a particularly intricate phenomenon, and explaining their emergence and sustainability is a challenging task that an entire body of literature—spanning information systems, organizational theory, and sociology—has sought to address (see Rijshouwer et al., 2023, Konieczny, 2010, Kane and Ransbotham, 2016).

The aim of this thesis, particularly in the third chapter, was to understand how interactions between individuals are influenced by the interactions of others. In the specific case of Wikipedia, the setting is unique in that the interactions necessary for coordination constitute the only source of contextual information. This exclusivity allows for the isolation of external

factors, enabling a clearer understanding of how one interaction influences another. As highlighted by (Arazy et al., 2016), the text itself in these contexts not only conveys information but also holds a normative power over user behavior. Our research aimed to demonstrate how the interactions within a page can impact the behavior and interactions of subsequent contributors. Specifically, in the case we analyzed, we observed that the emotions expressed in Wikipedia *Talk Pages* have an impact on subsequent emotions through mechanisms akin to mimicry.

What stands out as particularly relevant is the variation in the magnitude of the impact that contextually expressed emotions have across different levels of contextual granularity. The observation that emotions expressed within the *Thread* have a greater impact on subsequent emotions is understandable when considering the technical and procedural characteristics (see Rijshouwer et al., 2023) specific to Wikipedia. More clearly, emotions expressed within the *Thread* have a stronger influence because the *Thread* plays a critical role in the contribution process of the *Articles*. It contains the specific information regarding modifications made to a particular part of the article associated with the *Talk Page*.

In the other two studies, we focused on the role of the structural relationships among individuals and how these relationships, on one hand, influence overall processes and, on the other, are themselves influenced by these processes. These studies highlight the coexistence of an agentic dimension, in which individual actions play a significant role, and a systemic dimension, in which larger network structures shape and constrain individual behaviour.

The model generalizing DeGroot's opinion formation process demonstrated that relaxing the simplistic condition under which individuals interact leads to non-deterministic consensus outcomes. Additionally, the influence of individuals—the magnitude of their impact on the opinion formation process—is equally affected. Specifically, the introduction of stochastic interaction disproportionately affects individuals with high influence in deterministic models *à la* DeGroot. Notably, the inclusion of stochastic interactions diminishes the importance of structural characteristics of individuals (*e.g.*, centrality) or their ego networks (*Simmelian ties*) in determining their influence. As a result, sporadic interaction allows peripheral or less central individuals to exert a greater influence on the process.

This finding underscores how sporadic interactions can redistribute influence within a net-

work, highlighting the dynamic interplay between individual agency and systemic constraints.

Eventually, we demonstrated how a multiplex network approach can enrich the study of lobbying activities through a structuralist perspective. This research aimed to highlight the interconnectedness between the online and offline activities of individuals involved in legislative processes or matters of public interest. The absence of data often complicates the study of these phenomena, but a more advanced approach could help uncover greater complicity on the part of policymakers through their online interactions with lobbyists.

The strength of a network approach lies in its ability to reveal characteristics of the overall network—features that individual agents themselves cannot significantly influence. By focusing on these structural dynamics, researchers can gain deeper insights into the systemic interactions and dependencies that shape lobbying practices and their broader implications for public policy and transparency.

# References


Anderson, B. D. and Ye, M. (2019). Recent advances in the modelling and analysis of opinion dynamics on influence networks. *International Journal of Automation and Computing*, 16:129–149.

Arazy, O., Daxenberger, J., Lifshitz-Assaf, H., Nov, O., and Gurevych, I. (2016). Turbulent stability of emergent roles: The dualistic nature of self-organizing knowledge coproduction. *Information Systems Research*, 27(4):792–812.

Assenova, V. A. (2018). Modeling the diffusion of complex innovations as a process of opinion formation through social networks. *PloS one*, 13(5):e0196699.

Baumann, F., Lorenz-Spreen, P., Sokolov, I. M., and Starnini, M. (2020). Modeling echo chambers and polarization dynamics in social networks. *Physical Review Letters*, 124(4):048301.

Baumgartner, F. R. and Leech, B. L. (1998). *Basic Interests: The Importance of Groups in Politics and in Political Science*. Princeton University Press, Princeton.

Berdahl, J. L. and Martorana, P. (2006). Effects of power on emotion and expression during a controversial group discussion. *European journal of social psychology*, 36(4):497–509.

Bernhold, Q. S. and Giles, H. (2022). Emotional mimicry: a communication accommodation approach. *Cognition and Emotion*, 36(5):799–804.

Bleakley, A., Rough, D., Edwards, J., Doyle, P., Dumbleton, O., Clark, L., Rintel, S., Wade, V., and Cowan, B. R. (2022). Bridging social distance during social distancing: exploring social talk and remote collegiality in video conferencing. *Human–Computer Interaction*, 37(5):404–432.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Bolici, F., Howison, J., and Crowston, K. (2016). Stigmergic coordination in floss development teams: Integrating explicit and implicit mechanisms. *Cognitive Systems Research*, 38:14–22.

Boutyline, A. and Willer, R. (2017). The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political psychology*, 38(3):551–569.

Bradley, M. M. and Lang, P. J. (1999). Affective norms for english words (anew): Instruction manual and affective ratings. Technical report, Technical report C-1, the center for research in psychophysiology ....

Brandes, U. and Hildenbrand, J. (2014). Smallest graphs with distinct singleton centers. *Network Science*, 2(3):416–418.

Burfurd, I. and Wilkening, T. (2022). Cognitive heterogeneity and complex belief elicitation. *Experimental economics*, 25(2):557–592.

Burt, R. S. (2004). Structural holes and good ideas. *American journal of sociology*, 110(2):349–399.

Burt, R. S. and Celotto, N. (1992). The network structure of management roles in a large matrix firm. *Evaluation and Program Planning*, 15(3):303–326.

Chalmers, A. W. (2013). Trading information for access: informational lobbying strategies and interest group access to the european union. *Journal of European Public Policy*, 20(1):39–58.

Chartrand, T. L. and Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of personality and social psychology*, 76(6):893.

Chatterjee, R. A. and Eliashberg, J. (1990). The innovation diffusion process in a heterogeneous population: A micromodeling approach. *Management science*, 36(9):1057–1079.

Choi, Y.-P., Paolucci, A., and Pignotti, C. (2021). Consensus of the hegselmann–krause opinion formation model with time delay. *Mathematical Methods in the Applied Sciences*, 44(6):4560–4579.

Christensen, L. R. (2007). Practices of stigmergy in architectural work. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work*, GROUP '07, pages 11–20. ACM.

Christensen, L. R. (2013). Stigmergy in human practice: Coordination in construction work. *Cognitive Systems Research*, 21:40–51.

Christensen, L. R. (2014). Practices of stigmergy in the building process. *Computer Supported Cooperative Work (CSCW)*, 23(1):1–19.

Coe, K., Kenski, K., and Rains, S. A. (2014). Online and uncivil? patterns and determinants of incivility in newspaper website comments. *Journal of communication*, 64(4):658–679.

Cohen, W. M., Levinthal, D. A., et al. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, 35(1):128–152.

Cohen, W. W., Ravikumar, P., and Fienberg, S. E. (2003). A comparison of string distance metrics for name-matching tasks. In *Proceedings of the 2003 IJCAI Workshop on Information Integration on the Web (IIWeb)*. Citeseer.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American journal of sociology*, 94:S95–S120.

Colleoni, E., Rozza, A., and Arvidsson, A. (2014). Echo chamber or public sphere? predicting political orientation and measuring political homophily in twitter using big data. *Journal of communication*, 64(2):317–332.

Crivelli, A., Firinguetti, L., Montano, R., and Munóz, M. (1995). Confidence intervals in ridge regression by bootstrapping the dependent variable: a simulation study. *Communications in Statistics-Simulation and Computation*, 24(3):631–652.

Crowston, K. and Rezgui, A. (2020). Effects of stigmergic and explicit coordination on wikipedia article quality. In *Proceedings*. the Annual Hawaii International Conference on System Sciences, ACM.

Dahl, R. A. (1961). *Who Governs? Democracy and Power in an American City*. Yale University Press, New Haven.

Dang Nguyen, G., Dejean, S., and Jullien, N. (2017). Do open online projects create social norms? *Journal of Institutional Economics*, pages 1–26.

De Domenico, M. a. e. c., polarization dynamics in social networks Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M. A., Gómez, S., and Arenas, A. (2013). Mathematical formulation of multilayer networks. *Physical Review X*, 3(4):041022.

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical association*, 69(345):118–121.

Ding, Z., Chen, X., Dong, Y., and Herrera, F. (2019). Consensus reaching in social network degroot model: The roles of the self-confidence and node degree. *Information Sciences*, 486:62–72.

Dipple, A., Raymond, K., and Docherty, M. (2014). General theory of stigmergy: Modelling stigma semantics. *Cognitive Systems Research*, 31:61–92.

Dong, Y., Zha, Q., Zhang, H., Kou, G., Fujita, H., Chiclana, F., and Herrera-Viedma, E. (2018). Consensus reaching in social network group decision making: Research paradigms and challenges. *Knowledge-Based Systems*, 162:3–13.

Dorigo, M., Bonabeau, E., and Theraulaz, G. (2000). Ant algorithms and stigmergy. *Future generation computer systems*, 16(8):851–871.

Dunbar, R. I., Arnaboldi, V., Conti, M., and Passarella, A. (2015). The structure of online social networks mirrors those in the offline world. *Social networks*, 43:39–47.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative science quarterly*, 44(2):350–383.

Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Chapman and Hall/CRC.

Elliott, M. (2016). Stigmergic collaboration: A framework for understanding and designing mass collaboration. *Mass collaboration and education*, pages 65–84.

Faraj, S., Jarvenpaa, S. L., and Majchrzak, A. (2011). Knowledge collaboration in online communities. *Organization Science*, 22(5):1224–1239.

Fernández-Gavilanes, M., Juncal-Martínez, J., García-Méndez, S., Costa-Montenegro, E., and González-Castaño, F. J. (2018). Creating emoji lexica from unsupervised sentiment analysis of their descriptions. *Expert Systems with Applications*, 103:74–91.

Fischer, M., Varone, F., Gava, R., and Sciarini, P. (2019). How mps ties to interest groups matter for legislative co-sponsorship. *Social networks*, 57:34–42.

Fisher, D. R., Waggle, J., and Leifeld, P. (2013). Where does political polarization come from? locating polarization within the u.s. climate change debate. *American Behavioral Scientist*, 57(1):70–92.

Forte, A., Larco, V., and Bruckman, A. (2009). Decentralization in Wikipedia Governance. *Journal of Management Information Systems*, 26(1):49–72.

Freeman, L. C. (1978). Centrality in social networks: Conceptual clarification. *Social networks*, 1:215–239.

Freeman, L. C., Borgatti, S. P., and White, D. R. (1991). Centrality in valued graphs: A measure of betweenness based on network flow. *Social networks*, 13(2):141–154.

Friedkin, N. E. (1982). Information flow through strong and weak ties in intraorganizational social networks. *Social networks*, 3(4):273–285.

Friedman, R. A. and Podolny, J. (1992). Differentiation of boundary spanning roles: Labor negotiations and implications for role conflict. *Administrative science quarterly*, pages 28–47.

Gallois, C., Ogay, T., and Giles, H. (2005). Communication accommodation theory. *Theorizing about intercultural communication*, pages 121–148.

Gallus, J. and Bhatia, S. (2020). Gender, power and emotions in the collaborative production of knowledge: A large-scale analysis of Wikipedia editor conversations. *Organizational Behavior and Human Decision Processes*, 160:115–130.

Gargiulo, M., Ertug, G., and Galunic, C. (2009). The two faces of control: Network closure and individual performance among knowledge workers. *Administrative Science Quarterly*, 54(2):299–333.

Giráldez-Cru, J., Chica, M., and Cordón, O. (2021). A framework of opinion dynamics using fuzzy linguistic 2-tuples. *Knowledge-Based Systems*, 233:107559.

Golub, B. and Jackson, M. O. (2010). Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–149.

Gómez-Gardeñes, J. and Moreno, Y. (2006). From scale-free to erdos-rényi networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 73(5):056124.

Gómez-Solórzano, M., Tortoriello, M., and Soda, G. (2019). Instrumental and affective ties within the laboratory: The impact of informal cliques on innovative productivity. *Strategic Management Journal*, 40(10):1593–1609.

Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, 78(6):1360–1380.

Grassé, P.-P. (1959). La reconstruction du nid et les coordinations interindividuelles chez bellicositermes natalensis et cubitermes sp. la théorie de la stigmergie: Essai d'interprétation du comportement des termites constructeurs. *Insectes sociaux*, 6:41–80.

Greineder, M. J. and Blohm, I. (2020). A process theory on transformation of work through internal crowdsourcing. Academy of Management.

Haas, A. (2015). Crowding at the frontier: boundary spanners, gatekeepers and knowledge brokers. *Journal of Knowledge Management*, 19(5):1029–1047.

Hansen, M. T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative science quarterly*, 44(1):82–111.

Hargrave, L. and Langengen, T. (2021). The gendered debate: Do men and women communicate differently in the house of commons? *Politics & Gender*, 17(4):580–606.

Harrison, K. R., Ventresca, M., and Ombuki-Berman, B. M. (2016). A meta-analysis of centrality measures for comparing and generating complex network models. *Journal of computational science*, 17:205–215.

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

Hess, U. and Fischer, A. (2022). Emotional mimicry as social regulator: theoretical considerations. *Cognition and Emotion*, 36(5):785–793.

Heylighen, F. (2007). Why is open access development so successful? stigmergic organization and the economics of information. In Lutterbeck, B., Bärwolff, M., and Gehring, R. A., editors, *Open Source Jahrbuch 2007*. Lehmanns Media, Berlin.

Hofman, A. and Aalbers, M. B. (2017). Spaces of lobbying. *Geography Compass*, 11(3):e12309.

Hollis, G., Westbury, C., and Lefsrud, L. (2017). Extrapolating human judgments from skip-gram vector representations of word meaning. *Quarterly Journal of Experimental Psychology*, 70(8):1603–1619.

Hornsey, M. J. (2008). Social identity theory and self-categorization theory: A historical review. *Social and personality psychology compass*, 2(1):204–222.

Howlett, M. (2002). Do networks matter? linking policy network structure to policy outcomes: Evidence from four canadian policy sectors 1990-2000. *Canadian Journal of Political Science/Revue canadienne de science politique*, 35(2):235–267.

Humphreys, A. and Wang, R. J.-H. (2018). Automated text analysis for consumer research. *Journal of Consumer Research*, 44(6):1274–1306.

Hwang, K. (2013). Effects of the language barrier on processes and performance of international scientific collaboration, collaborators' participation, organizational integrity, and interorganizational relationships. *Science Communication*, 35(1):3–31.

Iosub, D., Laniado, D., Castillo, C., Fuster Morell, M., and Kaltenbrunner, A. (2014). Emotions under discussion: Gender, status and communication in online collaboration. *PLoS ONE*, 9(8).

Jäger, K. (2013). Sources of franco-german corporate support for the euro: The effects of business network centrality and political connections. *European Union Politics*, 14(1):115–139.

James, G., Witten, D., Hastie, T., Tibshirani, R., and Taylor, J. (2023). *An introduction to statistical learning: With applications in python*. Springer Nature.

Jarvenpaa, S. L. and Keating, E. (2021). When do good communication models fail in global virtual teams? *Organizational Dynamics*, 50(1):1–12.

Jarvenpaa, S. L. and Leidner, D. E. (1999). Communication and trust in global virtual teams. *Organization science*, 10(6):791–815.

Kane, G. C. and Ransbotham, S. (2016). Content as community regulator: The recursive relationship between consumption and contribution in open collaboration communities. *Organization Science*, 27(5):1258–1274.

Khvatova, T., Block, M., Zhukov, D., and Lesko, S. (2016). How to measure trust: the percolation model applied to intra-organisational knowledge sharing networks. *Journal of Knowledge Management*, 20(5):918–935.

Kilduff, M. and Lee, J. W. (2020). The integration of people and networks. *Annual Review of Organizational Psychology and Organizational Behavior*, 7(1):155–179.

Kim and Kunisky (2020). Mapping political communities: A statistical analysis of lobbying networks in legislative politics. *Legislative Studies Quarterly*, 45:315–335.

Kittur, A., Suh, B., Pendleton, B. A., and Chi, E. H. (2007). He says, she says: conflict and coordination in wikipedia. In *Proceedings*, pages 453–462. SIGCHI conference on Human factors in computing systems.

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). Multilayer networks. *Journal of complex networks*, 2(3):203–271.

Knoke, D. (1990). *Political Networks: The Structural Perspective*. Cambridge University Press, Cambridge.

Knoke, D., Pappi, F. U., Broadbent, J., and Tsujinaka, Y. (1996). *Comparing Policy Networks: Labor Politics in the U.S., Germany, and Japan*. Cambridge University Press, Cambridge.

Konieczny, P. (2010). Adhocratic governance in the internet age: A case of wikipedia. *Journal of information technology & politics*, 7(4):263–283.

Krackhardt, D. (1990). Assessing the political landscape: Structure, cognition, and power in organizations. *Administrative science quarterly*, pages 342–369.

Krackhardt, D. (1998). Simmelian ties: Super strong and sticky. *Power and influence in organizations*, 21:38.

Krackhardt, D. (1999). The ties that torture: Simmelian tie analysis in organizations. *Research in the Sociology of Organizations*, 16(1):183–210.

Kreps, D. M. et al. (1990). Corporate culture and economic theory. *Perspectives on positive political economy*, 90(109-110):8.

Kruse, L. M., Norris, D. R., and Flinchum, J. R. (2018). Social media as a public sphere? politics on social media. *The Sociological Quarterly*, 59(1):62–84.

Kruspe, A., Häberle, M., Kuhn, I., and Zhu, X. X. (2020). Cross-language sentiment analysis of european twitter messages duringthe covid-19 pandemic. *arXiv preprint arXiv:2008.12172*.

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

Larsen, J. B., Vrangbæk, K., and Traulsen, J. M. (2006). Advocacy coalitions and pharmacy policy in denmark—solid cores with fuzzy edges. *Social Science & Medicine*, 63(1):212–224.

Leifeld, P. (2017). *The Oxford handbook of political networks*. Oxford University Press Oxford.

Lenci, A. (2018). Distributional models of word meaning. *Annual review of Linguistics*, 4(1):151–171.

LePine, J. A. and Van Dyne, L. (1998). Predicting voice behavior in work groups. *Journal of applied psychology*, 83(6):853.

Lerner, J. and Lomi, A. (2020). The free encyclopedia that anyone can dispute: An analysis of the micro-structural dynamics of positive and negative relations in the production of contentious wikipedia articles. *Social Networks*, 60:11–25.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics doklady*, 10(8):707–710.

Lewis, S. C. (2015). Reciprocity as a key concept for social media and society. *Social Media+ Society*, 1(1):2056305115580339.

Liang, H., Yang, Y., and Wang, X. (2013). Opinion dynamics in networks with heterogeneous confidence and influence. *Physica A: Statistical Mechanics and its Applications*, 392(9):2248–2256.

Lindblom, C. E. (1982). The market as prison. *The journal of politics*, 44(2):324–336.

Lowi, T. J. (1969). *The End of Liberalism: Ideology, Policy, and the Crisis of Public Authority*. Norton, New York.

Lusher, D., Koskinen, J., and Robins, G. (2013). *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge University Press.

Mahoney, C. and Baumgartner, F. R. (2015). Partners in advocacy: Lobbyists and government officials in washington. *The Journal of Politics*, 77(1):202–215.

Marlow, S. L., Lacerenza, C. N., and Salas, E. (2017). Communication in virtual teams: A conceptual framework and research agenda. *Human Resource Management Review*, 27(4):575–589.

Marsh, L. and Onof, C. (2008). Stigmergic epistemology, stigmergic cognition. *Cognitive Systems Research*, 9(1-2):136–149.

McFarland, A. S. (2007). Neopluralism. *Annu. Rev. Polit. Sci.*, 10(1):45–66.

Mercken, L., Snijders, T. A., Steglich, C., Vertiainen, E., and De Vries, H. (2010). Smoking-based selection and influence in gender-segregated friendship networks: a social network analysis of adolescent smoking. *Addiction*, 105(7):1280–1289.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space.

Mohammad, S. M. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion measurement*, pages 201–237. Elsevier.

Morrison-Smith, S. and Ruiz, J. (2020). Challenges and barriers in virtual teams: a literature review. *Nature Applied Sciences*, 2:1–33.

Naurin, D. (2007). *Deliberation Behind Closed Doors: Transparency and Lobbying in the European Union*. ECPR press.

Opsahl, T., Agneessens, F., and Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social networks*, 32(3):245–251.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957). *The measurement of meaning*. Number 47. University of Illinois press.

Padgett, J. F. and Powell, W. W. (2012). *The emergence of organizations and markets*. Princeton University Press.

Panayiotou, A. (2004). Switching codes, switching code: Bilinguals emotional responses in english and greek. *Journal of multilingual and multicultural development*, 25(2-3):124–139.

Pang, B. and Lee, L. (2008). *Opinion Mining and Sentiment Analysis*, volume 2. Now Publishers Inc.

Parsegov, S. E., Proskurnikov, A. V., Tempo, R., and Friedkin, N. E. (2016). Novel multidimensional models of opinion dynamics in social networks. *IEEE Transactions on Automatic Control*, 62(5):2270–2285.

Prabhakaran, V. and Rambow, O. (2016). A corpus of wikipedia discussions: Over the years, with topic, power and gender labels. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2034–2038.

Raffo, J. and Lhuillery, S. (2009). How to play the "names game": Patent retrieval comparing different heuristics. *Research policy*, 38(10):1617–1627.

Redert, B. (2022). From clubs to hubs: Analysing lobbying networks in eu financial regulation after the crisis. *Journal of Public Policy*, 42:390–408.

Ricci, A., Omicini, A., Viroli, M., Gardelli, L., and Oliva, E. (2007). Cognitive stigmergy: Towards a framework based on agents and artifacts. In *Environments for Multi-Agent Systems III: Third International Workshop, E4MAS 2006, Hakodate, Japan, May 8, 2006, Selected Revised and Invited Papers 3*, pages 124–140. Springer.

Rico, R., Sánchez-Manzanares, M., Gil, F., and Gibson, C. (2008). Team implicit coordination processes: A team knowledge–based approach. *Academy of Management Review*, 33(1):163–184.

Rijshouwer, E., Uitermark, J., and de Koster, W. (2023). Wikipedia: a self-organizing bureaucracy. *Information, Communication & Society*, 26(7):1285–1302.

Risco, M. (2025). Network effects on information acquisition by degroot updaters. *Economic Theory*, 79:201–234.

Robertson, A., Magdy, W., and Goldwater, S. (2021). Black or white but never neutral: How readers perceive identity from yellow or skin-toned emoji. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–23.

Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks*, 29(2):173–191.

Ronzhyn, A., Cardenal, A. S., and Batlle Rubio, A. (2023). Defining affordances in social media research: A literature review. *New Media & Society*, 25(11):3165–3188.

Rösner, L. and Krämer, N. C. (2016). Verbal venting in the social web: Effects of anonymity and group norms on aggressive language use in online comments. *Social Media+ Society*, 2(3):2056305116664220.

Rowley, T., Behrens, D., and Krackhardt, D. (2000). Redundant governance structures: An analysis of structural and relational embeddedness in the steel and semiconductor industries. *Strategic management journal*, 21(3):369–386.

Rutten, W., Blaas-Franken, J., and Martin, H. (2016). The impact of (low) trust on knowledge sharing. *Journal of knowledge management*, 20(2):199–214.

Rutter, D. and Stephenson, G. M. (1979). The role of visual communication in social interaction. *Current Anthropology*, 20(1):124–125.

Sabbaghi, O. (2019). How do entrepreneurship rates vary across different races? *Journal of Small Business and Enterprise Development*, 26(3):325–341.

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

Scherer, K. R. (2021). Evidence for the existence of emotion dispositions and the effects of appraisal bias. *Emotion*, 21(6):1224–1238.

Schlager, E. (1995). Policy making and collective action: Defining coalitions within the advocacy coalition framework. *Policy sciences*, 28(3):243–270.

Scott, J. (1991). *Social Network Analysis: A Handbook*. Sage, London.

Scott, J. C. (2013). Social processes in lobbyist agenda development: A longitudinal network analysis of interest groups and legislation. *Policy Studies Journal*, 41(4):608–635.

Shen, Z., He, G., Wu, X., Zhu, Y., and Shen, M. (2023). Opinion dynamics with intermittent-influence leaders on the signed social network. *Engineering Reports*, 5(9):e12649.

Shirky, C. (2011). *Cognitive Surplus: How Technology Makes Consumers into Collaborators*. Penguin.

Sick, H. (2023). From rhetoric to regulation: inferring lobbying influence on eu efforts to regulate co2 emissions of cars using network analysis. *Interest Groups & Advocacy*, 12(4):388–412.

Simmel, G. (1950). *The Sociology of Georg Simmel*. Free Press.

Sluban, B., Mikac, M., Kralj Novak, P., Battiston, S., and Mozetič, I. (2018). Profiling the eu lobby organizations in banking and finance. *Applied network science*, 3:1–19.

Smith, E. A. (2010). Communication and collective action: language and the evolution of human cooperation. *Evolution and human behavior*, 31(4):231–245.

Snijders, T. A., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological methodology*, 36(1):99–153.

Sørensen, M. P. (2016). Political conversations on facebook–the participation of politicians and citizens. *Media, culture & society*, 38(5):664–685.

Susi, T. and Ziemke, T. (2001). Social cognition, artefacts, and stigmergy: A comparative analysis of theoretical frameworks for the understanding of artefact-mediated collaborative activity. *Cognitive Systems Research*, 2(4):273–290.

Swaab, R. I., Maddux, W. W., and Sinaceur, M. (2011). Early words that work: When and how virtual linguistic mimicry facilitates negotiation outcomes. *Journal of Experimental Social Psychology*, 47(3):616–621.

Tanis, M. and Postmes, T. (2003). Social cues and impression formation in cmc. *Journal of Communication*, 53(4):676–693.

Tanis, M. and Postmes, T. (2007). Two faces of anonymity: Paradoxical effects of cues to identity in cmc. *Computers in Human Behavior*, 23(2):955–970.

Tortoriello, M. and Krackhardt, D. (2010). Activating cross-boundary knowledge: The role of simmelian ties in the generation of innovations. *Academy of management journal*, 53(1):167–181.

Truman, D. B. (1951). *The Governmental Process: Political Interests and Public Opinion*. Knopf, New York.

Tsai, W. and Ghoshal, S. (1998). Social capital and value creation: The role of intrafirm networks. *Academy of management Journal*, 41(4):464–476.

Tsai, W.-C., Chi, N.-W., Grandey, A. A., and Fung, S.-C. (2012). Positive group affective tone and team creativity: Negative group affective tone and team trust as boundary conditions. *Journal of Organizational Behavior*, 33(5):638–656.

Tushman, M. L. (1977). Special boundary roles in the innovation process. *Administrative science quarterly*, pages 587–605.

Urman, A. and Katz, S. (2022). What they do in the shadows: examining the far-right networks on telegram. *Information, communication & society*, 25(7):904–923.

Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (easi) model. *Current directions in psychological science*, 18(3):184–188.

Veksler, A. (2016). Lobbying in the sunshine—hiding behind transparency? *Journal of Public Affairs*, 16(1):39–49.

Viégas, F. B. and Smith, M. (2004). Newsgroup crowds and authorlines: visualizing the activity of individuals in conversational cyberspaces. In *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*, pages 10–pp. IEEE.

Viégas, F. B., Wattenberg, M., Kriss, J., and Van Ham, F. (2007a). Talk before you type: Coordination in wikipedia. In *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, pages 78–78. IEEE.

Viégas, F. B., Wattenberg, M., and McKeon, M. M. (2007b). The hidden order of wikipedia. In *Proceedings 2*, pages 445–454, Beijing, China. Online Communities and Social Computing: Second International Conference, Held as Part of HCI International.

Vivian, R., Falkner, K., Falkner, N., and Tarmazdi, H. (2016). A method to analyze computer science students' teamwork in online collaborative learning environments. *ACM Transactions on Computing Education (TOCE)*, 16(2):1–28.

Walther, J. B. (1997). Group and interpersonal effects in international computer-mediated collaboration. *Human communication research*, 23(3):342–369.

Warriner, A. B., Kuperman, V., and Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, 45:1191–1207.

Watzek, V., Anselmann, V., and Mulder, R. H. (2019). Team learning and emotions during teamwork: A qualitative study. *Research Papers in Education*, 34(6):769–789.

Weltevrede, E. and Borra, E. (2016). Platform affordances and data practices: The value of dispute on wikipedia. *Big Data & Society*, 3(1):1121–1367.

Wubben, M. J., De Cremer, D., and Van Dijk, E. (2009). How emotion communication guides reciprocity: Establishing cooperation through disappointment and anger. *Journal of experimental social psychology*, 45(4):987–990.

Xia, W., Ye, M., Liu, J., Cao, M., and Sun, X.-M. (2020). Analysis of a nonlinear opinion dynamics model with biased assimilation. *Automatica*, 120:109113.

Yale, L. J. and Gilly, M. C. (1995). Dyadic perceptions in personal source information search. *Journal of Business Research*, 32(3):225–237.

Yarovoy, A., Nagar, Y., Minkov, E., and Arazy, O. (2020). Assessing the contribution of subject-matter experts to wikipedia. *ACM Transactions on Social Computing*, 3(4):1–36.

Yilmaz, K. E., Abul, O., and Hussain, S. A. (2024). From tweets to tensions: exploring the roots of political polarization in turkish constitutional referendum. *Journal of Information Technology & Politics*, pages 1–17.

Zeng, A. and Battiston, S. (2016). The multiplex network of eu lobby organizations. *PloS one*, 11(10):e0158062.

Zheng, L., Mai, F., Yan, B., and Nickerson, J. V. (2023). Stigmergy in open collaboration: An empirical investigation based on wikipedia. *Journal of Management Information Systems*, 40(3):983–1008.

Zhou, Q. and Wu, Z. (2022). Multidimensional friedkin-johnsen model with increasing stubbornness in social networks. *Information Sciences*, 600:170–188.

# 7 Appendix

## 7.1 Appendix to Chapter I

Table 9: Notation Table

| Notation | Name | Meaning |
|---|---|---|
| $G(\mathcal{N}, E)$ | Graph | The overall structure representing nodes $\mathcal{N}$ and links $E$ |
| $i \in \mathcal{N}$ | Node | Represents an individual or entity in the network |
| $(i, j) \in E$ | Tie or Link | Represents a relationship or interaction between nodes $i$ and $j$ |
| $\mathbf{T} \in \mathbb{R}^{n \times n}$ | Trust Matrix | A matrix representing the level of trust or influence each node places on the opinions of others |
| $\mathbf{p}(t) \in \mathbb{R}^n$ | Opinion Vector | A vector containing the opinions of all nodes at time $t$ |
| $\mathbf{T}^\infty \in \mathbb{R}^{n \times n}$ | Equilibrium Trust Matrix | The trust matrix as $t \to \infty$, capturing long-term influence |
| $\mathbf{p}^* \in \mathbb{R}^n$ | Consensus Opinion Vector | The stable opinion vector at consensus |
| $\mathbf{B}(t) \in \mathbb{R}^{n \times n}$ | Interaction Matrix at $t$ | A time-dependent matrix indicating whether nodes interact at time $t$ |
| $\tilde{\mathbf{T}}(t) \in \mathbb{R}^{n \times n}$ | Sporadic Trust Matrix | A modified trust matrix at time $t$ adjusting for sporadic communication |
| $\tilde{\mathbf{s}} \in \mathbb{R}^n$ | Social Influence Vector | A vector representing each node's influence in the final opinion distribution |
| $\xi_{i,j}(t)$ | Interaction Probability | The probability that node $i$ will interact with node $j$ at time $t$ |
| $\gamma_{i,j}$ | Weight of Opinion Distance | A parameter representing the weight node $i$ gives to the opinion distance with node $j$ |

```
def generate_gnp(n, p):
    while True:
        G = nx.fast_gnp_random_graph(n, p)
        if nx.is_connected(G):
            return G

# Function to calculate the vector s for a list of influence matrices
def vector_s_calculation(list_of_matrix):
    result = list_of_matrix[0]
    for matrix in list_of_matrix[1:]:
        result = matrix @ result
    return result

# Function to check for equilibrium in influence matrix
def check_equilibrium(influencematrix):
    import statistics
    primi = [matrix[0, 0] for matrix in influencematrix]
    if len(primi) < 2:
        return False
    return statistics.variance(primi) < 9e-10

# DeGroot model simulation
def degroot(G, trust_matrix, steps):
    event_matrices = []
    K = G.copy()
    t = 0

    def event():
        P = K.copy()
        for i in K.nodes:
            opchange = [P.nodes[i]['Opinion'] * trust_matrix[i, i]]
            for j in K.neighbors(i):
                opchange.append(P.nodes[j]['Opinion'] * trust_matrix[i, j])
            K.nodes[i]["Opinion"] = np.sum(opchange)
        event_matrices.append(trust_matrix.copy())
        return K

    while t < steps:
        t += 1
        event()
        if check_equilibrium(event_matrices):
            break
```

```
    return vector_s_calculation(event_matrices)

# Tilde probability model simulation
def tildeprob(G, trust_matrix, probability_matrix, steps):
    event_matrices = []
    num = len(G.nodes)
    K = G.copy()
    t = 0

    def event():
        P = K.copy()
        trust_matrix_real = np.zeros((num, num))
        trust_matrix_tilde = np.zeros((num, num))
        for i in K.nodes:
            trust_matrix_real[i, i] = trust_matrix[i, i]
            for j in K.neighbors(i):
                if random.random() < probability_matrix[i, j]:
                    trust_matrix_real[i, j] = trust_matrix[i, j]
        for i in K.nodes:
            trust_matrix_tilde[i, i] = trust_matrix_real[i, i] / np.sum(trust_matrix_real[i])
            for j in K.nodes:
                trust_matrix_tilde[i, j] = trust_matrix_real[i, j] / np.sum(trust_matrix_real[i])
        event_matrices.append(trust_matrix_tilde.copy())
        for i in K.nodes:
            opchange = [P.nodes[i]['Opinion'] * trust_matrix_tilde[i, i]]
            for j in K.neighbors(i):
                opchange.append(trust_matrix_tilde[i, j] * P.nodes[j]['Opinion'])
            K.nodes[i]["Opinion"] = np.sum(opchange)
        return K

    while t < steps:
        t += 1
        event()
        if check_equilibrium(event_matrices):
            break

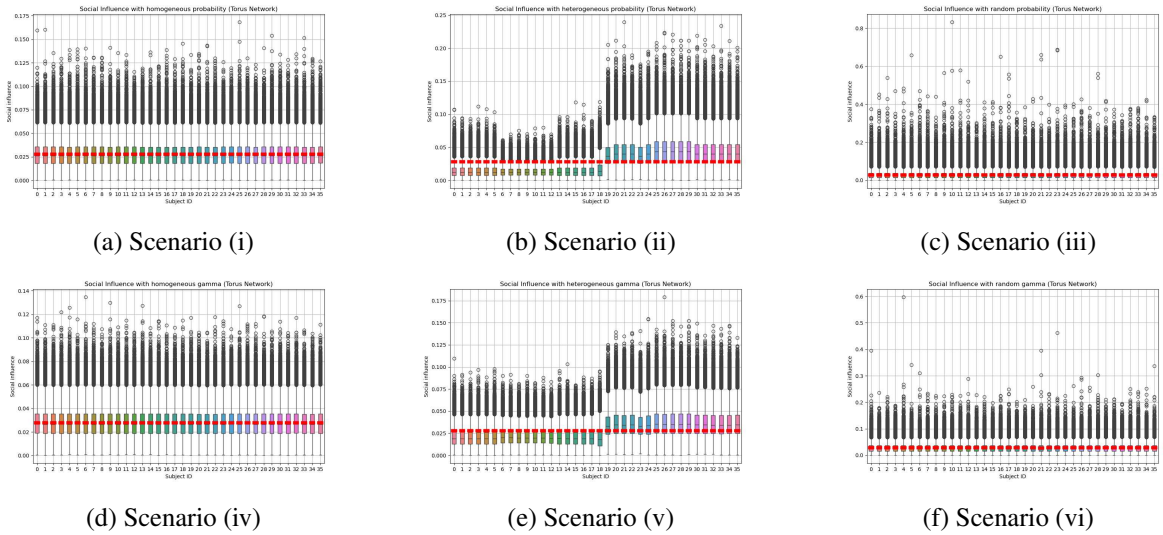
    return vector_s_calculation(event_matrices)

# Tilde probability with bound model simulation
def tildeprobbound(G, trust_matrix, gamma, steps):
    num = len(G.nodes)
    event_matrices = []
    K = G.copy()
    t = 0
```

```
def event():
    P = K.copy()
    trust_matrix_real = np.zeros((num, num))
    trust_matrix_tilde = np.zeros((num, num))
    for i in K.nodes:
        trust_matrix_real[i, i] = trust_matrix[i, i]
        for j in K.neighbors(i):
            if random.random() < (1 - ((abs(K.nodes[i]['Opinion']
            - K.nodes[j]['Opinion'])) ** gamma[i, j])):
                trust_matrix_real[i, j] = trust_matrix[i, j]
    for i in K.nodes:
        trust_matrix_tilde[i, i] = trust_matrix_real[i, i] / np.sum(trust_matrix_real[i])
        for j in K.nodes:
            trust_matrix_tilde[i, j] = trust_matrix_real[i, j] / np.sum(trust_matrix_real[i])
    event_matrices.append(trust_matrix_tilde.copy())
    for i in K.nodes:
        opchange = [P.nodes[i]['Opinion'] * trust_matrix_tilde[i, i]]
        for j in K.neighbors(i):
            opchange.append(trust_matrix_tilde[i, j] * P.nodes[j]['Opinion'])
        K.nodes[i]["Opinion"] = np.sum(opchange)
    return K

while t < steps:
    t += 1
    event()
    if check_equilibrium(event_matrices):
        break

return vector_s_calculation(event_matrices)
```




(a) Scenario (i) (b) Scenario (ii) (c) Scenario (iii)

(d) Scenario (iv) (e) Scenario (v) (f) Scenario (vi)

Figure 13: Distribution of $\tilde{s}$ in Torus Network

## 7.2 Appendix to Chapter II

Table 10: Tukey Test for Different Interaction Probabilities

| Interaction Probability | Groups | Diff. | Lower | Upper |
|---|---|---|---|---|
| Interaction Probability = 0.1 | | | | |
| $nonsimm_5$-$nonsimm_4$ | -3.414469e-04 | -0.001998144 | 0.001315250 | 0.9519480 |
| $Simm_4$-$nonsimm_4$ | -6.061764e-04 | -0.002262874 | 0.001050521 | 0.7832798 |
| $Simm_5$-$nonsimm_4$ | -3.777131e-04 | -0.002034410 | 0.001278984 | 0.9364273 |
| $Simm_4$-$nonsimm_5$ | -2.647296e-04 | -0.001921427 | 0.001391968 | 0.9766637 |
| $Simm_5$-$nonsimm_5$ | -3.626627e-05 | -0.001692964 | 0.001620431 | 0.9999362 |
| $Simm_5$-$Simm_4$ | 2.284633e-04 | -0.001428234 | 0.001885161 | 0.9847596 |
| Interaction Probability = 0.3 | | | | |
| $nonsimm_5$-$nonsimm_4$ | 1.268534e-03 | 0.0002068667 | 0.0023302009 | 0.0115145 |
| $Simm_4$-$nonsimm_4$ | 2.532278e-05 | -0.0010363443 | 0.0010869899 | 0.9999175 |
| $Simm_5$-$nonsimm_4$ | 1.971574e-03 | 0.0009099070 | 0.0030332412 | 0.0000109 |
| $Simm_4$-$nonsimm_5$ | -1.243211e-03 | -0.0023048781 | -0.0001815439 | 0.0139873 |
| $Simm_5$-$nonsimm_5$ | 7.030403e-04 | -0.0003586268 | 0.0017647075 | 0.3229678 |
| $Simm_5$-$Simm_4$ | 1.946251e-03 | 0.0008845842 | 0.0030079184 | 0.0000148 |
| Interaction Probability = 0.5 | | | | |
| $nonsimm_5$-$nonsimm_4$ | 0.0034081457 | 2.758658e-03 | 0.0040576338 | 0.0000000 |
| $Simm_4$-$nonsimm_4$ | -0.0001790860 | -8.285741e-04 | 0.0004704021 | 0.8938266 |
| $Simm_5$-$nonsimm_4$ | 0.0040738448 | 3.424357e-03 | 0.0047233329 | 0.0000000 |
| $Simm_4$-$nonsimm_5$ | -0.0035872317 | -4.236720e-03 | -0.0029377436 | 0.0000000 |
| $Simm_5$-$nonsimm_5$ | 0.0006656992 | 1.621107e-05 | 0.0013151873 | 0.0420760 |
| $Simm_5$-$Simm_4$ | 0.0042529309 | 3.603443e-03 | 0.0049024189 | 0.0000000 |
| Interaction Probability = 0.7 | | | | |
| $nonsimm_5$-$nonsimm_4$ | 0.0041940067 | 0.0038257202 | 0.0045622931 | 0.00e+00 |
| $Simm_4$-$nonsimm_4$ | -0.0007284150 | -0.0010967015 | -0.0003601285 | 2.20e-06 |
| $Simm_5$-$nonsimm_4$ | 0.0048435768 | 0.0044752903 | 0.0052118632 | 0.00e+00 |
| $Simm_4$-$nonsimm_5$ | -0.0049224217 | -0.0052907081 | -0.0045541352 | 0.00e+00 |
| $Simm_5$-$nonsimm_5$ | 0.0006495701 | 0.0002812836 | 0.0010178566 | 3.47e-05 |
| $Simm_5$-$Simm_4$ | 0.0055719918 | 0.0052037053 | 0.0059402782 | 0.00e+00 |
| Interaction Probability = 0.9 | | | | |
| $nonsimm_5$-$nonsimm_4$ | 0.0044322715 | 0.004263174 | 0.0046013686 | 0.00e+00 |
| $Simm_4$-$nonsimm_4$ | -0.0008825908 | -0.001051688 | -0.0007134937 | 0.00e+00 |
| $Simm_5$-$nonsimm_4$ | 0.0052216596 | 0.005052563 | 0.0053907566 | 0.00e+00 |
| $Simm_4$-$nonsimm_5$ | -0.0053148623 | -0.005483959 | -0.0051457652 | 0.00e+00 |
| $Simm_5$-$nonsimm_5$ | 0.0007893881 | 0.000620291 | 0.0009584851 | 0.00e+00 |
| $Simm_5$-$Simm_4$ | 0.0061042504 | 0.005935153 | 0.0062733474 | 0.00e+00 |

## 7.3 Appendix to Chapter III

Table 11: Model Estimations for Different Interaction Probabilities

| | Estimate | Std. Error | t value | Pr(>\|t\|) | |
|---|---|---|---|---|---|
| Interaction Probability = 0.1 | | | | | |
| (Intercept) | 2.790e-02 | 2.280e-04 | 122.391 | <2e-16 | *** |
| $\text{Nonsimm}_4$ | 3.313e-04 | 3.949e-04 | 0.839 | 0.401 | |
| $\text{Nonsimm}_5$ | -1.011e-05 | 3.949e-04 | -0.026 | 0.980 | |
| $\text{Simm}_4$ | -2.748e-04 | 3.949e-04 | -0.696 | 0.486 | |
| Interaction Probability = 0.3 | | | | | |
| (Intercept) | 0.0277760 | 0.0001461 | 190.106 | <2e-16 | *** |
| $\text{Nonsimm}_4$ | -0.0008164 | 0.0002531 | -3.226 | 0.00126 | ** |
| $\text{Nonsimm}_5$ | 0.0004522 | 0.0002531 | 1.787 | 0.07398 | . |
| $\text{Simm}_4$ | -0.0007910 | 0.0002531 | -3.126 | 0.00177 | ** |
| Interaction Probability = 0.5 | | | | | |
| (Intercept) | 2.779e-02 | 8.938e-05 | 310.89 | <2e-16 | *** |
| $\text{Nonsimm}_4$ | -1.826e-03 | 1.548e-04 | -11.79 | <2e-16 | *** |
| $\text{Nonsimm}_5$ | 1.582e-03 | 1.548e-04 | 10.22 | <2e-16 | *** |
| $\text{Simm}_4$ | -2.005e-03 | 1.548e-04 | -12.95 | <2e-16 | *** |
| Interaction Probability = 0.7 | | | | | |
| (Intercept) | 2.773e-02 | 5.068e-05 | 547.04 | <2e-16 | *** |
| $\text{Nonsimm}_4$ | -2.077e-03 | 8.779e-05 | -23.66 | <2e-16 | *** |
| $\text{Nonsimm}_5$ | 2.117e-03 | 8.779e-05 | 24.11 | <2e-16 | *** |
| $\text{Simm}_4$ | -2.806e-03 | 8.779e-05 | -31.96 | <2e-16 | *** |
| Interaction Probability = 0.9 | | | | | |
| (Intercept) | 2.783e-02 | 2.327e-05 | 1195.88 | <2e-16 | *** |
| $\text{Nonsimm}_4$ | -2.193e-03 | 4.031e-05 | -54.40 | <2e-16 | *** |
| $\text{Nonsimm}_5$ | 2.239e-03 | 4.031e-05 | 55.56 | <2e-16 | *** |
| $\text{Simm}_4$ | -3.075e-03 | 4.031e-05 | -76.30 | <2e-16 | *** |

[h!]

Table 12: Variable definition and description

| **Variable** | **Definition** | **Min** | **1$^{st}$qrt** | Mean | Median | **3$^{rd}$qrt** | Max | varnc | Obs. |
|---|---|---|---|---|---|---|---|---|---|
| EmbCommAro | It is the arousal of the comment evaluated through the embedded method. | 3,395 | 3,715 | 3,732 | 3,736 | 3,753 | 4,199 | 0,001 | 551087 |
| EmbCommVal | It is the valence of the comment evaluated through the embedded method. | 4,421 | 5,639 | 5,671 | 5,673 | 5,705 | 6,705 | 0,003 | 551087 |
| EmbSenAro | It is the arousal of the comments sent by the sender before the comment observed evaluated through the embedded method. | 3,395 | 3,735 | 3,740 | 3,741 | 3,746 | 3,885 | 0,000 | 448490 |
| EmbSenVal | It is the valence of the comments sent by the sender before the comment observed evaluated through the embedded method. | 4,921 | 5,667 | 5,677 | 5,677 | 5,688 | 5,978 | 0,000 | 448490 |
| EmbRecAro | It is the arousal of the comments sent by the receiver before the comment observed evaluated through the embedded method. | 3,527 | 3,736 | 3,740 | 3,741 | 3,746 | 3,849 | 0,000 | 456337 |
| EmbRecVal | It is the arousal of the comments sent by the receiver before the comment observed evaluated through the embedded method. | 5,240 | 5,667 | 5,677 | 5,678 | 5,688 | 5,911 | 0,000 | 456337 |
| EmbCommAroRel | It is the difference between the arousal of the comment and the arousal of past comments sent by the sender evaluated through the embedded method. | -0,315 | -0,024 | -0,007 | -0,003 | 0,013 | 0,457 | 0,001 | 447964 |

Continued on next page

**Table 12 – continued from previous page**

| Variable | Definition | Min | 1$^{st}$qrt | Mean | Median | 3$^{rd}$qrt | Max | varnc | Obs. |
|---|---|---|---|---|---|---|---|---|---|
| EmbCommValRel | It is the difference between the valence of the comment and the valence of past comments sent by the sender evaluated through the embedded method. | -1,473 | -0,037 | -0,005 | -0,003 | 0,028 | 1,008 | 0,003 | 447982 |
| EmbRecAro2 | It is the arousal of the comments sent by the receiver to the sender before the comment observed evaluated through the embedded method. | 3,541 | 3,730 | 3,739 | 3,741 | 3,750 | 3,844 | 0,000 | 208963 |
| EmbRecVal2 | It is the valence of the comments sent by the receiver to the sender before the comment observed evaluated through the embedded method. | 5,188 | 5,659 | 5,677 | 5,677 | 5,694 | 5,944 | 0,001 | 208963 |
| EmbRecAro2Rel | It is the difference between the arousal of the comments sent by the receiver to the sender before the comment observed and the arousal of the comments sent by the sender evaluated through the embedded method | -0,195 | -0,008 | -0,002 | -0,000 | 0,006 | 0,140 | 0,000 | 194916 |
| EmbRecVal2Rel | It is the difference between the valence of the comments sent by the receiver to the sender before the comment observed and the valence of the comments sent by the sender evaluated through the embedded method | -0,466 | -0,014 | -0,001 | -0,000 | 0,012 | 0,734 | 0,000 | 194916 |
| EmbPageAro | It is the arousal of the comments posted on the same *Talk Page* before the article observed evaluated through the embedded method | 3,551 | 3,732 | 3,739 | 3,740 | 3,747 | 3,943 | 0,000 | 422634 |
| EmbPageVal | It is the valence of the comments posted on the same *Talk Page* before the article observed evaluated through the embedded method | 5,188 | 5,663 | 5,676 | 5,677 | 5,691 | 5,908 | 0,000 | 422634 |

Continued on next page

**Table 12 – continued from previous page**

| Variable | Definition | Min | $1^{st}$qrt | Mean | Median | $3^{rd}$qrt | Max | varnc | Obs. |
|---|---|---|---|---|---|---|---|---|---|
| EmbPageAroRel | It is the difference between the arousal of the comments posted on the same *Talk Page* before the comment observed and the arousal of the comments sent by the sender before the comments observed evaluated through the embedded method. | -0,185 | -0,008 | -0,001 | -0,000 | 0,006 | 0,348 | 0,000 | 421326 |
| EmbPageValRel | It is the difference between the valence of the comments posted on the same *Talk Page* before the comment observed and the valence of the comments sent by the sender before the comments observed evaluated through the embedded method. | -0,538 | -0,0144 | -0,000 | -0,000 | 0,013 | 0,761 | 0,000 | 421374 |
| EmbThrAro | It is the arousal of the comments posted on the same thread before the article observed evaluated through the embedded method | 3,753 | 3,729 | 3,738 | 3,741 | 3,750 | 3,853 | 0,000 | 342138 |
| EmbThrVal | It is the valence of the comments posted on the same thread before the article observed evaluated through the embedded method | 5,105 | 5,658 | 5,677 | 5,678 | 5,697 | 5,968 | 0,001 | 342138 |
| EmbThrAroRel | It is the difference between the arousal of the comments posted on the same thread before the comment observed and the arousal of the comments sent by the sender before the comments observed evaluated through the embedded method. | -0,204 | -0,010 | -0,002 | -0,000 | 0,008 | 0,181 | 0,000 | 339668 |

Continued on next page

**Table 12 – continued from previous page**

| Variable | Definition | Min | $1^{st}$qrt | Mean | Median | $3^{rd}$qrt | Max | varnc | Obs. |
|---|---|---|---|---|---|---|---|---|---|
| EmbThrValRel | It is the difference between the valence of the comments posted on the same thread before the comment observed and the valence of the comments sent by the sender before the comments observed evaluated through the embedded method. | -0,579 | -0,018 | -0,001 | -0,000 | 0,017 | 0,750 | 0,001 | 339674 |
| WfaCommAro | It is the arousal of the comment evaluated through the WFA method. | 2.15 | 3.632 | 3.797 | 3.788 | 3.948 | 7.6 | 0.074 | 551087 |
| WfaCommVal | It is the valence of the comment evaluated through the WFA method. | 1.83 | 5.583 | 5.802 | 5.797 | 6.017 | 8.47 | 0.161 | 551087 |
| WfaSenAro | It is the arousal of the comments sent by the sender before the comment observed evaluated through the WFA method. | 2.785 | 3.769 | 3.817 | 3.813 | 3.862 | 5.245 | 0.009 | 448490 |
| WfaSenVal | It is the arousal of the comments sent by the sender before the comment observed evaluated through the WFA method. | 3.52 | 5.686 | 5.751 | 5.752 | 5.817 | 7.785 | 0.018 | 448490 |
| WfaRecAro | It is the arousal of the comments sent by the receiver before the comment observed evaluated through the WFA method. | 2.38 | 3.768 | 3.816 | 3.812 | 3.86 | 5.805 | 0.011 | 456337 |
| WfaRecVal | It is the valence of the comments sent by the receiver before the comment observed evaluated through the WFA method. | 2.61 | 5.686 | 5.75 | 5.75 | 5.815 | 7.89 | 0.02 | 456337 |
| WfaCommAroRel | It is the difference between the arousal of the comment and the arousal of past comments sent by the sender evaluated through the WFA method | -1.734 | -0.183 | -0.02 | -0.026 | 0.132 | 3.595 | 0.074 | 447964 |

Continued on next page

**Table 12 – continued from previous page**

| Variable | Definition | Min | $1^{st}$qrt | Mean | Median | $3^{rd}$qrt | Max | varnc | Obs. |
|---|---|---|---|---|---|---|---|---|---|
| WfaCommValRel | It is the difference between the valence of the comment and the valence of past comments sent by the sender evaluated through the WFA method | -4.007 | -0.166 | 0.052 | 0.044 | 0.263 | 3.75 | 0.159 | 447964 |
| WfaRecAro2 | It is the arousal of the comments sent by the receiver to the sender before the comment observed evaluated through the WFA method. | 2.15 | 3.711 | 3.807 | 3.801 | 3.896 | 7.24 | 0.03 | 208963 |
| WfaRecVal2 | It is the valence of the comments sent by the receiver to the sender before the comment observed evaluated through the WFA method. | 1.901 | 5.637 | 5.762 | 5.762 | 5.888 | 7.89 | 0.058 | 208963 |
| WfaRecAro2Rel | It is the difference between the arousal of the comments sent by the receiver to the sender before the comment observed and the arousal of the comments sent by the sender evaluated through the WFA method. | -1.875 | 1.804 | 1.945 | 1.953 | 2.094 | 4.185 | 0.074 | 194916 |
| WfaRecVal2Rel | It is the difference between the valence of the comments sent by the receiver to the sender before the comment observed and the valence of the comments sent by the sender evaluated through the WFA method. | -3.899 | -0.074 | 0.012 | 0.0 | 0.093 | 2.199 | 0.042 | 194916 |
| WfaPageAro | It is the arousal of the comments posted on the same *Talk Page* before the article observed evaluated through the embedded method | 2.3 | 3.745 | 3.815 | 3.812 | 3.881 | 5.805 | 0.015 | 422634 |
| WfaPageVal | It is the valence of the comments posted on the same *Talk Page* before the article observed evaluated through the WFA method | 2.43 | 5.665 | 5.753 | 5.755 | 5.843 | 7.785 | 0.027 | 422634 |

Continued on next page

**Table 12 – continued from previous page**

| Variable | Definition | Min | $1^{st}$qrt | Mean | Median | $3^{rd}$qrt | Max | varnc | Obs. |
|---|---|---|---|---|---|---|---|---|---|
| WfaPageAroRel | It is the difference between the arousal of the comments posted on the same *Talk Page* before the comment observed and the arousal of the comments sent by the sender before the comments observed evaluated through the WFA method. | -1.625 | -0.059 | -0.002 | -0.002 | 0.055 | 1.84 | 0.013 | 421326 |
| WfaPageValRel | It is the difference between the valence of the comments posted on the same *Talk Page* before the comment observed and the valence of the comments sent by the sender before the comments observed evaluated through the WFA method. | -3.393 | -0.072 | 0.003 | 0.001 | 0.075 | 2.226 | 0.024 | 421326 |
| WfaThrAro | It is the arousal of the comments posted on the same thread before the article observed evaluated through the WFA method. | 2.3 | 3.713 | 3.807 | 3.804 | 3.896 | 6.26 | 0.027 | 342138 |
| WfaThrVal | It is the valence of the comments posted on the same thread before the article observed evaluated through the WFA method. | 2.43 | 5.64 | 5.762 | 5.761 | 5.882 | 7.81 | 0.052 | 342138 |
| WfaThrAroRel | It is the difference between the arousal of the comments posted on the same thread before the comment observed and the arousal of the comments sent by the sender before the comments observed evaluated through the WFA method. | -1.625 | -0.086 | -0.01 | -0.007 | 0.064 | 2.38 | 0.023 | 339668 |

Continued on next page

**Table 12 – continued from previous page**

| Variable | Definition | Min | 1$^{st}$qrt | Mean | Median | 3$^{rd}$qrt | Max | varnc | Obs. |
|---|---|---|---|---|---|---|---|---|---|
| WfaThrValRel | It is the difference between the valence of the comments posted on the same thread before the comment observed and the valence of the comments sent by the sender before the comments observed evaluated through the WFA method. | -3.393 | -0.09 | 0.011 | 0.004 | 0.107 | 2.169 | 0.044 | 339668 |
| SenLonConDens | It is the amount of contributions made by the subject divided by the months passed from the first contribution (at the time of the observed comment). | 0,000 | 0,000 | 4,614 | 0,000 | 0,000 | 478 | 357,195 | 551130 |
| SenConDens | It is the amount of pages to which the subject has contributed divided by the number of contributions (at the time of the observed comment). | 0,000 | 0,000 | 0,592 | 0,000 | 0,000 | 132 | 4,604 | 551130 |
| SenConProd | It is the amount of contributions made by the user (at the time of the observed comment). | 0,000 | 0,000 | 26,826 | 0,000 | 0,000 | 488 | 5223,599 | 551130 |
| SanConAge | It is the time (months) passed since the first contribution (at the time of the observed comment). | 0,000 | 0,000 | 4,696 | 0,000 | 0,000 | 251 | 355,797 | 551130 |
| SenLonDens | It is the amount of comments made by the subject divided by the months passed from the first comment (at the time of the observed comment). | 0,000 | 0,000 | 0,725 | 0,008 | 0,454 | 77 | 6,432 | 551130 |
| SenDens | *User Contribution Density* is the ratio between the amount of contributions and the amount of *Articles* in which they were made (at the time of the observed comment). | 0,000 | 0,000 | 2,070 | 1 | 2,941 | 94 | 15,937 | 551130 |

Continued on next page

**Table 12 – continued from previous page**

| Variable | Definition | Min | $1^{st}$qrt | Mean | Median | $3^{rd}$qrt | Max | varnc | Obs. |
|---|---|---|---|---|---|---|---|---|---|
| SenProd | *User Comment Production* is the amount of comments posted by the user (at the time of the observed comment). | 0,000 | 0,000 | 9,560 | 1 | 8 | 532 | 632,517 | 551130 |
| SenAge | It is the time passed (months) since the first comment(at the time of the observed comment). | 0,000 | 0,000 | 4,696 | 0,000 | 0,000 | 251 | 355,797 | 551130 |