MODELAGEM E ANÁLISE DE COMPORTAMENTO COLETIVO EM REDES FORMADAS POR INTERAÇÕES ENTRE MÚLTIPLOS INDIVÍDUOS

CARLOS HENRIQUE GOMES FERREIRA

MODELAGEM E ANÁLISE DE COMPORTAMENTO COLETIVO EM REDES FORMADAS POR INTERAÇÕES ENTRE MÚLTIPLOS INDIVÍDUOS

Projeto de tese apresentado ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Doutor em Ciência da Computação.

Orientador: Jussara Marques de Almeida
Coorientador: Marco Mellia

Belo Horizonte
Março de 2022

CARLOS HENRIQUE GOMES FERREIRA

MODELING AND ANALYZING COLLECTIVE BEHAVIOR CAPTURED BY MANY-TO-MANY NETWORKS

Dissertation project presented to the Graduate Program in Computer Science of the Universidade Federal de Minas Gerais in partial fulfillment of the requirements for the degree of Doctor in Computer Science.

Advisor: Jussara Marques de Almeida
Co-Advisor: Marco Mellia

Belo Horizonte
March 2022

I dedicate this dissertation to all who one day believed in me.

Acknowledgments

I begin my acknowledgment by thanking God for giving me the strength and health to get here. Next, I would like to thank my supervisors, Professors Jussara Almeida and Marco Mellia, for their guidance, understanding, and patience during this doctoral process. I would also like to thank all the staff and professors of the Universidade Federal de Minas Gerais and the Politecnico di Torino for their support. In particular, I would like to mention Professors Fabrício Murai, Ana Paula Silva, Idilio Drago, Luca Vassio and Martino Trevisan for their collaboration and especially their friendship. I would also like to thank the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) and the *Compagnia di San Paolo* for their financial support. In addition, the Universidade Federal de Ouro Preto for the sabbatical leave.

I would also like to thank all my family, my parents Geralda and Francisco, my brothers Pedro and Francisco, my grandparents (in memoriam), uncles and cousins who were present in some way during this long journey. Furthermore, I would like to thank Mariana for all her love, understanding and support, my friends who have been fundamental in this journey, especially Alceu, Alexandre, Danilo, Leandro, Lourenço, Luiz, Rodolfo, and so many others to whom I am certainly grateful. Finally, I would like to thank all my colleagues at the Social Computing Laboratory, the SmartData Center, and the Universidade Federal de Ouro Preto.

"If a man does not know what port he is steering for, no wind is favorable to him."

(Seneca)

Resumo

Compreender o comportamento coletivo de (grupos de) indivíduos em sistemas complexos, mesmo em cenários em que as propriedades individuais de seus componentes sejam conhecidas, é um desafio. Do ponto de vista de modelos de redes, as ações coletivas desses indivíduos são, frequentemente, projetadas em um grafo formando uma rede de *co-interações*, aqui chamado de rede *many-to-many*. No entanto, o volume e a diversidade com que elas são observadas nos mais variados sistemas atuais como, por exemplo, aplicações de mídia social, transações econômicas e comportamento político em sistemas de votação, impõe desafios na extração de padrões (estruturais, contextuais e temporais) emergentes do comportamento coletivo e que estejam relacionados a um fenômeno alvo de estudo. Especificamente, a frequente presença de um grande número de co-interações fracas e esporádicas e que, portanto, não refletem necessariamente padrões relacionados ao fenômeno de interesse, acabam por introduzir "ruído" ao modelo de redes. A grande quantidade de ruído, por sua vez, pode mascarar os padrões de comportamento mais fundamentais capturados pelo modelo de rede, ou seja, os padrões que essencialmente são relevantes para o entendimento do fenômeno sob investigação. A remoção deste ruído é, portanto, um desafio importante.

Nesta tese, nosso objetivo é investigar a modelagem e análise do comportamento coletivo emergente de redes formadas por co-interações em diferentes contextos, visando extrair informação relevante e fundamental sobre um fenômeno alvo do estudo. Especificamente busca-se abordar a extração de propriedades estruturais, contextuais e temporais que emergem a partir de comportamento coletivo fundamentalmente representadas por comunidades extraídas da rede. Para tal, nós propomos uma estratégia geral que aborda os principais desafios mencionados acima. Em especial, esta estratégia contempla, como passo inicial, a identificação e a extração do *backbone* da rede, isto é, o subconjunto das arestas relevantes para o estudo alvo. Os próximos passos consistem na extração de comunidades deste *backbone*, como reflexo de padrões de comportamento coletivo presentes, e a caracterização das propriedades estruturais (topológicas), contextuais (relacionadas ao fenômeno de interesse) e temporais (dinâmica) destas comunidades. Tendo como base essa estratégia geral, nós produzimos artefatos específicos para as etapas que a compõe e avançamos o estado da arte, notavelmente, com um novo método para extração de *backbone*, um novo método capaz de representar temporalmente uma sequência de redes (*temporal node embedding*) possibilitando a extração de padrões temporais de interesse, e por fim, uma metodologia para auxiliar na seleção e avaliação de estratégias de extração de *backbones* do ponto de vista estrutural e contextual considerando o cenário mais comum, em que não há verdade fundamental (*ground truth*). Além disso, exploramos esses artefatos estudando três fenômenos que requerem diferentes estratégias de modelagem e análise. Especificamente, investigamos: (i) a formação de grupos ideológicos na Câmara dos Deputados do Brasil e dos Estados Unidos, (ii) discussões políticas *online* ocorrendo no Instagram no Brasil e na Itália e (iii) disseminação de informação no WhatsApp. Em suma, nossos resultados mostram que os artefatos propostos oferecem contribuições relevantes para o campo em que esta tese está inserida.

Palavras-chave: Redes Complexas; Comportamento Coletivo; Extração de *Backbones*; Detecção de Comunidades..

Sommario

Comprendere il comportamento collettivo di (gruppi di) individui all'interno di sistemi complessi è una sfida anche negli scenari in cui sono note le proprietà individuali dei loro componenti. Spesso questi fenomeni sono modellati come reti e le azioni collettive di questi individui sono proiettate su un grafo che rappresenta la rete di *co-interazioni*, che possiamo definire come una rete *many-to-many*. Il volume e la diversità di queste interazioni nei sistemi più grandi e complessi, come ad esempio, le piattaforme dei social media, l'insieme delle transazioni economiche o il comportamento politico nei sistemi di voto, complicano l'estrazione di schemi (strutturali, contestuali e temporali) dal comportamento collettivo e che sono fondamentalmente legati a un fenomeno di interesse. In particolare, la presenza di un numero elevato di co-interazioni occasionali che, quindi, non riflettono necessariamente pattern di interesse, introduce "rumore" nella rete modellata. Questo rumore può offuscare gli schemi comportamentali presenti nella rete in esame, rendendo di fatto impossibile la comprensione del fenomeno in esame. La rimozione di tale rumore diventa quindi una sfida chiave.

L'obiettivo in questa tesi è indagare la creazione e l'analisi dei modelli di comportamento collettivo in reti formate da co-interazioni, con l'obiettivo di estrarre informazioni rilevanti su un fenomeno d'interesse. Nello specifico, affrontiamo l'estrazione di proprietà strutturali, contestuali e temporali associate a modelli di comportamento collettivo, le quali sono tipicamente rappresentate da comunità estratte dalla rete stessa. A tal fine, proponiamo una strategia generale che affronti le suddette sfide. Questa strategia prevede, come passo iniziale, l'identificazione e l'estrazione della così detta rete *backbone*, ovvero il sottoinsieme degli archi che sono effettivamente rilevanti per l'obiettivo prefissato. Il passo successivo consiste nell'estrazione di comunità dalla *backbone* che manifestino i modelli di comportamento collettivo esistenti e permettano la caratterizzazione delle proprietà strutturali (topologiche), contestuali (relative al fenomeno d'interesse) e temporali (dinamiche) del fenomeno. Sulla base di questa strategia generale, proponiamo strategie specifiche che avanzano lo stato dell'arte per alcuni dei passaggi che lo compongono. In particolare, questa tesi fornisce approcci innovativi per l'estrazione della *backbone* e propone un nuovo metodo di temporali nodi *embedding* in grado di rappresentare ed estrarre differenti pattern temporali d'interesse da una sequenza di reti. Infine, definiamo una metodologia per supportare la selezione e la valutazione delle *backbone* da un punto di vista strutturale e contestuale, considerando il caso tipico di assenza di *ground truth*. In questa tesi, esploriamo questi approcci studiando tre diversi fenomeni che richiedono diverse strategie di modellazione e analisi. Nello specifico, indaghiamo: (i) la formazione di gruppi ideologici nella Camera dei rappresentanti Brasiliana e Statunitense, (ii) il dibattito online su Instagram in Brasile e in Italia e (iii) la disseminazione di informazioni via WhatsApp. Nel complesso, i nostri risultati mostrano che gli approcci proposti offrono contributi rilevanti nel campo in cui è inserita questa tesi.

Parole chiave: Sistemi complessi; Comportamento Collettivo; Estrazione del *Backbone*; Rilevamento della Comunità.

Abstract

Understanding the collective behavior of (groups of) individuals in complex systems, even in scenarios where the individual properties of their components are known, is a challenge. From the point of view of network models, the collective actions of these individuals are often projected on a graph forming a network of *co-interactions*, which we here refer to as a *many-to-many* network. However, the volume and diversity with which these co-interactions are observed in the most varied systems, such as, for example, social media platforms, economic transactions and political behavior in voting systems, impose challenges in the extraction of patterns (structural, contextual and temporal) emerging from collective behavior and that are fundamentally related to a phenomenon under study. Specifically, the frequent presence of a large number of weak and sporadic co-interactions, which, therefore, do not necessarily reflect patterns related to the phenomenon of interest, end up introducing "noise" to the network model. The large amount of noise, in turn, may obfuscate the most fundamental behavior patterns captured by the network model, that is, the patterns that are essentially relevant to the understanding of the phenomenon under investigation. Removing such noise becomes then a key challenge.

Our goal in this dissertation is to investigate the modeling and analysis of collective behavior patterns that emerge in networks formed by co-interactions in different contexts, aiming to extract relevant and fundamental information about a target phenomenon of interest. Specifically, we tackle the extraction of structural, contextual and temporal properties associated with patterns of collective behavior that are fundamentally represented by communities extracted from the network. To this end, we propose a general strategy that addresses the aforementioned challenges. In particular, this strategy includes, as an initial step, the identification and extraction of the network *backbone*, that is, the subset of the edges that are indeed relevant to the target study. The next steps consist of the extraction of communities from this backbone as a manifestation of the existing collective behavior patterns and the characterization of the structural (topological), contextual (related to the phenomenon of interest) and temporal (dynamic) properties of these communities. Based on this general strategy, we propose specific artifacts for some of the steps that compose it and advance the state-of-the-art, in particular with a new method for backbone extraction, a new temporal node embedding method capable of representing and extracting different temporal patterns of interest from a sequence of networks, and finally a methodology to support the selection and evaluation of backbones from a structural and contextual point of view, considering the most common scenario where there is no ground truth. Furthermore, we explore these artifacts by studying three different phenomena that require different modeling and analysis strategies. Specifically, we investigate: (i) the formation of ideological groups in the Brazilian and U.S. House of Representatives, (ii) online discussions on Instagram in Brazil and Italy, and (iii) information dissemination on WhatsApp. Overall, our results show that the proposed artifacts offer relevant contributions to the field in which this dissertation is inserted.

Keywords: Complex Networks; Collective Behavior; Backbone Extraction; Community Detection.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

A "complex system" is a system formed of many individual parts called "components" or "agents" interacting with each other and leading to large-scale behaviors. A key challenge when analyzing such complex systems is to uncover and understand the collective behavior of (groups of) individuals, even though the individual properties of its components may be known [Mitchell and Newman, 2002; Cohen and Havlin, 2010]. The notion of *collective behavior* has been widely studied in other domains such as Sociology and Psychology where it is defined in different ways [Rohall et al., 2013]. Here, we adopt the following definition, most related to the concepts explored in this dissertation: "Collective behavior refers to the kinds of *activities* engaged in by *sizable* but *loosely organized groups* of people." [Smelser, 2011; Turner et al., 2020].

We are surrounded by complex systems in both online and offline worlds. Many of them are of great interest to our society as they can strongly influence and drive social, cultural, economical and even political phenomena. Examples include (i) users helping to disseminate ideas and pieces of information as they share messages and comments on social media [Fraga et al., 2018; Paudel et al., 2019; Martin-Gutierrez et al., 2020]; (ii) members of a House of Representatives voting in a series of vote sessions and, from their voting patterns, forming ideological groups that, by crossing the formal boundaries of established political parties, more faithfully represent the political scenario of a country [Brito et al., 2020]; (iii) collective economic changes in a cryptocurrency market as result of successive changes in different financial assets such as stocks and commodities [Stosic et al., 2018; Papadimitriou et al., 2020]; and (iv) cultural mapping of a community by analyzing people's visits to different places driven by the need to pursue cultural interests [Yang et al., 2016a]. These are a few notable examples where loosely organized groups of people, acting individually, do interact with each other, driven by common interests, common goals or even hidden factors (e.g., coordinated behavior), and from such interactions a group/collective behavior may emerge without necessarily a previous social structure that explains it.

Network science has been a valuable field for modeling and studying these systems, as well as providing a set of theoretical tools that can be applied to describe and analyze the phenomena that lie behind many complex systems [Newman, 2003; Rossetti and Cazabet, 2018a]. The way interactions between the components of a complex system take place can be understood by first representing them in a mathematical model called a graph, and then forming is known as a network. In general terms, networks are a collection of nodes mapping the system components and interconnected by edges describing the relationship among them [David and Jon, 2010]. Once they have been modeled, a range of metrics such as density, diameter, clustering coefficients, number of connected components, among others can be applied to characterize the structural properties of the interactions under analysis, and from such characterization, obtain relevant knowledge about the underlying phenomenon being modeled [Barabási et al., 2016].

In addition to the aforementioned metrics that capture different aspects of the network topology, one of the most interesting problems in network characterization is the question of finding communities that are formed by the components of these networks [Fortunato, 2010; Rossetti and Cazabet, 2018a]. The main interest in community detection lies in the fact that communities typically display properties that are very particular and differ from the average properties of the complete network [Newman, 2006]. Thus, analyzing properties of different *communities* that build a global network may be a promising approach to reveal collective behavior patterns. This is because such communities naturally group users that are more "similar", with respect to common interactions and other behavior patterns. As such, by focusing on such communities, especially by exploring contextual properties associated with each such community, i.e., characteristics of the communities that are not explicitly captured by the network topology but rather are intrinsically related to the phenomenon being studied, one may be able to uncover properties that can help explain the emergence of collective behavior patterns, and, by doing so, gather a more clear glimpse of the driving factors behind the phenomenon under investigation [Yang et al., 2016a; Gao and Liu, 2016; Liu et al., 2018b; Lu et al., 2020].

Yet, many of the complex systems mentioned above (in both physical and online environments) are essentially structured by interactions occurring among multiple (potentially more than two) individuals simultaneously. For example, multiple users share the same piece of content or engage in a discussion on a social media platform. Similarly, multiple congressmen may vote similarly in a particular voting session. This contrasts with the traditional view of a network as a set of independent interactions among pairs of elements. The network that emerges from interactions among groups of elements simultaneously have attracted the interest of researchers in several areas such as Biology, Chemistry, Social Sciences and Economics [Hinds et al., 2000; Sporns and Kötter, 2004; Jin et al., 2007; Camacho et al., 2007; Lord et al., 2016], though only recently in the Computer Science community [Benson et al., 2018b; Meng et al., 2018; Guidotti et al., 2019]. It should be noted that the literature employs different terms with no clear consensus for this type of network as well as for the particular interactions driving its formation. Some of the terms used to refer to them are derived from networks formed by *sets* of interactions, sequences of *associations*, *cliques* (a particular type of motif) or *multi-actor* interactions [Benson et al., 2018a; Kumar et al., 2019; Lerner et al., 2019; Battiston et al., 2020]. We here adopt the terms *many-to-many* networks and *many-to-many* interactions, or simply *co-interactions*, as they clearly relate to the most basic concept of interactions occurring simultaneously among multiple elements.

Several studies have highlighted the effects of this sort of interaction on the topological structure of networks, notably when considering the aggregate effect of co-interactions occurring over time. In particular, they have shown how the competing dynamics behind them display a rich and varied pattern at different levels, including sequentiality, periodicity and sporadicity [Yin et al., 2017; Benson et al., 2018a; Guidotti et al., 2019; Coscia and Rossi, 2019; Yu et al., 2020; Sun et al., 2020; De Domenico and Altmann, 2020]. In particular, it is often the case that these many-to-many networks (as well as networks driven by traditional pairwise interactions) are overwhelmed by a large volume of edges representing random, sporadic and spurious interactions that, in essence are only weakly connected, if any at all, to the underlying phenomenon being investigated [Ghalmane et al., 2020]. As such, the large presence of these *noisy* edges adds further complexity to the study of collective behavior being modeled, as it requires one to first identify the (fewer) co-interactions that indeed are relevant to the particular target phenomenon.

In light of this, our aim in this dissertation is to investigate the modeling and analysis of *collective behavior* that emerges in *many-to-many networks* in different contexts, focusing on the *structural, contextual and temporal properties of the communities* that can be extracted from them.

1.1 Motivation

Several studies have investigated community detection in many-to-many networks projecting those co-interactions in an undirected and weighted graph [Coscia and Rossi, 2019; Pacheco et al., 2020; Brito et al., 2020; Neal, 2020; Uyheng and Carley, 2021]. In this way, the nodes represent the system components and an edge links two given components by the number of co-interactions in which they appear together. However, many-to-many networks display complex and diverse temporal properties. Specifically, some co-interactions may repeat consistently over time, while new ones emerge sporadically from partial copies or merge with previous ones. Thus, while persistent and repetitive co-interactions may occur, many random and sporadic ones are also present. In great volume, these weaker co-interactions may indeed hide the real underlying structure of the network representing the phenomenon under study, masking the true communities representing the collective behavior patterns that drive such phenomenon [Benson et al., 2018a; Coscia and Rossi, 2019; Guidotti et al., 2019]. We further elaborate on this issue, which is a key motivation to this dissertation, using a few concrete examples.

Let's start with the case of a *co-voting* network, i.e., a network modeling the voting patterns of a set of congressmen in a series of voting sessions. In such network, each node is a congressman and the weighted edges reflect the level of agreement between two congressmen along with those sessions. By analyzing this network, one can infer collective patterns that go beyond the traditional boundaries of political parties, revealing ideological similarities that can be useful to understand the political system of a country [Porter et al., 2005; Lee et al., 2017; Neal, 2020]. However, it is possible that a vote for a general-interest humanitarian cause may lead congressmen of different ideologies to adopt the same position, thus adding edges to the co-voting network. These edges represent sporadic interactions and are driven by a particular topic of large agreement, and thus do not necessarily reflect an ideological alignment. The presence of many such sporadic interactions, especially when considering sequences of voting sessions over time, may mask those interactions reflecting true ideological alignment, and by doing so, make it difficult to study, for example, the formation of the ideological groups (communities) that emerge in the network [Brito et al., 2020].

A similar issue may arise when studying information dissemination in social media applications. One common network model adopted in that case is to represent users by nodes and connect two users by an edge weighted by a number of pieces of content shared in common [Pacheco et al., 2020; Nobre et al., 2020; Uyheng and Carley, 2021]. In such context, it is possible that a very popular (or viral) piece of content is shared by a large number of different users, generating a large number of edges in the network. Yet, these users are simply reacting to a popular content, mostly independently from each other. This can hardly be seen as true *co-interactions* if one aims at investigating patterns of collective behavior relevant to the information dissemination process, that is, groups of users (communities) driving the information spread in the system [Bogdanov et al., 2014; Nobre et al., 2020]. Moreover, one may consider that measurement errors may occur in some settings. For example, errors are common in wearable sensor networks, used for human behavior modeling [Yang et al., 2015], as they are often subject to corruption, delay or loss of information caused by the wireless communication and the presence of hardware inaccuracies in the nodes [Ni et al., 2009]. This suggests that many co-interaction may indeed be *noisy* in the sense that they are little value (if any at all) to the given phenomenon being studied.

Yet, most algorithms for community detection in networks are designed under the assumption that the network structure modeled from the individuals' interactions faithfully represents the studied phenomenon [Coscia and Rossi, 2019]. In other words, all existing edges are taken into consideration in the process of uncovering communities. As such, in face of large volumes of noisy and sporadic edges, these algorithms are susceptible to misinterpretations, producing misleading conclusions. Indeed, there has been a widespread debate about the implications of ignoring data quality in network [Coscia and Neffke, 2017; Leão et al., 2018; Newman, 2018]. The selection of edges that are important to the phenomenon under study, referred to as *salient edges*, is tackled by algorithms that filter out noisy edges and provide a reduced (representative) version of the network that only contains those salient edges. Such reduced version of the network is called the *network backbone*. The definition of edge salience is based on an ensemble of problem-specific and node-specific perspectives of the network and quantifies the extent to which there is a consensus among the nodes with regard to the importance (representativeness) of an edge [Grady et al., 2012]. Often, a statistical edge property is defined and, then, used as the criterion to determine whether edges should be preserved or discarded [Qian et al., 2015].

Although a number of methods for network backbone extraction are available in the literature [Slater, 2009; Serrano et al., 2009; Radicchi et al., 2011; Grady et al., 2012; Silva et al., 2014b; Jacobs, 2015; De Melo et al., 2015; Rahimi et al., 2015; Dianati, 2016; Coscia and Neffke, 2017; Newman, 2018; Valverde-Rebaza et al., 2018; Yan et al., 2018; Coscia and Rossi, 2019; Kobayashi et al., 2019; Marcaccioli and Livan, 2019], there is still a lack of studies that evaluate how they perform on many-to-many networks, when applied with the purpose of uncovering collective behavior patterns represented by communities of nodes that are both tightly connected (in a topological sense) as well as cohesive and meaningful for the particular context under consideration (e.g., ideological behavior of congressmen), considering more than one method to choose the most appropriate one for the given phenomenon. In other words, the removal of noisy and sporadic edges considering the phenomenon to be studied leaving only the network backbone, particularly in the case of many-to-many networks, is a necessary step prior to applying community detection algorithms to reveal collective behavior patterns.

Another issue that arises is the frequent lack of a ground truth for evaluating the quality of an extracted backbone, which makes it difficult to validate the obtained results. Usually, authors evaluate the results only on topological metrics, such as community modularity, density, clustering coefficient, arguing that the extracted backbone has more clearly defined substructures than the original network [Serrano et al., 2009; Grady et al., 2012; Dianati, 2016; De Melo et al., 2015; Newman, 2018; Valverde-Rebaza et al., 2018; Dai et al., 2018; Neal et al., 2021; Mukerjee et al., 2022]. While topological properties are important to study the quality of the extracted backbone, they are only one aspect. Contextual criteria that relate the backbone properties (e.g., identified communities) to the characteristics of the phenomenon under study (e.g., amount of information shared by a community) should also be considered as part of the evaluation to provide a clearer picture of whether the backbone actually captures the driving factors behind the phenomenon [Yang et al., 2016a; Gao and Liu, 2016; Coscia and Neffke, 2017; Dai et al., 2018; Liu et al., 2018b; Marcaccioli and Livan, 2019; Lu et al., 2020]. Yet, to our knowledge, there is no previous work that addresses both topological and contextual metrics for evaluating backbones in terms of collective behavior represented by communities and, in particular, highlights how the properties of both the phenomenon and the method should be considered.

Moreover, from a temporal perspective, there are several studies that address the dynamics of co-interactions focusing on the challenges associated with prediction tasks [Yu et al., 2016; Benson et al., 2018a; Sun et al., 2020; Kumar et al., 2020; Yu et al., 2020; Zeno et al., 2020], models for temporal representation learning [Rossi et al., 2018; Lee et al., 2019; Hu and He, 2019], and mining of special graph substructures (e.g., motifs) over time [Paranjape et al., 2017; Liu et al., 2019; Fournier-Viger et al., 2020; Wang et al., 2020]. The different contexts evaluated by these studies leave no doubt that these systems display distinct and valuable, from the perspective of understanding system dynamics, temporal properties. However, these previous studies have mostly not addressed the temporal dynamics of communities as well as contextual information from the emerging backbone. Understanding such temporal patterns, in turn, can offer valuable insights into the dynamics of collective behavior in the system and, as such, a better understanding of the phenomenon under investigation [Del Re, 2013; Mitra et al., 2017; De Domenico and Altmann, 2020].

In sum, all of these issues must be considered and properly tackled when one aims at investigating collective behavior modeled in many-to-many interactions. As argued, no prior work has jointly addressed them all. Yet, such endeavor has the potential to reveal fundamental knowledge that is crucial to understand the phenomenon under study. For example, in the context of co-voting networks, it may reveal which ideological groups dominate the political system and how they evolve over time. Similarly, when analyzing networks of information dissemination, community analysis may help explain how a particular content becomes viral, whether particular groups of users are responsible for dictating the online discussions or even by coordinating the spread of misinformation [Nobre et al., 2020; Pacheco et al., 2020; Cruickshank and Carley, 2020; Uyheng and Carley, 2021].

Having stated the challenges and motivations driving this dissertation, we next define our guiding question and introduce the problem we address.

1.2 Problem Statement

In this dissertation, we are interested in modeling and analyzing collective behavior captured by many-to-many networks of user co-interactions in different contexts. To that end, we face the challenges of identifying and removing random, sporadic and weak (i.e., noisy) edges from the network as a step to uncover the communities that emerge from the remaining salient edges. In particular, we aim at characterizing structural, contextual and temporal properties of such communities as a means to reveal fundamental knowledge about collective behavior patterns driving the phenomenon of interest. Overall, the work developed in this dissertation is driven by the following guiding question:

> *Given a particular phenomenon of interest to be studied in the light of collective user behavior in a complex system, and given the (noisy) many-to-many network model built from a set of user co-interactions collected from that system, how can we reveal structural (topological), contextual and temporal properties of cohesive groups of users (communities) that can help shed light into how collective behavior emerges and evolves, driving the phenomenon under investigation?*

Our goal is to tackle this guiding question in different contexts, by exploring the general steps presented in Figure 1.1. We start with a target phenomenon to be investigated and a corresponding dataset gathered from a system where such phe-

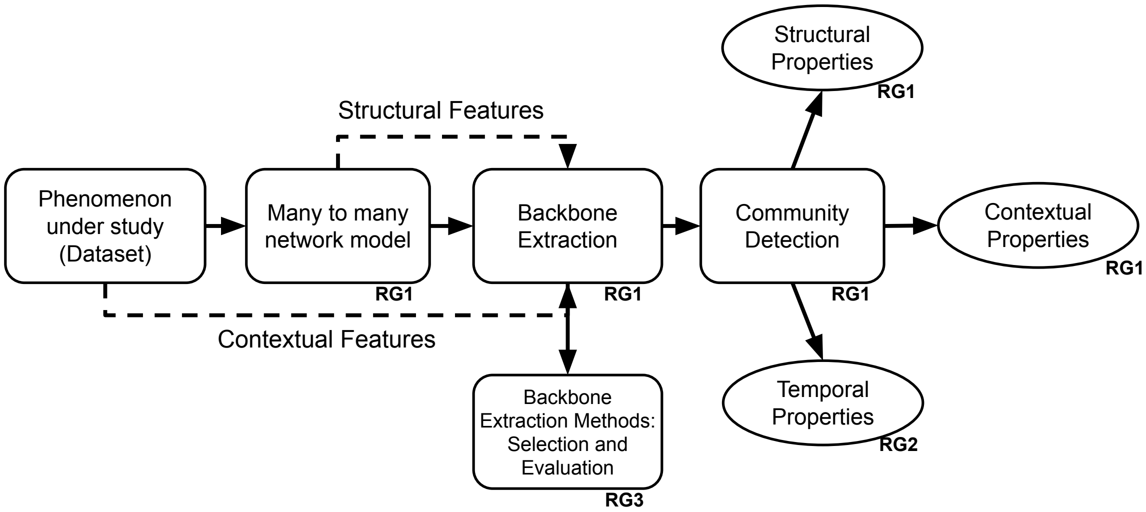


Figure 1.1. A visual representation of the scope of this dissertation.

nomenon will be studied. This dataset should contain a sequence of timestamped co-interactions among different individuals of the system, covering a period of interest. The first step is to build a many-to-many network model of these co-interactions which, as argued, may carry a large volume of noisy edges. Thus, as a next step, we must extract the backbone of this network, i.e., the set of salient edges with respect to the phenomenon of interest. For this purpose, we either rely on existing algorithms or propose new algorithms that exploit the peculiarities and requirements of the system under study. In addition, we can choose a backbone method whose assumptions are appropriate for a given phenomenon, as well as a set of candidate methods with the goal of finding out which of them best captures the underlying phenomenon of the network. Finally, community detection algorithms are used to uncover groups of users representing different collective behavior patterns influencing the system. We then aim at analyzing such communities, focusing on topological (community structure), contextual (system-related community attributes) and temporal (community dynamics) properties, attempting to draw fundamental knowledge about the target phenomenon.

1.3 Research Goals

The challenges associated with our target problem defined in Section 1.2 have led to the definition of the following research goals that we explore in this dissertation:

- **RG1: Uncovering topological and contextual properties of communities in many-to-many networks**: Our first goal consists of identifying communities representative of collective behavior in the target system and characterizing structural and contextual properties of such communities that are fundamentally related to the phenomenon under investigation. As mentioned, one key challenge to be addressed is the identification of the salient edges that compose the network backbone. We adopt different methods of backbone extraction, existing ones as well as new ones, based on the specificities of the system and the phenomenon under study.

- **RG2: Modeling the temporal dynamics of communities in many-to-many networks**: We are interested in analyzing the temporal dynamics of the identified communities by examining how the structural and contextual properties of the backbones evolve over time. From the structural perspective, we are interested in understanding and quantifying the dynamics at individual and community membership levels. With the contextual perspective, we can in turn examine the contextual properties of the phenomenon behind these communities (e.g., the discussion topics, co-interactions patterns) as they evolve over time.

- **RG3: Establishing a methodology for selecting and evaluating network backbone extraction methods in the face of a phenomenon modeled in many-to-many networks:** As mentioned earlier, several methods for backbone extraction in the literature may be used for our purposes in RG1 and RG2. However, it is challenging to select and evaluate the most appropriate method in scenarios for which often there is no ground truth. This largely depends on a comprehensive knowledge of the assumptions of both methods and phenomena. Our ultimate research goal, therefore, is to identify the key properties of such methods and potential phenomena to guide the selection, use, and evaluation of methods for the study of a particular phenomena.

In Figure 1.1, we label the steps that compose each research goal by labeling them with RG1, RG2, and RG3. Aiming at investigating collective behavior in different contexts, we examine the research goals 1 and 2 in different case studies. Then, motivated by the possibility of using more than one method and proposing alternatives for selecting and evaluating backbone methods, we approach our RG3 with a range of methods and structural and contextual metrics, finally testing it on prior and new case study.

1.4 Contributions

Building on the motivation and research question presented in the previous section, we offer our contributions to each of our RG below.

RG1 and RG2: Our contributions to RG1 and RG2 are explored through two case studies of very different domains, notably political ideologies of voting congressmen and online discussions in a popular social media application. As our first case study, we have investigated the emergence and the dynamics of ideological communities in political co-voting networks. In that front, our main contributions can be summarized as:

- We propose a methodology to uncover and analyze dynamic ideological communities and their polarization in party systems using historical data from the House of Representative of two different countries, namely Brazil and the United States;
- As part of our methodology, we propose to extract the network backbone by sequentially employing two approaches. This first one is driven by contextual information of the target phenomenon, that is the formation of ideologically aligned communities. The second approach is based on structural information of the initially extracted backbone, aiming at revealing *polarized* ideological communities;
- We investigate the dynamics of community membership by quantifying the extent to which community membership changes over time. We also propose a new method to jointly learn temporal node embeddings for multiple networks representing the target system in different periods of time. This method allows us to track the shifting of individual members over time in the political space defined by the identified ideological communities.
- We offer an extensive characterization of the properties of ideological communities, in particular polarized communities, in the Brazilian and American party systems over a long period of 15 years. Our results reveal strikingly different patterns, in terms of both structural and dynamic properties, and help understand the dynamics of ideological groups in distinct political systems.

As our second case study, we study the emergence of communities of discussions that may drive the information dissemination on a currently very popular social media platform, namely Instagram. Our main contributions in that direction are the following:

- We model the interactions that occur among groups of users commenting on the same post by an influencer. For the sake of comparison, we use two groups of influencers, one made up of politicians and political figures and the other composed by celebrities in general. Moreover, our study considers data from two different countries, namely Brazil and Italy, aiming at identifying cultural-driven similarities and difference. To drive this study we gathered a very large dataset for each country, covering, in both cases, a 10-week period around a major political election campaign. In total, our dataset contains the activity of approximately 1.8 million unique commenters on almost 37 thousand posts by 320 influencers in Brazil and Italy;

- Aiming at extracting the backbone of the modeled networks, we propose TriBE, a novel backbone extraction method based on a probabilistic reference network model where the edges are built on the assumption of *independent behavior* of commenters. The main idea is to focus on commenter co-interactions that deviate significantly (in a probabilistic sense) from this assumption, as these may more faithfully represent the ongoing discussions and driving information spread in the system. TriBE takes into account particular characteristics of user behavior in social media, notably the popularity of posts and commenters' engagement towards each influencer;

- We analyze the structure of the backbones and the communities that compose them in terms of their topological structure, textual properties of the discussions carried out by their members and temporal evolution. Our results reveal rich and distinct characteristics in terms of political and non-political discussions in both countries.

RG3:: Towards addressing RG3, we propose a methodology for selecting and evaluating methods for extracting networks based on a phenomenon under study. Specifically, our contributions are:

- We review nine methods for extracting backbones, characterize their assumptions and requirements, and discuss aspects to consider for their applicability in practice. We identify the network properties that these approaches exploit by showing how they can be used to study various phenomena. Compared to previous works, we offer a thorough and reasoned investigation covering a wide range of state-of-the-art methods, including recently proposed ones.

- We propose a methodology for applying, evaluating, and selecting the best method(s) for a given target phenomenon. Our methodology builds on the existing literature by bringing together metrics for backbone quality that capture both structural and contextual (i.e., phenomenon-specific) aspects. This allows us to evaluate the resulting backbone from the perspective of the emerging structure and the extent to which it captures the phenomenon under study. In addition, our methodology explicitly considers each method's matching properties and requirements with the key characteristics of the phenomenon under study as a step towards method selection.

- We apply the proposed methodology to two large-scale case studies (i.e., online discussions on Instagram and information dissemination on WhatsApp) related to phenomena with different requirements. For each case study, we show that different methods can lead to very different results and that the choice of the most appropriate method is of paramount importance to reveal knowledge about the phenomenon under study.

1.5 Outline

The remainder of this dissertation is organized as follows: Chapter 2 discusses previous work in areas closely related to the topic of this dissertation. Chapter 3 states our target problem, presents the associated challenges, and provides our general approach to tackle it. Chapters 4 and 5 present our investigation of the first two research goals in two different case studies, while Chapter 6 describes our methodology for selecting and evaluating network extraction methods. Finally, Chapter 7 concludes this dissertation and provides an outlook on future work.

Chapter 2

Background and Related Work

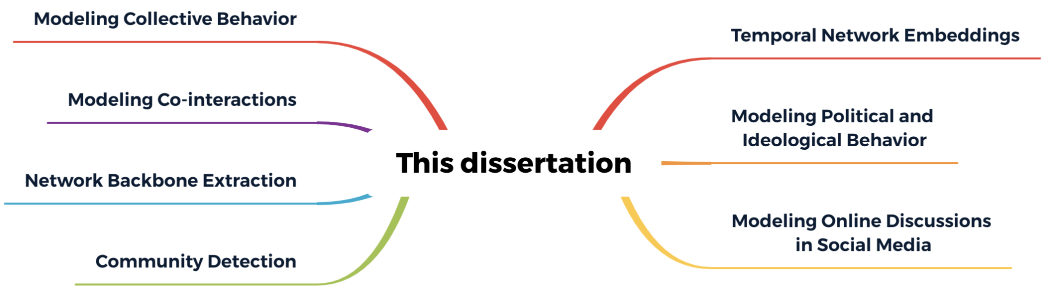


Figure 2.1. Mind mapping of the topics related to this dissertation.

In this chapter, we present a summary of background knowledge and related work that is essential to the understanding of this dissertation. Figure 2.1 shows the main topics of prior work related to this dissertation are organized. In light of this, this chapter is organized as follows:

- Section 2.1 discusses prior studies on collective behavior focusing on distinct models, their applications and outstanding problems;
- Section 2.2 provides an overview prior strategies to model user co-interactions;
- Section 2.3 offers a brief overview of existing network backbone extraction techniques and also focuses on previous efforts that attempt to systematize the use of a set of backbone extraction methods for a particular purpose. We then discuss their advantages, limitations, and opportunities for new contributions;
- Section 2.4, in turn, provides an overview on community detection in complex network, here taken as a fundamental representation of collective behavior;

- Section 2.5 discusses an important alternative for extracting patterns in networks from a temporal perspective, notably embeddings;
- Finally, Sections 2.6 e 2.7 present representative prior studies on the modeling of political ideological behavior and online discussions, respectively, which are topics that we tackle in our two case studies.

2.1 Modeling Collective Behavior

Collective behavior is a wide and complex concept covering different kinds of behaviors, structures, processes in varied contexts. In essence, collective behavior is a multidisciplinary topic of investigation, being studied by researchers in areas such as Sociology, Anthropology, Psychology, Political Science, Economics and, more recently, Computer science [Del Re, 2013]. Most of the existing theories about collective behavior are based on specific phenomena. For instance, protests, riots or panic, fads and trends in which a large number of people are obsessed with an object or idea for a period of time [Smelser, 2011; Mackay, 2012; Rohall et al., 2013]. In light of that, different definitions of collective behavior are presented in the literature. However, they all have something in common, namely the focus on the behavior of people in groups, usually in response to an event or to express a common feeling.

The concept of collective behavior has changed over time, broadening in response to changes in social relations and to the evolution of society. One of the main reasons was technological advances, which allows collective behavior to reach global scale [Del Re, 2013]. Popularity of online content [Lu et al., 2019b, 2020; De Domenico and Altmann, 2020], human mobility [Silva et al., 2014b; Yang et al., 2016a], market transaction [Peron and Rodrigues, 2011; Mateo et al., 2019; Saeedian et al., 2019] are notable examples where collective behavior has been studied. Here, we consider the definition adopted by Neil Smelser in 1962:

Definition: "Collective behavior refers to the kinds of *activities* engaged in by *sizable* but *loosely organized groups of people*." [Smelser, 2011; Turner et al., 2020].

In other words, according to this definition, collective behavior is essentially associated with a considerable number of people acting in a given context in general. Although there may be any small structure for a subset of these people who act collectively, it is extrapolated by the scale of joint actions considering all individuals. Therefore, they still constitute a loosely organized group.

In Computer Science, the modeling of collective behavior has been explored across different frontiers. The combination of new technologies and methodologies enables the crawling, storage and processing of large volumes of data that are sources of information and can contribute to a multidisciplinary sphere. In such a context, distinct computational models have been used to study collective behavior. In the following, we discuss some prior efforts in that direction.

For example, the adoption of models based on time series has allowed studies on how the preference for content generated by influential users favors bursts of engagement on Twitter [De Domenico and Altmann, 2020]. In [Mitra et al., 2017], the authors showed a relationship between the dynamics of news spreading on Twitter and their level of credibility. They observed that news with lower levels of credibility tends to attract more users to spread them. Lehmann et al. [Lehmann et al., 2012] showed that the diffusion of hashtags on Twitter analyzed from a collective perspective is naturally driven by exogenous factors. That is, co-interactions through hashtags posts are mostly expressed by common feelings that instigate users' actions.

In a different direction, prior studies explored machine learning techniques to analyze collective behavior patterns. For example, the authors of [Yang et al., 2016a] proposed to extract cultural similarities in cities by analyzing patterns of collective behavior through the use of Location-Based Social Networks data. By generating an affinity matrix between cities based on the features as daily activity, mobility, and linguistic perspectives of groups of peoples, the authors used spectral clustering techniques to discover cultural clusters around the world.

In a similar direction, Silva et al. proposed to map people's mobility between places using Location-Based Social Networks (LBSN) data to build a transition graph model. Using such model, the authors analyzed patterns of human mobility in different cities, discussing how to identify similarities and differences in human dynamics by grouping cities according to characteristics of people's mobility [Silva et al., 2014b]. Other prior studies by the same authors investigated the emergence of gender preference for venues in a given region in the real world [Mueller et al., 2017] and user preferences regarding eating and drinking habits across populations at different scales, e.g., countries, cities, or neighborhoods [Silva et al., 2017]. In [Hamedmoghadam et al., 2019], the authors modeled a network of taxi travel demands to discover latent collective mobility patterns. The proposed model allows identifying, for example, points of origin and destination that are more influential and groups of people who constantly present similar demands in their travels.

Conversely, Belhadi et al. [Belhadi et al., 2021] focused on identifying patterns of collective abnormal human behavior in pedestrians using images. The authors referred to abnormality as a set of pedestrians that are highly correlated, i.e, with a large number of shared locations. By proposing the use of machine learning models based on deep learning for image recognition, the authors showed that it is possible to learn different characteristics about mobility through historical data to extract collective abnormal behavior patterns. Taking a different approach, Barros et al. propose a methodology for modeling and analyzing node mobility in networks based on a node embedding method that models and reveals the importance of nodes in mobility and connectivity patterns while maintaining their spatial and temporal properties [Barros et al., 2021b].

Other approaches propose the use of alternative models to analyze collective behavior. As an example, Lu et al. [Lu et al., 2019b] proposed a *survival model* to identify factors that motivate and drive collective attention under content generated on online social networks. Orthogonally, Martin-Gutierrez et al. [Martin-Gutierrez et al., 2020] proposed a probabilistic framework to analyze how actions performed by individuals embedded in a social system trigger collective reactions (or responses). In [Gleeson et al., 2014], the authors propose a generative model to analyze how users collectively interact with Facebook applications taking into account the history of recent decisions and the cumulative popularity of each application. The results show that the future popularity of applications on Facebook is strongly associated with recency factors that are cumulative, suggesting that the adoption of an application among a set of users follows a collective trend.

Focusing on network-oriented models, closer to the goals pursued in this dissertation, a number of studies target the analysis of collective behavior notably in social media applications. For example, Lu et al. [Lu et al., 2020] proposed a framework for modeling collective behavior in cascading social systems (e.g., Twitter and Weibo). The authors found that users following the same profile can be organized into different groups, each one with particular characteristics driving such collective behavior, e.g., timing, structure, posts topic, and user interests. The author of [Liu et al., 2018a] used the Facebook and Wiki-talk (A dataset build on edits on user talk pages on Wikipedia) aimed at understanding how user's *social signature* changes over time considering their ego network. The results, obtained on Facebook and Wiki-talk, show that there are strong and temporally stable social signatures built around co-interactions on such platforms. In the same direction, Xu et al. [Xu et al., 2018] aimed at developing strategies to predict user behavior, given some knowledge of the behavior adopted by the user's neighbors in the network.

By associating users with common interests, Awal et al. [Awal and Bharadwaj, 2019] examined collective preferences by adopting an overlapping community detection approach. Their goal was to extract information and predict which categories of articles a user would read or would be interested in reading, based on his/her social collective actions on a consumer review site. Gao et al. modeled human behavior during different extreme events by correlating data from Twitter and GoogleTrends. The goal of the authors was to assess how human risk and emotional intensity generate collective responses in different regions and how these responses evolve over time. The results of a case study of the 2011 Japanese earthquake show different communities emerging with different perceptions and actions about the event studied [Gao and Liu, 2016].

Network-based modeling has also been used in other complex systems in addition to social media platforms. For example, Mateo et al. showed how network topology has a significant impact on collective behavior in the study of swarm robots [Mateo et al., 2019]. The authors proved the existence of optimal network topology to produce the most effective collective response. With a focus on final markets, Peron et al. [Peron and Rodrigues, 2011] aimed at identifying the emergence of collective behavior when stock prices exhibit a similar tendency, defining the market's direction synchronization. Similarly, the authors of [Saeedian et al., 2019] presented a broader analysis of the degree of collective behavior among the markets and the share of each market in the world global network. Stosic et al., in turn, studied studies the presence of communities in the context of cryptocurrency price changes, highlighting distinct community structures built on price variation [Stosic et al., 2018].

In short, there is a large number of studies in different areas but notably in Computer Science, focused on modeling collective behavior. Similar to some of the aforementioned prior efforts, we here also propose a network-oriented approach. Naturally, the collective actions of groups of individuals constitute what we here refer to as *co-interactions*, which can be modeled by a network. Consequently, this favors the extraction of patterns of different natures (e.g., structural, contextual and temporal). However, most of the previous efforts neglect some aspects of paramount importance that we will discuss and address throughout this dissertation. In particular, many such co-interactions occur from random and sporadic behavior, thus reflecting very weak relations from the perspective of analyzing collective behavior to understand the target phenomenon. Thus, it is important to filter out such noise, focusing only on those network edges that are really relevant (or salient) for the purpose of characterizing collective behavior [Benson et al., 2018a; De Domenico and Altmann, 2020; Ghalmane et al., 2020]. Although some prior studies have addressed this issue, they are limited to specific strategies to the systems studied [Peron and Rodrigues, 2011; Silva et al., 2014b; Stosic et al., 2018].

Similarly, focusing on the extraction of communities of people exhibiting collective behavior, the adoption of community detection approaches seems a natural strategy to identify structural properties of how individuals relate and provide a view of how the system is structured. Moreover, mapping the dynamics of these groups over time converge or contrast is something seldom explored by previous studies, although it constitutes a fundamental property of any collective behavior [Del Re, 2013; Gleeson et al., 2014; Mitra et al., 2017].

Here, our main hypothesis is that considering these previously neglected elements together can provide valuable insights about collective behavior in a given system. This imposes a series of challenges, starting with the possibility of proposing a generalized and agnostic modeling approach. In the following sections, we discuss other prior studies that are also related to this dissertation, including prior techniques employed in the analysis of collective behavior. We start by discussing prior studies on modeling user co-interactions, which, though closely related to our current goal, differs by not necessarily aiming at uncovering patterns of collective behavior.

2.2 Modeling User Co-Interactions

A plethora of phenomena related to human interactions online and in the real world have been analyzed using concepts from complex networks. Yet, many of these interactions occur among multiple (potentially two or more) entities simultaneously, here referred to as *co-interactions*. Acknowledging the distinct properties of co-interactions, some recent studies have empirically analyzed them through the lens of *higher-order* models. Such models focus on preserving the different structures of connectivity that can occur as two or more individuals interact with each other [Battiston et al., 2020; Yoon et al., 2020].

In the most basic form, these structures are represented by different motifs (e.g., triangular motifs, star, structural hubs, etc) [Benson et al., 2016; Rossi et al., 2018]. Indeed, Benson et al. [Benson et al., 2016] explored the diversity of such co-interactions to analyze how they build networks with diverse structural patterns. Similarly, Rossi et al. [Rossi et al., 2018] proposed a general framework to learn higher-order structures embedding representations based on distinct motifs, while others have used neural networks to extract latent patterns from signals to forecast new co-interactions [Meng et al., 2018; Hu and He, 2019]. Also, alternative methods have been proposed to represent different settings of higher-order structures [Wehmuth et al., 2016, 2017]. However, analyzing such co-interactions on large scale is, often, computationally intractable despite recent advances in high-performance data processing, which limits the use of many of the recently developed models [Wang et al., 2020; Yoon et al., 2020].

In this dissertation, we focus on the densest and most uniform form of co-interactions, where all individuals interact with each other, a particular type of motif known as clique [Battiston et al., 2020]. For such case, the literature adopts different nomenclatures, including *sets* of interactions, sequences of *associations*, *cliques*, and *multi-actor* interactions [Kumar et al., 2019; Lerner et al., 2019; Battiston et al., 2020; Yoon et al., 2020]. These co-interactions constitute an important substructure for some complex systems, exhibiting diverse properties that are relevant to the study of the (often global) phenomenon of interest [Benson et al., 2018a; Meng et al., 2018]. Examples of such co-interactions are online shopping carts with a set of items being purchased together, co-authors of scientific publications, co-interactions among proteins, people co-visiting the same places driven by cultural interests, congressmen taking the stand to vote during a voting session and groups of users commenting on the same topic on a social media application [Yang et al., 2016a; Benson et al., 2018b; Newman, 2018; Nobre et al., 2020; Brito et al., 2020].

Moreover, we are interested in looking at how a sequence of such co-interactions, driven by actions of interest of multiple individuals build a network. To that end, we model them by projecting them into a weighted and undirected graph $G = (V, E)$, where nodes (or vertices) in set V correspond to individuals and an edge $e_{ij} = (i, j)$, with weight w, is added to set E linking the components i and j if they have already interacted. The weight w corresponds to the number of interactions both individuals shared in common during the period under analysis. From this network model, the structural topology can be characterized based on traditional metrics such as density, diameter, clustering coefficient and number of connected components, among others [David and Jon, 2010].

However, as already argued [Liebig and Rao, 2016; Benson et al., 2018b; Coscia and Rossi, 2019; Cao et al., 2019; Kumar et al., 2020], such projected networks may include a large number of random, sporadic and weak edges that are not really part of the fundamental underlying network component representing the target phenomenon. Indeed, although this practice of projecting the original network into a graph of pairwise connections has been widely adopted, little attention has been paid to understanding how the projection considering all co-interactions may obfuscate relevant structural properties in the projected network.

For example, Cruickshank et al. [Cruickshank and Carley, 2020] proposed to analyze discussion topics on Twitter by modeling a sequence of networks of co-interaction built on co-occurrences of hashtags used by tweet users. Moreover, other information is considered to model the network, for instance, the textual similarity of Tweets using the same hashtag. Then, a community detection method is applied to identify topical clusters of hashtags. In another work, the authors analyzed the online hate of such communities [Uyheng and Carley, 2021]. Similarly, Pacheco et al. used contextual metrics, for instance, activity synchronization and text similarity to model the network representing Twitter users, aiming at uncovering coordinated behavior in Twitter [Pacheco et al., 2020, 2021].

More recently, some studies have shown the importance of removing random or weak edges to reveal an underlying structure (i.e., the backbone) in networks formed by co-interactions [Grinberg et al., 2019; Malang et al., 2020; Coscia and Rossi, 2019; Coscia et al., 2020; Mattsson, 2020; Mattsson and Stuart, 2020; Jaffe et al., 2020; Jiang et al., 2020]. For example, Leão et al. showed how removing random edges in co-authoring networks converge to a topology with more pure social relationships and better quality community structures, compared to the original complete network [Leão et al., 2018]. Another example is the study of information dissemination on the WhatsApp platform [Nobre et al., 2020], where the authors explored a network of co-sharing patterns (i.e., a network connecting users who shared the same piece of content) to reveal important properties of information spread on the platform.

Thus, as part of our approach to model co-interactions, we must tackle the challenges of using the projected network, notably the presence of a potentially large number of weak and possibly irrelevant edges. Specifically, we must investigate strategies that remove such noise and reveal edges in the projected network that, in fact, contribute to the study of a given phenomenon emerging from collective behavior pattern. In other words, we must investigate strategies to extract the backbone of the original network. Moreover, as our focus is on *collective behavior*, we must also extract communities from the backbone, as each community may represent an important pattern of collective behavior. Ultimately, our goal is to provide a methodology by combining these steps to gain relevant insight into the phenomenon under study. Such methodology should consider backbone extraction as a crucial step in addressing this problem. In short, the final goal is to use the best method for extracting backbones from a set of candidates to capture patterns of collective behavior. In the next sections, we discuss previous work on the two main components that make up such endeavor: network backbone extraction and community detection.

2.3 Network Backbone Extraction

Starting with applications, multiple works in various fields have shown the importance of backbone extraction methods to deal with random, sporadic, and weak edges that may obfuscate the phenomenon under study. For example, several studies applied early proposed methods to study phenomena in biological networks [Perkins and Langston, 2009; Bordier et al., 2017], transportation networks [Wu et al., 2006; Dai et al., 2018], economic networks [Namaki et al., 2011; Mattsson, 2020; Mattsson and Stuart, 2020], co-authoring networks [Leão et al., 2018; Galuppo Azevedo and Murai, 2021], human mobility networks [Silva et al., 2014b; Coscia et al., 2020; Bonaventura et al., 2021] as well as congressional voting networks [Brito et al., 2020]. More recently, some studies have highlighted the importance of this task in social media applications [Olson and Neal, 2015; Abbar et al., 2016; Pacheco et al., 2020, 2021].

Moving to the body of work that focuses on the proposal of new methods, they compared methods to alternatives in light of specific phenomena of interest in various domains, such as transportation, finance, and ecology [Serrano et al., 2009; Radicchi et al., 2011; Grady et al., 2012; Dianati, 2016]. Most of these prior studies rely on structural/topological properties, including node and edge coverage, clustering coefficient, centrality measures, and community quality measures, to evaluate different backbones extracted from the same network (thus comparing alternative extraction methods). As such, they offer only a partial view of the quality of the backbones. Contextual (i.e., phenomenon-specific) criteria, capturing the extent to which the extracted backbone represents the phenomenon under study, are not considered. More recently, some studies have proposed and compared backbone extraction methods based on regression-models as a means to capture contextual attributes specific to the phenomenon, relating them to topological properties of the backbone [Coscia and Neffke, 2017; Marcaccioli and Livan, 2019; Coscia, 2021]. However, these studies, as well as the aforementioned ones, do not provide a clear rationale as to why the subset of used methods fits the given phenomenon and therefore whether they are adequate to the study. Such reasoning is of utmost importance as different methods have different assumptions and properties, which may constrain their use or introduce unwanted biases to the study of specific phenomena.

We then focus on methods that are considered state of the art and that have already been used to model collective behavior in the context of many-to-many to give a brief overview of them. In addition, in Chapter 6, we propose a categorization that highlights their assumptions, requirements and various relevant aspects of these methods that have not been clearly stated in previous work as part of our methodology for selecting and evaluating backbone methods. In this way we make our own contribution to a more fundamental understanding of the methods. We describe here naive threshold-based backbone extraction, High Salient Skeleton [Grady et al., 2012], Random rElationship ClAssifier STrategy (RECAST) [De Melo et al., 2015], Disparity Filter [Serrano et al., 2009], Polya Urn Filter [Marcaccioli and Livan, 2019], Marginal Likelihood Filter [Dianati, 2016], Noise Corrected (NC) [Coscia and Neffke, 2017], and Global Statistical Significance (GloSS) Filter [Radicchi et al., 2011], and finally, Tripartite Backbone Extraction (TriBE) [Ferreira et al., 2020], which is a particular contribution of this dissertation and is described in detail in Chapter 5.

Threshold-based backbone extraction: one of the simplest, most intuitive and most used methods [Rahimi et al., 2015; Yan et al., 2018; Soro et al., 2020]. It consists on removing edges whose weights are smaller (or higher) than a pre-defined threshold τ, that is, edge saliency refers simply to edge weight. This method is adequate to studies where the salient edges are those with higher (or lower) weights. Otherwise, as previously argued [Tsur and Lazer, 2017], thresholds may bias the analysis and lead to misinterpretation of the results.

High Salient Skeleton (HSS) [Grady et al., 2012]: the backbone is extracted by first normalizing the edge weights and then performing a sampling process to define the shortest-path trees from each node to all other nodes in the network. Edge saliency is defined based on the frequency of its occurrence in the shortest path trees: edges with frequency below a pre-defined threshold τ are disregarded. In doing so, this method attempts to capture edges that simultaneously have heavy weights and are fundamental for keeping nodes connected. As such, the notion of edge saliency is inherently connected to network topology. Moreover, like for the threshold-based method, the use of a global threshold may lead to biases and misinterpretation [Grady et al., 2012].

***Random rElationship ClASsifier sTrategy* (RECAST) [De Melo et al., 2015]:** it assumes that a salient edge connecting two nodes must have at least one of two properties that differ significantly from random graphs: (i) the common neighborhood of two adjacent nodes (your friends are my friends) and (ii) the regularity of interactions (persistence) within the observed time period. Following this rationale, each edge is classified into one of four possible classes as follows. First, a single reference network model is created, as a random graph with the same numbers of nodes and edges and the same node degree of the original network. The distributions of the two aforementioned properties – common neighborhood and persistence – are then computed for the random graph. The following four classes are then obtained depending on whether each of the two properties significantly deviates (according to a pre-defined p-value) from the random graph: *Friends* (both common neighborhood and persistence deviate from random), *Bridge* (only persistence deviates), *Acquaintance* (only common neighborhood deviates), and *Random* (neither deviates). These classes provide a flexible concept of edge saliency as they can be employed differently, depending on the phenomenon being studied, to remove edges with particular properties.

Disparity Filter (DF) [Serrano et al., 2009]: it assumes that an edge connecting a given pair of nodes is salient if it has a disproportionate weight compared to the other edges leading from the nodes to their respective neighbors. In other words, salient edges are those whose weights deviate significantly from the null hypothesis that the weights of all edges incident to a given node are uniformly distributed.

Polya Urn Filter [Marcaccioli and Livan, 2019]: similarly to DF, this method assumes that edge weights emerge from the aggregate process of individual nodes' preferences to interact with each other over time. It also assumes that interactions between nodes are maintained and reinforced, such that the larger the number of interactions between the same two nodes, the higher the probability of they interacting again. A reference model is built for each edge, using the Polya Urn model [Hoppe, 1984] which captures the reinforcement of existing interactions by examining the degree and strength (the sum of the weights of all edges incident to the node) of each node incident to this edge. This reinforcement mechanism can be regulated and estimated by the system through a fine-tuning process. Salient edges are those that deviate significantly from such reference model (according to a given p-value).

Marginal Likelihood Filter (MLF) [Dianati, 2016]: assumes that edge saliency should be analyzed in light of the strengths of the two nodes the edge connects. The higher the strengths the larger the edge weight must be to be considered salient. Specifically, the method builds a reference edge weight distribution model for each edge: the probability that edge between nodes i and j ends up with weight w_{ij} is based on a Binomial distribution with parameters n defined by the total strengths of all nodes in the network and p computed based on the strengths of nodes i and j. An edge is considered salient if the observed weight deviates significantly from the one predicted by the reference model.

Noise Corrected (NC) [Coscia and Neffke, 2017]: similarly to DF and the Polya Urn Filter methods, NC also assumes that edge saliency arises from the cooperation between nodes. However, unlike those methods, NC preserves peripheral-peripheral connections, which is crucial for capturing edges that, despite having small weights, may still be considered relevant for the phenomenon under study. These connections may be preserved by estimating the expectation and variance of edge weights using a hypergeometric distribution, taking into account the propensity of both nodes to send and receive edges. It also provides a direct approximation through a per-edge reference Binomial distribution (similarly to the MLF method). The main advantage of NC, though, is the ability to estimate an error for the expectation of the weights. As in the other methods, an edge is considered salient if its observed weight significantly exceeds the expected weight (given the strengths of both nodes).

Global Statistical Significance (GloSS) Filter [Radicchi et al., 2011]: it assumes that salient edges cannot be identified independently of the overall network topology, once nodes have different degrees. As such, it builds a single (null) reference model that preserves the edges between nodes as well as the overall edge weight distribution. Yet, when selecting salient edges, i.e., edges whose observed weights significantly deviate from the reference model, the method estimates the probability of observing an edge weight between two given nodes considering the nodes' observed degrees and strengths as constraints.

Tripartite Backbone Extraction (TriBE) [Ferreira et al., 2020, 2021]: this method was proposed to study phenomena driven by user interactions in social media applications. It exploits the tripartite structure commonly found in such platforms, that is, a piece of content, the content creator, and the other users (e.g., the followers) who interact with each other in reaction to that content (e.g., by commenting on a post, retweeting the same tweet, etc). As such, the method addresses the heterogeneity in user activity level and content popularity typically observed in social media applications. Specifically, it builds a reference weight distribution model for each edge, based on a Poisson binomial distribution, whose parameters are computed based on the distributions of content popularity and user engagement towards content from the same creator (as estimated by prior interactions). Once again, salient edges are those whose observed weights significantly deviate from their corresponding reference models. This particular method is proposed in this dissertation and explained in more detail in Chapter 5.

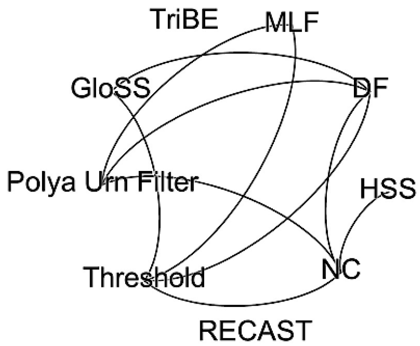


Figure 2.2. Selected backbone extraction methods: edges connect methods already compared to each other in prior work.

The aforementioned methods have been analyzed in the context of various phenomena of interest. Yet, no prior study is available in the literature where all nine methods are evaluated under the same analysis framework. For illustration purposes, Figure 2.2 shows which methods have been compared to each other in at least one of the studies aforementioned. Several methods have not been compared to each other (e.g, NC and MLF). Some of them (e.g., RECAST and TriBE) have not been compared to alternatives at all. Clearly not all methods are adequate to all studies, which justifies the lack of some comparisons.

Yet, the literature lacks an approach for selecting backbone extraction alternatives and evaluating them for a given study, as we note that few works evaluate existing solutions for a target study. Notably, Dai *et al.* have evaluated six methods for extracting the salient edges from transportation networks, i.e., those edges that are critical in the network [Dai et al., 2018]. As most prior studies, the authors considered only topological properties in such evaluation, which seems adequate given the interest in network connectivity (i.e., paths). Similarly, Mukerjee *et al.* investigated the impact of method parameters on network connectivity [Mukerjee et al., 2022]. They proposed choosing the best method and its parameters based on topological properties, notably, by maximizing the number of edges while maintaining the network's connectivity. Last, Zachary *et al.* evaluated existing methods for the specific case of bipartite networks, but once again considering only topological properties [Neal et al., 2021].

In contrast, our focus is on collective human behavior in many-to-many networks, which are expected to have more nuanced aspects (compared to, for example, transportation networks). These aspects should be investigated from both topological and contextual perspectives. Thus, we see a gap in terms of a principled methodology for selecting and evaluating the best method among alternatives for a given target phenomenon, taking into account whether the assumptions and requirements of each method are appropriate to the characteristics of the phenomenon.

2.4 Community Detection

Detecting communities (or clusters) in networks has been a widely studied problem for many decades due to its considerable range of applications. There is no universal definition for community detection [Leskovec et al., 2010]. Often, the concept of community has been defined as a group of nodes that have a higher likelihood of connecting (or similarity) to each other than to nodes from other communities [Barabási et al., 2016]. To extract such communities, it is then necessary to define some measure of connectivity or similarity that captures such intuition [Leskovec et al., 2010]. Given such a loose definition, a number of strategies have been proposed, each one targeting a somewhat different goal depending on the particular system and problem under study [Fortunato and Hric, 2016].

Existing solutions can be classified according to different criteria [Porter et al., 2005; Labatut and Balasque, 2012; Yang et al., 2016b; Fortunato and Hric, 2016]. For example, a community detection method can be classified as to whether or it allows overlap among communities, whether it considers static or dynamic communities, or even whether it considers node attributes or only topological properties [Xie et al., 2013; Jia et al., 2017; Rossetti and Cazabet, 2018a]. Here, we focus on detecting static and non-overlapping communities. In other words, we aim primarily at extracting communities composed of disjoint sets of nodes in a given network, to be applied in a sequence of networks representing the system in consecutive non-overlapping time windows. Even in this particular context, the diversity of existing methods is quite rich. In the following, we offer a brief discussion of the most important methods, following the categorization proposed by Fortunato et al. [Fortunato and Hric, 2016].

Some of the most widely used methods for detecting communities in networks are based on the *optimization* of some quality measure of the partitions representing the identified communities [Fortunato and Hric, 2016]. One such measure of quality is *modularity* [Newman and Girvan, 2004]. There are different methods to compute modularity, depending on the properties of the input graph. For a weighted graph $G(V, A)$, the modularity Q is defined as:

$$Q = \frac{1}{2M} \sum_{i,j \in V} \left[A(i,j) - \frac{k(i)k(j)}{2M} \right] \delta(c(i), c(j)) \tag{2.1}$$

where

- $A(i, j)$ is the weight of edge connecting vertices i and j;
- $k(i)$ and $k(j)$ are the sum of the weights of the edges attached to vertices i and j, respectively;
- M is the sum of the weights of all edges in graph G;
- $c(i)$ and $c(j)$ are the communities assigned to i and j, respectively; and
- δ is the Kronecker delta function, i.e., $\delta(c(i), c(j)) = 1$ if $c(i) = c(j)$, 0 otherwise.

Since exploring all possible partitions of the graph into communities is computationally impractical (it is NP-hard), several heuristic algorithms have been proposed. The Louvain method is one of these [Blondel et al., 2008]. One such heuristic that has been widely used in the literature is the Louvain method [Blondel et al., 2008]. This method has been applied in different domains, from biological networks [Rubinov and Sporns, 2010; Han et al., 2016] to social media applications [Vosoughi et al., 2018; Nobre et al., 2020].

It starts by finding first small (i.e., single-node) communities, optimizing the modularity locally on all vertices. It then proceeds iteratively: each small community is grouped into one (meta-)node and the first step is repeated. At each step, the resulting network partition is evaluated by the modularity metric. The process is repeated until no modularity increase can occur. In other words, given the graph $G = (V, E)$, the Louvain algorithm extracts the set of communities that provides the highest modularity value. Modularity lies in the range [-0.5,1], although, in practice, values between 0.3 and 0.7 are can be taken as evidence of well structured communities [Newman and Girvan, 2004].

The Louvain method does not require defining the number of partitions (i.e., communities) in advance and does not depend on other representations of the graph (e.g., to encode it in a dimensional space), but the use of modularity as a measure of quality suffers from a problem known as *Resolution Limit* [Fortunato and Barthelemy, 2007]. In its original formulation [Newman and Girvan, 2004], the modularity metric tends to increase as very small communities are merged into larger ones. Thus, methods that aim at optimizing this metric tend to favor such merges, which ultimately yields results that lose inherently small communities. This is directly related to the number of links on the network. Communities whose the sum of its nodes' degree is smaller than $\sqrt{(2 * E)}$ are invisible for the method and may be merged with other communities [Menczer et al., 2020]. There have been proposals to change the modularity metric by incorporating a parameter γ in its definition so as to make it robust to such cases, effectively allowing small communities to be identified [Lambiotte et al., 2008]. As in other studies [Han et al., 2016; Kaalia and Rajapakse, 2019], we here set γ to its maximum value ($\gamma = 1$). The revised definition of modularity, which is the one we use in this dissertation, is as follows:

$$Q = \frac{1}{2M} \sum_{i,j \in V} \left[A(i,j) - \gamma \frac{k(i)k(j)}{2M} \right] \delta(c(i), c(j)) \tag{2.2}$$

where

- $A(i, j)$ is the weight of edge connecting vertices i and j;
- γ is the resolution parameter;
- $k(i)$ and $k(j)$ are the sum of the weights of the edges attached to vertices i and j, respectively;
- M is the sum of the weights of all edges in graph G;
- $c(i)$ and $c(j)$ are the communities assigned to i and j, respectively; and
- δ is the Kronecker delta function, i.e., $\delta(c(i), c(j)) = 1$ if $c(i) = c(j)$, 0 otherwise.

However, it has been recently argued that the Louvain method has an inherent limitation that may lead to arbitrarily badly connected communities being extracted, regardless of the specific quality metric adopted. Such limitation gave rise to the Leiden algorithm [Traag et al., 2019]. This method works similarly to Louvain's with some modifications. Briefly, the Leiden algorithm consists of three steps. First, it starts by creating singletons partition from the whole network locally moving nodes from one community to another to find a partition. Unlike the Louvain algorithm that visits all the nodes in the network after the first visit, the Leiden algorithm only visits those nodes whose neighborhood has changed in each interaction, making the local moving step more efficient. In the second step, the algorithm tries to identify refined partitions from the ones created in the first step. When refining such partitions, they can be divided into more communities as long as their nodes are better connected. Nodes may be merged with any community within its partition for which the quality function increases. This step prevents possible problems of badly connected communities. The last step is to aggregate the network based on the refined partitions until no further improvements can be made [Traag et al., 2019].

Moreover, the Leiden algorithm can use the Constant Potts Model (CPM) as quality metric, which has some important differences compared to modularity [Traag et al., 2011]. One such difference is that, unlike modularity, CPM does not suffer from the resolution limit problem. Essentially, it increases as the intra-community edges are maximized and the inter-community edges are minimized. To achieve this, a parameter γ is established to force that communities should have a density of at least γ, while the density between communities should be lower than γ. CPM is defined as follows:

$$\mathcal{H} = \sum\nolimits_c [e_c - \gamma \binom{n_c}{2}], \tag{2.3}$$

where

- e_c is the actual number of edges in community c;
- n_c is the number of nodes in community c
- γ is parameter defining the community density of at least γ, while the density between communities should be lower than γ.

Other methods exploit *statistical inference* to fit a generative network model on the data [Shuo and Chai, 2016]. One of the most widely used approaches in this category builds a generative stochastic block model (SBM) [Lee and Wilkinson, 2019]. In essence, this method works as follows. Given two components, the vector of community memberships and the block matrix, each entry of such matrix represents the edge probability of two nodes be connected conditioned on their group membership. This makes it possible to assess the probability of the observed data, for modeling purposes. To find the latent groups of nodes in a network, it is necessary to infer the parameters of the model that provide the best fit for the observed network [Rosvall et al., 2019].

However, there is one problem: for the observed graph, neither component is known *a priori*. Thus, the objective of fitting a random graph constructed by SBM to an observed graph is to infer these two components simultaneously according to a function [Rosvall et al., 2019]. Some more recent approaches consider identifying the number of groups as well as other properties, such as, degree-corrected which is a new parameter that controls the expected degree of each node [Lee and Wilkinson, 2019]. An advantage of these methods is that they discover not only communities but also other properties such as, for example, disassortativity, core-periphery structure and the hierarchy between communities [Fortunato and Hric, 2016].

Other methods of community detection are based on *spectral clustering*. The fundamental idea behind these approaches is to identify communities using spectral properties of the graph, for example, a Laplacian matrix [Porter et al., 2005]. In brief terms, these methods map modes in a space using their respective eigenvectors and eigenvalues to define the coordinates [Fortunato and Hric, 2016]. Finally, the resulting points can be grouped into clusters using standard clustering techniques such as k-means [Shiga et al., 2007]. Methods based on spectral clustering are agnostic of the particular algorithm applied to cluster the point and, unlike those that rely on statistical inference, do not depend on any random graph model. In addition, they do not depend on any random graph model. Another advantage is that they tend to be quite scalable, for example, when matrix factorization is adopted [Bhattacharyya and Chatterjee, 2017]. Conversely, these methods may not work well for sparse networks or networks with very heterogeneous degrees, which is quite common in many domains [Krzakala et al., 2013].

Yet another class of community detection methods allows communities to be identified based on particular *patterns*, for example, by defining structural similarity measures [Labatut and Balasque, 2012; Fortunato and Hric, 2016]. For instance, Structural Clustering Algorithm for Networks (SCAN) is one of them that is able to identify and isolate community that are similar based on inter-community density, while also including two kinds of nodes that play special roles, i.e., vertices that bridge clusters (hubs) and vertices that are marginally connected to clusters (outliers) [Xu et al., 2007].

In contrast, other approaches are based on running a *dynamic processes* on the network. network. For example, some methods, such as the Walktrap [Pons and Latapy, 2005], rely on running random walks in the network. The assumption behind these methods is that if there is a reasonably strong community structure in the network, the random walkers tend to be trapped in nodes within a community before moving to other communities. Though presenting some satisfactory results, Walktrap may have a high computational cost, especially on denser networks [Fortunato and Hric, 2016].

Conversely to the aforementioned methods, which aim to identify communities in projected networks, others are concentrated in community detection on higher-order models. For instance, Pizzuti et al. employed a genetic algorithm to detect communities in networks formed by motifs [Pizzuti and Socievole, 2017], while Yin et al. adopted a local graph clustering approach based on motif conductance [Yin et al., 2017]. Tsourakakis et al. demonstrated the potential of methods based on motifs for tackling clustering problems and graph mining [Tsourakakis et al., 2017]. Huang et al., in turn, proposed a motif-based community detection algorithm for high-order multilayer networks [Huang et al., 2019]. All these methods aim at uncovering communities that are based on particular structural patterns (e.g., motifs) and, therefore, are not applicable for the modeling adopted here.

In this dissertation, our focus is on exploring algorithms for detecting existing communities, like the ones presented here, as a fundamental step to capture patterns of collective behavior. As already mentioned, our primary focus is on detecting static non-overlapping communities, by applying the selected method on snapshots of the network corresponding to different time windows. Once communities are detected in each time window, we are able to analyze their temporal dynamics by contrasting community membership across different time windows. To that end, different metrics and strategies can be applied, as will be discussed in Chapters 4 and 5. In particular, one such strategy is the use of temporal network embeddings to model the dynamics of a network. We review prior work in this area next.

2.5 Temporal Network Embeddings

Another body of work that relates to this dissertation is the use of alternative strategies for temporal modeling, such as temporal embeddings used to model dynamics of behavior in different contexts. Despite the rich literature on the use of embeddings to extract latent signals in various domains (e.g., word embeddings in textual documents [Kusner et al., 2015; Bamler and Mandt, 2017] and node embeddings in networks [Cui et al., 2018]), the study of *temporal* embeddings is relatively new. Specifically, in the network context, embeddings offer an important tool to network analysis due to their capability of encoding the structures and properties of networks with latent representations [Lu et al., 2019a]. Some efforts have proposed temporal latent space models by exploiting network embeddings [Zhu et al., 2016; Nguyen, 2018; Xie et al., 2020], in some cases jointly with node attributes [Li et al., 2017; Huang et al., 2017]. More recently, various dynamic graph embedding techniques have been proposed for diverse purposes and applications. We refer to [Barros et al., 2021a] for a complete review of such approaches and their applications.

Recall that, in this dissertation, we build sequences of networks (backbones and corresponding communities) for non-overlapping time windows. Thus, we are interested in methods that learn embeddings for such time windows independently. This ultimately leads to an "alignment problem" [Yao et al., 2018]. In simple terms, this means that while learning different embeddings for different networks, it may not be possible to place all learned embeddings in the *same* latent space, since the learning is done independently. Thus, it may be hard to track an element (i.e., a node) across time windows. One challenge in the "alignment problem" is to preserve similarities and to reveal differences of the neighborhood across time in the *same* latent space, which is a seldom addressed problem.

Recent work has tackled this problem, yet the proposed solution has some key properties that may not hold in several practical scenarios [Goyal et al., 2018b; Mahdavi et al., 2018; Goyal et al., 2018a]. Specifically, it assumes that the temporal changes in the networks are of short duration since it only considers the network of the previous time window to learn the next time embedding. Also, it uses the learned embedding from the previous time window to initialize the new one. These two properties implicitly keep the new embedding (time t) close to the immediately previous one (time t-1). Thus, the approach is unstable in sparse networks, when not all nodes are present in all time windows.

In a completely different context, the authors of [Yao et al., 2018] proposed a method to model word semantic evolution which simultaneously learns time-aware word vector representations and effectively solves the aforementioned "alignment problem". The method (presented in details in Chapter 4) tackles the problem of inferring how word semantics evolve over time by proposing a dynamic statistical model for learning time-aware word embeddings using all time windows simultaneously. The main advantage is that it reaches robustness for scenarios with both smooth and rough changes, thus being more flexible. It is also more robust to data sparsity and more scalable in terms of memory usage, which is important for large networks. Inspired by this work, we here adapt the proposal to the context of *network* embeddings and apply it to analyze the dynamics of individual nodes over time. We describe how we have performed this adaptation and the results obtained with it in Chapter 4.

2.6 Modeling Political and Ideological Behavior

A number of studies on political ideologies and behavior are based on the analysis of user behavior in online social media applications [Brady et al., 2019; Oliveira et al., 2020] as well as roll call vote networks [de Melo, 2015; Brito et al., 2020]. In particular, roll call votes may be used to build networks such that the nodes represent people (e.g., politicians), and two nodes are connected if they have voted similarly in one (or more) voting session. Using these networks, Andris et al.. studied committees' formation in the US House of Representatives, concluding that, despite the recent increase in polarization, there are moderate members in both parties who cooperate with each other [Andris et al., 2015]. Similarly, Porter et al.. studied the committees and subcommittees of the same chamber, exploiting the network connections that are built according to common membership [Porter et al., 2005]. Analogously, the polarization in the US Senate was evaluated using a network defined by the similarity of senators' votes [Moody and Mucha, 2013].

In [Dal Maso et al., 2015], the authors studied the relations between members of the Italian parliament according to their voting behavior, analyzing the community structure with respect to political coalitions and government alliances over time. Similarly, the cohesiveness of members of the European parliament was investigated through the analysis of network models combining roll call votes and Twitter data [Cherepnalkoski et al., 2016]. Others studied the behavior of political members, modeling roll call votes using signed networks. For example, Levorato et al. used signed networks to evaluate aspects related to political governance and party behavior in the Brazilian House of Representatives [Levorato and Frota, 2017]. The results revealed inefficient coalitions with the government as parties that make such coalitions have members distributed in different ideological communities over time. Mendonça et al., in turn, proposed an algorithm to evaluate signed networks using the European parliament network as case study [Mendonça et al., 2017]. Orthogonally, others have investigated the ideology of political members and users through profiles of social networks [Agathangelou et al., 2017; Darwish et al., 2017; Oliveira et al., 2018, 2020].

A closely related body of work has used roll call votes to measure *latent ideological patterns*. One such family of procedures is known as NOMINATE, whose variants are D-NOMINATE (originally called 'NOMINATE') [Poole and Rosenthal, 1985], W-NOMI-NATE [Poole and Rosenthal, 2000] and DW-NOMINATE [Poole and Rosenthal, 2001]. NOMINATE procedures assume a spatial model where each member has an ideal position in a space, while 'yea' and 'nay' votes on each roll call take on two positions in that space. Both D-NOMINATE and W-NOMINATE assume a multidimensional space (typically bidimensional), where errors (i.e., a member closer to a certain vote decides to vote the opposite way) follow a logit model. Unlike the former, W-NOMINATE assumes a distance model where dimensions are weighted differently, allowing for more flexible utility functions. DW-NOMINATE builds and improves upon W-NOMINATE by letting errors be normally distributed.

In [Bateman and Lapinski, 2016], the authors discussed some key shortcomings of methods based on ideal positions such as DW-NOMINATE and why they are not used more often in the American Political Development literature. One such limitation is the assumption of linear change in a member's ideal position over time. Moreover, these methods disregard important data. For instance, such methods cannot leverage information from unanimous votes – a typical situation in less polarized and fragmented political systems – which are discarded before parameter estimation [Poole and Rosenthal, 1985]. Similarly, the identities of members who have changed parties during the period of analysis are also disregarded. For instance, in [Poole and Rosenthal, 2011], the voting behavior of a member who changes parties once is considered as two independent sequences. While this is not a severe issue in non-fragmented party systems, it can introduce a large amount of noise when analyzing fragmented systems, such as the case of Brazil's party system, where such changes occurs with greater frequency.

Clinton et al. [Clinton et al., 2004] proposed a Bayesian simulation approach that improves existing methods by allowing the inclusion of ancillary information (e.g., the location of extremist members, member-specific covariates, or the evolution of the legislative agenda) in the model. The proposed framework also allows estimating changes of ideal positions over time by modeling the process associated with that change (e.g., members switching political parties). Although this approach offers a number of advantages over the aforementioned point estimate models, it also retains some statistical issues in relation to Bayesian ideal point estimation, such as proper variance estimates, scale and translation invariance, reflection invariance and outliers [Bafumi et al., 2005].

Some other prior efforts used alternative methods to network models. For instance, the approach proposed by Vaz de Melo [de Melo, 2015] addresses the problem of party fragmentation in Brazil by proposing an analytical method to identify the ideal number of parties that the country should have. The results show that party fragmentation is a reality in Brazil and that the number of parties that the country should have is much smaller than the existing one. Motivated by the problem of frequent party migration in Brazil, Desposato et al. [Desposato, 2006] proposed a model based on game theory to identify reasons behind the high migration of members among political parties.

Most of the aforementioned studies are based on scenarios of non-fragmented party systems (e.g., the United States) in which ideologies are clearer and, therefore, easier distinguished. However, moving to scenarios that suffer from party fragmentation (composed of multiple political parties), there is the challenge of dealing with the ideological overlap of these various political parties before analyzing ideologies. Although some studies based on alternative models to complex networks are robust to this characteristic, they also have limitations in terms of extracting temporal patterns. For example, quantifying the extent to which ideological groups (communities) change over time may be quite challenging. Moreover, network models can be considered for more than two positions (i.e., 'yea' and 'nay'), eliminating the need to collapse 'absence' and 'nay' as a single opposition category, as is the case for the United Nations General Assembly, where abstention is a milder form of disapproval than a 'nay' vote [Rosenthal, 2018].

In this dissertation, we tackle the modeling of ideological and political behavior in roll call votes as a one case study of our proposed general approach to model and analyze collective behavior. In this context, the collective behavior emerges from loosely organized individuals who, despite particular party boundaries, may form ideological groups that cross such boundaries. In this case, co-interactions among individuals (i.e., politicians) occur as they vote similarly, leading to a network model that exhibits, in essence, a noisy nature. For example, co-interactions may arise as highly consensual topics generate similar votes from many individuals, even though such behavior is no reflection of individual ideologies. Thus, this network model is a natural candidate to be analyzed using our proposed general approach. Specifically, we contribute to the aforementioned prior studies by using a network-oriented approach to identify and characterize ideological communities in both fragmented and non-fragmented party systems and extract relevant properties of them. The results from this case study are presented in Chapter 4.

2.7 Modeling Online Discussions in Social Media

The growing use of social media applications in recent years has attracted the attention of researchers to the analysis of online user discussions, considering different objectives and methodologies [Tang, 2017]. Examples include studying online discussions aiming at characterizing the presence and properties of hate speech [Mondal et al., 2017; ElSherief et al., 2018; Saha et al., 2019], trolls [Cheng et al., 2017], conflicts [Kumar et al., 2018], abuse [Waseem et al., 2017; Zampieri et al., 2019; Founta et al., 2019], toxicity [Gehman et al., 2020; Rajadesingan et al., 2020], cyberbullying [Raisi and Huang, 2017; Yao et al., 2019; Mukhopadhyay et al., 2020] and mental health issues [McClellan et al., 2017; Silveira et al., 2020].

In this dissertation, we use online discussions in social media applications as a case study of collective behavior, focusing on *political* discussions. Indeed, social media applications have been largely studied from the perspective of platforms for political debate. Twitter, in particular, has been the target of a large number of studies. As a summary of prior efforts, Nguyen [Nguyen, 2018] presented a literature review of the role of Twitter on politics, discussing prior findings in terms of the effectiveness of the platform to help politicians win elections, political polarization, and the benefits of using Twitter in the political arena. Indeed, many studies have already argued for the increasing polarization in the online political debates [Gruzd and Roy, 2014; Vergeer, 2015], whereas others have explored the benefits that politicians can have from using Twitter to reach their supporters. For example, Chi and Yang suggested that politicians can significantly benefit from using Twitter, as they are able to establish networks with their peers and acquire their support [Chi and Yang, 2010].

In [Gorkovenko and Taylor, 2017], the authors studied user behavior on Twitter during live political debates. They observed that people often use the social network to share their opinions, experiences, make provocative or humorous statements, and interact and inform others. In [Badawy et al., 2018], the authors found evidence of the use of Twitter for political manipulation. Caetano et al., in turn, analyzed the behavior of politically engaged user groups on Twitter during the 2016 US presidential campaign [Caetano et al., 2018]. Using information from user profiles, contact networks and tweet content, the authors identified four different groups, namely advocates for both main candidates, bots and regular users, and analyzed their behavior in terms of language patterns, popularity and how tweets from each candidate affected the mood expressed in their messages.

The online discussions have also been studied in the context of other social media platforms. For example, Tasente et al. [Tasente, 2020] analyzed the political debate around the *Brexit* on Facebook. The author focused on the frequency at which European institutions spoke about Brexit on their Facebook pages and on identifying and analyzing the messages that generated higher engagement from users. In [Silva et al., 2020], the authors developed a system to detect political ads on Facebook and used it to present evidence of misuse during the Brazilian 2018 elections. In that direction, WhatsApp has also been the target of recent studies as an important platform for political debate and information dissemination, notably for the spread of fake news during political elections [Resende et al., 2019a; Caetano et al., 2019; Nobre et al., 2020; Maros et al., 2020].

Considering Instagram, which is the platform used in our second case study, the literature on user behavior and interactions is reasonably recent and somewhat restricted. Some studies focused on how different types of content or profiles attract user engagement, notably in the political context. For example, Zarei et al. [Zarei et al., 2019] analyzed user engagement of twelve Instagram profiles divided into different categories (namely, sports, news and politics), searching for *impersonators* – i.e., users who simulate others' behavior to perform specific activities, such as spreading fake news. Muñoz et al. [Muñoz and Towner, 2017] studied image content posted by candidates during the 2016 US primary elections, highlighting combined factors that attract user engagement, whereas Trevisan et al. [Trevisan et al., 2019] performed a quantitative study of the political debate on Instagram, highlighting that politician's profiles tend to have significantly more interactions than others.

Moreover, other studies have also analyzed user engagement on online discussions from different perspectives on Instagram. For example, Kang et al. [Kang et al., 2020] studied a well-known strategy to trigger user interactions in social media, namely mentioning a friend in a comment. The authors proposed a model capable of classifying mentions into three categories based on its motivation: information-oriented, relationship-oriented, and discussion-oriented. Jaakonmäki et al. [Jaakonmäki et al., 2017] analyzed the influence of the content posted for social media marketing. They used machine learning algorithms to extract textual and visual content features from posts, along with creator and context features, to model their influence on user engagement. In another direction, Yang et al. [Yang et al., 2019] studied the brand mentioning practices of influencers, finding that audience has highly similar reactions to sponsored and non-sponsored posts. They also proposed a neural network model to classify the sponsorship of posts combining network embedding with features related to the posts and followers. Similarly, Kim et al. [Kim et al., 2020] proposed a multimodal deep learning model that uses contextual information of posts, including textual and image content, to classify influencers as well as their posts into specific topics, such as food, fashion and traveling.

Other studies have analyzed properties of the textual content shared by Instagram users. For example, Zhan et al. [Zhan et al., 2018] analyzed the sentiment of captions of Instagram posts to provide a preview of the content to the reader. Arslan et al. [Arslan et al., 2019] also used sentiment analysis tools to develop a message-level emotion classifier, with the goal of detecting cyberbullying. With a similar goal, Gupta et al. [Gupta et al., 2020] focused on the temporal perspective, showing that, cyberbullying activities exhibit recurrent temporal patterns such as the occurrence of bursts. In the same direction, Kao et al. [Kao et al., 2019] proposed a social role detection framework to analyze cyberbullying on social media platforms, taking Instagram as one of their case studies. The framework considers the roles of victim, bully and supporter, which are automatically identified by analyzing comment network and linguistic properties.

In contrast to prior studies, we here take a completely different perspective by analyzing online discussions on Instagram, notably discussions on political subjects, through the less of a many-to-many network. In particular, we focus on discussions triggered by posts of particular influencers. As such, co-interactions occur among users who comment on the same post. The network that emerges from such co-interactions, just as the co-voting network discussed in the previous section, does suffer from the presence of noisy edges. Thus, our proposed general approach to study collective behavior may be applied in this context as well: by taking the online discussions and political debates as expressions of collective behavior, we are able to uncover fundamental properties governing their dynamics.

Our present effort is orthogonal to the aforementioned prior studies as it focuses on the *network* that emerges as users engage in online debates by commenting on the same posts, an aspect mostly neglected so far, especially in the political context. Indeed, the case study we address here largely complements prior studies of user interactions around political subjects [Zarei et al., 2019; Muñoz and Towner, 2017; Trevisan et al., 2019] by offering a much broader analysis of the dynamics of communities of users who engage in political discussions. Thus, we are able to offer a more comprehensive investigation by delving into the structural and contextual properties of the discussions these communities engage in as well as on their temporal dynamics. We further elaborate on this case study, presenting and discussing our main results in Chapter 5.

2.8 Summary

Recall that our main goal in this dissertation is to develop a general approach to study collective behavior through the lens of many-to-many networks, focusing on structural, contextual and temporal properties, aiming at uncovering fundamental knowledge about a phenomenon of interest. In this chapter, we have presented a review of the literature related to the main topics covered in our work. In particular, we discussed prior studies and findings in the areas of (i) modeling of collective behavior, (ii) modeling of user co-interactions, (iii) network backbone extraction, (iv) community detection, (v) temporal network embeddings, as well as prior findings related to our two case studies, notably the modeling and analysis of (vi) political ideological behavior and (vii) online discussions.As we discussed, our present effort builds on prior work to provide an original and novel approach to modeling and analyzing collective behavior. In the next chapter, we formally define our target problem and present an initial description of a general solution to tackle it.

Chapter 3

Modeling and Analyzing Collective Behavior

This chapter takes the first step towards modeling and analyzing collective behavior in many-to-many networks. The chapter is divided into four sections. First, we revisit the problem statement in Section 3.1. Then, in Section 3.2, we discuss some of the key challenges related to our target problem. In Section 3.3, we present our general solution to tackle this problem, focusing in particular on our first two research goals (defined in Section 1.3). We explore such solution through two case studies of interest, notably the emergence of ideological groups in political systems and the study of online discussions in a social media application in Chapters 4 and 5, respectively. Finally, we provide a summary in Section 3.4.

3.1 Problem Statement

This dissertation focuses on the following setting of investigation. Consider a particular *phenomenon* of interest that emerges or is driven by the collective actions of a number of individuals. Such individuals, acting either independently or partially coordinated, produce patterns of *collective behavior* that favor or leverage such a phenomenon in a *complex system*. Our general goal is to uncover fundamental knowledge that helps in understanding the dynamics of the given phenomenon, from the *lens of these collective behavior patterns*.

Examples of such phenomena include *information dissemination* in social media platforms, driven by groups of users sharing ideas and pieces of information; the emergence of *ideological groups in a political system*, as politicians cross the formal boundaries of established political parties by voting together and forming short or long term alliances; and a *cultural description of a geographical region* built from the individual choices made by its inhabitants in terms visitation patterns [Yang et al., 2016a], eating habits [Silva et al., 2014a; Cao et al., 2018] or any other expression of their cultural interests[1].

The collective behavior driving such phenomena (e.g., content sharing, voting, visitation or eating patterns) emerges from a sequence of individual actions that, intentionally or not, coincide from the perspective of the target phenomenon. For example, multiple users may share the *same* piece of information thus favoring its dissemination on the platform; multiple politicians consistently vote similarly despite belonging to different political parties, thus revealing fundamental ideological similarities; multiple inhabitants may exhibit similar interests in terms of places of visitation in a city as well as similar trajectories, revealing particular cultural patterns (e.g., popularity of particular types of restaurants or attractions [Silva et al., 2014a; Cao et al., 2018]). We here refer to a collection of such coincident actions, involving two or more individuals as a *co-interaction*. Note that individuals participating in these co-interactions may not necessarily have any previous social structure connecting them; rather, they are guided by common goals or interests and driven by hidden contextual elements. In the light of the aforementioned example phenomena, we can cite as potential sources of contextual elements, respectively, posts containing specific topics attracting interested users to online discussions in social media platforms, voting sessions with specific themes, and cultural interests that drive human actions in a particular geographic region.

It is important to emphasize that the collective behavior patterns driving the phenomenon of interest are often fundamentally very dynamic, as the contextual elements driving human behavior change over time. For example, user discussions on social media platforms naturally evolve over time, covering different topics and different user groups. Similarly, politicians in some political systems, notably those that are more fragmented, may also change their party memberships and, ultimately, their political ideologies, as time passes.

As described, investigating the properties of those co-interactions is a key step to reveal the collective behavior patterns that fundamentally influence and drive the target phenomenon. One approach to pursue such investigation is to encode these co-interactions into a graph model that captures the intensity between co-interactions. As the co-interactions may involve more than two individuals, we refer to such graph model as a *many-to-many* network to emphasize the multi-peer nature of these interactions.

[1] In that matter, *cultural mapping* [Silva et al., 2014b] is a practical and participatory tool to build such descriptions.

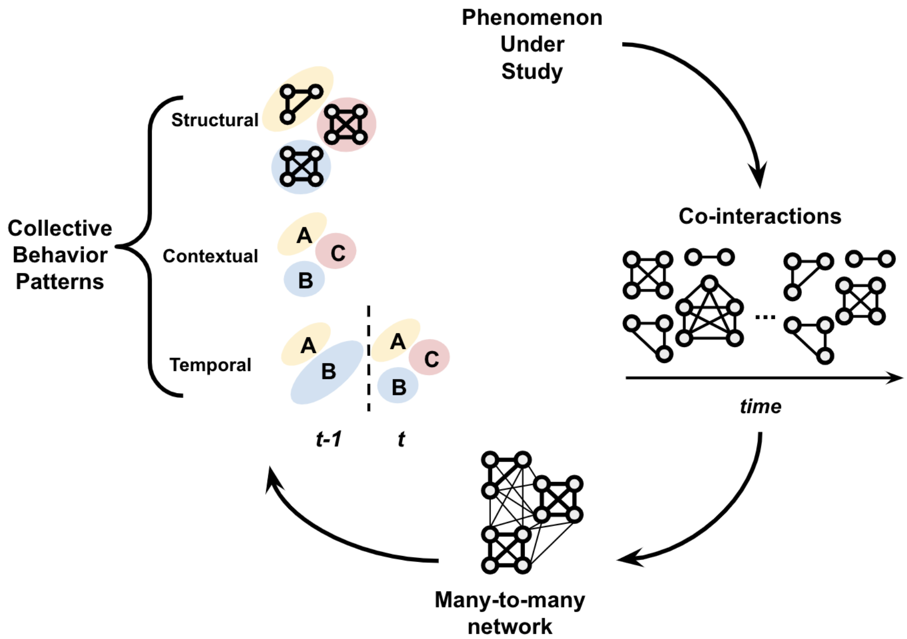


Figure 3.1. Overview of the problem statement.

In such a setting, the problem this dissertation aims to tackle is to *reveal key properties associated with the collective behavior patterns driving a phenomenon of interest, as a means to uncover fundamental knowledge about this phenomenon.* We consider as key: (1) *topological properties*, associated with the connectivity of individuals in the network of co-interactions, notably *communities* of individuals that exhibit common (collective) behavior; (2) *contextual properties*, associated with the contextual elements driving individual and collective behavior; and (3) *temporal properties*, reflecting the network dynamics.

Figure 3.1 illustrates the key elements that compose the problem this dissertation aims to tackle.

3.2 Challenges

The modeling and analysis of collective behavior, notably in many-to-many networks, raise a number of challenges in different domains. In this section, we discuss some of these challenges, notably those related to: (i) the presence of noise in the network; (ii) the identification of the network component that is more relevant to the target phenomenon (backbone extraction); (iii) the identification of groups of users exhibiting common (collective) behavior patterns (community detection); and (iv) the characterization of specific patterns of interest.

3.2.1 Presence of Noise in the Network

The multitude of different co-interaction patterns influencing a particular phenomenon of interest can be quite large. Indeed, as already shown [Benson et al., 2018b; Cao et al., 2019; Coscia and Rossi, 2019; Fu et al., 2019; Kumar et al., 2020; Battiston et al., 2020], they often exhibit a richness of patterns, including sequentiality, periodicity and sporadicity. Moreover, it is often the case that to study the target phenomenon, one must consider large volumes of co-interactions covering a reasonably long period of time of co-interactions. However, not all co-interactions are equally important to the study of the target phenomenon. As a matter of fact, it is often the case that many such co-interactions occur only sporadically or as result of pure chance and, as such, have weak relation, or no relation at all, to the target phenomenon.

To give an illustration, let us consider some possible scenarios when co-interactions take place in a given system where a phenomenon is to be investigated. Figure 3.2 shows the case of five co-interactions represented in dashed circles involving, in total, seven different individuals (numbered 1 to 7). Although a co-interaction may not be an atomic action (e.g., individuals co-interacting in an online discussion post their comments at different moments in time), for the sake of simplicity, we show them in some chronological order from left to right in the figure.

The simple example shown in Figure 3.2 illustrates different patterns, including: i) the presence of individuals with a high level of activity participating in all observed co-interactions (e.g., individual 2); ii) individuals participating in a few co-interactions possibly due to particular interests (e.g., particular topics of discussion) driving their choices (e.g., individual 5), and iii) individuals who sporadically co-interact (e.g., individuals 6 and 7); and iv) co-interactions involving a large fraction (or all) individuals (e.g., the 3^{rd} co-interaction from left to right). The last case may suggest a common goal driving individual behavior that is so general and broad that might not be of great interest to the study, that is, it might reflect a general trend and, as such, is not very relevant to understand the specific collective patterns driving the phenomenon under investigation.

Given such richness of patterns, it is unclear the extent to which the aforementioned diversity of co-interactions affects the study of the phenomenon under consideration. For example, the presence of a large number of random, sporadic and thus weak edges, may introduce excessive noise to the study of the phenomenon from the network perspective. As consequence, the most fundamental underlying network substructure,

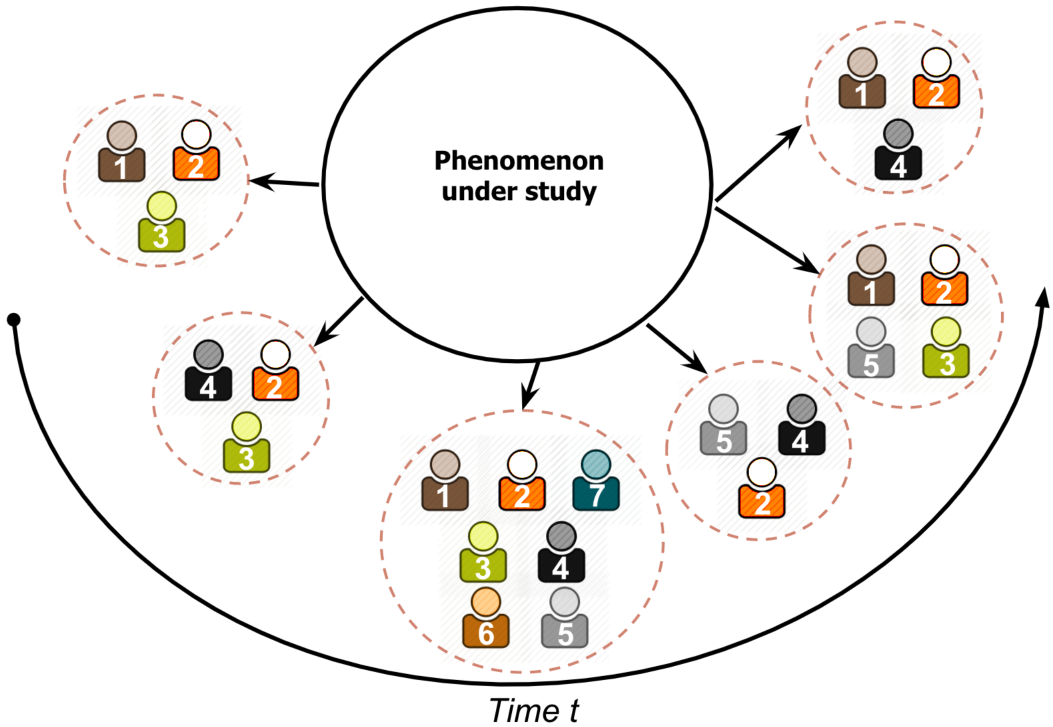


Figure 3.2. Example of diversity of co-interactions in a given system.

built from edges that are indeed relevant to the phenomenon, is may be obfuscated [Cao et al., 2019; Coscia and Rossi, 2019; Fu et al., 2019]. This situation gives rise to the following question:

> *What makes a particular edge relevant (or salient) to the study of the given phenomenon under consideration?*

As argued by Grady et al. [Grady et al., 2012], the definition of edge salience is based on an ensemble of node-specific perspectives in the network and quantifies the extent to which there is a consensus among the nodes with regard to the importance (representativeness) of a link. Hence, there is a wide range of different factors associated with the studied phenomenon and the target system that can define whether an edge is salient or not, which makes the identification of a salient edge a challenging issue.

Another challenge is to operationalize the identification and extraction of these edges in the input network, once an appropriate definition of edge salience has been found. The set of salient edges is referred to as the network *backbone* In the following section, we discuss some challenges related to the design and use of algorithms for extracting the backbone in the following section.

3.2.2 Network Backbone Extraction

The identification of salient edges that form the backbone in noisy networks is widely discussed in the literature, and several algorithms to perform this task have been proposed as presented in Section 2.3. However, each method is designed based on specific assumptions and, as such, reveals a very particular underlying structure. Thus, one challenge is identifying the method that should be applied for a target study given a pre-defined definition of edge salience.

Some methods provide a clearer definition of salience criteria to be captured, making them easier to apply to specific phenomena. Examples include the use of a threshold-based approach in some biological networks, where a minimum (or maximum) level of interaction is expected in a protein-protein interaction network [Milenković and Pržulj, 2008], financial market networks where a minimum correlation should be observed between stocks connected by edges [Namaki et al., 2011], and co-voting networks where congressmen with similar ideologies should have a minimal agreement during voting sessions captured by the edge (as we present in Chapter 4).

On the other hand, other methods are more complex and require closer examination to capture their assumptions and main characteristics. This is the case with most of the methods mentioned in Section 2.3. Once this is accomplished, it is possible to categorize them according to similar assumptions and characteristics. While this is quite a challenge, it does provide the opportunity to apply a set of methods to a particular phenomenon to achieve a better result. Nevertheless, this leads us to another challenge in evaluating and validating the best strategy. Newman argues that it is impossible to evaluate the quality of the backbone extracted by a particular method under normal circumstances because the true structure is unknown by definition [Newman, 2018]. Therefore, a major challenge is to investigate alternatives to evaluate their quality conditioned on the phenomenon's characteristics. There is evidence in the literature of the construction of models that exploit available contextual information and certain assumptions, and characteristics of the phenomenon under study, which therefore could be useful for the present work [Coscia and Neffke, 2017; Marcaccioli and Livan, 2019; Coscia, 2021].

Finally, we note the possibility of developing new approaches for backbone extraction. While innovative strategies can be used to study new phenomena, the challenge is primarily to identify relevant characteristics of one or more phenomena on different systems and understand how they affect the properties of the network model to rationalize the development of such new approaches.

3.2.3 Community Detection

Having discussed the challenges of backbone extraction, the next step is to identify patterns of collective behavior in it. The graph concept that directly maps into this idea is that of *community*. The literature in community detection is quite vast, as different approaches capture different concepts of communities for different network models [Labatut and Balasque, 2012; Yang et al., 2016a]. The greatest challenge in community detection is that there is a lack of a universal definition of community structure [Fortunato and Hric, 2016]. One definition of community that is widely adopted is a group of nodes that are more densely interconnected among themselves than with those in the rest of the network. Strictly speaking, according to this definition, a community is a cohesive subset of nodes that is distinctly separated from the rest of the network. A number of formal interpretations of this definition have been made in an attempt to formalize and combine both the aspects of cohesion and separation [Abraham and Hassanien, 2012; Labatut and Balasque, 2012]. For instance, algorithms like Louvain [Blondel et al., 2008] and Leiden [Traag et al., 2019] are driven by the *density* of the edges. In general terms, these approaches rely on metrics (e.g., modularity, coverage, and conductance) computed over intra-community and inter-community edges to assess the cohesion and separation of the detected communities [Fortunato and Hric, 2016].

Moreover, other approaches are concerned with specific patterns in the network. For instance, Louvain or Leiden adopting the Constant Potts Model (CPM) to establish a minimal intra-community density [Traag et al., 2011]. Remember, that this quality function allows a parameter to be set that makes it possible to define the communities with a particular density. The definition of a community is, to some extent, independent of the actual graph since the nature of this separation and the notion of cohesion depend on the selected density pattern [Rossetti et al., 2019]. On the other hand, some approaches involve looking for communities formed by a group of nodes that are similar to each other, but different from the rest of the network by employing a similarity measure. The strong point of this approach is that it goes beyond structural analysis and allows contextual information to be considered in the definition of the similarity function [Labatut and Balasque, 2012]. The SCAN (Structural Clustering Algorithm for Networks) algorithm is an example of this kind of approach [Xu et al., 2007].

Given the diversity of community definitions and corresponding community detection methods, another challenge is how to choose the best method for a target study. Once again, this choice should consider the specificities of the target phenomenon and the perspective taken in the investigation as well as how the definition of community associated with each method maps into the types of collective behavior of interest.

3.2.4 Collective Behavior Pattern Extraction

As discussed in the last sections, the tasks of extracting the network backbone and detecting communities exhibit particular challenges, mostly because a proper solution should consider the phenomenon under study closely. Having overcome such challenges, the communities found in the backbone offer an interpretable summary of the patterns closely related to this phenomenon, both from a structural point of view and from a contextual perspective of the system under analysis. For example, one can analyze the structural property of communities using metrics such as modularity [Blondel et al., 2008], degree distribution and clustering coefficient [David and Jon, 2010]. Moreover, context-oriented information associated with the nodes in each community may be used to aggregate external and domain-specific knowledge about it. If structural and contextual analyses are reasonably straightforward, investigating the temporal dynamics of these communities may offer some challenges.

As mentioned in Section 2.5, there are different approaches to model temporal networks. One such approach, adopted in this dissertation, is by building a sequence of networks representing a sequence of snapshots of the *real* (dynamic) network. However, mapping the structural and contextual properties of communities across different snapshots involves facing some particular challenges. Fundamentally, tracking a particular community over successive snapshots may be quite hard, as nodes dynamically change community membership and some nodes may leave or join the network as time passes.

There is a variety of metrics that can be applied to characterize the dynamics of communities. Examples include the persistence (or coverage) of nodes or edges, Jaccard index, degree correlation, among others [Fortunato and Hric, 2016; Rossetti and Cazabet, 2018b]. By applying these metrics across different snapshots, one may obtain a general assessment of the amount of change in the communities across consecutive snapshots. For example, Normalized Mutual Information (NMI) [Vinh et al., 2010; Wei and Carley, 2015], which will be formally defined in Chapter 4, is an information theoretic metric that can be used to quantify the extent to which the community structure identified in snapshot Δ_t changes in snapshot Δ_{t+1}. However, this metric requires the same set of nodes to be present in both snapshots, thus disregarding the arrival of new nodes and the disappearance of other nodes. Moreover, this metric, as others, offers a general perception of change but does not allow to zoom in which particular members have changed and how they changed.

Thus, it may be worth analyzing the dynamics of individual nodes, as a step towards a more clear understanding of how the communities are evolving. Embedding representations of nodes and graphs [Grover and Leskovec, 2016] have been proposed as an efficient means to uncover the network structure and perform various network mining tasks at node level [Nguyen, 2018; Xie et al., 2020]. As such, embeddings may be a useful framework for analyzing the dynamics of nodes. As we argue in Section 2.5, there is a challenge to track a given node across embeddings learned for different snapshots, as such learning is performed independently. Therefore, it is not possible to map a node (or group of nodes) across different embeddings, leading to an "alignment problem" [Yao et al., 2018].

In summary, the modeling and analysis of collective behavior driving a target phenomenon of interest raise a number of challenges. In this dissertation, we seek to tackle such challenges by offering a general approach that can be instantiated in different case studies. Our ultimate goal is to design a unifying framework that combines all the elements of a general solution while also offering a discussion on the issues one must consider when adopting it. In the following section, we offer a sketch of the first steps towards a general solution, which will be instantiated, in the following chapters, in two case studies.

3.3 A General Approach

In this section, we present our general approach to study collective behavior in many-to-many networks, which is depicted in Figure 3.3. Starting with a phenomenon to be studied in a given system, we assume the existence of a sequence of timestamped user actions covering a period of interest and gathered from that system. These actions represent expressions of user behavior that are fundamentally related to the phenomenon that will be studied (e.g., comments posted on a social media application, votes during a voting session).

In general terms, we propose to divide the period of interest into adjacent, non-overlapping and fixed-duration windows (snapshots). For each such snapshot, we first identify co-interactions from the set of user actions. This is done by grouping together actions that coincide from the perspective of the phenomenon under investigation. By coincident actions we mean that, collectively, they represent the same perspective from the study of the target phenomenon (e.g., the same ideology in a study of ideological groups). We then build a many-to-many network by projecting the co-interactions

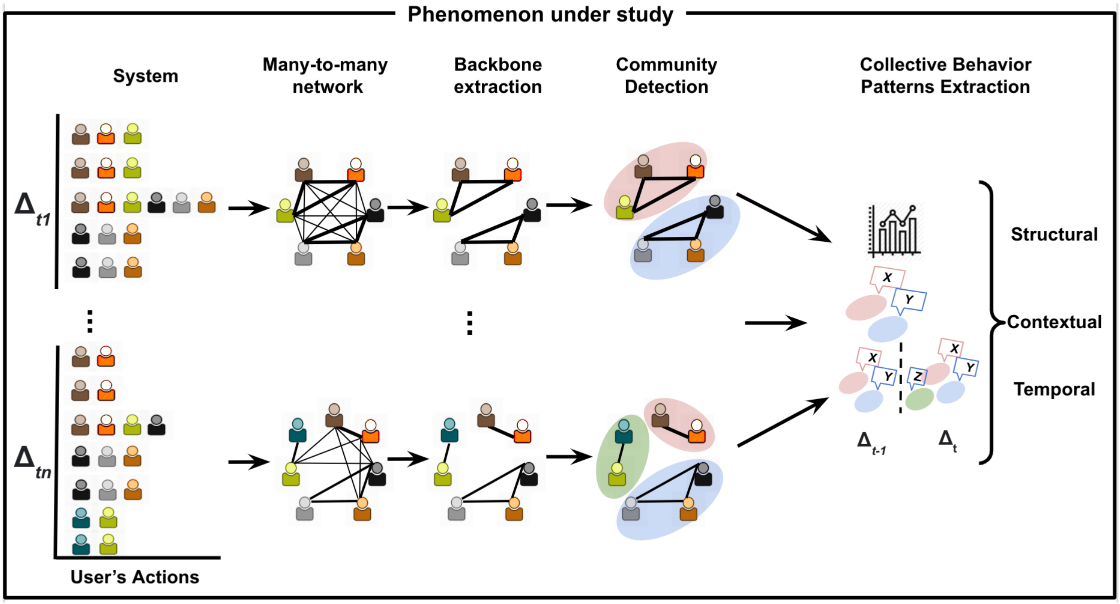


Figure 3.3. Overall solution for modeling and analyzing collective behavior.

into a weighted graph where weights express the number of times two individuals participated in a co-interaction (or any other contextual metric) during the respective time window. Then, we must extract the backbones of these networks by identifying the subset of edges that are salient with respect to the phenomenon under investigation. For that purpose, we may either rely on an existing algorithm or propose a new approach, exploiting singularities and constraints of the target problem and system. Yet, we note the possibility of using more than one method for the target phenomenon in the backbone extraction step. However, as we mentioned in the last section, a particular challenge in this step is to select and evaluate a set of methods to choose the one that best captures the phenomenon. Above all, this requires knowledge and categorization of the assumptions and properties of a set of methods under consideration. Therefore, we address this challenge later, in particular in Chapter 6, where we present a methodology for applying and evaluating a set of methods compatible with the same phenomenon.

Next, a community detection algorithm should be employed to uncover groups of individuals representing different collective behavior patterns influencing the system. We then aim at analyzing such communities, focusing on topological (community structure), contextual (system-related community attributes) and temporal (community dynamics) properties, aiming at uncovering fundamental knowledge about the target phenomenon.

Formally speaking, we propose to model the system as follows. Given a sequence $T = (\Delta_{t1}, \Delta_{t2}, ..., \Delta_{tn})$ of non-overlapping time windows of fixed duration, consider $I_{\Delta_t} = \{i_1, i_2, ..., i_j\}$ a set of individuals who interact with the system and among themselves during a given time window Δ_t, collectively driving the dynamics of the target phenomenon during that period. We consider that the system may define, explicitly or not, *opportunities* for individuals to interact, expressing their interests and personal goals with respect to that particular *opportunity* and, at the same time, the system may also impose restrictions on which interactions may occur at any given time (or in response to specific opportunities). Specifically, we define a set $O_{\Delta_t} = \{o^1_{\Delta_t}, o^2_{\Delta_t}, ..., o^m_{\Delta_t}\}$ of opportunities for individuals to interact during a given time window Δ_t. A group of individuals who interact in response to a given *opportunity* $o^k_{\Delta_t}$ is said to form a *co-interaction*. In this fashion, the panorama between individuals and opportunities in a given system is tied as two or more individuals in I_{Δ_t} choose to co-interact, driven by a particular opportunity in O_{Δ_t}, during a time window in $\Delta_t \in T$.

For instance, consider the study of online discussions in social media applications such as YouTube and Instagram. In that case, a user who shares content on a given topic (e.g., a user who shares a post on Instagram or a video on YouTube) may trigger comments from others, starting a thread of discussion, which is the object of investigation. Thus, the content initially posted opened an *opportunity* for users to comment on the topic, co-interacting with each other. In this particular case, the original post is the *opportunity*, whereas a co-interaction is said to occur among those users who, attracted by the original post, choose to comment on it. Note that users act individually by commenting on a given post; the co-interactions happen as multiple users choose to interact with each other by commenting on the same post. Another example, consider now the case of a study based on a co-voting network. In that case, co-interactions occur during voting sessions when congressmen express their votes with respect to pre-defined bills. In such case, the voting sessions are the *opportunities* whereas co-interactions occur as different congressmen vote similarly in the same session.

As defined, a co-interaction represents multiple individuals simultaneously interacting with the system (and with each other) in response to a given opportunity, taking actions that impact the system and, thus, reflect on the phenomenon. Although a given co-interaction is driven by a single *opportunity*, it should be noticed that an *opportunity* may generate multiple co-interactions. For instance, during a voting session (*opportunity*), there might be two co-interactions, one among congressmen who voted in favor of the specific bill being analyzed (*yes*), another among those who voted against it (*no*).

Moreover, *opportunities* may represent particular points in time when co-interactions can occur (e.g., voting sessions), or alternatively, specific events (e.g., a user post in a social media application) that drive users to co-interact, though such co-interactions may occur at any time after the *opportunity* happens. In the latter case, we may or may not define a limit on the user interactions following a particular *opportunity* that form a co-interaction (e.g., a maximum time period or even a maximum number of users participating in a co-interaction), depending on the phenomenon under investigation. For example, in a social media application, a given post may continue receiving new comments even after the end of the time window during which the post was shared. Still, these comments may be included in the co-interaction triggered by that post. In a co-voting network, in turn, a particular voting session restricts congressmen to take a position immediately. Similarly, the number of interactions by the same user (e.g., number of comments by the same user) in response to the same *opportunity* may or may not be explicitly considered, depending on the study.

Given the above description, we can see that each opportunity $o^k_{\Delta_t}$ has an associated set of co-interactions $C(o^k_{\Delta_t})$ that occurred in response to it. Each co-interaction c in set $C(o^k_{\Delta_t})$, in turn, is a set of users who participated in the co-interaction. In other words, $C(o^k_{\Delta_t}) = c^1_{k,\Delta_t}, c^2_{k,\Delta_t}, ...c^q_{k,\Delta_t}$ such that $c^j_{k,\Delta_t} \subseteq I_{\Delta_t}$. We also define the set of all co-interactions associated with time window Δ_t, C_{Δ_t}, as all co-interactions associated with opportunities that occurred during Δ_t. In other words:

$$C_{\Delta_t} = \cup\, C(o^k_{\Delta_t}),\ \forall o^k_{\Delta_t} \in O_{\Delta_t} \tag{3.1}$$

Given a set of co-interactions C_{Δ_t}, we build a many-to-many network model for time window Δ_t by projecting such co-interactions into an undirected and weighted graph $G_{\Delta_t} = (V, E)$ such that:

- V: is the set of vertices represented by I_{Δ_t}. In other words, all individuals who co-interacted in the system during a time window Δ_t.

- E: is the set of undirected and weighted edges, such that the weight of edge e_{i_1,i_2} connecting two individuals $i_1, i_2 \in I_{\Delta_t}$ is $\gamma_{\Delta_t}(i_1, i_2) = f(C_{\Delta_t}, i_1, i_2)$, where $f(C_{\Delta_t}, i_1, i_2)$ may be any aggregation function (e.g, count) that takes into account the co-interactions between i_1 and i_2 that happened during window Δ_t and/or any contextual information associated to them available during that period Δ_t[2].

Having defined the graph G_{Δ_t}, we are interested in extracting its backbone, $B_{\Delta_t} = (V_b, E_b)$, such that $V_b \subseteq V$ and $E_b \subseteq E$. As discussed in Section 3.2.2, choosing the best algorithm to extract B_{Δ_t} is a challenge as characteristics of the system and phenomenon under study may impose constraints on how co-interactions among individuals occur.

[2]As an example of contextual information being explored to build the graph G_{Δ_t}, the authors of [Pacheco et al., 2020, 2021] used several contextual metrics to connect nodes representing Twitter users, aiming at uncovering coordinated behavior in Online Social Networks.

For example, it is possible that an individual $i \in I_{\Delta_t}$ may not be particularly interested in an *opportunity* $o^k_{\Delta_t} \in O_{\Delta_t}$. Yet, a stronger constraint forces i to take part in a co-interaction around this opportunity. For example, a congressman feels compelled to voice her vote in a voting session due to her duty to the constituents, rather than by a particular interest in the bill being voted. Similarly, it might be the case that a particular user cannot, by system constraints, react to a particular *opportunity* o. For example, a congressman cannot vote in a voting session if she does not attend it. Equivalently, in a social media application, a user may not be able to comment on a post if the post is not visible to her (e.g., she does not follow the post's author). In other words, for any given system and phenomenon of interest, there may be a number of factors that impact the possibilities for co-interactions among individuals. Consequently, such specificities should be reflected in different edge weight probability distribution in G_{Δ_t}. Thus, not only structural features but also contextual features can be used to assist in extracting the backbone.

From the extracted backbone, a community detection algorithm should be applied in B_{Δ_t} to reveal a set P_{Δ_t} of communities (partitions) during a time window Δ_t. The communities unveiled should then be analyzed with respect to their structural and contextual properties, as well as with respect to their temporal dynamics. Lastly, we can define universal sets for the components considered during the period analyzed as: $\mathcal{I} = \{I_{\Delta t_1} \cup I_{\Delta t_2} \cup ... \cup I_{\Delta t_n}\}$ the universal set of *individuals*; $\mathcal{O} = \{O_{\Delta t_1} \cup O_{\Delta t_1} \cup ... \cup O_{\Delta t_n}\}$ the universal set of *opportunities*; $\mathcal{C} = \{C_{\Delta t_1} \cup C_{\Delta t_2} \cup ... \cup C_{\Delta t_n}\}$ the universal set of *co-interactions*; $\mathcal{G} = \{G_1, G_2, ..., G_{\Delta t_n}\}$ and $\mathcal{B} = \{B_1, B_2, ..., B_{\Delta t_n}\}$ the universal sets of *networks* and backbones, respectively; and $\mathcal{P} = \{P_1, P_2, ..., P_{\Delta t_n}\}$ the universal set for partitions;

After formalizing our general approach, we present two case studies of interest in Chapters 4 and 5. Next, we dive into the backbone extraction step of such approach by proposing a methodology for selecting and evaluating different backbone strategies in Chapter 6.

3.4 Summary

In this chapter, we have set out the preliminaries and fundamental concepts for the work developed in this dissertation. We started by revisiting our problem statement. We then discussed several key challenges associated with the target problem. In particular, we discussed challenges related to: (1) how we filter out noise from the network by defining and identifying only salient edges with respect to the problem under investigation; (2) how the extraction of these salient edges are operationalized in different backbone extraction algorithms and issues that must be considered when selecting one approach; (3) alternative definitions of *community* – a graph concept representing collective behavior – with possible implications on the study; and (4) how we extract relevant patterns associated with the identified communities, notably temporal patterns. We then proceeded to introduce a general solution that forms the skeleton of this dissertation. In the next two chapters, we use it to examine two case studies.

Chapter 4

Ideological Groups in Co-voting Networks

In this chapter, we present our first case study tackling the modeling of ideological groups in co-voting networks. It is organized as follows: Section 4.1 presents a brief contextualization as well as the research questions related to the phenomenon here studied; Section 4.2 describes our methodology according to the aspects described in Chapter 3; Sections 4.3-4.6 show the results while Section 4.7 discusses our findings on the phenomenon in question; Finally, the Section 4.8 summarizes the contributions and implications obtained.

4.1 Contextualization

Party systems can be classified with respect to fragmentation and polarization [Sartori, 2005]. Fragmentation corresponds to the number of parties existing in a political system (e.g., a country), while polarization is related to the multiple opinions that lead to the division of congressmen into groups with distinct political ideologies [Sartori, 2005; Fiorina and Abrams, 2008]. In countries where the party system has low fragmentation, the polarization of political parties can be more clearly observed since one party tends to occupy most seats supporting the government than the others opposing it [Mann and Ornstein, 2016]. Conversely, in fragmented systems, the many political parties often create coalitions, a inter-party alliance, to raise their influence in the political system and reach a common end [Ames, 2001; Budge and Laver, 2016]. Thus, a great deal of ideological similarity, as expressed by voting decisions, is often observed across different parties.

Previous work has analyzed the ideological behavior of political party congressmen by the modeling of voting data in signed and weighted networks [Andris et al., 2015; Cherepnalkoski et al., 2016; Arinik et al., 2017; Levorato and Frota, 2017; Mendonça et al., 2017] These prior efforts tackled topics such as community detection, party cohesion and loyalty analysis, governance of a political party and congressman influence in such networks. Yet, the identification and characterization of ideological communities, particularly in fragmented party systems, require observing some issues, such as: (i) presidents may define coalitions to strengthen the implementation of desired public policies, which may be ruptured after some time [Mainwaring and Shugart, 1997; Budge and Laver, 2016]; (ii) political members have different levels of partisanship and loyalty, and their political preferences may change over time [Baldassarri and Gelman, 2008; Andris et al., 2015]; and (iii) different parties may have the same political ideology, being redundant under a party system [de Melo, 2015].

In such context, we focus on the *phenomenon* related to the formation of ideological groups in political systems that go beyond pre-existing party boundaries. We study the dynamic behavior of political party members aiming at identifying how ideological communities are created and evolve over time and how individual members change their ideological behavior with respect to others. For the sake of comparison, we consider two very diverse political *systems*: the House of Representative of Brazil, which is characterized as a highly fragmented political system (i.e., several political parties occupying the seats) [de Melo, 2015], and the House of Representative of the United States, which is mostly dominated by two main parties – Republicans and Democrats [Dal Maso et al., 2015].

Using public voting data of the House of Representatives of both countries, covering a 15-year period, we model and characterize the emergence and evolution of communities of House members with similar political ideology and ideological changes of individual members over time. We study group and individual ideological behavior, as captured by their voting decisions, aiming at tackling four research questions (RQs):

- **RQ1: How are ideological communities in governments with different (fragmented and non-fragmented) party systems characterized?** We model the voting behavior of each House member during a given time period using a network, where nodes represent congressmen of the same House of Representatives, and weighted edges are added if two congressmen voted similarly. We use the Louvain algorithm [Blondel et al., 2008] to detect communities in each network and characterize structural properties of such communities. Unlike the aforementioned prior analyses in the political domain, we compare the properties of these communities in fragmented and in non-fragmented party systems.

- **RQ2: How can we identify polarization in the ideological communities?** We use neighborhood overlap to estimate the tie strength associated to each network edge, characterizing it as either strong or weak [David and Jon, 2010]. This approach to estimate tie strength has been employed in several contexts [Granovetter, 1973; Jones et al., 2013; McGee et al., 2013; Wiese et al., 2015] and to a short extent in the political domain [Waugh et al., 2009]. But unlike prior studies, we use strong ties to identify polarized communities in each network, comparing distinct political systems with respect to polarization.

- **RQ3: How do polarized communities evolve over time?** We analyze how polarized communities evolve over the years of a government, characterizing how the membership of such communities change over time.

- **RQ4: How can we assess the ideological changes of individual House congressmen over time?** We capture ideological changes, as expressed by members' voting behavior, by mapping the network into a temporal latent ideological space. Building upon a recent work [Yao et al., 2018], we learn temporal *vertice embeddings* for consecutive networks (representing consecutive years) jointly, so that we can track individual congressmen over time in the defined space. By doing so, we are able to analyze how the locations of individual congressmen in this space change and thus measure ideological shifts over time. Unlike prior studies that use contextual information (e.g., topics of voting sessions [Nguyen et al., 2015; Kornilova et al., 2018], prior speeches of individual congressmen [Lauderdale and Herzog, 2016; Sakamoto and Takikawa, 2017; Eidelman et al., 2018]) to build an ideological space, we use only the topological structure of the networks (which come from the voting data itself) to build such space, being thus a more general approach.

4.2 Methodology

This section describes the methodology used in our study, starting with basic concepts (Section 4.2.1) and our case studies (Section 4.2.2). We then present our modeling of voting behavior (Section 4.2.3) and the time-aware node embedding approach used to model an ideological latent space (Section 4.2.6.1).

4.2.1 Basic Concepts

The House of Representatives is composed of several congressmen who occupy the seats during each government period. House members participate in a series of voting sessions, when bills, amendments, and propositions are discussed and voted. Thus, attending such sessions is the most direct way for congressmen to express their ideologies and opinions. When these congressmen are associated with a large number of political parties, the party system in question is regarded as fragmented. In this case, during a term of office, coalitions are often established, leading political parties to organize themselves into ideological communities, defending together common interests during voting sessions [Mainwaring and Shugart, 1997; Sartori, 2005].

One can evaluate the behavior of parties and their congressmen in terms of how cohesive they are as an ideological community by analyzing voting data using widely disseminated metrics, such as Rice's Index [Rice, 1925]. Yet, the use of Rice Index has been shown to be problematic when there are more than two voting options (other than only *yes* and *no*) [Hix et al., 2005], as, for example, in the European parliament and in our study, as we will see.

Instead, we here employ the *Partisan Discipline* and *Party Discipline* metrics, proposed in [de Melo, 2015]. The former captures the ideological alignment of a member to her party (estimated by the behavior of the majority), and the latter expresses the ideological cohesiveness of a party. Given a member m, belonging to party p_m, the *Partisan Discipline* of m, pd_m, is given by the fraction of all voting sessions to which m attended and voted similarly to the majority of p_m's members. That is, let n be the number of voting sessions attended by m and $I(m, p_m, i)$ be 1 if m voted similarly to the majority of congressmen of p_m in voting session i ($i = 1..n$) and 0 otherwise. Then,

$$pd_m = \frac{\sum_{i=1}^{n} I(m, p_m, i)}{n} \tag{4.1}$$

We note that pd_m ranges from 0 to 1, where 1 indicates that member m voted similarly to the majority of p_m's congressmen in all voting sessions, and 0 indicates the opposite behavior. We note also that the *Partisan Discipline* can be generalized to assess the discipline and ideological alignment of a member to any community (not only his original party).

The *Party Discipline* of a party p is computed as the average *Partisan Discipline* of all of its members, that is,

$$PD(p) = \frac{\sum_{m=1}^{M} pd_m}{M} \tag{4.2}$$

where M is the number of congressmen of p. *Party Discipline* captures how cohesive a party (or community) is in a set of votes. That is, a $PD(p)$ value of 1 (maximum) indicates that party p is totally disciplined (or cohesive).

4.2.2 Case Studies

We consider two case studies: Brazil and the United States (US). In Brazil, the House of Representatives consists of 513 seats. A member vote can be either *Yes*, *No*, *Obstruction* or *Absence* in each voting session. A *Yes* or *No* vote expresses, respectively, an agreement or disagreement with the given proposition. Both *Absence* and *Obstruction* mean that the member did not participate in the voting, although an *Obstruction* expresses the intention of the member to cause the voting session to be canceled, for instance, due to insufficient quorum. Similarly, the US House of Representatives includes 435 seats, and a member vote can be *Yes*, *No* or *Not Voting*, whereas the last one indicates the member was not present in the voting session. In our study, we disregard *Absence* and *Not Voting* votes, as they do not reflect any particular inclination of the congressmen with respect to the topic under consideration. However, we do include *Obstructions* as they reflect an intentional action of the congressmen and a clear opposition to the topic. Thus, for Brazil, three different voting options were considered.

For both case studies, we collected voting data from public sources. The plenary roll call votes of Brazil's House of Representatives are available through an application programming interface (API) maintained by the government[1]. We collected roll call votes from 2003 to 2017 (4 legislatures). US voting data covering the same 15-year period (i.e., between the 108^{th} and 115^{th} congresses) was collected through the ProPublica API[2]. Each dataset consists of a sequence of voting sessions; for each session, the dataset includes date, time and voting option of each participating member.

In a preliminary analysis of the datasets, we noted that some congressmen had little attendance to the voting sessions, especially in Brazil. Thus, we chose to filter our datasets to remove congressmen with low attendance as they introduce noise to our analyses. Specifically, we removed congressmen that had not attended (thus had not associated vote) to more than 33% of the voting sessions during each year[3]. On

[1] `http://www2.camara.leg.br/transparencia/dados-abertos/dados-abertos-legislativo` (in Portuguese).

[2] `https://projects.propublica.org/api-docs/congress-api/`

[3] This threshold was chosen based on Article 55 of the Brazilian Constitution that establishes that a deputy or senator will lose her mandate if she does not attend more than one third of the sessions.

average, 19% and 1.98% congressmen were removed from the Brazilian and US datasets for each year, respectively.

Table 4.1. Datasets overview (PD: party discipline, SD: st. dev.)

Brazil (52^{nd} to 55^{th} legislatures)							
Year	**President (Party)**	**# of Sessions**	**# of Votes**	**# of Parties**	**# of Members**	**Avg. PD(%)**	**SD PD**
2003	Lula (PT)	150	106755	23	435	88.23	0.08
2004	Lula (PT)	118	71576	23	377	87.43	0.08
2005	Lula (PT)	81	50616	24	382	88.91	0.07
2006	Lula (PT)	87	62358	24	419	91.12	0.05
2007	Lula (PT)	221	190424	31	478	92.45	0.07
2008	Lula (PT)	157	122482	31	452	92.34	0.07
2009	Lula (PT)	156	125759	30	465	91.87	0.06
2010	Lula (PT)	83	63255	29	452	92.46	0.05
2011	Dilma (PT)	98	78662	29	481	89.34	0.08
2012	Dilma (PT)	79	60219	28	454	89.56	0.05
2013	Dilma (PT)	158	115751	29	451	88.70	0.06
2014	Dilma (PT)	87	66154	28	451	92.93	0.04
2015	Dilma (PT)	273	231031	28	502	85.84	0.06
2016	Dilma (PT) Temer (PMDB)	218	156006	28	452	90.12	0.05
2017	Temer (PMDB)	230	159704	29	435	89.76	0.08
United States (108^{th} to 115^{th} congresses)							
Year	**President (Party)**	**# of Sessions**	**# of Votes**	**# of Parties**	**# of Members**	**Avg. PD(%)**	**SD PD**
2003	Bush (R)	623	258867	3	432	95.76	0.03
2004	Bush (R)	502	203557	3	427	95.11	0.03
2005	Bush (R)	637	264735	3	432	95.02	0.03
2006	Bush (R)	511	210592	3	428	94.98	0.04
2007	Bush (R)	956	297957	2	414	92.23	0.04
2008	Bush (R)	605	244734	2	426	92.73	0.04
2009	Obama (D)	929	385344	3	431	93.78	0.02
2010	Obama (D)	631	253296	3	422	95.34	0.01
2011	Obama (D)	908	377601	2	428	91.98	0.01
2012	Obama (D)	621	253812	2	425	91.50	0.01
2013	Obama (D)	594	245430	2	427	93.04	0.01
2014	Obama (D)	531	217822	2	426	93.24	0.01
2015	Obama (D)	662	277732	2	432	94.87	0.01
2016	Obama (D)	588	241263	2	427	95.11	0.01
2017	Trump (R)	708	292503	2	427	95.99	0.00

Table 4.1 shows an overview of both (filtered) datasets, with Brazil on the top part of the table and the US on the bottom. The table presents, for each year, the acting president[4] and his/her party[5], total number of voting sessions, total number of member votes, as well as numbers of parties and congressmen occupying seats in the House of Representatives during the year. The two rightmost columns, *Avg. PD* and *SD PD*, present the average and standard deviation of the *Party Discipline* computed

[4]Brazilian president Dilma Rousseff was impeached from Office in 2016 and, therefore, Brazil had two Presidents that year.

[5]For Brazil: Worker's Party (PT) and Democratic Movement Party (PMDB). For the US: Democratic (D) and Republican (R).

across all parties. We show data for different Brazilian legislatures and US congresses in separate blocks of the table.

Starting with the Brazilian dataset, we can see that the number of parties occupying seats has somewhat grown in recent years, characterizing an increasingly fragmented party system. Yet, in general, average *PD* values are very high (ranging from 85% to 92%), with small variation across parties, indicating that, despite the fragmentation, most party congressmen have high partisan discipline. Regarding the American dataset, Table 4.1 shows that the number of voting sessions is much larger than in Brazil. This is because the API of the Brazilian House of Representative provides only data related to votes in plenary, while the US dataset covers all votes. Moreover, although the numbers of congressmen are comparable to those in the Brazilian dataset, the number of parties occupying seats in each year is much smaller. Indeed, only two parties, namely Republican (R) and Democrat (D), fill all available seats since the 112$^{\text{th}}$ Congress. Thus, unlike the Brazilian case, party fragmentation is not an issue in the US system. Nevertheless, parties have a high party discipline in both systems.

4.2.3 Network Model

We model the dynamics of ideological communities in voting sessions in each country using graphs as follows. We discretize time into non-overlapping windows of fixed duration. Since in Brazil, government coalitions are usually made every year, we choose one year as the time window for analyzing community dynamics. Thus, we build a set $T_{BR} = \{\Delta_{t1}, \Delta_{t2}, ..., \Delta_{t15}\}$ and the other for the United States $T_{US} = \{\Delta_{t1}, \Delta_{t2}, ..., \Delta_{t15}\}$.

We define *individuals* as congressmen that had seats in their respective House of Representatives during the time window Δ_t. We then define the sets of *individuals* for each time window Δ_t as $I_{BR,\Delta_t} = \{i_1, i_2, ..., i_{j1}\}$ for Brazil and $I_{US,\Delta_t} = \{i_1, i_2, ..., i_{j2}\}$ for the United States, where $j1$ and $j2$ are the total number of congressmen in the House for Brazil and the United States, respectively, during window Δ_t. The *opportunities* whose those congressmen may interact are defined as voting sessions that take place during a given time window Δ_t (i.e., during a year). Thus, we define the sets $O_{BR,\Delta_t} = \{o^1_{\Delta_t}, o^2_{\Delta_t}, ..., o^{m1}_{\Delta_t}\}$ and $O_{US,\Delta_t} = \{o^1_{\Delta_t}, o^2_{\Delta_t}, ..., o^{m2}_{\Delta_t}\}$ formed by all the voting sessions that were opened, respectively, in Brazil and in the United States during window Δ_t. Notice that the numbers of voting sessions $m1$ and $m2$ may be different.

Having defined that, a co-interaction c built around opportunity $o^k_{\Delta_t} \in O_{BR,\Delta_t}$ (i.e., $c \in C(o^k_{\Delta_t})$) is formed by Brazilian congressmen who took the same position (*Yes*, *No* or *Obstruction*) at the House of representative of Brazil. The collection of co-interactions that took place during window Δ_t in Brazil is referred to as C_{BR,Δ_t}. Similarly, we define a co-interaction $c \in C(o^k_{\Delta_t})$, for $o^k_{\Delta_t} \in O_{US,\Delta_t}$, as a set of American congressmen who voted alike in the k^{th} voting sessions of the American House of Representatives. Also we define set C_{US,Δ_t} of all co-interactions that occurred during all opportunities in O_{US,Δ_t}.

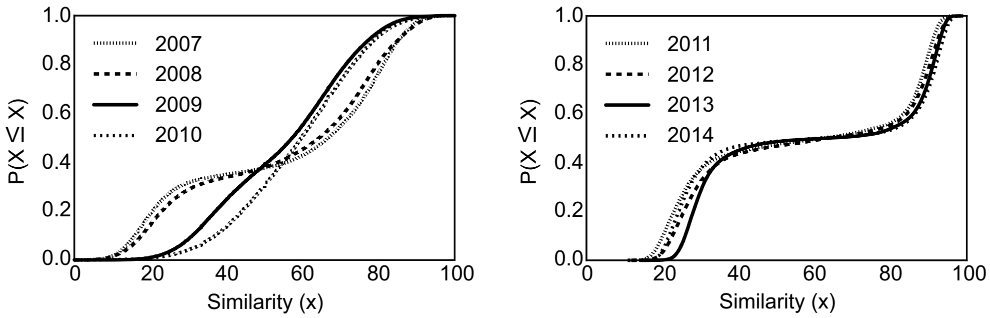
We then move towards the network modeling. For each scenario under study (Brazil and the United States) and for each time window Δ_t in T_{BR} and T_{US} we build an undirected and weighted graph. Recall that our goal is to analyze the formation of ideological groups, that is, congressmen who have the same ideological alignment during the analyzed voting sessions. Thus, for each scenario s ($s = BR, US$) and time window Δ_T we build a co-voting network G_{s,Δ_t} as follows:

- V_{s,Δ_t}: is the set of vertices represented by all congressmen. That is, all individuals in I_{s,Δ_t} who participated in at least one co-interaction in $C_{\cdot,\Delta_t}$.
- E_{s,Δ_t}: is the set of undirected and weighted edges, such that the weight of edge e_{i_1,i_2} linking two congressmen i_1 and i_2 is $\gamma_{\Delta_t}(i_1, i_2) = sim(i_1, i_2)$, where $sim(i_1, i_2)$ is given by the ratio of the number of sessions in which both congressmen voted similarly to the total number of sessions in O_{s,Δ_t} to which both congressmen attended.

As a result, we have the following sets of networks for Brazil and the US: $\mathcal{G}_{\mathcal{BR}} = \{G_{BR,\Delta_{t_1}}, G_{BR,\Delta_{t_2}}, \ldots, G_{BR,\Delta_{t15}}\}$ and $\mathcal{G}_{\mathcal{US}} = \{G_{US,\Delta_{t_1}}, G_{US,\Delta_{t_2}}, \ldots, G_{US,\Delta_{t15}}\}$. Given such networks, our goal is to infer patterns of common (collective) behavior in terms of voting choices that extrapolate the traditional boundaries political parties, thus revealing ideological similarities among congressmen. Based on these patterns, we intend to study how ideological groups are formed and evolve over time in the two – very different – political systems analyzed.

4.2.4 Network Backbone Extraction

After building each graph sequence, we noted that all pairs of congressmen voted similarly at least once in all years analyzed and in both countries, and therefore all graphs built are complete and do not allow that patterns be extracted. This reflects the fact that some voting sessions are not discriminative of ideology or opinion, as most congressmen (regardless of party) voted similarly. For instance, it is possible that some voting sessions target general-interest or humanitarian causes. As a consequence, many different congressmen, despite having different ideological beliefs and behaviors,



(a) Brazil **(b)** United States

Figure 4.1. Cumulative Distribution Function of Edge Similarity.

may still take a similar voting position with respect to the particular bill being voted, adding edges to the co-voting network. These edges represent sporadic interactions and are driven by a particular topic of large agreement, and thus do not necessarily reflect an ideological alignment. Thereby, it is necessary to filter out edges that do not contribute to the detection of ideological communities. By doing so, we intend to reveal the *salient* edges to identify the backbone B_{s,Δ_t} for each time graph G_{s,Δ_t}.

Intending to remove such edges and extract the backbone, we employ the *threshold-based* approach. To define the right threshold for our problem, we follow previous works that use *contextual information* to address this point [Perkins and Langston, 2009; Bordier et al., 2017; Yan et al., 2018]. We begin by analyzing the distributions of edge similarity for all the networks that capture agreement between voting congressmen. Representative distributions for specific years are shown in Figures 4.1a and 4.1b for Brazil and US, respectively.

We note that the distributions for the U.S. show clear concentrations around very small (roughly 30%) and very large (around 85%) similarity values, while the distributions for Brazil exhibit greater variability, which is consistent with the greater fragmentation of the party system. Aiming at filtering out non salient edges from the networks, in line with the phenomenon, we argue that the threshold should not be much smaller than the average partisan discipline of the individual members. That is, two congressmen that have a much lower similarity (ideological agreement) than their partisan disciplines observed in the political system in question should not be considered part of the same ideological community. Therefore, these interactions can be taken out as they do not discriminate ideology, as mentioned earlier.

Based on these assumptions, we chose to remove all edges with weights below the 90$^{\text{th}}$ percentile of the similarity distribution for the Brazilian graphs. For the US, we removed edges with weights below the 55$^{\text{th}}$ percentile of the similarity distribution.

Both percentiles correspond roughly to a similarity value of 80%, which is not much smaller than the average partisan disciplines in both countries (see Table 4.1). We removed nodes that become isolated after the edge filtering. We found that these thresholds yield a good trade-off between removing less discriminative connections and graph sparsity. Specifically, the fraction of nodes and edges removed from the Brazilian networks fall in the 0-24% and 86-93% ranges, respectively, across all years analyzed. For the US, the fractions are much lower, varying from 0% to 11% for nodes and from 54% to 56% for edges. Therefore, we have the following sets of backbone for Brazil and the U.S., respectively, $\mathcal{B_{BR}} = \{B_{BR,\Delta_{t_1}}, B_{BR,\Delta_{t_2}}, \ldots, B_{BR,\Delta_{t15}}\}$ and $\mathcal{B_{US}} = \{B_{US,\Delta_{t_1}}, B_{US,\Delta_{t_2}}, \ldots, B_{US,\Delta_{t15}}\}$.

4.2.5 Community Detection

After extracting the backbones, we use the Louvain algorithm [Blondel et al., 2008] to identify ideological communities P_{s,Δ_t} in each backbone B_{s,Δ_t}. The goal is to find communities formed by congressmen with closer ideology according to the edges that directly encode voting similarity between congressmen. As explained in Chapter 2, Louvain is based on the optimization of *modularity* [Newman, 2006], a metric to evaluate the structure of clusters in a network. *Modularity* is large when the clustering is good, with a maximum value of 1. Thus, we have for each scenario a set of communities $\mathcal{P}_s = \{P_1, P_2, ..., P_{\Delta t_n}\}$.

4.2.6 Community Characterization

Once the communities are extracted, we characterize them in terms of their structural, contextual, and temporal dynamics. For the first two, we use *modularity* along with *party discipline* as the main metrics to assess the cohesiveness of the communities found. Modularity captures the quality of the result in terms of the topological structure of the communities in the network, while *party discipline*, which is computed for the communities (rather than for individual parties), captures the quality in terms of context semantics.

In terms of temporal properties,, we compute complementary metrics, namely *persistence* and *normalized mutual information* (NMI) [Vinh et al., 2010; Wei and Carley, 2015], for each pair of consecutive years. We define the **persistence** of two consecutive years Δ_{t1} and Δ_{t2} as the fraction of all members of polarized communities in Δ_{t1} who remained in some polarized community in Δ_{t2}. A *persistence* equal to 100% implies that all members of polarized communities in Δ_{t1} remained in some polarized community in Δ_{t2}. Yet, the membership of individual communities may have changed as members switched communities. To assess the extent of change in community membership over consecutive years, we compute the normalized mutual information over the communities, taking only members who persisted over the two years.

NMI is based on Shannon entropy of information theory [Shannon, 2001]. Given two sets of partitions $X \in P_{s,\Delta_{t1}}$ and $Y \in P_{s,\Delta_{t2}}$[6], defining community assignments for nodes, the mutual information of X and Y can be thought as the informational "overlap" between X and Y, or how much we learn about Y from X (and about X from Y). Let $P(x)$ be the probability that a node picked at random is assigned to community x, and $P(x, y)$ the probability that a node picked at random is assigned to both x in X and y in Y. The NMI of X and Y is defined as:

$$NMI(X,Y) = \frac{\sum_x \sum_y P(x,y) \log \frac{P(x,y)}{P(x)P(y)}}{\sqrt{H(X)H(Y)}} \tag{4.3}$$

where $H(X) = -\sum_x P(x) \log P(x)$ is the Shannon entropy for X. NMI ranges from 0 (all members changed their communities) to 1 (all members remained in the same communities).

4.2.6.1 Ideological Space Model

In order to model how the ideological behavior of individual party congressmen evolves over time, we start from the networks defined in Section 4.2.3, which capture each individual's behavior in terms of how a member voted relative to others during a given time window. We then build a network representation that embeds vertices into a low-dimensional vector space, which preserves properties of the network's topological structure. Since the ideological behavior of an individual member is here captured by how she voted relatively to her peers (i.e., by her neighborhood in the network), we consider the low-dimensional latent space produced by the graph embedding technique an *ideological space*. One key challenge is how to track individual congressmen over time in this ideological space so as to identify changes in their behavior. This is difficult because there are multiple networks (and thus network embeddings), one for each time window under consideration. In this section, we describe our approach to address this challenge and build a consistent time-aware ideological space.

We build upon node2vec, a popular graph embedding technique [Grover and Leskovec, 2016]. Node2vec learns low-dimensional representations for vertices in

[6]For the sake of simplicity, we reduce the notation used so far to explain this metric.

a single graph by performing biased random walks, and using them as input to word2vec [Mikolov et al., 2013], a widely used word embedding technique. Word2vec receives as input a textual corpus and produces as output a vector space. Each word in the input corpus is mapped into a point in the vector space such that words that share common contexts in the input corpus fall close to each other in the vector space. In the case of node2vec, assuming that random walks are input sentences and visited vertices represent individual words, vertices are mapped into the low-dimensional latent space so as to maximize the likelihood of preserving the network neighborhoods. Grover *et al.* defined a flexible notion of neighborhood [Grover and Leskovec, 2016], which can be instantiated differently by carefully choosing the parameters of the biased random walk procedure (see more details below).

However, like word2vec, node2vec also suffers from the "alignment problem" when applied to a temporal sequence of networks. That is, the embeddings generated by the successive application of node2vec to networks for consecutive time windows are *not* mapped onto the same latent space. Thus, a vertex representation in one embedding has no correspondence to its representation in the next embedding (i.e., the one generated from the next time window).

Yao *et al.* tackled the problem of inferring how word semantics evolve over time by proposing a dynamic statistical model for learning time-aware word embeddings [Yao et al., 2018]. The proposed solution, which we refer to as DynamicWord2Vec, effectively addresses the "alignment problem" in the context of word embeddings. Inspired by that work, we build a temporally-consistent ideological space to represent parties and their political congressmen by adapting DynamicWord2Vec to the network domain, combining it with node2vec. That is, we modify the node2vec implementation so that it uses DynamicWord2Vec (instead of word2vec) to generate an embedding from the sampled walks. Next, we briefly review how node2vec works and how we combine it with DynamicWord2Vec. We refer the reader to [Grover and Leskovec, 2016; Yao et al., 2018] for further details on each technique.

Node2vec [Grover and Leskovec, 2016] uses a strategy of neighborhood sampling through a biased random walk which behaves, at each step, either as breadth-first sampling (BFS) or as depth-first sampling (DFS). In BFS, the neighborhood of a given source vertex v_s is restricted to vertices that are immediate neighbors of the source, while DFS consists of vertices sequentially sampled at increasing distances from v_s. We here want the walk to enforce BFS more often than DFS to better capture the similarities in the ideological space, rather than structural equivalences in the network [Grover and Leskovec, 2016]. To control this behavior, node2vec has two parameters, p and q. Parameter p determines the likelihood of immediately going back to an already visited vertex. Parameter q allows us to control whether the walk stays close to the source vertex, exploring the same neighborhood (i.e., corresponding to BFS), or whether it should walk further away, exploring other vertices (i.e., corresponding to DFS). We here are focused on the former, i.e., sampling immediate neighbors of the source more often. Thus, we set the parameter values according to the authors' recommendations for such case [Grover and Leskovec, 2016], i.e., $p = 1$ and $q = 0.5$. By doing so, we skew the random walks towards the immediate neighborhood of each source vertex.

In addition to p and q, node2vec allows us to define the number of walks per vertex and the length of each walk (i.e., number of vertices visited in each walk). These parameters directly determine the sampling process, which tends to saturate at a certain point as they increase [Grover and Leskovec, 2016]. In our experiments, we found that 16 walks per vertex, each with length 40, are sufficient to perform the sampling process in our case studies. Increasing either the number of walks or the length of each walk further caused a proportional increase in the co-occurrence of vertices in the walks, without bringing further information.

After computing the probabilities of the possible paths according to p and q and sampling the walks, node2vec builds a walk matrix S of size $k \times l$, where k is the product of number of walks and number of vertices and l is the length of each walk. S contains all vertices visited in all walks performed, starting from all vertices in the graph as sources. In the original node2vec algorithm, given a matrix S, the representations of the vertices are optimized using stochastic gradient descent so that vertices in the same neighborhood appear more closely in the generated latent space. Instead, we here use the DynamicWord2Vec technique as follows.

For the sake of simplicity and generalization, hereinafter we describe our approach independent of the considered scenarios while in Section 4.6 we instantiate it according to our case studies. We want to learn a single latent space covering Δ_t successive time windows, given any set of graphs $\mathcal{G} = \{G_{\Delta_{t_1}}, G_{\Delta_{t_2}}, \ldots, G_{\Delta_{t_n}}\}$[7] representing the networks produced for each windows Δ_t in T. Let $\mathcal{S} = \{S_{\Delta_{t_1}}, S_{\Delta_{t_2}}, \ldots, S_{\Delta_{t_n}}\}$ be the set of matrices generated by node2vec for each graph in $\mathcal{G}$, and $V = \{v_1, v_2, \ldots, v_j\}$ be the set of all vertices that appear at least once in any graph in $\mathcal{G}$.

Then, we use DynamicWord2Vec to observe the association of vertices over time according to the sampled walks, mapping them to a temporal ideological latent space. To do this, for each matrix S_{Δ_t} and each pair of vertices $v_1, v_2 \in V$ (representing two party congressmen in one of the case studies), we count: (1) the number $\#(v_1)$ of individual occurrences of v_1 in the walks represented by rows of S_{Δ_t} ; (2) the number

[7] In particular, we here use the backbones. But for generalization purposes, here we present it to any set of graphs.

$\#(v_2)$ of individual occurrences of v_2; and (3) the number of co-occurrences of vertices v_1 and v_2, restricted within a window of size L from v_1 (either before or after v_1), denoted as $\#(v_2, v_2)$. Typically, L is set between 5 to 10 as proposed in [Mikolov et al., 2013]. Here, we use L=5, resulting in a window containing 10 vertices in addition to the middle vertex. The degree of association between v_1 and v_2 is captured by the pointwise mutual information (PMI) [Levy and Goldberg, 2014], defined as a function of the empirical probabilities of occurrences of v_1, occurrences of v_2 and co-ocurrences of v_1 and v_2 in matrix S_{Δ_t}. Specifically, given $|S_{\Delta_t}| = k \times l$, the PMI matrix entry corresponding to (v_1, v_2) is given by:

$$\mathrm{PMI}(S_{\Delta_t}, L)_{v_1,v_2} = \log_2\left(\frac{\#(v_1,v_2)\cdot|S_{\Delta_t}|}{\#(v_1)\cdot\#(v_2)}\right), \ \forall v_1, v_2 \in V. \tag{4.4}$$

When v_1 and v_2 co-occur very frequently in the sampled walks, the corresponding PMI is high, indicating high proximity between them. On the other hand, when the argument inside $\log_2(.)$ is very small, PMI tends to take on negative values. According to [Levy and Goldberg, 2014; Yao et al., 2018], the pairs (v_1, v_2) with more representative association have PMI values greater than 1, that is, they co-occur more than twice in the walks sampled. Thus, considering only the positive values of PMI does not significantly affect the solution while providing better numerical stability to matrix factorization. Thus, given a walk matrix S_{Δ_t}, we define a positive PMI matrix, referred to as PPMI(S_{Δ_t}, L), whose entry for given two vertices v_1 and v_2 is defined as:

$$\mathrm{PPMI}(S_{\Delta_t}, L)_{v_1,v_2} = \max(\mathrm{PMI}(S_{\Delta_t}, L)_{v_1,v_2}, 0) \coloneqq Y(\Delta_t). \tag{4.5}$$

Given the PPMI matrix $Y(\Delta_t)$, DynamicWord2Vec learns the embedding vectors $\mathbf{u}_{v_1}$ and $\mathbf{u}_{v_2}$ for vertices v_1 and v_2, respectively, by applying a low-rank factorization such that, for any pair v_1 and v_2, $\mathbf{u}_{v_1}^\top \mathbf{u}_{v_2} \approx \mathrm{PPMI}(\Delta_t, L)_{v_1,v_2}$. Each $\mathbf{u}_{v_1}$ has length $d \ll |V|$. Thus, for each time window Δ_t, a temporal embedding $U(\Delta_t)$=$\{\mathbf{u}_{v_1}, \ldots, \mathbf{u}_{v_j}\}$ must satisfy $U(\Delta_t)U(\Delta_t)^\top \approx Y(\Delta_t)$.

This low-rank factorization is obtained by solving an optimization problem. Two regularization terms are added to the objective function in order to address, respectively, overfitting and alignment issues. To avoid overfitting, a typical penalty term based on the Frobenius norm[8] of each low-rank matrix $U(\Delta_t)$, such that, Δ_t in T is added [Davenport and Romberg, 2016]. To enforce alignment, a penalty term that assumes some smoothness between subsequent time windows is added. Also, this term is based on the Frobenius norm of the differences between matrices $U(\Delta_{t-1})$ and $U(\Delta_t)$

[8] The Frobenius Norm of a given matrix $M_{m\times n}$ is defined by: $\|M\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2}$.

for each Δ_t in T. The function to be minimized is

$$\min_{U(\Delta_{t_1}),\ldots,U(\Delta_{t_n})} \frac{1}{2}\sum_{n=1}^{|T|} \|(Y(\Delta_{t_n}) - U(\Delta_{t_n})U(\Delta_{t_n})^\top\|_F^2 + \frac{\lambda}{2}\sum_{n=1}^{|T|} \|U(\Delta_{t_n})\|_F^2 + \frac{\tau}{2}\sum_{n=2}^{|T|} \|U(\Delta_{t_{n-1}}) - U(\Delta_{t_n})\|_F^2, \quad (4.6)$$

where λ, $\tau > 0$. Observe that each embedding $U(\Delta_t)$ depends, indirectly, on all other Δ_{t-1} embeddings. The smoothing term $\|U(\Delta_{t-1}) - U(\Delta_t)\|_F^2$ enforces alignment across embeddings. Parameters λ and τ control the degree of the regularization and smoothness, respectively. Specifically, parameter τ controls the alignment of the embeddings for successive windows Δ_t: τ=0 implies no alignment, whereas $\tau \rightarrow \infty$ produces a static embedding with $U(\Delta_{t_1}) = U(\Delta_{t_2}) = \ldots = U(\Delta_{t_n})$. We discuss how to set parameters λ and τ in Section 4.6. In order to solve Equation (4.6), DynamicWord2Vec uses the block coordinate descent [Yu et al., 2012] obtaining a representation vector $\mathbf{u}_{v_1}(\Delta_t)$ for each vertex $v_1 \in V$ and for each time window Δ_t.

Given the embedding vectors, we can compute the change of a given member v_1 in the defined ideological space between two time windows Δ_{t_1} and Δ_{t_2} using a metric of distance between vectors. We here use the widely adopted cosine distance:

$$\cos(v_{1(\Delta_{t_1})}, v_{1(\Delta_{t_2})}) = 1 - \frac{\mathbf{u}_{v_1}(\Delta_{t_1}) \cdot \mathbf{u}_{v_1}(\Delta_{t_2})}{\|\mathbf{u}_{v_1}(\Delta_{t_1})\|\|\mathbf{u}_{v_1}(\Delta_{t_2})\|}. \quad (4.7)$$

Cosine distance ranges from 0 to 1. Values close to 0 indicate that the two vertices $\mathbf{u}_{v_1}(\Delta_{t_1})$ and $\mathbf{u}_{v_2}(\Delta_{t_2})$ coincide, i.e., the corresponding party member did not change ideologically between windows Δ_{t_1} and Δ_{t_2}. Values close to 1 indicate that the member drastically shifted his ideology within the period.

In the next four sections, we discuss the results of our analyses when tackling the research questions posed in Section 4.1.

4.3 Identifying Ideological Communities

We start by tackling our first research question (RQ1) and characterizing the ideological communities discovered in both Brazilian and US networks. Table 4.2 shows an overview of all networks for both countries, presenting some topological properties [David and Jon, 2010], i.e., numbers of vertices (*# of nodes*) and edges (*# of edges*), number of connected components (*# of CC*), average shortest path length (*SPL*), average degree, clustering coefficient and density[9]. The difference between the number of nodes in this table and the number of members in Table 4.1 corresponds to nodes that were removed after the edge filtering.

Table 4.2. Statistics of Networks and Ideological Communities (CC: connected components, SPL: shortest path length, Mod: modularity)

					Brazil						
Year	**# of Nodes**	**# of Edges**	**# of CC**	**Avg. SPL**	**Avg. Degree**	**Avg. Clust.**	**D**	**# of Comm.**	**Mod.**	**Avg. PD(%)**	**SD PD**
2003	342	9329	5	1.83	55.01	0.65	0.16	8	0.11	95.48	2.22
2004	326	7079	2	1.90	43.43	0.62	0.13	4	0.14	92.68	3.36
2005	359	7211	1	3.18	40.17	0.59	0.11	5	0.21	88.32	3.64
2006	419	8613	1	2.47	41.11	0.61	0.09	4	0.36	90.50	2.36
2007	427	11394	3	1.77	53.37	0.67	0.12	6	0.14	95.97	1.26
2008	400	10180	2	1.62	50.90	0.70	0.12	5	0.08	95.78	1.94
2009	434	10784	2	1.92	49.70	0.66	0.11	4	0.18	91.45	3.49
2010	446	10151	1	2.42	45.52	0.64	0.10	4	0.19	92.01	1.29
2011	408	11519	2	1.89	56.47	0.60	0.13	6	0.12	93.69	3.76
2012	345	6527	3	2.47	46.11	0.48	0.11	4	0.33	87.00	4.25
2013	449	10094	1	2.21	44.96	0.61	0.10	4	0.38	86.51	4.18
2014	450	10036	1	2.18	44.60	0.58	0.09	3	0.43	91.14	1.79
2015	490	12563	1	2.90	51.28	0.69	0.10	5	0.60	85.90	3.11
2016	425	10159	2	1.44	47.81	0.66	0.11	4	0.38	92.62	1.83
2017	396	9434	4	1.64	47.65	0.72	0.12	6	0.24	90.25	3.16

					United States						
Year	**# of Nodes**	**# of Edges**	**# of CC**	**Avg. SPL**	**Avg. Degree**	**Avg. Clust.**	**D**	**# of Comm.**	**Mod.**	**Avg. PD(%)**	**SD PD**
2003	431	41892	2	1.11	194.39	0.95	0.45	2	0.48	93.60	1.03
2004	426	40928	2	1.10	192.15	0.95	0.45	2	0.48	92.97	0.55
2005	431	41892	2	1.10	194.39	0.95	0.45	2	0.48	92.60	0.79
2006	426	41112	2	1.10	193.01	0.95	0.45	2	0.49	91.45	0.33
2007	414	38471	2	1.12	185.85	0.94	0.45	2	0.44	91.55	3.78
2008	424	40729	2	1.11	192.12	0.94	0.45	2	0.46	95.45	1.97
2009	429	41698	2	1.15	194.40	0.94	0.45	2	0.40	93.86	2.42
2010	420	39969	1	3.06	190.33	0.95	0.45	3	0.43	94.92	1.86
2011	426	41119	2	1.18	193.05	0.96	0.45	3	0.44	90.31	1.91
2012	417	40545	3	1.17	194.46	0.96	0.46	3	0.44	91.63	1.86
2013	423	40921	2	1.11	193.48	0.96	0.45	2	0.47	93.23	1.03
2014	418	40735	2	1.08	194.90	0.96	0.46	2	0.48	94.37	0.34
2015	427	41890	2	1.09	196.21	0.95	0.46	2	0.47	94.40	1.36
2016	423	40927	2	1.11	193.51	0.95	0.45	2	0.48	94.70	1.36
2017	423	40928	2	1.09	193.51	0.95	0.45	2	0.46	96.02	0.44

Table 4.2 also summarizes the characteristics of the ideological communities identified using the Louvain algorithm. In the four rightmost columns, it presents the number of communities identified, their *modularity* (*Mod.*) as well as average and standard deviation of the party discipline (*Avg PD* and *SD PD*), computed with respect to the ideological communities.

Starting with the Brazilian networks (top part of Table 4.2), we can observe great fluctuation in most topological metrics over the years, but, overall, the networks are

[9] The *density* of a network is the ratio of the total number of existing edges to the maximum possible number of edges in the graph. The *clustering coefficient* measures the degree at which nodes tend to group together to form triangles, and is defined as the ratio of the number of existing closed triplets to the total number of open and closed triplets. A triplet is three nodes that are connected by two (open triplet) or three (closed triplet) undirected ties.

sparse: the average shortest path length is long, the average clustering coefficient is moderate and the network density is low. Also, the number of communities identified is much smaller than the total number of parties (see Table 4.1) confirming the fragmentation and ideological overlap of multiple parties. Yet, the *party discipline* of these communities is, on average, very close to, and, in some cases, slightly larger than the values computed for the individual parties, despite a somewhat greater standard deviation observed across communities. Thus, these communities are indeed very cohesive in their voting patterns.

In contrast, the topological structure of the identified communities, as expressed by the *modularity* metric, is very weak, especially in the former years. That is, there is still a lot of similarity across members of different ideological communities. We note that in the former years the government had greater support from most parties, as their members voted similarly in most sessions. Such approval dropped during a period of political turmoil that started in 2012, when the distinction of ideologies and opinions become more clear[10,11]. This may explain why the *modularity* starts low and increases in the most recent years, when there is greater distinction between different communities. This occurs despite the large average party discipline maintained by the communities. Thus, these two metrics offer complementary interpretations of the political scenario.

Considering the greater variation in modularity, we take a step further and investigate a possible explanation for it. We note that, in the early years of our analysis, the Presidents had higher governance by observing the number of victories obtained in the voting sessions and assuming that the position of their respective party's congressmen represents the position of the government. In this way, we look at the majority positioning of the president's party in each voting session, obtaining the percentage of voting sessions whose final result is aligned with the government's positioning. In other words, we look at whether most of the congressmen followed the position of those congressmen that represent the government party.

By doing so, we compute the Pearson's correlation between the percentage of voting sessions where the governing party obtained the majority per year and compare to the modularity of communities for the respective year. The correlation between such metrics is -0.58 considering the 15 years analyzed. Such moderate negative correlation indicates that the higher the percentage of votes in the year in which the government obtained the majority, the lower the modularity and, therefore, the weaker is the community structure. If we disregard the last two years analyzed (2016-2017), where

[10] `https://www.vox.com/2016/4/21/11451210/dilma-rousseff-impeachment`
[11] `http://www.bbc.com/news/world-latin-america-19359111`

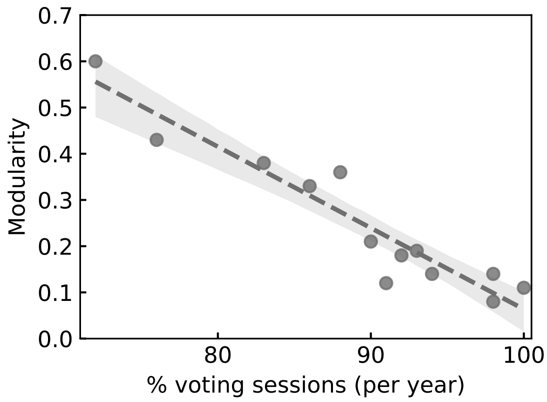
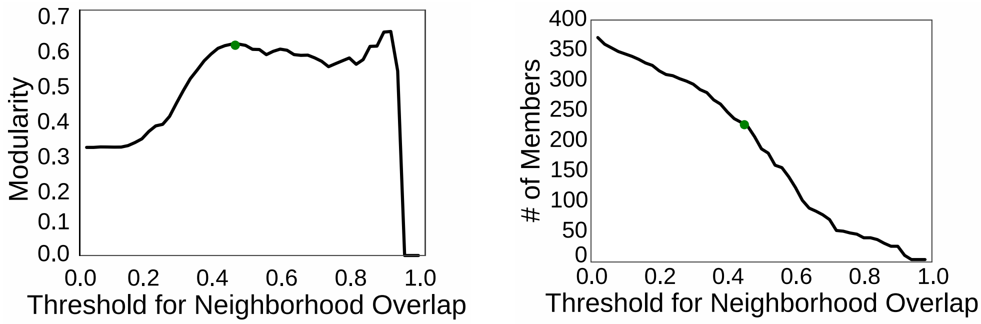


Figure 4.2. Correlation between the percentage voting sessions that the government obtained the majority and the the modularity obtained. Each point represents a year analyzed in Table 4.2 disregarding the period 2016-2017.

there was an impeachment process and the change of President that resulted in an even more obscure period in terms of support from his own party, the negative correlation increase to -0.94. The relation between both variables is presented in Figure 4.2. Indeed, it is possible to observe a strong influence between the positioning and influence of the government with the formation of such ideological groups, suggesting that high governability makes it difficult to identify distinct ideological groups for some periods.

Turning our attention to the US (bottom part of Table 4.2), we note that, unlike in Brazil, most metrics remain roughly stable throughout the years. The networks are much more dense, with higher average clustering (Avg. Clust.) coefficient and density and shortest path length. The number of identified communities coincides with the number of connected components as well as with the number of political parties (see Table 4.1) in most years. These communities are more strongly structured, despite some ideological overlap, as expressed by moderate-to-large *modularity* value. They are also consistent in their ideologies, as expressed by large party disciplines, comparable to the original (party-level) ones. These metrics reflect the political behavior of a non-fragmented and stronger two-party system, quite unlike the Brazilian scenario.

In sum, in Brazil, the several parties can be grouped into just a few ideological communities, with strong disciplined members, although the separation between communities is not very clear. In the US, on the other hand, ideological communities are more clearly defined, both structurally and ideologically, though some inter-community similarity still remains.



(a) *Modularity* **(b)** Number of Members

Figure 4.3. Modularity values for different thresholds choices on Brazil's 2017 data. Green dot indicates selected threshold, 0.42.

4.4 Identifying Polarized Communities

As mentioned, the ideological communities identified in the previous section still share some similarity, particularly for the Brazilian case. In this section, we address our second research question (RQ2), with the aim of identifying polarized communities, i.e., communities that have a more clear distinction from the others in terms of voting behavior. To that end, we take a step further and consider that members of the same polarized community should not only be neighbors (i.e., similar to each other) but should also share most of their neighbors. Thus two members that, despite voting similar to each other, have mostly distinct sets of neighbors should *not* be in the same group.

To identify polarized communities, we start with the networks used to identify the ideological communities and compute the *neighborhood overlap* for each edge. The neighborhood overlap of an edge (v_1, v_2) is the ratio of the number of nodes that are neighbors of both v_1 and v_2 to the number of neighbors of at least one of v_1 or v_2 [David and Jon, 2010]. The neighborhood overlap of v_1 and v_2 is taken as an estimate of the strength of the tie between the two nodes. Edges with tie strength below a given threshold are considered as *weak* ties, whereas edges with tie strength above that threshold are classified as *strong* ties. We consider that weak ties come from overlapping communities, and strong ties are edges within a polarized community. Thus, edges representing weak ties are removed. As before, nodes that become isolated after this new filtering are also removed.

Once again the selection of the best neighborhood overlap threshold is not straightforward as it involves a complex trade-off: larger thresholds lead to more closely connected communities and higher *modularity*, which is the goal, but also produce sparser graphs, resulting in a larger number of isolated nodes, which are disregarded. Thus, for each network, we selected a threshold that produced a good trade-off between the two metrics (i.e., the lowest threshold yielding modularity close to maximum). Figure 4.3 shows an example of this trade-off for one specific year (2017) in Brazil, with the selected threshold value shown in green. For Brazil, the selected threshold fell between 0.40 and 0.55, while for the US it was from 0.1 to 0.28. We then re-executed the Louvain algorithm to detect (polarized) communities in the new networks.

Table 4.3 presents the topological properties of the networks as well as the structural and ideological properties of the identified polarized communities, for both Brazil and US. Focusing first on the Brazilian networks (top part of the table), we see that the number of nodes with strong ties decreases drastically (by up to 66%) as compared to the networks analyzed in Section 4.3. This indicates the large presence of House members that, despite great similarity with other members, are not strongly tied (as defined above) to them, and thus do not belong to any polarized community. The number of connected components dropped for some years and increased for others, suggesting that some components in the first set of networks were composed of structurally weaker communities or of multiple smaller communities. Network density, average shortest path length, and clustering coefficient also dropped, indicating sparser networks, as expected.

The number of polarized communities somewhat differs from the number of communities obtained when all (strong and weak) ties are considered, increasing in most years. This suggests that some ideological communities identified in Section 4.3 may be indeed formed by multiple more closely connected subgroups. Yet, those numbers are still smaller than the number of parties in each year (Table 4.1). Moreover, compared to the ideological communities first analyzed, the polarized communities are stronger both structurally and ideologically, as expressed by larger values of *modularity* and average party discipline.

For the US case, the numbers in Table 4.3 are very similar to those in Table 4.2. Less than 2% of the nodes have only weak ties and were removed from the networks in all years. Thus, almost all members have strong ties to each other, building ideological communities that are, in general, very polarized.

In summary, in this section we analyzed ideological communities and showed that in Brazil, there is a high volatility in the formation of these communities within a government, changing in recent years. Meanwhile, in the United States, there is a clearly polarized party system and a third community within one of the parties.

Table 4.3. Statistics of Strongly Tied Networks and Polarized Communities in Brazil (CC: connected comp., SPL: shortest path length, Mod: modularity)

Brazil											
Year	**# of Nodes**	**# of Edges**	**# of CC**	**Avg. SPL**	**Avg. Degree**	**Avg. Clust.**	**D**	**# of Comm.**	**Mod.**	**Avg. PD (%)**	**SD PD**
2003	186	1436	1	1.48	15.44	0.38	0.08	4	0.35	97.78	0.86
2004	154	866	1	1.52	11.25	0.33	0.07	5	0.36	97.11	0.57
2005	119	1210	2	1.19	20.34	0.59	0.17	4	0.37	95.40	0.93
2006	136	590	10	1.37	8.68	0.52	0.06	12	0.57	96.62	2.16
2007	175	977	3	1.68	11.17	0.32	0.06	6	0.44	97.31	1.36
2008	216	1019	2	1.94	9.44	0.23	0.04	5	0.42	97.11	0.46
2009	209	1217	1	1.30	11.65	0.41	0.05	5	0.56	94.57	1.67
2010	225	726	6	1.45	6.45	0.22	0.02	11	0.51	94.31	1.80
2011	250	1891	1	1.78	15.13	0.31	0.06	4	0.40	96.56	0.86
2012	145	1151	3	1.84	29.82	0.48	0.11	6	0.37	94.42	1.98
2013	318	4437	5	1.77	27.91	0.58	0.08	9	0.47	91.30	2.17
2014	287	1672	3	1.37	11.65	0.41	0.04	5	0.63	94.04	1.28
2015	372	6290	6	1.41	33.82	0.64	0.09	9	0.64	93.93	1.70
2016	269	1726	3	1.43	12.83	0.44	0.04	8	0.63	95.08	1.21
2017	227	1631	5	1.58	14.37	0.44	0.06	6	0.60	95.25	2.01
United States											
Year	**# of Nodes**	**# of Edges**	**# of CC**	**Avg. SPL**	**Avg. Degree**	**Avg. Clust.**	**D**	**# of Comm.**	**Mod.**	**Avg. PD (%)**	**SD PD**
2003	431	41872	2	1.11	194.30	0.95	0.45	2	0.47	93.60	1.03
2004	426	40741	2	1.12	191.27	0.95	0.45	2	0.48	92.97	0.55
2005	431	41886	2	1.11	194.37	0.95	0.45	2	0.47	92.60	0.79
2006	426	41073	2	1.10	192.83	0.95	0.45	2	0.48	91.45	0.33
2007	414	38462	2	1.12	185.81	0.94	0.44	2	0.42	91.55	3.78
2008	423	40708	2	1.11	192.47	0.95	0.45	2	0.43	95.49	1.93
2009	428	41690	2	1.15	194.81	0.94	0.45	2	0.40	93.89	2.45
2010	418	39958	2	1.13	191.19	0.95	0.45	3	0.43	94.86	1.97
2011	422	41112	2	1.15	194.84	0.97	0.46	3	0.45	90.01	3.16
2012	413	40529	2	1.07	196.27	0.97	0.47	3	0.44	91.70	2.17
2013	421	40910	2	1.10	194.35	0.96	0.46	2	0.46	93.32	0.94
2014	417	40717	2	1.08	195.29	0.96	0.46	2	0.48	94.40	0.38
2015	424	41759	2	1.08	196.98	0.95	0.46	2	0.47	94.53	1.41
2016	418	40890	2	1.08	195.65	0.96	0.46	3	0.46	95.67	0.80
2017	421	40923	2	1.08	194.41	0.95	0.46	2	0.48	95.37	0.11

4.5 Temporal Analysis of Polarized Communities

We now turn to RQ3 and investigate how the polarized communities evolve over time. Table 4.4 shows *persistence* (*Pers*) and NMI (See Section 4.2.6) results for all pairs of consecutive years and both countries. For Brazil (BR), the values of *persistence* varied over the years, ranging from 46% to 80%. Thus, a significant number of new nodes join polarized communities every year. Indeed, in most years, roughly half of the members of polarized communities are newcomers. The values of NMI are also small,

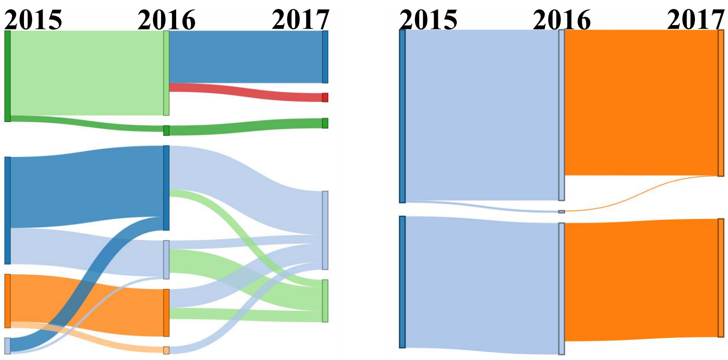
Table 4.4. Temporal Analysis of Polarized Ideological Communities (NMI: normalized mutual information)

Consecutive Years	Brazil		United States	
	Persistence	NMI	Persistence	NMI
2003 - 2004	58.24%	0.14	98.13%	0.97
2004 - 2005	46.30%	0.16	90.80%	0.97
2005 - 2006	53.04%	0.20	98.36%	1.00
2007 - 2008	68.26%	0.22	97.57%	1.00
2008 - 2009	63.80%	0.18	86.74%	1.00
2009 - 2010	61.38%	0.26	96.24%	0.94
2011 - 2012	80.08%	0.14	96.18%	0.96
2012 - 2013	67.87%	0.59	96.76%	0.80
2013 - 2014	61.23%	0.56	97.85%	1.00
2015 - 2016	57.85%	0.65	97.63%	0.97
2016 - 2017	57.47%	0.58	86.26%	0.98

especially in the earlier years, reflecting great change also in terms of nodes switching communities. This is consistent with a period of less clear distinction between the communities and weaker polarization, as discussed in the previous sections. Since 2012, the values of NMI fall around 0.6, reflecting greater stability in community membership. For the US, in contras, both *persistence* and NMI are very large, approaching the maximum of 1. Almost all members persist in their polarized communities over the years.

A visualization of some of these results is shown in Figure 4.4 which presents the flow of nodes across polarized communities over the years of 2015 to 2017 in Brazil and in the US. Each vertical line represents a community, and its length represents the number of members belonging to that community who persisted in some polarized community in the following year. Thus, communities for which all members did not persist in any polarized community in the following year are not represented in the figure. Recall that, according to Table 4.3, the number of polarized communities in Brazil in 2015, 2016 and 2017 was 9, 8 and 6, respectively. A cross-analysis of these results with Figure 4.4a indicates that members of only 4 out of 9 polarized communities in 2015 persisted polarized in the following year. Moreover, two polarized communities in 2016 were composed of only newcomers and both communities disappeared in 2017 (as they do not appear in the figure). Similarly, one polarized community in 2017 was composed of only newcomers. The figure also shows a great amount of switching, merging and splitting across communities over the years. Figure 4.4b, on the other hand, illustrates the greater stability of community membership in the US.

4.6 Evaluating Ideological Changes



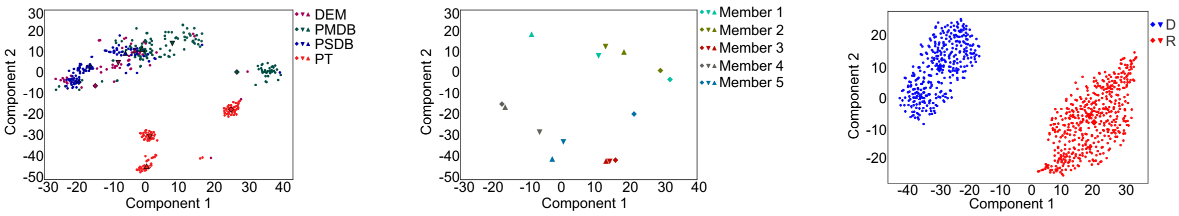
(a) *Brazil* **(b)** *United States*

Figure 4.4. Dynamics of Polarized Communities over 2015-2017.

We now turn to our final analyses of changes in ideological behavior. We employ the strategy described in Section 4.2.6.1 to model an ideological space and track individual members over time in this space. For analysis purposes, we focus on changes during a period Δ_t equal to the duration of a term of office (time period during which elected members should serve), divided in yearly time windows. In Brazil, party members are elected for a 4-year term (named *legislature*), whereas in the US, they are elected for a 2-year period (called *Congress*).

We start by defining, for each case study and each term of office specified in Table 4.1, a corresponding a set of backbones $\mathcal{B}_{\mathcal{BR}} = \{B_{BR,\Delta_{t_1}}, B_{BR,\Delta_{t_2}}, \ldots, B_{BR,\Delta_{t15}}\}$ and $\mathcal{B}_{\mathcal{US}} = \{B_{US,\Delta_{t_1}}, B_{US,\Delta_{t_2}}, \ldots, B_{US,\Delta_{t_{15}}}\}$ representing the networks produced for windows Δ_t in T, as described in Section 4.2.3. For each such sequence $\mathcal{B}$ we then produce a single latent ideological space following the method in Section 4.2.6.1. Specifically, for each window Δ_t (year), we obtain a matrix of embedding vectors $U(\Delta_t) = \{\mathbf{u}_{v_1}(\Delta_t), \mathbf{u}_{v_2}(\Delta_t), \ldots, \mathbf{u}_{v_n}(\Delta_t)\}$ where $V = \{v_1, v_2, \ldots, v_n\}$ is the set of vertices in $\mathcal{G}_{s,\Delta T}$ ($s = \{BR, US\}$). Recall that our model is robust to missing values, allowing us to infer an ideological representation of a member v_1 in Δ_t from (Δ_{t-1}) and (Δ_{t+1}). Nevertheless, we choose to include in V only members who appeared in $\mathcal{B}$ in at least two years. This choice is based on a conservative approach to improve robustness, particularly for the Brazilian case, which, as already discussed, has greater instability and a longer term of office (4 years).

We train our ideological space model for a given term of office by carrying out a grid search to determine the best values of parameters λ and τ, as proposed in [Yao et al., 2018]. We consider various combinations of parameter values, varying λ in [0;100] and τ in [0;100]. For each combination, we first generate our latent space model and the vertice representations (embeddings) for each window Δ_t. We then evaluate



(a) Brazil, 55^{th} Leg. (top-4 parties) **(b)** Brazil, 55^{th} Leg. (5 selected members) **(c)** United States 114^{th} Congress

Figure 4.5. 2-D representation of latent ideological space. Symbols $\diamond$, $\triangledown$ and $\triangle$ represent party centroids (a,c) or members (b) respectively during 1^{st}, 2^{nd} and, in case of Brazil, 3^{rd} year of legislature as well.

the goodness of these embeddings (and correspondingly of the generated latent space) as follows. We apply the spherical k-means algorithm [Banerjee et al., 2005], which uses cosine similarity as distance metric, to group the vertice embeddings produced for window Δ_t, $\mathbf{u}_v(\Delta_t)$, into k clusters, where k is the number of ideological communities detected for the same window Δ_t (see Section 4.3). We then calculate the Normalized Mutual Information (NMI) (Equation 4.3) between the ideological communities and the clusters detected by the spherical k-means on the embeddings yielded by our model. The most representative latent space model (i.e., the best parameter values) is the one that best recovers the originally defined ideological groups, thus yielding a higher NMI result.

The results of the grid search were very consistent across most terms of office for each case study. For the US, the same values of $\lambda = 15$ and $\tau = 20$ were found to be the best in all cases. Also, the NMI values were very high, being at least 0.97 and very often reaching the maximum of 1, reflecting the clear structural and ideological separation of the networks. For Brazil, the best values are $\lambda = 5$ and $\tau = 10$ for all but the last term of office for which the best parameter values are $\lambda = 10$ and $\tau = 5$. The NMI values are lower than in the US, yet still reasonably high, especially in the most recent terms (the NMI reached 0.85 in the last 55^{th} legislature), reflecting the stronger community structure and more clear ideological separation of party members in the period (as discussed in Section 4.3).

For the sake of visualization, we select one example term of office from each case study and plot the generated latent ideological space in a low-dimensional 2-D view using the t-distributed Stochastic Neighbor Embedding (t-SNE) [Maaten and Hinton, 2008]. Figures 4.5a and 4.5c show the representations obtained for the 55^{th} legislature in Brazil[12] and the 114^{th} congress in the United States, respectively. Each color corre-

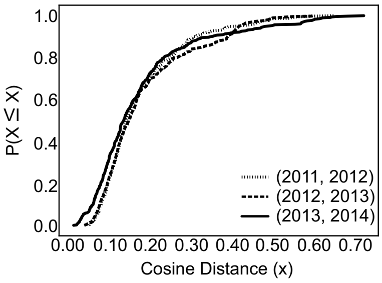
[12]Recall that our dataset covers only 3 years of 55^{th} legislature.

sponds to one political party and each point corresponds to one party member in one of the covered years in the latent ideological space. For the sake of graph readability, Figure 4.5a shows only the four brazilian parties (the largest ones). Also, the figures do not distinguish between different members of the same party nor different locations of the same member over the years (in case the member changed position over time): all of them are represented by points of the same color. Yet, to illustrate changes in ideological behavior, we plot the centroids of each party, distinguishing its location in each year of the analyzed term by using different representations. Each centroid is represented by a *diamond* in the first year, by an *upside down triangle* in the second year, and, in case of Brazil, by a regular *triangle* in the third year.

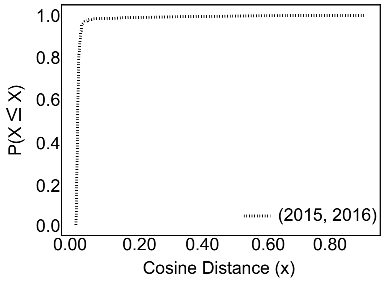
We further illustrate changes in individual ideological behavior by focusing now on 5 selected Brazilian party members. Figure 4.5b shows the locations of their corresponding vertice embeddings in the years of the 55^{th} legislature in the same low-dimensional 2-D view. Each member is shown in a different color and, once again, we use a *diamond*, an *upside down triangle* and a regular *triangle* to represent their locations in the first, second and third years, respectively.

As Figure 4.5a shows, all four Brazilian parties have changed their ideological behavior over the years, as illustrated by the changes in the locations of their centroids in the ideological space. However, some of them remained quite cohesive throughout the period, that is, the changes were mostly in group. For example, despite individual changes, the distinction between the Work Party (PT) and the Brazilian Social Democracy Party (PSDB), represented in red and blue, respectively, is clear in all three years. These two parties have faced each other for over twenty years in the presidential elections in Brazil. In any of the years, the cosine distance of any given two members (one from each party) is close to 1, indicating great ideological distinction. On the other hand, the distance between any two members from the same party tends to be close to 0, indicating strong ideological alignment. Another interesting example is the Brazilian Democratic Movement Party (PMDB) which started the 55^{th} legislature ideologically aligned with PT, but approached the opposition, composed of PSDB and DEM (among others), as the years went by. The change in Figure 4.5a reflects what happened in reality as the second government of president Dilma Rousseff (PT) started with the support of PMDB. However, the party of the vice-president decreased its support to the left-wing president and shifted towards center, more aligned with PSDB and DEM. Such movement, which was replicated by other supporting parties, culminated with the presidential impeachment in 2016.

The changes in individual ideological behavior over the three analyzed years can be more clearly visualized in Figure 4.5b. Note that some members, such as member 3,



(a) Brazil



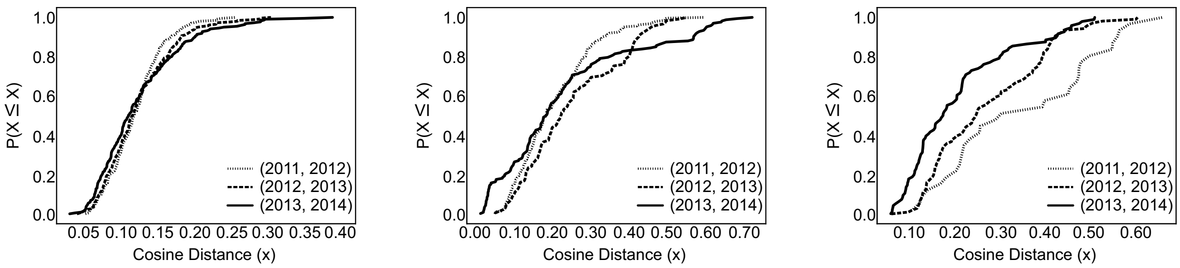
(b) United States

Figure 4.6. CDF of ideological shift of members over consecutive years (measured w.r.t. cosine distance).

exhibit very small changes in the ideological space, whereas others have a much more dynamic behavior, falling into different regions of the space over the years. Also note that while some seem to be converging to the same region of the space (e.g., members 4 and 5), others are drifting away (e.g., member 1). In contrast, Figure 4.5c shows that, in the US, party members have quite stable and distinct ideological behavior.

We delve deeper into the Brazilian scenario, by looking into the 54^{th} legislature, encompassing years 2011-2014, which, as already mentioned, consisted of a period of great political turmoil during which ideological distinctions became more clear. Using the temporal embedding obtained for this term of office, we compute, for each individual member, her *ideological shift*, i.e., the cosine distance between her embedding representations, in consecutive years. Figure 4.6a shows the cumulative distributions of the ideological shift of individual members for each pair of consecutive years. The three distributions are similar, but we can see a trend towards larger distances in the more recent years, in alignment with our discussion in the previous sections. Also, although most members exhibit some ideological shift over consecutive years, there is great variability across members. For comparison purposes, Figure 4.6b shows the distribution for the two years of the 114^{th} US congress, when practically all members remained unchanged, confirming the consistency of ideologies over time.

The greater variability in the Brazilian case can be explained, to some extent, by the heterogeneity in ideological behavior between polarized and non-polarized party members. To further analyze this issue, we separate, for each pair of consecutive years, the party members into *polarized*, i.e., members who persisted in some polarized community in the two years (as in Table 4.4), and the other, *non-polarized members*. Figures 4.7a and 4.7b show the distributions of ideological shift for each group and for each pair of consecutive years of the 54^{th} legislature. Clearly, non-polarized party members exhibit much greater changes (larger cosine distances), while polarized members do exhibit a more consistent ideology over time.



(a) Members of polarized communities

(b) Non-polarized members

(c) Members who left polarized communities

Figure 4.7. CDF of ideological shift of Brazilian members over consecutive years (w.r.t. cosine distance) grouped by polarization.

Yet, even polarized members do experience changes over time, which indirectly affect the membership of the polarized communities. Indeed, as already discussed in Section 4.5, polarized members often switch between polarized communities, especially in the earlier years. To get a hint of the extent to which such polarized members shift in the ideological space but still remain polarized, we compare them with members who started polarized but left the polarized ideological communities (i.e., became non-polarized) in the following year. Figure 4.7c shows the distributions of ideological shift for the latter for the same period. Clearly, members who ceased being polarized tend to exhibit greater changes in ideological behavior. Thus, the changes in membership of polarized communities are mostly due to members changing to nearby (polarized) communities, as they slowly shift in the ideological space. Moreover, as shown in Figure 4.7c, the shift in the ideological space of members who ceased being polarized decreases over the years. This suggests that the polarized communities and, thus, their members, strengthen polarization over time.

4.7 Discussion

In this case study, we have proposed a methodology to analyze the formation and evolution of ideological and polarized communities in party systems, applying it to two strikingly different political contexts, namely Brazil and the US. Our analyses showed that the large number of political parties in Brazil can be reduced to only a few ideological communities, maintaining their original ideological properties, that is well disciplined communities, with a certain degree of redundancy. These communities have distinguished themselves both structurally and ideologically in the recent years, a reflection of the transformation that Brazilian politics has been experiencing since 2012. For the US, the country's strong and non-fragmented party system leads to the identification of ideological communities in the two main parties throughout the analyzed period. However, there are still some highly similar links crossing the community boundaries. Moreover, for some years, a third community emerged, without however affecting the strong discipline, ideology and community structure of the American party system.

We then took a step further and focused on polarized communities by considering only tightly connected groups of nodes. We found that in Brazil, despite the party fragmentation and the existence of some degree of similarity even across the identified ideological communities, it is still possible to find a subset of members that organize themselves into strongly polarized ideological communities. However, these communities are highly dynamic, changing a large portion of their membership over consecutive years. In the US, on the other hand, most ideological communities identified are indeed highly polarized and their membership remain mostly unchanged over the years.

Finally, we delved deeper into the individual ideological behavior of party members by proposing a temporal ideological space model. Based on temporal vertice embeddings, our model allows to analyze the ideological shift of individual members over the period of a term of office. We observed that in Brazil, the vast majority of party members did exhibit some change in ideological behavior over time, though the extent of which varies greatly across members. Whereas members of polarized communities had a somewhat more consistent ideological behavior, members of non-polarized communities fluctuated much more ideologically, especially after 2012. In contrast, the representations of US party members in an ideological space confirm much greater stability over time.

4.8 Summary

In this chapter, we have addressed our RG1 and RG2 defined in Section 1.3 by applying our general approach to a specific phenomenon. First, to extract the backbones from the complete networks, an extreme case of noise, we relied on contextual information, often overlooked, and showed it could be used to identify salient edges. In particular, we employed the partisan discipline to estimate the expected agreement between two congresses belonging to the same ideological group. Then, from the assumption that salient edges are those whose agreement is consistent with the party discipline of each political system, we observed different community structures merging over the analyzed period of fifteen years in both scenarios (BR and USA). Specifically for Brazil, we found that some communities capturing the collective behavior of congressmen beyond their original parties were weakly structured due to governability issues.

We then went a step further and proposed a second strategy that extended the first assumption. The idea is to require that two congressmen not only have a strong ideological alignment but are also part of a common group of congressmen that exhibit similar ideological behavior. Therefore, we focused on ideological communities that are strongly polarized. To satisfy the new assumption, we proposed the use of neighborhood overlap as a second backbone extraction strategy and showed that it is possible to combine different backbone extraction approaches. Finally, we presented our contribution and progress in the field of temporal embedding of nodes. In particular, we proposed a new technique based on the two state-of-the-art techniques of static node embedding and temporal word embedding.

In the next Chapter, we use our general approach to examine a new case study with completely different characteristics, which requires, above all, a more sophisticated strategy for extracting the backbone.

Chapter 5

Online Political Discussions on Instagram

In this chapter, we present our second case study focused on modeling political discussions on social media platforms, notably Instagram. It is organized as follows: Section 5.1 provides a contextualization of such study and particular questions that we tackle; Section 5.2 describes our methodology based our overall solution presented in Section 3; Section 5.3 describe our dataset; Sections 5.4-5.6 show the results while Section 5.7 discusses our main findings on the phenomenon in question; Finally, the Section 5.8 summarizes the implications and contributions obtained.

5.1 Contextualization

Social media applications are a major forum for people to express their opinions and information. By interacting with such applications, users build complex networks that favor the dissemination of information [Al-Garadi et al., 2018]. Indeed, social media has become an important source of information for a large fraction of the world population [Shearer, 2018; Watson, 2020; Newman et al., 2019]. It has been shown to play an important role in social mobilization and political engagement [Resende et al., 2019a; Muñoz and Towner, 2017], notably during major political events [Pierri et al., 2020].

Instagram has observed a surge in popularity in recent years [Lerman, 2020], especially among youth. Use of Instagram for consuming news has doubled since 2018, and the platform is set to overtake Twitter as a news source [Newman et al., 2020]. It is no surprise that political figures are increasingly using this platform to reach the population at large. Therefore, it is of utmost importance to understand how users interact with each other to find out how information is disseminated on the platform and how online debate affects our society. Prior studies of user behavior on Instagram have mainly focused on user engagement according to content type [Reece and Danforth, 2017; Jaakonmäki et al., 2017; Garretón et al., 2019; Kao et al., 2019; Kim et al., 2020; Weerasinghe et al., 2020], general characteristics of comments related to political messages [Trevisan et al., 2019; Zarei et al., 2019], and the impact of posted content on marketing contexts [Jaakonmäki et al., 2017; Yang et al., 2019; Kang et al., 2020]. However, the literature lacks an investigation of the networks that emerge from users' interactions, particularly in the context of political content that fosters the spread of information.

In Instagram jargon, a *profile* is followed by a set of *followers*. A profile with a large number of followers is called an *influencer*. Influencers post content, i.e., *posts*, containing a photo or a video. Followers – or any registered user in the case of public profiles – can view the profile's posts and comment on them, becoming *commenters*. We here refer to multiple users who comment on the same post as *co-commenters*, and to the interactions that occur among multiple users (often more than two) when they comment on the same post as *co-interactions*. Co-commenters may form *communities* that arise either naturally (i.e., based on common interests on specific topics) or driven by hidden efforts (e.g., ad-campaign or coordinated behavior). By feeding the discussions, these communities may favor the spread of specific ideas or opinions while also contributing to increase the visibility of particular influencers. Thus, revealing how such communities emerge and evolve over time is key to understanding information dissemination on the system.

In light of this, the *phenomenon* we investigate here is the analysis of online discussions, notably on political themes on Instagram. Our goal in studying online discussions is to analyze how discussion groups that emerge from collective behavior around political and non-political personalities can be characterized by identifying similarities and differences, as well as triggers that attract them. As case studies, we analyze a large dataset of Instagram comments containing the activity of approximately 3 million commenters on 36 824 posts of 320 influencers. These influencers include profiles of popular political figures – 80 from Italy and 80 from Brazil, as well as 160 top-influencers in other categories (e.g., sportsmen, celebrities, musicians), 80 from each country, which we use as a baseline. We focus on two months surrounding elections that took place in each country.

Our ultimate goal is to analyze how the collective behavior of such commenters favor the dissemination of information at large. In particular, we address the following three research questions (RQs):

- **RQ1: What are the characteristics of the network backbones emerging from *salient co-interactions* on Instagram?** First, we model co-commenters' activity as a network in which nodes represent commenters and edge weights indicate the number of posts commented by both users. To filter out uninteresting edges and reveal the underlying network backbone, we propose TriBE, a new backbone extraction built on a probabilistic reference network model where edges are based on the assumption that commenters behave independently from each other. Our model primarily considers two factors: the popularity of posts and the engagement of commenters with each influencer. By comparing the network observed in the real data with our reference model, we prune edges whose weights are within the expected range under the assumption of independent user behavior, revealing the backbones with *salient* for all scenarios and thus constraining them.

- **RQ2: What are the distinguishing properties of the communities that compose such backbones, notably communities formed around political content?** We delve further into the components of the network backbone by using Louvain's algorithm to extract communities. We then analyze the structural and contextual properties of these communities, in particular, how their structure emerges and the interests of their profiles, textual features of their comments, including sentiment analysis, topics they discuss, and the psycholinguistic properties.

- **RQ3: How do community properties evolve over time?** By observing 10 weeks around political elections, we characterize communities temporally in both their structure and context. We analyze how their characteristics evolve over time in terms of membership and discussions held.

5.2 Methodology

In this section we formally define the network of co-commenters on Instagram and describe the probabilistic network model used as reference to uncover salient interactions. We then describe how we extract communities from the network backbone and present the techniques employed to characterize these communities.

5.2.1 Network Model

We model the interactions among users who comment on the same Instagram post as a sequence of snapshots of fixed time window Δ_t, where each snapshot aggregates posts of a selected set of influencers and their associated comments. We here consider Δ_t equal to one week as a reasonable period to cover discussions around posts. For a given scenario s (our scenarios is detailed in Section 5.3), we define a set of time windows of one-week duration as $T_s = \{\Delta_{t1}, \Delta_{t2}, ..., \Delta_{t10}\}$. In our solution, we model *individuals* as being *commenters*, i.e., users commenting on posts made by influencers analyzed in a given scenario and in a given time window Δ_t, thus forming the sets of *individuals* for each time window Δ_t and scenario as follow: $I_{s,\Delta_t} = \{i_1, i_2, ..., i_j\}$. Here, *opportunities* are taken as users' posts created during a given time window Δ_t. In this way, we define the set of opportunities containing all posts in each scenario in Δ_T as $O_{s,\Delta_t} = o^1_{\Delta_t}, o^2_{\Delta_t}, ...o^m_{\Delta_t}$. The total j of *commenters* and opportunities m varies across scenarios and time windows. In Section 5.3 we provide a overview of our dataset.

Here, we consider all users who commented in the same post at twice once to be a co-interaction. In that manner we choose to disregard commenters whose activities were concentrated on a single post, and thus reflect sporadic behavior.[1] Thus each opportunity (post) $o^k_{\Delta_t}$ leads to a single co-interaction in set $C(o^k_{\Delta_t})$. Collectively, all opportunities during time window Δ_t leads to a set C_{s,Δ_t} of co-interactions associated with that time period for scenario s. Having defined these elements, we then model a network of co-commenters G_{s,Δ_t} for each scenario s and time window Δ_t, defined as:

- V_{s,Δ_t}: is the set of vertices representing all commenters in I_{s,Δ_t} who participated in at least on co-interaction in C_{s,Δ_t}.

- E_{s,Δ_t}: is the set of undirected and weighted edges, such that the weight of edge e_{i_1,i_2} linking two commenters i_1 and i_2 is $\gamma_{\Delta_t}(i_1, i_2) = count(i_1, i_2)$, where $count(i_1, i_2)$ is the number of times both commenters commented in the same post shared during time window Δ_t.

In such network, the presence of noise occurs due to different circumstances. First, many users become co-commenters incidentally because of the high popularity of some posts and/or influencers. Equally, very active commenters naturally become a co-commenter of many other users as a side effect of their great frequency of commenting activity. Yet, such co-interactions are mostly driven by chance, as opposed to true group behavior. As such, they do not reflect fundamental properties (if any at all) of the ongoing online debate and, more broadly, of the ongoing information dissemination process. Moreover, often happening in large volumes, those casual and incidental

[1]Note that, by doing so, commenters who commented multiple times on a *single* post, but did not comment on other posts, are removed.

co-interactions may lead to the formation of networks of co-interactions with lots of sporadic, uninteresting or weak edges.

In other words, such co-interactions are to some extent *expected* given users' activity and post popularity. Thus, to analyze interactions among co-commenters, we filter out such expected edges and focus on those whose frequencies of occurrence are large enough to allow us reject, with some confidence, the assumption of independent behavior. That is, we focus on *salient* edges that most probably reflect real online discussions, forming the underlying fundamental network backbone.

In this context, we present our method in the next section. Next, Section 5.3 describes our scenarios of study. In sum, we have four scenarios built on disjoint sets of influencers in the following fashion: *Brazil* and *Italy* and, for each one of them, two groups of influencers, *Politics* and *General*. Each scenario s ($s = \{BR_{Politics}, BR_{General}, IT_{Politics}, IT_{General}\}$) builds the following universal sets of networks, backbone and communities, respectively: $\mathcal{G}_s = \{G_{s,\Delta_{t_1}}, G_{s,\Delta_{t_2}}, \ldots, G_{s,\Delta_{t10}}\}$, $\mathcal{B}_s = \{B_{s,\Delta_{t_1}}, B_{s,\Delta_{t_2}}, \ldots, B_{s,\Delta_{t10}}\}$ and $\mathcal{P}_s = \{P_{s,\Delta_{t_1}}, P_{s,\Delta_{t_2}}, \ldots, P_{s,\Delta_{t10}}\}$.

5.2.2 Network Backbone Extraction

A fundamental question that arises when studying complex networks is how to quantify the statistical significance of an observed network property [Grady et al., 2012; Coscia and Neffke, 2017]. To that end, reference models are often used to determine whether networks display certain features to a greater extent than expected under a null hypothesis (e.g., independent behavior) [Newman, 2018]. A reference (or *null*) model matches some of the features of a graph and satisfies a collection of constraints, but is otherwise taken to be an unbiased random structure. It is used as a baseline to verify whether the object in question displays some non-trivial features (i.e., features that would not be observed as a consequence of the constraints assumed). An appropriate reference model behaves according to a reasonable null hypothesis for the behavior of the system under investigation. One strategy to build a reference model is by employing generative growing networks [David and Jon, 2010; De Melo et al., 2015; Newman, 2018].

For the sake of simplicity, let's disregard the subscript s used to refer to a scenario. We here employ a reference generative model $\widehat{G}_{\Delta_t}$ for each network G_{Δ_t} that is based on the hypothesis that commenters during Δ_t behave independently from each other. That is, edge weights in $\widehat{G}_{\Delta_t}$ are defined under a generative process in which commenters act independently from each other, although their interactions with influencers' posts (i.e., which post each user comments on) are not identically distributed. We can then observe which edges of the real network G_{Δ_t} do not behave in accordance with the reference model $\widehat{G}_{\Delta_t}$ – i.e., reflect interactions that significantly deviate from an independent behavior. Such edges will compose the network *backbone*. Intuitively, we want to highlight co-interactions that occurred more often than what would be expected if commenters behaved independently.

To achieve this, we proposed TriBE (Tripartite Backbone Extraction), a novel backbone extraction method that uses as input the popularity of each post (number of unique commenters) and the engagement of commenters towards each influencer (number of posts by the influencer on which each commenter writes) to create a null model. Using these statistics, comments are randomly assigned to commenters while preserving: i) the set of influencers on which each commenter writes a comment; ii) the popularity of each post, and iii) the engagement of each commenter towards each influencer. The model assigns commenters to each post using independent and identically distributed (i.i.d.) draws from a distribution where the probability is proportional to the commenter's engagement towards the target influencer. By doing so, we prevent the backbone from being dominated by very active commenters or by those engaged in highly popular posts.

More specifically, let U_{Δ_t} be the set of all influencers who wrote a set of posts considered as being opportunities O_{Δ_t}. Let $I_o \subseteq I_{\Delta_t}$ be the set of unique commenters in post $o \in O_{\Delta_t}$ and $O_{\Delta_t,u} \subseteq O_{\Delta_t}$ be a partitioning of O_{Δ_t} based on the influencer $u \in U_{\Delta_t}$ who created the post. We define the engagement of commenter $i_1 \in I_{\Delta_t}$ towards influencer u (measured by the total number of posts in $O_{\Delta_t,u}$ commented by i_1) as

$$x_u(i_1) = \sum_{o \in O_{\Delta_t,u}} \mathbb{1}\{i_1 \in I_o\}, \tag{5.1}$$

where $\mathbb{1}\{.\}$ is the identity function. We then define i_1's *relative engagement* towards u w.r.t. other commenters as:

$$f_u(i_1) = \frac{x_u(i_1)}{\sum_{i_2 \in I_{\Delta_t}} x_u(i_2)} = \frac{x_u(i_1)}{\sum_{o \in O_{\Delta_t,u}} |I_o|}. \tag{5.2}$$

In this way, we can describe in details the three steps of the generative process to build our reference model $\widehat{G}_{\Delta_t}$:

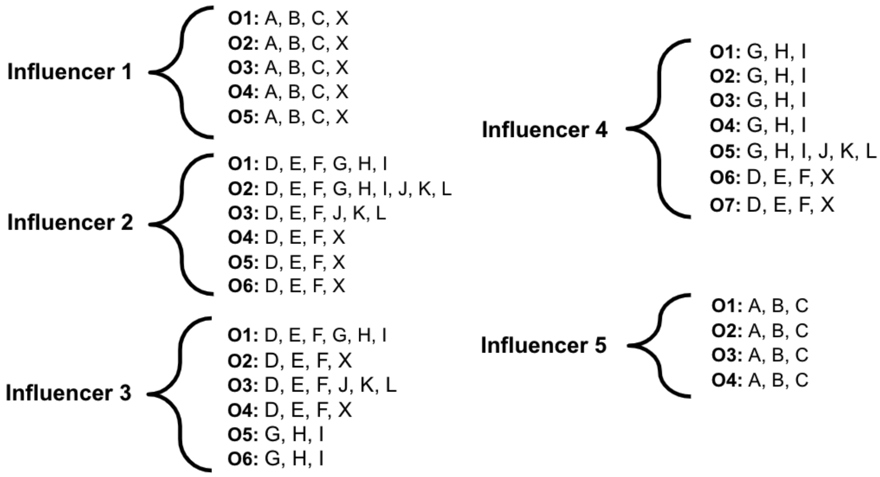
1. For each post $o \in O_{\Delta_t}$, we consider a random assignment of each of the $|I_o|$ (unique) commenters to a commenter $i_1 \in I_{\Delta_t}$ with probability $f_u(i_1)$, where i_1 is the author of o. Specifically, under the assumption of independent behavior, we consider each such assignment as a Bernoulli random variable with parameter

$f_u(i_1)$. The probability that commenter i_1 is not assigned to o is thus a Binomial random variable, with 0 successes in $|I_o|$ experiments. Conversely, under the assumption of independent behavior, the probability that i_1 has commented (at least once) on a post $o \in O_{\Delta_t,u}$ is $r_o(i_1) = 1 - (1 - f_u(i_1))^{|I_o|}$.

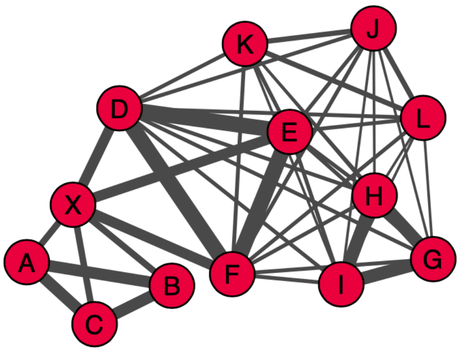
2. For each pair of commenters i_1 and i_2, we denote by $r_o(i_1, i_2)$ the probability that both get assigned to post o and by $r_o(i_1|i_2)$ the probability that i_2 gets assigned to o given that i_1 is assigned to o. The conditional probability $r_o(i_1|i_2)$ is necessary because, strictly speaking, although we are drawing commenters independently, when i_1 is drawn, it decreases the number of chances i_2 has for being drawn (since $|I_o|$ is fixed). Hence, $r_o(i_1, i_2) = r_o(i_1) \times r_o(i_1|i_2)$. We approximate $r_o(i_1, i_2) \approx r_o(i_1) \times r_o(i_2)$, for each post $o \in O_{\Delta_t}$. Intuitively, this approximation works well when $|I_o|$ is large (as in the case of most influencers' posts), because drawing i_1 decreases by only one the number of draws that can be used to draw i_2. Then, for each post $o \in O_{\Delta_t}$, our model defines a distribution over the set of vertices corresponding to I_o, where the value of the random variable $\widehat{\Gamma}_o(i_1, i_2) \in \{0, 1\}$ indicates the existence of an edge between commenters c and d, and is given by a Bernoulli trial with parameter $r_o(i_1, i_2)$, i.e. $\widehat{\Gamma}_o(i_1, i_2) \sim \text{Bernoulli}(r_o(i_1, i_2))$.

3. The reference model $\widehat{G}_{\Delta_t} = (\widehat{V}_{\Delta_t}, \widehat{E}_{\Delta_t})$ is composed by the superposition of all the edges created for all posts $o \in O_{\Delta_t}$. Hence, an edge $\widehat{e}_{i_1,i_2} \in \widehat{E}_{\Delta_t}$ will have a weight distribution described by a random variable $\widehat{\Gamma}(i_1, i_2) = \sum_{o \in O_{\Delta_t}} \widehat{\Gamma}_o(i_1, i_2)$. Therefore, it will be a sum of Bernoulli random variables with distinct probabilities [Wang, 1993], which follows a Poisson Binomial distribution with parameters $r_1(i_1, i_2), r_2(i_1, i_2), \ldots, r_{|O_{\Delta_t}|}(i_1, i_2)$.

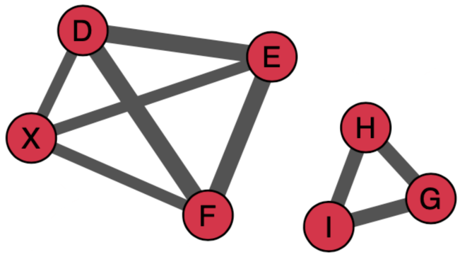
We can then compare the reference model $\widehat{G}_{\Delta_t}$ with the observed network G_{Δ_t} to extract the backbone B_{Δ_t} of the latter. We do so by keeping in B_{Δ_t} only edges of G_{Δ_t} whose weights have values exceeding the ones expected in $\widehat{G}_{\Delta_t}$ by a large margin. Specifically, for each edge $\widehat{e}_{i_1,i_2}$ we compute the $(1-\alpha)^{th}$ percentile, denoted by $\widehat{\gamma}_{1-\alpha}(i_1, i_2)$, of the distribution of edge weight $\widehat{\Gamma}(i_1, i_2)$, and compare it with the observed edge weight $\gamma(i_1, i_2)$. We keep edge e_{i_1,i_2} if $\gamma(i_1, i_2) > \widehat{\gamma}_{1-\alpha}(i_1, i_2)$. Intuitively, we keep only edges between co-commenters who interacted much more often than expected under the assumption of independent behavior. That is, edges for which the chance of such frequency of interactions being observed under the independence assumption is below α. We here set $\alpha = 5\%$, as done in prior studies [Serrano et al., 2009; Kobayashi et al., 2019]. Note that the $(1-\alpha)^{\text{th}}$ percentile is computed separately for each edge $e_{i_1,i_2} \in E_{\Delta_t}$ from random variable $\widehat{\Gamma}(i_1, i_2)$. For such a Poisson



(a) Raw Data



(b) Original Network G_{Δ_t}



(c) Network Backbone B_{Δ_t}

Figure 5.1. Illustration of the backbone extraction process in a simplistic graph. The isolated vertices are removed from the final B_{Δ_t} used in our analysis.

binomial distribution, there is a closed form for computing a given percentile [Hong, 2013], which, however, is expensive to compute. Instead, we here use the Refined Normal Approximation (RNA) [Hong, 2013], a method that proved very good performance with low computational complexity.

After filtering out edges, isolated vertices are also removed. At the end, we extract from the network G_{Δ_t} its backbone $B_{\Delta_t} = (V_{\Delta_t,b}, E_{\Delta_t,b})$ where $V_{\Delta_t,b} \subseteq V_{\Delta_t}$ and $E_{\Delta_t} \subseteq E_{\Delta_t}$. Hence, we obtain a universal set of backbones being defined as $\mathcal{B} = \{B_1, B_2, ..., B_{\Delta t_n}\}$.

In the next section, we present a toy example to show how TriBE works.

5.2.2.1 Backbone Extraction Exemplified

We illustrate how the backbone is extracted from a given input network G_{Δ_t} by means of the toy example shown in Figure 5.1. Figure 5.1a shows a total of five influencers, each with a different number of posts (o_1, o_2, etc), and each post with associated commenters (A, B, etc). Posts have different popularity, and commenters have different activity levels and engagement towards each influencer. The projected graph G_{Δ_t} is depicted in Figure 5.1b, whereas the extracted backbone B_{Δ_t} is shown in Figure 5.1c. In both networks, line thickness is proportional to the edge weight.

The question that arises is: *why did we extract only the edges shown in Figure 5.1c to compose the network backbone?* Recall that our model selects as salient edges those that have weights large enough so that we can reject the assumption of independent user behavior. Thus, for each edge in G_{Δ_t}, we ask ourselves: is there enough evidence to reject the assumption of independent behavior? If so, the edge is kept; otherwise, it is removed.

Let's illustrate our decisions regarding four groups of edges, focusing first on edges incident to commenters A, B, C. Note that all three commenters commented on posts only by influencers 1 and 5 and they commented on *all* posts by both influencers. These commenters are thus quite active, and the popularity of these posts is actually high, considering the population of users who commented on them. As such, it is possible that A, B and C are driven by their individual interests on these two influencers, and, as such, most probably would comment in most (if not all) posts by them. Thus, based on the observed data, we cannot reject the assumption of independent user behavior when considering co-interactions among A, B and C and the corresponding edges are not kept as part of the network backbone in Figure 5.1c. For example, the edge e_{AB} has weight $\gamma(A, B) = 9$ which is below or equal to the 95^{th} percentile of the corresponding edge weight distribution $\widehat{\gamma}_{0.95}(A, B) = 9$. The same reasoning applies to commenter X, who only commented on posts by influencer 1. Thus, the co-interactions of X with A, B and C are *not* considered salient and the corresponding edges are not kept.

Let's consider now the edges incident to commenters J, K and L. These users co-comment with low frequency in posts by influencers 2 (o_2 and o_3), 3 (o_3) and 4 (o_5). These posts are the most popular posts by such influencers, receiving comments from several other users as well. It is therefore expected that commenters active on these posts will have several co-commenters, as we can observe in Figure 5.1b. However, when it comes to J, K and L, the weights of these edges are small, as the co-interactions are somewhat sporadic. Moreover, note that the posts on which these users commented are among the most popular ones by the corresponding influencers, attracting most of their commenters. For example, o_2 by influencer 2 received comments by 9 out of all 10 users who commented on her posts. Co-interactions built around such highly popular post are *not* considered salient as one cannot tell whether commenters are truly interacting with each other or simply reacting independently to a quite attractive content. From an operational perspective, recall that, when building the reference model $\widehat{G}_{\Delta_t}$ we do need to assign commenters to comments associated with each post. In the case of such very popular posts, most if not all potential commenters are assigned, thus raising the chance of the edge being added to $\widehat{G}_{\Delta_t}$, and thus of the edge being considered expected under the assumption of independent behavior.

We now turn our attention to the edges incident to two groups of commenters: i) D, E, F and X; and ii) G, H, I. In both cases, the commenters co-interact on posts by influencers 2, 3 and 4, and the co-interactions occur very often on different posts by these influencers. However, unlike the case of A, B and C, discussed above, there are other users who also commented on the same posts. Compared to these other commenters, D, E, F, and X (as well as G, H, I) clearly stand out as frequent co-commenters. That is, taking the overall behavior of the commenters of these posts, we find that the co-interactions among D, E, F, and X (as well as G, H, I) are more frequent than expected if these users were being driven by independent behavior. For example, the weight of edge e_{DE} is $\gamma(D, E) = 12$ which is larger than the 95^{th} percentile of the corresponding edge weight distribution $\widehat{\gamma}_{0.95}(D, E) = 10$. We consider this evidence strong enough to reject the assumption of independent behavior. The same holds for the other aforementioned commenters. As consequence, the corresponding edges are maintained in the backbone (see Figure 5.1c).

Finally, we note that all isolated nodes are removed from the final network backbone (see, for example, nodes A, B, C, K, J, and L, no longer present in Figure 5.1c).

5.2.3 Community Detection

Once extracted the backbone B_{Δ_t}, our next step consists of identifying communities in B_{Δ_t}. Qualitatively, a community is defined as a subset of vertices such that their connections are denser than connections to the rest of the network. To extract communities from B_{Δ_t}, we adopt the Louvain algorithm [Blondel et al., 2008; Newman and Girvan, 2004] explained in Chapter 2. By doing so, we expect to capture communities of commenters *acting* on distinct sets of posts and Influencers represented by the universal set $\mathcal{P} = \{P_1, P_2, ..., P_{\Delta t_n}\}$.

5.2.4 Community Characterization

Once communities are extracted, we characterize them in terms of the textual properties of the content shared by their members as well as their temporal dynamics.

5.2.4.1 Content Properties

We analyze the discussions carried out by each community by focusing on the textual properties of the comments shared by its members. In particular, we employ three complementary textual analysis approaches.

First, we perform sentiment analysis using SentiStrength,[2] a lexical dictionary labeled by humans with multi-language support, including Portuguese and Italian. Given a sentence, SentiStrength classifies its sentiment with a score ranging from -4 (extremely negative) to +4 (extremely positive) [Thelwall et al., 2010]. SentiStrength has been widely applied to analyze the sentiment of social media content, notably short texts (e.g., tweets), for which identifying sentiment is usually harder [Ribeiro et al., 2016; Thelwall, 2017].

Second, we use *Term Frequency - Inverse Document Frequency* (TF-IDF) [Jones, 1972] to reveal terms that characterize each community. TF-IDF is traditionally used to describe *documents* in a collection with their most representative terms. Given a particular term and a document, the TF-IDF is computed as the product of the frequency of the term in the given document (TF) and the inverse of the frequency at which the term appears in distinct documents (IDF). Whereas TF estimates how well the given term describes the document, IDF captures the term's capacity to discriminate the document from others. To apply TF-IDF in our context, we represent each community as a *document* consisting of all comments of the community members. We pre-process the comments to remove emojis, stopwords, hashtags, punctuation and mentions to other users, perform stemming, as well as remove the overall top-1% most popular terms and rare terms (less than 10 occurrences).[3]

Each community is then represented by a vector d with dimension equal to the number of unique terms in the collection. The element $d[i]$ is the TF-IDF of term i. We here use a modified version of IDF, called probabilistic inverse document frequency [Baeza-Yates et al., 1999], which is more appropriate when the number of documents is small (as is our case). It is defined as $IDF(i) = \log \frac{N-n_i}{n_i}$, where N is the total number of communities and n_i is the number of communities using the term i.

[2] http://sentistrength.wlv.ac.uk/index.html

[3] The former are words whose frequency is extremely high and would not help to characterize the communities, while the latter are mostly typing errors or grammar mistakes.

We manually evaluate the terms with large TF-IDF of each community searching for particular subjects of discussion.

Last, we delve deeper into community contents using LIWC [Tausczik and Pennebaker, 2010], a lexicon system that categorizes text into psycholinguistic properties. LIWC organizes words of the target language as a hierarchy of categories and subcategories that form the set of LIWC attributes. Examples of attributes include linguistic properties (e.g., articles, nouns and verbs), affect words (e.g., anxiety, anger and sadness) and cognitive attributes (e.g., insight, certainty and discrepancies). The hierarchy is customized for each language, with 64 and 83 attributes for Portuguese and Italian, respectively. We apply LIWC to each comment of each community to quantify the fraction of words that falls into each attribute. We search for statistical differences across communities based on the average frequencies of their respective attributes. We first use Kruskal's non-parametric test to select only attributes for which there is a significant difference across communities [Kruskal and Wallis, 1952]. Then, we rank attributes with significant differences to select the most discriminative ones using the Gini Coefficient [Yitzhaki, 1979].

5.2.4.2 Temporal Properties

Finally, we analyze how communities evolve over time, both in terms of their memberships and the main topics of discussion. To analyze the dynamics of community membership, we use the two metrics, in particular *persistence* and *normalized mutual information*, explained in Section 4.2.6. For topic analysis, we start by focusing on the most representative terms used by each community, as captured by the TF-IDF metric, to examine the extent to which communities use the same lexicon in successive time windows. To that end, we first generate, for each time window, the vector representation of each identified community (as described in the previous section). Given the large size of the vocabulary, we consider only the top-100 words with the highest TF-IDF scores in each document, zero-ing other entries in the TF-IDF vectors. Next, we need to match the communities found in week Δ_{t2} to the communities found in week Δ_{t1} so as to be able to follow users commenting on the same topics across windows. Rather than doing so by using the structural information, we match them based on the topics or, more precisely, on the set of terms they used in each window.

Specifically, we use the cosine similarity [Baeza-Yates et al., 1999] of the TF-IDF vectors[4] to compute the pairwise similarity between all pairs of communities in windows

[4] The similarity between communities p_1 and p_2 is defined as $sim(p_1, p_2) = d_1 \times d_2$, where d_1 and d_2 are the TF-IDF vector representations of communities c_1 and p_2, respectively. Note that $sim(p_1, p_2)$ ranges from 0 (maximum dissimilarity) to 1 (maximum similarity).

Table 5.1. Dataset Overview (weeks including election dates are shown in bold in the respective country). The number of posts and commenters (comm.) by each scenario and week.

Week	Politics				General			
	Brazil		Italy		Brazil		Italy	
	# Posts	# Comm.	# Posts	# Comm.	# Posts	# Comm.	# Posts	# Comm.
1	1 487	37 406	779	17 427	746	172 454	733	54 407
2	1 648	67 799	739	20 873	778	180 711	703	49 290
3	1 798	103 506	742	20 876	719	164 040	594	52 052
4	1 951	94 327	907	21 402	854	186 333	649	54 677
5	**2 307**	**145 618**	1 080	22 029	680	125 414	683	52 318
6	958	184 993	1 240	22 890	771	158 522	720	69 066
7	1 195	123 797	**1 316**	**26 600**	723	131 563	657	61 168
8	**1 400**	**145 499**	701	31 308	798	152 705	635	66 337
9	799	191 282	762	17 171	733	146 128	540	31 520
10	606	50 546	656	19 926	763	159 628	507	33 781

Δ_{t1} and Δ_{t2}, matching each community $p_1^{\Delta t1}$ in window Δ_{t1} with the most similar one in window Δ_{t2}, provided that this similarity exceeds a given criterion of significance.

The criterion we adopt consists of comparing the similarity between two communities $p_1^{\Delta t1}$ and $p_2^{\Delta t2}$ and the similarity between $p_1^{\Delta t1}$ and an "average" community in window Δ_{t2}. Let $\mathbf{d}_1^{\Delta t1}$ be the TF-IDF vector representation of community j in window Δ_{t1}, we use *all* comments associated with window Δ_{t2} to compute its TF-IDF vector $\mathbf{d}_*^{\Delta t2}$ using the term frequencies in the complete *document* (i.e., all comments) but the IDF values previously computed considering individual communities in Δ_{t2}. In practice, the cosine similarity between the TF-IDF vectors $\mathbf{d}_1^{\Delta t1}$ and $\mathbf{d}_*^{\Delta t2}$ gives us a significance threshold for matching the communities, i.e., when $\text{sim}(\mathbf{d}_1^{\Delta t1}, \mathbf{d}_2^{\Delta t2}) > \text{sim}(\mathbf{d}_1^{\Delta t1}, \mathbf{d}_*^{\Delta t2})$, the similarity between $p_1^{\Delta t1}$ and $p_2^{\Delta t2}$ is larger than the similarity between p_1^w and an "average community" in window Δ_{t2}. In case no community $p_2^{\Delta t2}$ satisfies that condition, we deem that no match was found for $p_1^{\Delta t1}$. Instead, if we find a match, it means that we have a significant mapping between two communities in different windows.

5.3 Dataset

We now describe the dataset used in our study, which consists of over 39 million comments produced by over 1.8 million unique commenters, participating in discussions triggered by 320 top influencers over two countries (Brazil and Italy).

5.3.1 Dataset crawling

We collected data from Instagram profiles in Brazil and Italy. Our collection targets electoral periods to capture the political debate taking place on the social network. For

Brazil, we focus on Instagram posts submitted during the national general elections of October 7^{th} (first round) and October 28^{th} (second round), 2018. Our dataset covers 10 weeks (from September 2^{nd} until November 10^{th}, 2018) which includes weeks before and after the election dates. Similarly, for Italy we observed the European elections held on May 26^{th}, 2019, collecting data published from April 7^{th} to June 15^{th} (also 10 weeks). We monitor posts shared by selected profiles (see below), gathering all comments associated with those posts.

We use a custom web crawler to scrape data from Instagram that relies on the Instaloader library[5]. We performed the crawling in September 2019. Given a profile i, the crawler looks for posts i created during the predefined period. For each post, the crawler downloads all comments associated with it. As the interest in posts on Instagram tends to decrease sharply with time [Trevisan et al., 2019], we expect that our dataset includes almost all comments associated with posts created during the period of analysis. We focus only on *public* Instagram profiles and posts, collecting all visible comments they received. We performed the crawling respecting Instagram rate policies to avoid overloading the service. We did not collect any sensitive information of commenters, such as display name, photos, or any other metadata, even if public.

For each country, we monitor two groups of influencers:

- *Politics*: the most popular Brazilian and Italian politicians and official political profiles. We consider 80 profiles for each country. In total, the Brazilian politics profiles created 14 149 posts and received more than 8 million comments by 575 612 unique commenters during the monitored period. Similarly, the Italian profiles created 8 922 posts, which received more than 1.9 million comments by 94 158 distinct commenters.

- *General*: non-political influencers used as a control group. We rely on the HypeAuditor[6] rank to obtain the list of most popular profiles for the Sport, Music, Show, and Cooking categories in each country. Similarly to the *Politics* group, we pick 80 profiles for each country. The Brazilian *general* profiles created 7 565 posts and received 15 million comments by 295 753 distinct commenters during the monitored period. Similarly, the Italian general profiles created 6 421 posts and received 14 million comments carried out by 897 421 commenters.

[5] `https://instaloader.github.io`
[6] `https://hypeauditor.com/`

5.3.2 Data pre-processing

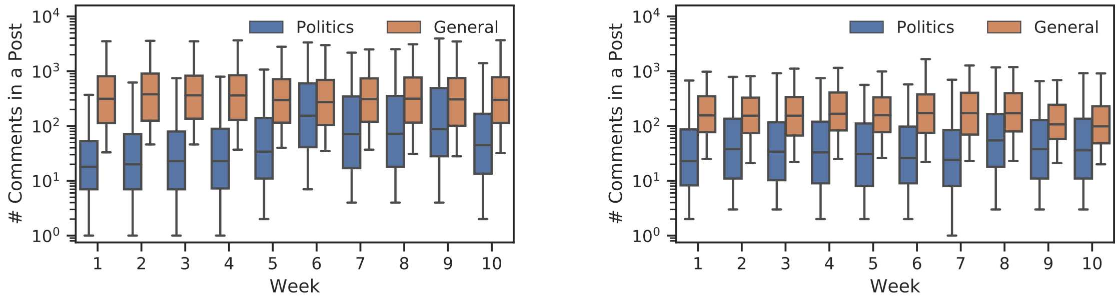
We only consider commenters who commented on more than one post when building the network for a given week Δ_t. This step removes 70–85% of the commenters. We observe that 95% of removed commenters commented fewer than three times when considering all period of the dataset. All results presented in the following refer to the dataset after removing these occasional commenters. To build the network of co-commenters, we aggregate posts by week separating data by country (Brazil and Italy) and category of influencers (general and politics).[7] We then use the comments these posts received to build the co-commenter network. This procedure generates 40 weekly-snapshots, here called *week* for simplicity: one for each of the 10 evaluated weeks, for 2 countries and 2 categories.

5.3.3 Dataset overview

Table 5.1 presents an overview of our dataset, showing the numbers of posts and distinct commenters per week. Election weeks are shown in bold. In Brazil, elections were on Sunday of the 5^{th} and 8^{th} weeks (1^{st} and 2^{nd} rounds, respectively), whereas the election in Italy took place on Sunday of the 7^{th} week. Focusing first on politics, we observe that the number of posts tends to steadily increase in the weeks preceding elections, reach a (local) maximum on the week(s) of the election, and drop sharply in the following. Interestingly, the largest number of commenters appears on the week immediately after the elections. Manual inspection reveals this is due to celebrations by candidates and supporters. Regarding the general category, we observe that the number of posts and commenters is rather stable, with a slight decrease in the last two weeks for Italy due to the approaching of summer holidays.

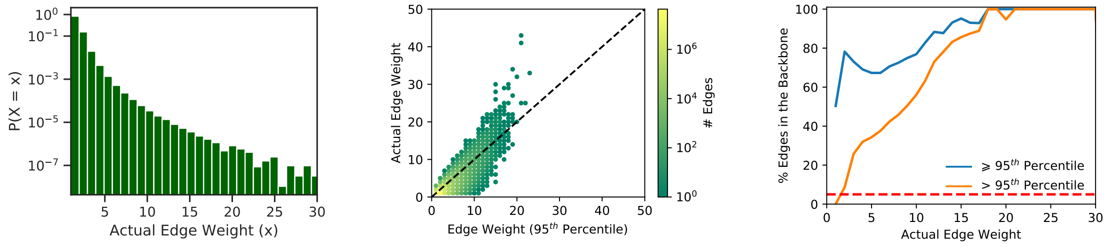
We complement the overview with Figure 5.2, which shows the distributions of the number of comments per post during each week. We use boxplots to ease visualization. The black stroke represents the median. Boxes span from the 1^{st} to the 3^{rd} quartiles, whiskers mark the 5^{th} and the 95^{th} percentiles. For politics, the median is a few tens of comments per post, while general posts receive 10 times as much (notice the log y-axes). Recall that the number of distinct commenters is similar on both cases (see Table 5.1), thus commenters are more active in the general profiles. Yet, posts of the main political leaders attract thousands of comments, similar to famous singers or athletes (holding for both countries). Considering time evolution, the number of comments on politics increases by an order of magnitude close to elections, with a sharper increase in Brazil.

[7]We consider weeks starting on Monday and ending on Sunday.



(a) Brazil. **(b)** Italy.

Figure 5.2. Distributions of number of comments per post (notice the log scale in y-axis).



(a) Probability distribution of edge weights $\gamma(cd)$ in the original graph G_P.

(b) Actual edge weight $\gamma(cd)$ compared with null model edge weight $\hat{\gamma}_{95}(cd)$.

(c) Fraction of edges included in the backbone B_P, separately for different values of γ.

Figure 5.3. Network characteristics for posts of influencers for Brazil - Politics (Week 1).

5.4 Structural analysis

We here describe the network structure emerging from our data. We first illustrate characteristics of the original and network backbones. Then, we characterize the communities and highlight insights emerging from the co-commenters backbones.

5.4.1 The Network Backbones

We first show an example of network backbone, using the 1st week of the Brazilian Politics scenario as case study.

Figure 5.3a depicts the histogram of the edge weights in the original graph G_P. Notice that 82% of edges have weight equal to 1, i.e., the majority of co-commenters co-comment in a single post. Higher weights are less frequent (notice the log scale on the y-axis). Yet, some co-commenters interact on more than 20 posts. In the following, we assess whether these weights are expected – i.e., their weights agree with

Table 5.2. Characteristics of the original network G_P and network backbone B_P for Brazil - Politics (Week 1).

Network	# Nodes	# Edges	# Comm	Modularity
Original	37 k	74.09 M	6	0.22
Backbone	26 k (70.7%)	1.06 M (1.4%)	19	0.59

the assumption of independent user behavior.

The scatter plot in Figure 5.3b compares the observed weight in G_P and the 95$^{\text{th}}$ percentile of weight estimated by our reference model $\hat{G}_P$. Colors represent the number of edges, and lighter colors indicate larger quantities. Most edges have very low value for both observed and estimated weights – notice the lightest colors for weights 1 and 2 in the bottom left corner. We are interested in the edges in which weights exceed the 95$^{\text{th}}$ percentile of the expected weight – i.e., those above the main diagonal. The fraction of edges over the diagonal is higher for larger weight values. This indicates that co-commenters interacting on many posts deviate from the expectation.

Figure 5.3c digs into that by showing the percentage of edges that are included in the network backbones separately by observed edge weight. If the null model held true, 5% of the edges would be included (those exceeding the 95$^{\text{th}}$ percentile) – highlighted by the red dotted line. But in G_P, edges weights do not always follow the null hypothesis of independent behavior, especially for edges with large weights.

It is also important to remark that G_P edge weights are integer numbers, and our generative model provides discrete distributions. Therefore, the computation of percentiles is critical since the same value can refer to a *range* of percentiles. This causes a rounding issue that is critical for low values. Filtering weights *greater than* or *greater or equal to* particular values results in significant differences for low weights. Figure 5.3c illustrates it by reporting the fraction of edges that would be included in the backbone in the two cases. Using *greater than* corresponds to a conservative choice since we include only edges for which the expected weight is strictly higher than the 95$^{\text{th}}$ percentile (orange curve). Notice how the number of edges in the backbone is reduced for low weights. Conversely, *greater or equal to* would preserve more edges, including those whose weight possibly corresponds to a lower percentile (blue curve). We here maintain a *conservative* choice and keep edges whose actual weight is strictly greater than the 95$^{\text{th}}$ percentile.

Table 5.2 describes the resulting network backbone B_P after filtering, comparing it with the original graph G_P. We focus on week 1 here, but results are consistent for all weeks. Our approach discards 98.6 % of the edges – i.e., the vast majority of them is not salient. We remove 29% of nodes, which remain isolated in B_P. To highlight the

Table 5.3. Breakdown of backbone and communities over different weeks for Brazil, Politics. In bold, the weeks of the elections.

Week	**% Nodes**	**% Edges**	**% Edges** $\gamma(cd) > 1$	**# Comm**	**Mod.**
1	70.69	1.40	11.43	19	0.59
2	93.36	2.11	12.19	27	0.64
3	73.81	1.01	4.75	20	0.52
4	93.63	2.23	15.10	32	0.69
5	**94.30**	**2.65**	**19.36**	**17**	**0.61**
6	91.49	2.36	19.37	31	0.66
7	94.05	1.87	15.45	31	0.66
8	**95.40**	**2.13**	**15.29**	**27**	**0.64**
9	68.01	0.62	4.06	24	0.59
10	71.33	1.11	7.21	29	0.61

benefits of the approach, we include the number of communities and the modularity in the original and backbone graphs. The Louvain algorithm identifies only 6 communities with very low modularity in the original graph. On the backbone, it identifies more communities, and modularity increases from 0.22 to 0.59.

Table 5.3 summarizes the main characteristics of the network backbones obtained on each week for Brazil, Politics. Focusing on the first four columns, notice that we still include the majority of nodes, with percentages ranging from 68% to 95%. Considering edges, the percentage is always low (0.6–2.6%). The fourth column reports the fraction on edges in the backbone having weight larger than 1. Remind that, by design, a random behavior would lead to 5% of edges in the backbone, while here we observe up to 19%, despite our conservative filtering criteria. Results are rather stable and consistent over time.

5.4.2 Communities of Commenters

We now study the communities obtained from the backbone graphs. The last two columns of Table 5.3 show that we obtain from 19 to 32 communities, depending on the week. Modularity values are high (always above 0.5), meaning that the community structure is strong.

We summarize results for the other scenarios in Table 5.4, reporting only average values across the 10 weeks. First, focusing on Politics and comparing Brazil and Italy (first two rows), we observe similar percentages of nodes in the network backbones. For Italy a larger fraction of edges are retained, potentially because of the smaller volume of profiles and comments (see Section 5.3). For Brazil, we obtain a larger number of communities with higher values of modularity than in Italy.

Moving to the General scenarios (3rd and 4th rows), we notice that fewer nodes and edges are in the backbones compared to Politics. Interestingly, we identify more and stronger communities. We root this phenomenon in the heterogeneity of the General scenarios that include influencers with different focuses, potentially attracting commenters with different interests. Manual inspection confirms the intuition – i.e., we find some communities interested in sports, others on music, etc. For politics, instead, we find a more tangled scenario. Even if communities are rather strong, some of them include profiles commenting on politicians of different parties and embracing different topics. Next, we evaluate communities in the Politics scenario.

5.4.3 Analysis of Political Communities

We now focus on Politics and show how the activity of commenters spreads across political profiles of different parties. Here we focus on the election week for both countries to better capture the peak of the political debate on Instagram.

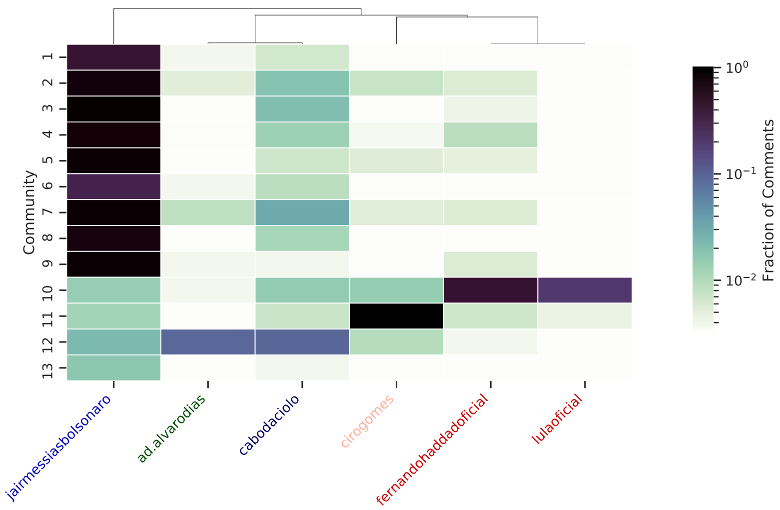
We first focus on the main political leaders of the two countries and study how the communities of co-commenters distribute their interests among their posts. We consider six politicians in each country. Figure 5.4 shows how the commenters of each community are spread among posts of each politician using a heatmap. Columns represent politicians and rows represent communities. The color of each cell reflects the fraction of the comments of the community members that are published on the posts of the politician.

To gauge similarity of profiles, the top of the heatmaps report a dendrogram that clusters politicians based on the communities of their commenters. We define as similarity metric of politicians the Pearson correlation among the activity of communities on their posts. In other words, we compare them by computing the correlation between the corresponding columns of the heatmap. Hence, two politicians that receive comments from the same communities have high similarity.

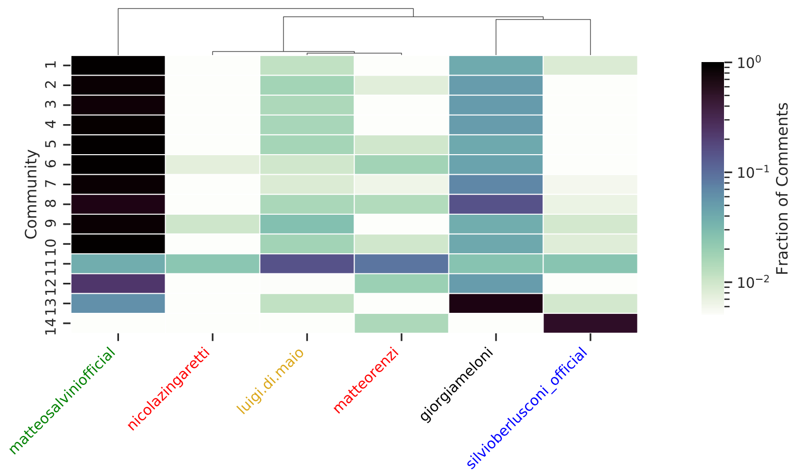
Looking at the Brazilian case (Figure 5.4a), we notice that most communities are interested in a single candidate - Jair Bolsonaro (jairmessiasbolsonaro), with the large majority of comments focused on his posts. This behavior is expected given his large number of followers and popularity. Indeed, communities $1-9$ comment

Table 5.4. Networks backbone and identified communities for Brazil (BR) and Italy (IT). We show average values over the 10 weeks.

Scenario	% Nodes	% Edges	% Edges $\gamma(cd) > 1$	# Comm	Mod.
BR Politics	84.61	1.81	12.42	26	0.62
IT Politics	87.33	3.39	21.79	11	0.44
BR General	65.35	0.82	8.83	81	0.79
IT General	60.03	2.23	12.57	48	0.72



(a) Brazil (1st round)



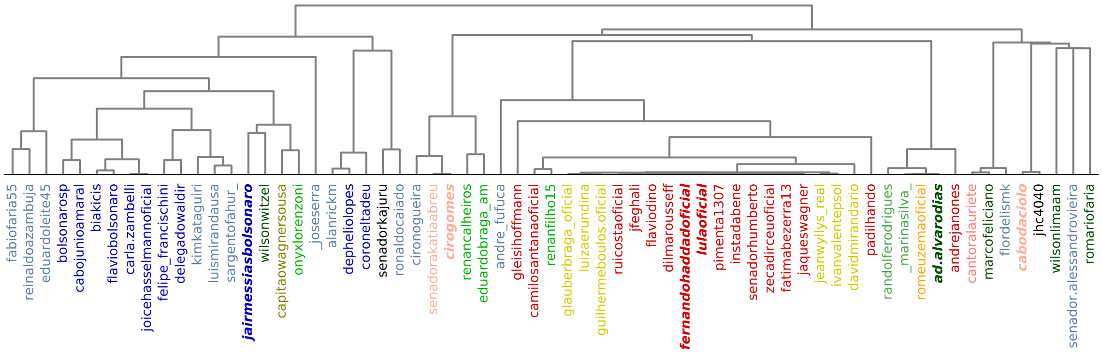
(b) Italy

Figure 5.4. Distribution of comments among political leaders for each community during the main election weeks.

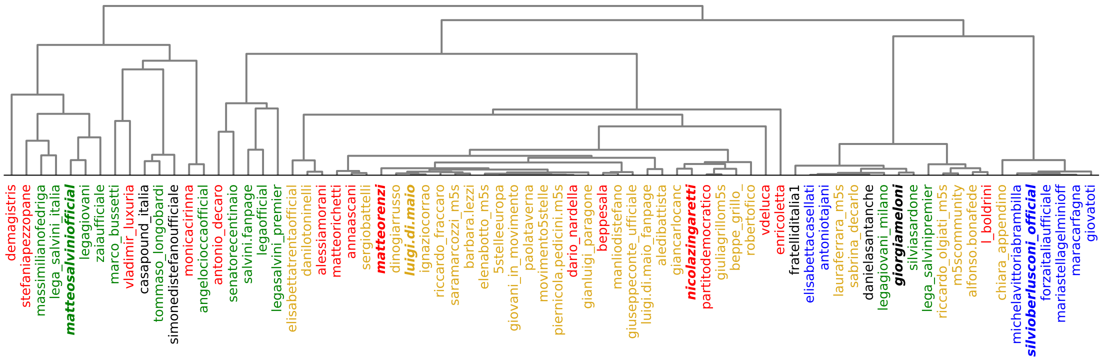
almost uniquely on Bolsonaro. Focusing on the dendrogram on the top of the figure, Bolsonaro has the highest dissimilarity from the others, i.e., he is the first candidate to be separated from others. Other clusters reflect accurately the candidates' political orientation. Left-leaning candidates (Ciro Gomes, Fernando Hadaad and Luiz Inacio Lula[8]) are close, as well as the ones leaning towards the right-wing parties (Alvaro Dias, Cabo Daciolo and Jair Bolsonaro).

Similar considerations hold for the Italian case (Figure 5.4b). Communities 1 – 10 focus on Matteo Salvini (matteosalviniofficial). He is the only one for which we

[8]Haddad replaced Lula, who was barred by the Brazilian Election Justice.



(a) Brazil.



(b) Italy.

Figure 5.5. Dendogram of political influencers clustered according to commenter communities. Influencers are colored according to their political coalition.

identify multiple and well-separated communities. The other right-wing leaders have communities active almost exclusively on their posts, e.g., communities 13 and 14 for Silvio Berlusconi and Giorgia Meloni. Other leaders (e.g., Matteo Renzi and Nicola Zingaretti for the Democratic Party and Luigi Di Maio for the Five Star Movement) share a large fraction of commenters in community 11. This suggests these commenters are almost equally interested in the three leaders. Indeed, looking at the dendrogram, these last three profiles are close to each other. Matteo Salvini (leader of the most popular party) has the maximum distance from others. Similar to the Bolsonaro's case, Salvini is a single leader who polarizes communities, thus well-separated from others.

We now broaden the analysis to all politicians. We label each politician according to his/her political *coalition* using available public information.[9] For Brazil, we rely on

[9]Differently from e.g., the US or UK, in both Brazil and Italy the political system is fragmented into several parties that form coalitions during and after elections.

the Brazilian Superior Electoral Court,[10] while for Italy we use the official website of each party. Rather than reporting the activity of each community on all politicians, we show only the dendrograms that cluster them, following the same methodology used in Figure 5.4.

Figure 5.5 shows the results, where the party leaders/candidates shown in Figure 5.4 are marked in bold. Politicians of the same parties appear close, meaning that their posts are commented by the same communities. For Brazil, the higher splits of the dendrogram roughly create two clusters, for left and right-wing parties. In Italy, we can identify three top clusters, reflecting the tri-polar system. Less expected are the cases in which politicians from distant political leanings attract the interest of the same communities and are close in the dendrogram. For example, in Italy, we find the profile of Monica Cirinnà (left-wing) very close to Angelo Ciocca (right-wing). Manual inspection reveals a considerable number of disapproving comments to posts of the first politician that are published by commenters supporting the second. The same happens for Vladimir Luxuria, whose some supporters disapprove Marco Bussetti's posts (and vice-versa). The structure of the backbone graph reflects the presence of profiles that bridge communities.

In sum, our methodology uncovers the structure of communities, which reflect people's engagement to politicians over the spectrum of political orientation. Most communities are well-shaped around single profiles, but sub-communities emerge too, possibly around particular topics, as we will check next. In some cases, commenters cross the expected political divide, commenting on profiles from different political orientations.

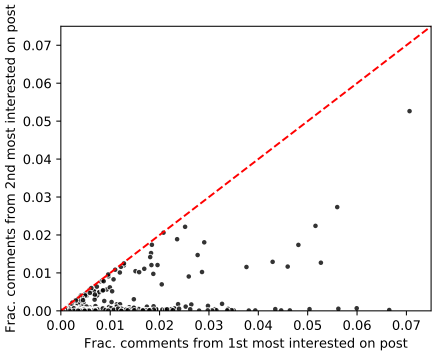
5.5 Textual Properties of Discussions

We now focus on how the communities differ in terms of textual, sentiment and psychological properties of comments.

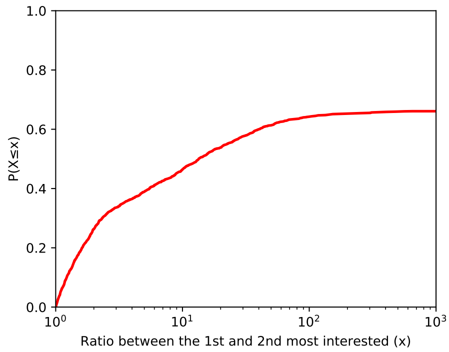
5.5.1 Political Communities' Interests

We now look into how communities in politics are attracted by different posts. Since communities differ in the number of members and in the number of comments they post, we consider a relative *interest index* of the community in a post, given by the

[10] http://divulgacandcontas.tse.jus.br/divulga/#/estados/2018/2022802018/BR/candidatos



(a) Index of interest of the 1^{st} and 2^{nd} most interested communities. Each point represents a post.



(b) Ratio between the index of the 1^{st} and 2^{nd} most interested community on posts.

Figure 5.6. Interest of communities on posts.

fraction of the community's comments going to the post. We use data from the week of the main elections in Brazil (week 5).

Figure 5.6a quantifies, for each post, the two most interested communities. The x-axis reports the index for the community with the highest interest on the post, while the y-axis reports the index for the second most interested community in the post. We observe that, in all cases, the most interested community leaves less than 7% of its comments in a unique post (see the x-axis). Given there are 2 144 posts in this snapshot, even a relative interest of 1% could be considered highly concentrated attention, suggesting that communities are built around specific posts. In $\approx$ 40% of the posts, the relative interest of the second most interested community (y-axis) is very low compared to the most interested one. We quantify this in Figure 5.6b, which reports the ratio between the relative interests of the first and the second most interested communities. We observe that, in the 55% of cases, the most interested community has at least 10 times higher index than the second one – notice the x-axis log-scale. Hence, we have strong evidences that communities are attracted by specific posts.

Figure 5.7 shows posts that attracted high interest from communities 3 and 7, which we use as running examples along with communities 10 and 11. Community 3 comments mostly on posts related to public events Bolsonaro promoted via Instagram (as in Figures 5.7a and Figures 5.7b), while community 7 comments on posts where the candidate tries to show his proximity with black people to debunk his associations with racism (Figures 5.7c and Figures 5.7d).

(a) Community 3 - Post about a rally in São Paulo. `www.instagram.com/p/BoXpvV6Hrkk`

(b) Community 3 - Post about a rally in Vitória. `www.instagram.com/p/BoXMwvwn6xj`

(c) Community 7 - Post discussing racism. `www.instagram.com/p/BomRItfH9p8`

(d) Community 7 - Another post discussing racism. `www.instagram.com/p/Boe7fQcHfJB`

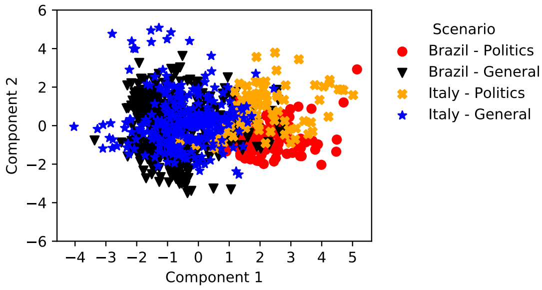
Figure 5.7. Examples of posts by Jair Bolsonaro (jairmessiasbolsonaro) in which two communities show high interest.

5.5.2 Properties of Communities' Comments

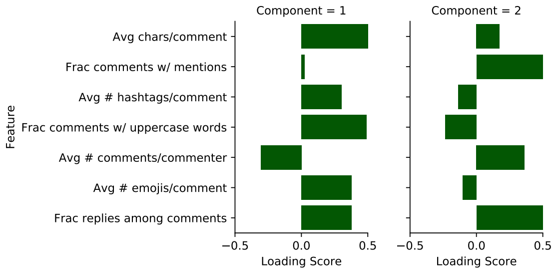
We now take all communities found in each week and extract properties of the comments of their members, namely: i) Average comment length (in characters); ii) Fraction of comments that include at least one mention; iii) Average number of hashtags per comment; iv) Fraction of comments with at least one uppercase word; v) Average number of comments per commenter; vi) Average number of emojis per comment; and vii) Fraction of replies among comments. Together these metrics capture important aspects of the communities' behavior. For example, the comment length, the number of emojis per comment and the use of uppercase words (commonly associated with a high tone) can describe the way the communities interact on Instagram. Mentions, the use of hashtags and replies are strongly associated with engagement, information spreading and direct interaction of commenters, respectively.

We study the communities by applying Principal Component Analysis (PCA) to vectors that represent communities using the seven previously described metrics. PCA is a well-known method for dimensionality reduction in multivariate analysis. It projects the data along its principal components (PCs), i.e., axes that capture most of the variance in the data [Tipping and Bishop, 1999]. Figure 5.8a shows the representation obtained for each community using the two principal components, where the color represents the pair country-scenario. The 2-D representations of communities for both politics scenarios are more tightly clustered and overlapping than for the general scenario. This behavior suggests that, when considering the given features, communities on politics are more homogeneous than the communities on the general scenario.

To understand which metrics best distinguish the communities in Figure 5.8a,



(a) PCA



(b) Description of the 2 first principal components

Figure 5.8. (a) 2-D representation of communities based on seven metrics using PCA. (b) Description of the two principal components in terms of the original metrics; the bar represents the loading scores for the components (positive or negative).

we study the *loading scores* for the two principal components. The loading score quantifies the contribution of each metric to a principal component. The largest the score (in absolute value) the more the metric contributes to the component (positively or negatively).

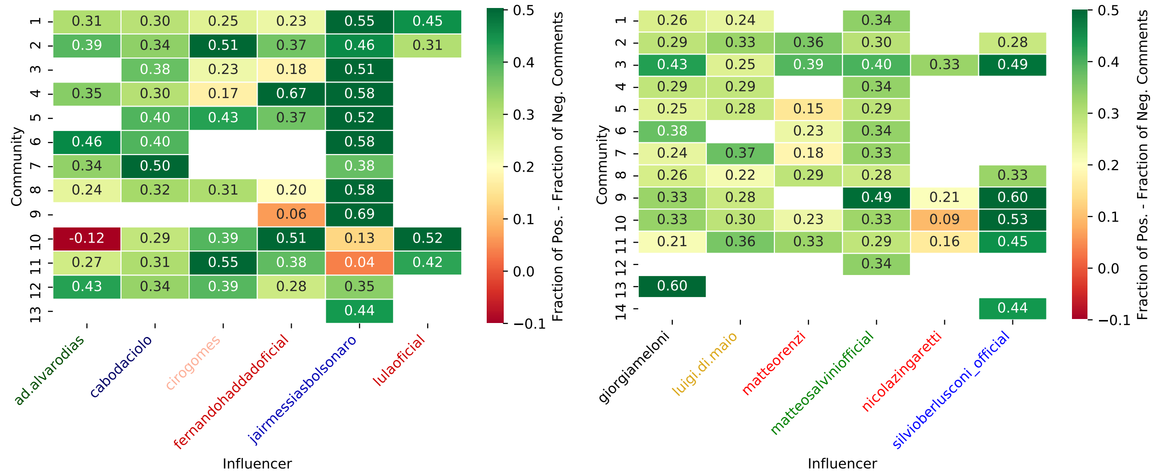
In Figure 5.8b bars represent the magnitude of loading scores for each metric for the PC 1 (left) and PC 2 (right). The PC 1 (left) can be associated with lengthy comments, high usage of uppercase, emojis, replies and hashtags, and a low number of comments per commenter. From Figure 5.8a, we see that high values for PC 1 is more common for communities in the politics scenarios. Conversely, most communities of the general scenario have negative x coordinates, thus pointing to the opposite behavior.

A less clear picture emerges for PC 2. Large values for PC 2 are associated with high number of replies, mentions and comments per commenter (see Figure 5.8b, right plot). For the politics scenario in Figure 5.8a, communities are concentrated in the $y \in [-2, 3]$ range, with those for Italy being slightly more positive than those from Brazil. In the general scenario, however, points are spread out along the y axis.

We conclude that commenters of politics influencers exhibit a more homogeneous behavior than commenters of other influencers. Particularly, commenters on politics leave larger comments and use higher tone. They also often rely on typical online social mechanisms, such as replies, mentions and emojis.

5.5.3 Sentiment Analysis

Although communities grow around particular posts and influencers, their members do comment on posts from other influencers. Here, we analyze whether there is a differ-



(a) Brazil (week 5) (b) Italy (week 7)

Figure 5.9. Contrastive sentiment score (difference between fraction of positive and negative comments) of communities towards political leaders during the main election week.

ence between the sentiment expressed in comments across influencers. As explained in Section 5.2, we use SentiStrength to extract the sentiment of each comment. SentiStrength provides an integer score ranging from -4 (strongly negative) to +4 (strongly positive). Score 0 implies a neutral sentiment. We here consider as *negative*, *neutral* and *positive* comments with scores smaller than 0, equal to 0, and greater than 0, respectively.

Table 5.5 shows fraction of positive, neutral and negative comments. We notice that positive comments are generally more common (between 47% and 58%), followed by neutral comments (between 32% and 41%). We look into the neutral comments to understand why they represent a significant fraction and observe a large number of short comments, misspelled words, abbreviations etc, which seem to complicate sentiment extraction by SentiStrength. Negative comments are the minority in our data, but they are more prevalent in the politics scenarios for both countries.

We now analyze how the communities' sentiment varies towards profiles of different politicians. More specifically, we compute the breakdown of positive, neutral and negative comments of each community on posts of each influencer. To summarize dif-

Table 5.5. Fraction of sentiment captured in comments using SentiStrenght.

Scenario	**Sentiment**		
	Negative	**Neutral**	**Positive**
BR Politics	0.10	0.32	0.58
IT Politics	0.13	0.40	0.47
BR General	0.06	0.41	0.53
IT General	0.04	0.36	0.57

ferences, we report in Figure 5.9a and Figure 5.9b a *contrastive score* calculated as the difference between the fractions of positive and negative comments for the particular community and influencer. We ignore cases where a community has made less than 100 comments on a given influencer's posts to ensure that samples are representative. These cases are marked as white cells in the heatmaps.

In Figure 5.9a we consider the six political leaders already used for Figure 5.4. We focus on the week of the first election round in Brazil (week 5). Predominantly, communities make positive comments on the profiles in which they are more active, i.e., their "referring candidate". More negative comments are seen on "opposing candidates". For instance, communities 1 to 9, highly active on Jair Bolsonaro's posts, display a more positive contrastive score on them. Analogously, communities 10 to 12, mostly formed by commenters very active on the profiles of left-wing influencers such as Ciro Gomes (cirogomes) and Fernando Haddad (fernandohaddadoficial), tend to write negative comments on their opponents, such as Jair Bolsonaro. This behavior appears on all weeks and suggests that communities in politics tend to promote their central influencers while trying to demote others.

Considering the Italian case, we observe similar results in Figure 5.9b. Communities exhibit positive contrastive scores towards candidates in general, but with higher scores for the referring candidate.

5.5.4 Main Topics of Discussion

We now turn our attention to the analysis of the main topics around discussions. As before, we focus on politics, during the election weeks. To summarize the overall behavior of each community, we group together all their respective comments in one document. As explained in Section 5.2, the documents build a corpus on which we then use the TF-IDF metric to identify the most representative words of each document (i.e., community), henceforth called *top words*.

We show in Table 5.6 the top-10 words (translated to English) for communities yielding the most interesting observations. We manually inspect the comments and related posts, providing a reference context as the last column of the table. The manual inspection suggests that these words give a good overview of what the communities discuss.

Table 5.6. Example of words with the highest TF-IDF for some communities in the politics scenario in the main election week.

Scenario	Comm.	Key Words	Context
BR	3	‘Anapolis’, ‘Orla’, ‘Righteousness’, ‘Constant’, ‘Natal’, ‘Paulista’, ‘Spontaneous’, ‘JB17’, ‘Gesture’, ‘Avenue’	It refers to several places where pro-Bolsonaro rallies took place during the election campaign.
BR	7	‘Nazi’, ‘Jew’, ‘Hitler’, ‘Black People’, ‘Anonymity’, ‘Bozonazi’, ‘Distrust’, ‘Jerusalem’, ‘Homosexual’	It refers to Bolsonaro’s posts about specific social groups in an attempt to show he has no prejudice against such groups.
BR	10	‘Manuela’, ‘Haddad, ‘Scammer’, ‘Lulalivre’, ‘Birthday’, ‘Guilherme’, ‘Dilma’, ‘Gratefulness’, ‘Lula’	It refers to left-wing names, such as Fernando Haddad, his deputy Manuela, Dilma Rousseff and Lula (ex-presidents).
BR	11	‘Ciro’, ‘Experience’, ‘Political Activism’, ‘Polarization’, ‘Brazil’, ‘Second Round’, ‘Turn’, ‘Prepared’, ‘Project’	It refers to the center-left candidate Ciro Gomes who arrived close to reach the second round of the elections.
IT	3	‘Gooders’, ‘Big ciao”, ‘Captain’, ‘Crime’, ‘Good night’, ‘Polls’, ‘Never Give Up’, ‘Electorate’, ‘Lampedusa’, ‘Riace’	General Salvini’s jargon, as well as places related to the arrival of immigrants in Europe (e.g., Lampedusa).
IT	4	‘Monetary’, ‘Elite’, ‘Unity’, ‘Budget’, ‘Fiscal’, ‘Colonial’, ‘Equalize’, ‘Yellow Vests’, ‘Masonic’, ‘Store’, ‘IVA’,	Generic taxes and monetary issues.
IT	10	‘Consumption’, ‘Fuel’, ‘Insurance’, ‘Traffic’, ‘Helpless’, ‘Vehicular’, ‘Taxes’, ‘Redundancy’, ‘Veterinary’, ‘Animal rights’, ‘Cats’, ‘Abuse’, ‘Cruelty’, ‘Breeding’	A combination of terms related to taxes, vehicles and animals’ rights.
IT	11	‘5S’, ‘Toninelli (ex-Transport Minister)’, ‘Corruption’, ‘Zingaretti (PD’s leader)’, ‘Calenda (ex-PD politician)’, ‘Honesty’, ‘Election list’, ‘Coalition’, ‘Budget’, ‘Growth’	Debate on Five Stars Movement (a government party at the time) and Democratic Party (the main opposition party at the time)

Matching with Figure 5.7, communities 3 and 7 for Brazil are associated with rallies in different locations in the country, and with debunking Bolsonaro's prejudice against ethnic and racial groups. The terms highlighted by TF-IDF reflect quite accurately the respective topics, reporting locations of rallies and words linked to racism and Nazism. Similarly, the top words for communities 10 and 11 are focused on the names of the candidates whose profiles they mostly comment on. For Italy, community 3 reflects the typical jargon used by Salvini's supporters. Community 4 debates on taxes and monetary issues. Community 10's comments refer to provoking posts that mix taxes, car costs and animals' rights. Last, community 11 seems to debate over the left-wing party (the main opposition party at the time) and the 5-Stars movement (the governing party at the time).

In a nutshell, the TF-IDF is instrumental to analyze what the communities are discussing. The analysis demonstrates that communities are well-formed around the topics they discuss, even if they have been built solely on the network of commenters' interactions.

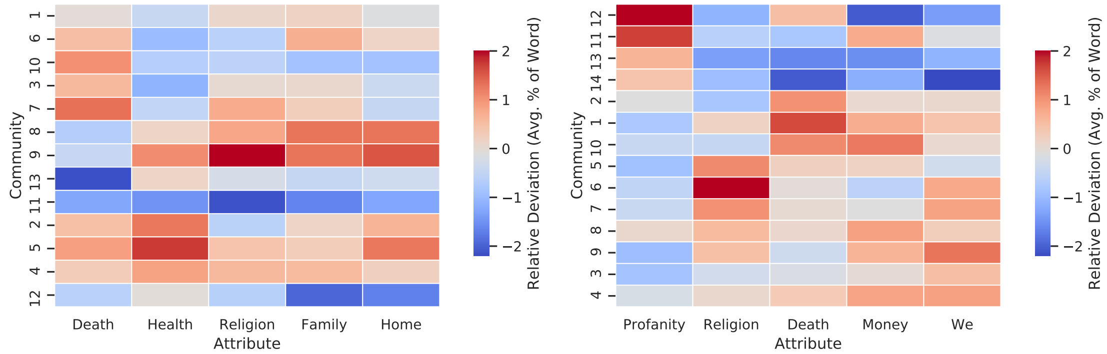
5.5.5 Psycholinguist Properties

In this section, we study the psycholinguistic properties of comments, aiming at finding similarities and differences in the way commenters of communities communicate. We rely on the Linguistic Inquiry and Word Count (LIWC) tool to calculate the degree at which various categories of words (called attributes in the LIWC terminology) are used in a text (see Section 5.2). For example, attribute *Home* includes the words "Kitchen" and "Landlord, and attribute *Family* "Daughter", "Dad" and "Aunt".

For each community, we run LIWC on the comments and compute the average frequency of the attributes. We then look for statistical differences between communities based on the average values of the attributes. For Brazil, we identify 62 attributes (from the 64 available in LIWC's Portuguese dictionary) for which differences across communities are statistically significant[11]. For Italy, we identify 77 (from 83 available in the LIWC Italian dictionary). From those, we select the five attributes that exhibit the largest variability across communities in terms of Gini index and use them characterize the psycholinguistic of communities.

Figure 5.10 shows heatmaps for the top-five attributes found for the Brazilian (top) and Italian (bottom) politics scenarios. The heatmap cells in a column indicate the relative deviation of the given attribute for the given community from the other

[11] We used the Kruskal non-parametric test to select attributes with a significant difference between the distribution of community comments considering $p - value = 0.01$.



(a) Brazil, Politics. **(b)** Italy, Politics.

Figure 5.10. Top 5 LIWC attributes and their relative difference between communities.

communities. In other words, each column (attribute) is z-score normalized – i.e., $z = (x - mean)/std$. Thus, each value gets subtracted the average of the column, then divided by the standard deviation of the column. The results show how communities are different in terms of the LIWC selected attributes. For instance, for Brazil, Politics, communities 6, 10, 3 and 7 frequently use words regarding *death*, but seldom words related to health. Communities 2, 5 and 4 show positive scores on almost all attributes. Community 13 focuses mostly on *health*. In Italy, community 6 is very focused on *religion* (commenters debated Salvini's post that depicts a Rosary). Community 12 and 13 exhibit some hate speech.

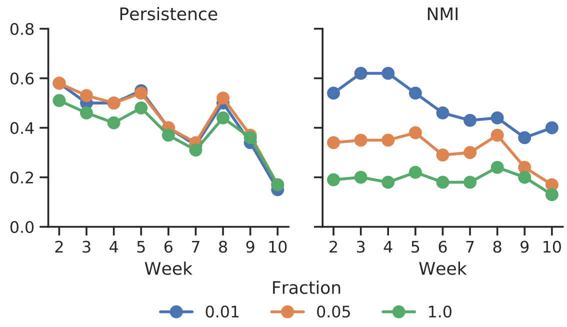
In summary, LIWC is a useful tool to analyze the content of Instagram comments, complementing the TF-IDF analysis with information on the topics being debated. We find that communities debate on different topics and using different lexicon.

5.6 Temporal Analysis

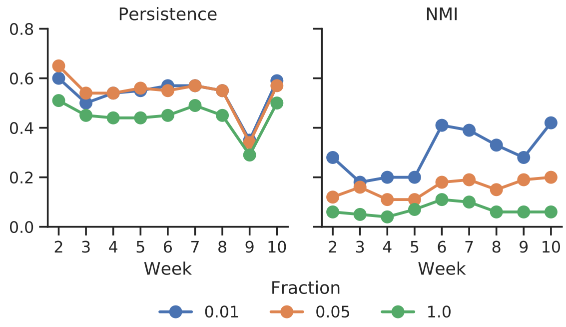
In this section, we focus on the dynamics of communities during the 10 weeks of observation. First, we analyze the community membership, studying to what extent commenters persist in the network backbone and are found in the same communities across weeks. Next, we characterize the dynamics of the content, i.e., the topics that these communities are engaged in.

5.6.1 Community Membership Persistence

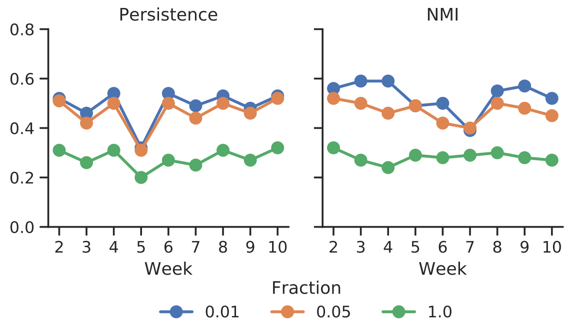
We start our analysis by studying the persistence of commenters inside the network backbone and to what extent these commenters end up in the same community week



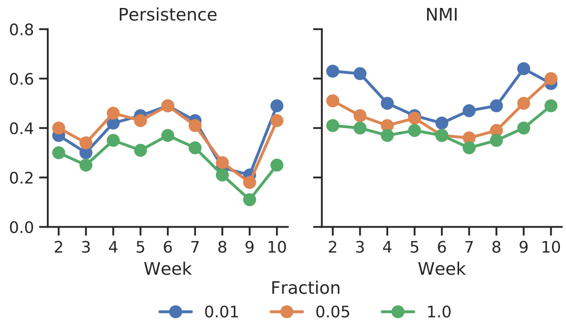
(a) Brazil, Politics.



(b) Italy, Politics.



(c) Brazil, General.



(d) Brazil, General.

Figure 5.11. Temporal evolution of commenters in communities. Blue: top 1%, Orange: top 5%, Green: all commenters.

by week. We also want to check if the most engaged commenters exhibit a different behavior – i.e., tend to persist more than those who are less engaged. To this end, we perform a separate analysis selecting the top-1% and top-5% commenters in terms of number of comments in week Δ_t and Δ_{t+1}. Then, we compute the persistence and NMI score (see Section 5.2.4.2), restricting to these commenters and comparing the results with those obtained with the full set of commenters.

We report results in Figure 5.11 separately by country and for Politics and General. Considering Politics (Figures 5.11a and 5.11b), we note that the persistence in Brazil is moderately high, regardless the subset of commenters. Around 50-60% of commenters remain in the backbone week after week until the first round of elections (week 5). Since then, we observe a decrease (also due to the drop of commenters in general) until the second round election (week 8), followed by a significant drop after. This trend shows that commenters were very engaged in the election period, mostly in the first round when the debate included more politicians, senators, congressmen and governors. In the second round, fewer candidates faced – yet people were consistently engaged before finally plumbing two weeks after elections. These results corroborate the first intuition we observed in Table 5.1 – where the number of commenters varied over time. Since persistence is similar for all subsets of commenters, we can conclude that all commenters in the backbone are persistently engaged. That is, the backbone members are quite stable.

Considering the membership of commenters within the same community, the NMI shows that the top-1% and top-5% most active commenters (blue and orange curves) are considerably more stable in their communities during the whole time. When considering all commenters in the backbone, the NMI is significantly lower. This is due to the birth and death of new communities, centered around specific topics, where the debate heats up and cools down. These dynamics attract new commenters that afterward disappear or change community.

For Italy, Politics (Figure 5.11b) different considerations hold. The constant persistence suggests a stable engagement of commenters in the backbone. We just observe a sudden drop the week after the election, where the interest in the online debate vanished. On the other hand, the NMI is rather low, revealing more variability in community membership, even if we restrict our attention to the most active commenters. Despite commenters in the backbone tending to be the same (persistence is typically above 0.5), they mix among different communities. Considering the low modularity of communities for this scenario (see Table 5.4), we conclude that the community structure is weaker in this case, indicating overlapping among communities that favor membership changes. This result is also visible from the dendrogram in Figure 5.5, where we observe that influencers receive comments from similar communities making the latter also more clustered.

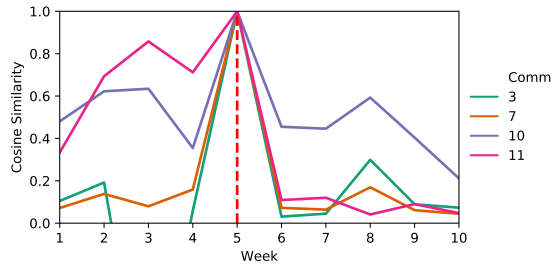
Moving to General (Figures 5.11c and 5.11d), we observe slightly lower persistence than in Politics, but more stable over time. NMI instead often results higher for General than Politics, reflecting better separation between communities, which persist over time. More in detail, for Brazil (Figure 5.11c) we observe that persistence and NMI are high and stable – especially for the most active users. This suggests that the most engaged commenters have diverse, specific and stable interests. Indeed, here there is no exogenous event that pushes a temporal dynamic, like elections do for politics. Again, this result reflects the high heterogeneity of posts and influencers in the General category. Moving to Italy, Figure 5.11d shows that persistence is small and varies over time. Here the lower popularity of Instagram in Italy than in Brazil may play a role, coupled with the smaller number of comments (see Table 5.1). However, NMI is high and stable. We conclude that although many users do not persist in the backbone, the remaining are very loyal to their interests.

5.6.2 Topic Persistence

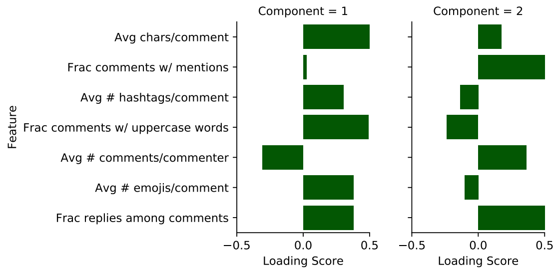
We now discuss how the topics discussed by communities evolve over time. To that end, we take as reference weeks 5 for Brazil and 7 for Italy in the political scenario, being the weeks of elections in each country. We compute the cosine similarity between the communities in the reference weeks (illustrated in Table 5.6) and the communities extracted in all other weeks, for each country. That is, for a given week, we identify whether there exists a document/community that is significantly similar to those found in the reference week, following the steps presented in Section 5.2.4.2.

Figures 5.12 show examples for both scenarios. In weeks 5 and 7 for Brazil and Italy, respectively, the cosine similarity is 1 since the documents are compared with themselves. Focusing on Brazil first, we observe very distinct behaviors among the picked-up examples. Remember that communities 3 and 7 are focused, mainly, on Bolsonaro's profile and comment on posts related to *rallies* and *racism*, respectively. In both cases, we can observe that discriminating terms for these communities are momentary and sporadic, with some communities using terms about *rallies* that appear in some weeks, still with a very similarity low. Conversely, the set of significant terms representing community 10 and related to candidate Fernando Haddad. At last, consider community 11, focused on Ciro Gomes. Again, we can observe that terms used by his community in the election week were used in some communities earlier, exhibiting high similarity. However, immediately after the first round (week 5) when Ciro Gomes lost, the similarity drops significantly. Indeed, Ciro Gomes was excluded from the run-off and the online debate (and community) suddenly vanished. In Italy (Figure 5.12b) we observe a similar behavior. Community 3, mostly consisting of Salvini's supporters, uses very specific jargon and are always present. Community 4 debates around taxes and monetary issues were already debated during week 3. The same considerations hold for community 10, in which the fight between democrats and five stars supporters heats more frequently.

In summary, people commenting in politics are more volatile than those commenting on general topics, with debates suddenly growing and cooling down. Some sets of terms remain "alive" throughout the observation period, while others include communities born around short events such as rallies, which take place on a specific date.



(a) Brazil - Politics.



(b) Italy - Politics.

Figure 5.12. Example of how communities' comments change over time. We set weeks 5 and 7 as reference, being the election weeks in Brazil and Italy, respectively.

5.7 Discussion

In this chapter, we studied the political discussion on Instagram, focusing on the co-interactions of users co-commenting on posts of influencers covering ten weeks of data from Brazil and Italy during major electoral periods in both countries to study politicians and other influencers. We first introduced TriBE, a novel method for backbone extraction based on a probabilistic model that considers features particular to online social media applications. Our model targets commenters' noisy and sporadic nature, removing network edges that emerge by side effects while revealing the salient co-interactions that compose the underlying network, i.e., the network backbone.

Then, we performed our study on a large dataset of Instagram comments, including approximately 1.8 *million* unique commenters on 36 824 posts by 320 influencers in two countries (Brazil and Italy). From a structural view, the analyses of the extracted backbones revealed the existence of stronger well-structured communities, especially around politics. We observed that communities built the same influencers, although communities grew around specific topics. Also, those communities around politicians have distinguishing textual properties that reflect more assertive and engaged discussions such as emotional content, longer comments containing more emojis, hashtags, and uppercase words. Finally, we analyzed the temporal evolution of the communities. In general, we observed that communities formed around political influencers are more dynamic than those formed around non-political influencers, which may be related to the topics associated with the posts and the evolution of the electoral process.

Moreover, we observed that communities in politics are more dynamic than non-political influencers regarding temporal evolution. Notably, we showed that the interest in particular discussions changes drastically over successive weeks, possibly reflecting shifts of interest occurring in society as the electoral process evolves. This observation contrasts with the communities in non-political cases, which are more stable over time.

In addition, we observed great variation in community membership over successive weeks in politics, although the most active commenters tend to remain consistently active in the same communities over time. Finally, concerning discussion topics, we observed great diversity in dynamics: Whereas some topics attract attention momentarily, others remain active over time.

5.8 Summary

In this chapter, we revisited our RG1 and RG2 established in section 1.3 by applying our general approach to the study of online discussions on Instagram. In contrast to our first case study, we addressed a more challenging phenomenon from a modeling perspective, whose study in four different scenarios resulted in a large amount of data over a long period of time. First, we studied the interactions between users on this platform, for which there are no antecedents in the literature, by modeling them as co-commenters networks. In addition, we proposed TriBE, a novel backbone extraction method, to reveal how the underlying structure of co-commenters networks facilitates the dissemination of information through online discussions. Assuming independent user behavior, TriBE was able to capture the interactions that act as triggers to encourage users to comment on posts. Furthermore, by analyzing the properties of the communities that emerged on the backbones, we found that the communities that emerge from a stronger and clearer structure unveiled by TriBE do indeed capture the collective behavior of co-commenters.

In summary, TriBE takes into account fundamental characteristics of the phenomenon and the analyzed system (social media platforms) such as the tripartite structure (content creators, content sets, and users interested in a particular subset of content), the heavy-tail character of the content, and users' popularity which generates a large number of edges that are not necessarily relevant to such a phenomenon. Therefore, our analyzes also open up several avenues for further research in this area, such as extending the study to other platforms like Facebook and Twitter that have a similar structure, which could uncover both consistent and different patterns than those found here.

In the next chapter, we delve into the central point of our work, backbone extraction, and propose a methodology that extends our general approach for selecting, experimenting, and evaluating different backbone extraction methods (including TriBE) for a given phenomenon.

Chapter 6

Selecting and Evaluating Backbone Extraction Methods

In this chapter, we dive into the core of our general approach - backbone extraction - and present a methodology for selecting and evaluating backbone extraction methods for a given phenomenon. We first introduce the motivation for this methodology in Section 6.1, followed by an overview of the problem and its formal statement in Section 6.2. We then describe our proposed methodology in Section 6.3 and show how it can be applied to two different case studies in Sections 6.4 and 6.5. We discuss our findings in Section 6.6. Finally, Section 6.7 summarizes the implications and contributions.

6.1 Motivation

In the last chapters, we have shown that the complexity and diversity of interactions between users in *many-to-many* networks pose some challenges. In other words, the presence of a large amount of sporadic interactions affects the understanding and interpretability of the phenomenon in question. To address this problem, we have proposed and applied a general approach based on smart algorithms whose goal is to select the *salient* edges to a given phenomenon in order to obtain a reduced and representative version of the network, the *network backbone*. Thus, it is intended to be a more representative model of the collective behavior driving the phenomenon under study. However, the definition of edge salience is highly subjective and several methods for extracting the network backbone are available in the literature, each containing specific assumptions about edge salience [Slater, 2009; Radicchi et al., 2011; Grady et al., 2012;

Coscia and Neffke, 2017; De Melo et al., 2015; Dianati, 2016; Marcaccioli and Livan, 2019; Ferreira et al., 2020, 2021]. Therefore, it is often quite difficult to decide which of these methods should be applied to a particular phenomenon.

In addition, the general lack of ground truth for evaluating the quality of an extracted backbone challenges the analysis of certain methods. In most previous studies, authors evaluated results based on topological metrics, such as community modularity, density, clustering coefficient, arguing that the extracted backbone has more clearly defined substructures than the original network [Dai et al., 2018; Neal et al., 2021; Mukerjee et al., 2022]. More recently, a few studies have looked at regression models that relate topological properties of the network backbone to phenomenon-specific attributes [Coscia and Neffke, 2017; Marcaccioli, 2020; Coscia, 2021]. In general, however, the existing literature lacks a principled methodology for selecting the most appropriate method for extracting backbones for a given phenomenon, taking into account both topological and contextual aspects.

In this chapter, we take a step towards filling this gap by presenting *a methodology for selecting and evaluating methods for extracting networks based on a phenomenon*, which extends a fundamental step of our general solution. Unlike most previous work, we argue here for a more *principled approach to selecting the most appropriate backbone extraction method*, in which the characteristics of the phenomenon under study should be aligned with the key assumptions and properties of the method. We begin with a discussion of nine backbone extraction methods that we reviewed in Chapter 2. We then challenge them by highlighting their key assumptions and characterizing them in terms of the salience criteria they capture. Finally, we consider two case studies to validate our methodology: (i) online discussions on Instagram (as studied in the previous chapter) and (ii) information spreading in social media applications on WhatsApp.

6.2 Problem Statement

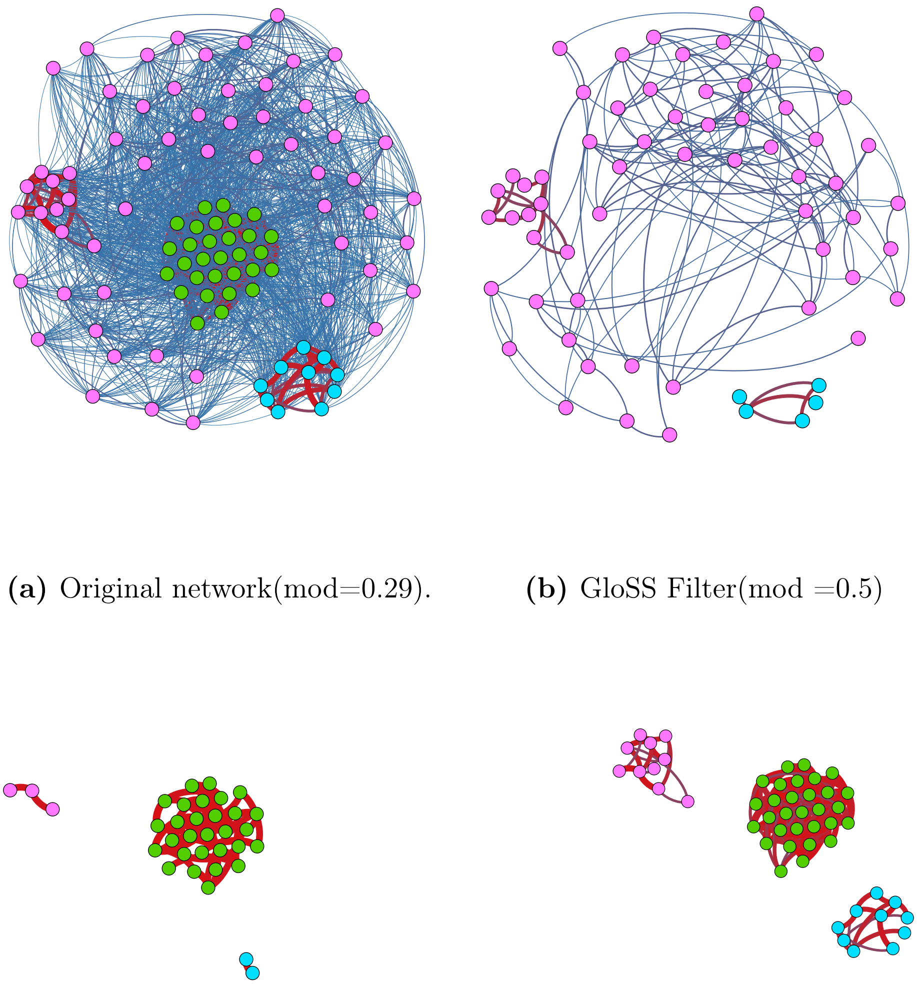
We tackle the challenge of selecting and evaluating network backbone extraction methods available in the literature given a target phenomenon to be studied. Inherent to such problem is the use of a potentially *noisy* network to model interactions driving the given phenomenon. Recall that by *noisy* we mean a network that may contain a large number of spurious edges that are not relevant for understanding the phenomenon at hand, and, even more, may obfuscate the relevant ones (i.e., the salient edges), jeopardizing the understanding of the phenomenon and the validity of conclusions drawn from the study.

Problem Statement: Given (i) a particular phenomenon of interest driven by collective behavior, and (ii) a dataset capturing real interactions that represent manifestations of such phenomenon, *how can we evaluate alternative network backbone extraction methods and select the one that, when applied to a network model of the input interactions, is able to accurately reveal key properties associated with the phenomenon of interest?*

One major assumption that guides our present effort is that not every network backbone is adequate to study the given phenomenon. Rather, key characteristics of such phenomenon must be matched to the assumptions and requirements of each method. Thus, a characterization of these properties is of utmost importance to drive the analysis. A mismatch between those characteristics, assumptions and requirements may lead to biases and misinterpretations.

Specifically, we are interested in identifying the methods that provide the best agreement between topological properties associated with the connectivity of users in the network and the contextual properties associated with factors driving the phenomenon that emerges from those patterns. Since our interest is in collective behavior patterns, the topological properties of interest are mostly associated with *communities* representing tightly connected groups of users who exhibit common (collective) behavior. One key challenge we must face is that each backbone extraction method removes edges and, consequently, nodes from the network based on its own definition of edge salience and noise. Thus, backbones extracted by different methods may reveal different topological structures, with properties that, though possibly strong and clear, may not be relevant (or related) to the phenomenon being studied.

Let us start presenting a simple case to exemplify the complexity of the problem. Consider the network in Figure 6.1-a) built by connecting different users (nodes) who shared the same piece of content on WhatsApp. Edges are weighted by the number of times they shared the same content. This network, consisting of 100 nodes and 2 522 edges, is a subgraph of the network analyzed in Section 6.5. Suppose we build this network to investigate evidence of users coordination to speed up content spreading dissemination on the platform. Figures 6.1-b), 6.1-c) and 6.1-d) show the backbones extracted from the network by three different methods, namely a threshold-based method, Gloss Filter and Disparity Filter. Note that nodes that end up isolated after edge removal are also removed from the backbones. Thus, each backbone contains a different subset of the original edges and nodes. The question that arises is: *Which backbone is the best one to study our phenomenon of interest, i.e., coordinated behavior*?



(a) Original network(mod=0.29). **(b)** GloSS Filter(mod =0.5)

(c) Threshold(mod=0.46) **(d)** Disparity Filter(mod=0.36)

Figure 6.1. Example network and the backbones extracted from it by three different methods (modularity values presented within parentheses). Edge thickness represents edge weight and nodes' color possible coordinated users' communities.

Note first that all methods remove a large fraction of the original edges, underlining the large presence of spurious noisy edges in the original network. Indeed all three methods reveal clear topological structures in terms of communities, as can be seen visually and by the modularity [Newman and Girvan, 2004] computed for the complete network and backbones (see values in captions of the figures). The backbone extracted by Gloss Filter (Figure 6.1-b) is quite different from the other two: it completely misses the connections among the green nodes in the center of the network, which clearly form a very tightly connected community. These users shared the same content many times, which is a strong evidence of coordination. Moreover, it keeps a large number of (weaker) edges among the various pink nodes: it is not clear whether these edges are indeed relevant to study coordination. The backbones extracted by the threshold-based and Disparity Filter methods (Figures 6.1-c) and 6.1-d)) look more similar. Both reveal three tightly connected groups of users. However, the threshold-based model misses some strong edges among users and the resulting backbone ends up with fewer nodes and smaller communities.

Intuitively, this example suggests that Disparity Filter was able to reveal a more complete view of these communities, thus allowing a more thorough study of the phenomenon under investigation[1]. Note that such conclusion cannot be based solely on topological/structural metrics. Indeed the modularity of the backbone extracted by Disparity Filter is the lowest among the three backbones (but higher than the modularity of the original network). However, the presence of communities formed by users tied by edges with higher weights is more visible in the backbone extracted by the Disparity Filter. For this, we claim that the method selection must also consider contextual aspects related to the phenomenon under study.

To generalize and tackle our target problem, we propose a methodology consisting of four steps, which we detail in the next section. We explicitly target the studies of collective behavior emerging from social media applications. We validate our methodology by applying it to two case studies, which are presented in Sections 6.4 and 6.5. Although our case studies are drawn from the context of social media, the methodology we propose is general enough to be applied to phenomena in other online and offline domains that are also modeled by noisy networks (e.g., co-authorship networks [Galuppo Azevedo and Murai, 2021]).

6.3 Methodology

Our proposed methodology consists of the 4 steps shown in Figure 6.2, a natural extension of our overall solution presented in Chapter 3. Given a dataset capturing user interactions that represent manifestations of collective behavior patterns driving a phenomenon of interest, we start by building a network model of these interactions (step 1). This network potentially contains spurious and random edges with little relevance

[1] We cannot claim it was able to reveal all communities of users acting in coordination as no ground truth is available. We can only claim result quality in comparison with backbones extracted by other methods.

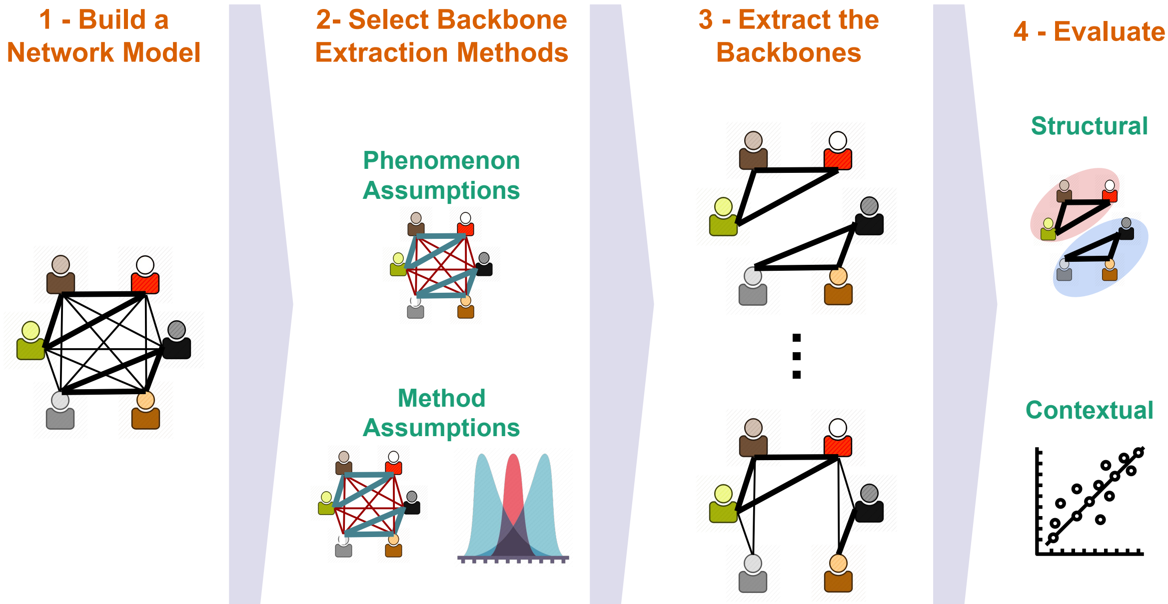


Figure 6.2. Overall methodology.

(if any) to the target study. We then need to choose a method to identify and remove such non-salient edges, thus revealing the network backbone. We here argue that the choice of possible strategies to perform this task should be guided by a fundamental understanding and careful matching between characteristics of the phenomenon and the assumptions of each method (step 2). Having identified adequate candidates among alternative backbone extraction methods, the next two steps consist of applying such methods to the network (step 3) and evaluating the quality of the extracted backbones with respect to both topological and contextual (phenomenon-specific) criteria (step 4). The backbone with the highest quality with respect to the considered criteria is then used to investigate the collective patterns driving the phenomenon under study. These four steps are described in detail next.

6.3.1 Step 1 – Building a Network Model

We assume the availability of a dataset containing a temporal sequence of user interactions taking place over a period of interest, gathered from the target system. In essence, these interactions may occur among multiple users simultaneously, being thus referred to as *many-to-many* interactions and are observable actions (e.g., comments posted in a social media application) reflecting different user behavior patterns. We are interested in revealing those patterns that are fundamentally related to (and drive) the phenomenon that will be studied.

As in our general approach, we also start building a network model from the input dataset by projecting the user interactions into an undirected and weighted graph $G = (V, E)$ such that:

- V is the set of users who interacted at least once during the period of interest;
- E is the set of undirected and weighted edges connecting pairs of users, such that the weight of edge e_{i_1,i_2} connecting users $i_1, i_2 \in V$ is $\gamma(i_1, i_2) = f(i_1, i_2)$, where $f(i_1, i_2)$ is any aggregation function (e.g., count) defined over the set of interactions between i_1 and i_2 and/or any contextual information available associated to them. Examples include the sharing of similar content (e.g., same URLs, same hashtags, or same messages) and/or temporarily synchronized activities [Pacheco et al., 2020; Cruickshank and Carley, 2020; Uyheng and Carley, 2021; Weber and Neumann, 2020; Cann et al., 2021; Weaver et al., 2019].

Before investigate strategies for extracting the *backbone* of the original network where noisy edges are filtered out, we note that one might be interested in the dynamics of such backbone over different periods of time covered by the input dataset. In that case, one strategy is to break the original data into subsets covering non-overlapping and consecutive time windows (e.g., weeks or months) and build one network model for each such window (see Chapter 3). Given that the same phenomenon under study is the same, it is reasonable to assume that the contextual criteria impacting the selection of the best backbone extraction method would remain the same for all network models. In that case, the methodology could be applied to one such network model (e.g., the one corresponding to a period of peak user interactions) so as to identify the most adequate backbone extraction method to the study. Such method could then be used to extract different backbones (one for each window) allowing an assessment of the temporal evolution of their properties (e.g., as done in our prior case studies).

6.3.2 Step 2 – Selecting Candidate Backbone Extraction Methods

In principle, any backbone extraction method could be applied to a given network model, and the backbones extracted by different methods may be quite different (as illustrated in Figure 6.1). Some backbones may miss a few important edges while still offering important insights, whereas others may be composed mostly of edges of little relevance to the study. Detecting the latter is not always easy, especially for large-size networks. Thus, we argue that a careful and principled selection of candidate methods must be performed *before* evaluating the extracted backbones to avoid misinterpretations and facilitate evaluation. To that end, our goal in this step is to shortlist backbone extraction methods that are adequate to study the given phenomenon. By adequate we mean that their assumptions and requirements are in alignment with key characteristics of the phenomenon, at the cost of generating completely unrelated backbones otherwise.

In the following, we offer a characterization of nine alternative methods, including state-of-the-art solutions, and discuss issues one must consider to study a target phenomenon. These are the same methods introduced in Section 2.3. The discussion below reflects our analyses of the methods' applicability to different scenarios. To guide this discussion, in Table 6.1 we present a summary with some key properties of each method. Specifically, we categorize the methods along with three aspects that are important to assist one in determining the suitable methods for a given case study.

6.3.2.1 Global vs. Local Methods

The first aspect (2^{nd} column of Table 6.1) is inherently related to how the method determines whether an edge is salient or not. While some methods apply a single criterion to all edges, others may use different criteria for different edges. Thus, we propose to classify each method as either *local* or *global*. The former refers to methods that determine the salience of each edge based on local information associated with the neighborhood of the edge, thus capturing aspects that are specific to the edge (and adjacent nodes) being analyzed. Global methods, instead, use the entire graph or a single global property for all edges in the graph. As such, the same (global) criterion is applied to all edges. As shown in Table 6.1, the simple threshold-based backbone extraction, HSS and RECAST are global methods. All other six methods are local. It is important to note that besides of GloSS uses a single reference model, the selection of salient edges is based on local information about the degree and strength of adjacent nodes [Radicchi et al., 2011].

Table 6.1. Our characterization of selected backbone extraction methods.

Method	Edge Salience Criteria		Parameters
	Local vs. Global	**Structural vs. Statistical**	
Threshold-based Backbone Extraction	Global	Structural	Threshold
High Salient Skeleton (HSS) [Grady et al., 2012]	Global	Structural	Threshold
RECAST [De Melo et al., 2015]	Global	Statistical Reference model: Two global distribution for all edges from random graphs with the same topology as the original network	p-value
Disparity Filter (DF) [Serrano et al., 2009]	Local	Statistical Reference model: Uniform distribution of edge weight per node	p-value
Polya Urn Filter [Marcaccioli and Livan, 2019]	Local	Statistical Reference model: Beta-Binomial distribution of edge weight per edge	p-value, α (reinforcement learning)
Marginal Likelihood Filter (MLF) [Dianati, 2016]	Local	Statistical Reference model: Binomial distribution of edge weight per edge	p-value
Noise Corrected (NC) [Coscia and Neffke, 2017]	Local	Statistical Reference model: Binomial distribution or a Hypergeometric distribution (obtained by a Bayesian Framework) of edge weight per edge	p-value
Global Statistical Significance (GloSS) [Radicchi et al., 2011]	Local	Statistical Reference model: A single null model considering both the edges between nodes and the weight distributions of the original network. Each edge is evaluated under its end nodes' properties using a proposed Bayesian Approach	p-value
Tripartite Backbone Extraction (TriBE) [Ferreira et al., 2020, 2021]	Local	Statistical Reference model: Poisson-Binomial distribution of edge weight per edge	p-value

The choice between a local or a global method should take into account whether the phenomenon exhibits an inherent heterogeneity or possible biases across different edges that are relevant to the understanding of the phenomenon. For example, it is well-known that several attributes related to user behavior in social media applications (e.g., content popularity, content sharing, etc.) are very heterogeneous, resulting in heavy-tailed distributions [Crovella et al., 1998; Ratkiewicz et al., 2010]. Such distributions naturally lead to network models with edge weights, node strengths and other properties that are widely distributed, often over different scales [Ahn et al., 2007; Grabowicz et al., 2014; Newman, 2003]. If the phenomenon under investigation is inherently related to a single (dominant) scale (e.g., revealing the most frequent interactions – the heaviest edges in the network) or to properties that go beyond single edges and their adjacent nodes (e.g., revealing users who can easily reach all others in the network), then a global method could be adequate.

Otherwise, if the phenomenon occurs at all scales defined by the heterogeneous structure of the network, a local method is probably more adequate. By exploring local information to define the salience of an edge, such methods might be able to retain edges that are representative of multiple scales, thus being relevant to the phenomenon. One such example is the study of online discussions in social media. Participation in such discussions is naturally highly heterogeneous reflecting the differences in user behavior. Yet, to get a clear picture of what is being discussed, one must capture the contributions of users with different levels of activity. Applying a global method may bias the extracted backbone to the interactions among the most active users or the most popular content, which would offer only a partial view of the discussions. A local method, instead, would be able to retain interactions among users with different levels of activity, thus offering a more complete and accurate representation of the interactions driving the phenomenon. We further elaborate on this particular study in Section 6.4.

6.3.2.2 Structural vs. Statistical Methods

A second aspect to be considered (3^{rd} column of Table 6.1) is whether edge salience is based on structural properties or on a statistical reference model. The former relates to methods that determine whether an edge is salient based solely on topological attributes of the network (e.g., edge weights, neighborhood overlap, paths, etc.), thus relying only on the empirical distributions of these attributes. These distributions are often evaluated via thresholds. As shown in Table 6.1, both the threshold-based and HSS methods fall into this category. For the former, salient edges are those whose weights are above (or below) a given threshold. For the latter, the frequency of occurrence in the shortest path trees is used as attribute. Structural methods are more adequate if the phenomenon being studied is inherently related to the network topology or connectivity, as represented by the used attribute. Examples include revealing the interactions among users/nodes with the largest number of neighbors in common (highly neighborhood overlap), or revealing users who are sources of information with greater reach in the network [Csermely et al., 2013].

In contrast, other phenomena may be studied in more detail by examining statistical deviations from an expected reference behavior. In such cases, one should consider methods that build statistical reference models for edge weights. These methods consider as salient the edges whose weights deviate significantly (according to a given p-value) from the reference model. The idea is that such reference model reflects the random network structure that would emerge if the phenomenon would not be taking place. As such, it is built based on network properties (e.g., distribution of node degrees, node strengths, or edge weights) often under the assumption of independent user behavior. By looking at edges that statistically deviate from the reference, these methods avoid uninteresting (common) behaviors, thus focusing on the edges that have greater chance of reflecting uncommon interactions that drive the phenomenon under investigation.

Different methods employ different reference models, thus directly impacting the definition of salience. To select a method, one should consider whether the employed reference model reflects a baseline for analysis of the phenomenon being studied. Consider, for instance, the study of coordination among users to spread information where interaction occurs when two users share the same piece of content. A strategy to model this phenomenon is to consider that users should have similar sharing patterns with their neighbors in the network if no coordination is taking place. This behavior leads to a uniform distribution of edge weights for all edges incident to the same node as the reference model.Edges with weights that significantly deviate from such reference offer potential evidence of coordination and, thus, should be retained as part of the backbone. We further elaborate on this study in Section 6.5.

6.3.2.3 Parameters to Filter

The third aspect (4^{th} column) relates to parameters employed by each method. As shown in Table 6.1, all structural methods rely on a threshold parameter to determine salient edges. As mentioned in Section 2.3, the use of such approach may lead to biases in the analyses. To avoid such problems, a threshold can be set contextually, i.e., based on an expected value for an edge according to the phenomenon. Since setting the threshold based on a contextual decision may be quite complex, prior work has proposed to consider a percentile of the empirical distribution of weights, analyzing the impact of this value on topological properties, e.g., density and community quality [Coscia and Rossi, 2019; Soro et al., 2020].

Conversely, all statistical methods make use of a p-value for statistical testing to identify salient edges. Typically, the literature uses classical values (i.e., 0.1, 0.05, 0.01, or 0.0001) [Radicchi et al., 2011; Grady et al., 2012; Marcaccioli and Livan, 2019]. However, some studies have argued that such classical values do not always yield the best topological structure of the network [Coscia and Rossi, 2019; Mukerjee et al., 2022]. We here propose to test a range within these values looking for the impact on both topological and contextual properties. In addition to a p-value, the Polya Urn filter also requires a parameter α that governs the process of reinforcement of existing interactions [Marcaccioli and Livan, 2019]. The higher the value of α, the larger the weight of an edge between two nodes must be, compared to the weights of the other edges adjacent to those nodes, in order for the edge to be considered salient.

6.3.2.4 Additional Considerations

Having discussed three aspects that must be considered when selecting backbone extraction methods, we complete this step with some general considerations and insights about specific methods that may also help guide the selection. First, we note that some of the local statistical methods, notably TriBE, MLF and NC, use binomial or Poisson binomial distribution as reference model for edge weights. These statistical distributions assume – by design [Ehm, 1991] – that each unit of edge weight is assigned to a pair of nodes under the assumption of *independence*. Deviations from this assumption are considered relevant evidence of salience in the context of social media applications, as they suggest that the weights are generated by hidden effects, e.g., when users are attracted to certain content and therefore interact around them [Ferreira et al., 2020; Coscia, 2021]. The Polya-Urn filter, on the other hand, assumes the beta-binomial distribution which breaks with the assumption of independence since each assignment is not independent of the others and changes from trial to trial (see Section 2.3). Also, as already mentioned, social media applications are characterized by a great degree of heterogeneity in user activity and content popularity. TriBE, being designed for this context, captures such heterogeneity directly, by using these two factors to build the reference model. In contrast, MLF and NC capture it indirectly, by considering node degrees and node strengths to build the reference models. Intuitively, these node attributes are closely related to user activity and content popularity. On one hand, as very active users tend to interact more with others, the degrees and strengths of the corresponding nodes in the network tend to be larger. Similarly, more popular content tends to attract more users, thus contributing to increasing the strengths and degrees of the corresponding nodes. GloSS filter also uses the same attributes to determine whether an edge is salient, though using a somewhat different approach. Therefore, all these four methods share similarities in terms of the definition of edge salience, producing backbones that include edges with great variety of weights.

In contrast, the other methods here referenced explore network heterogeneity in the sense that edges with larger weights, either from a local (Polya Urn and DF) or a global (RECAST, HSS and threshold-based) perspective, are more likely to be considered salient. Both Polya Urn and DF build different reference models to seek edges that stand out (from a local point of view) by their weights considering a subset of nodes/edges. HSS and the threshold-based method, instead, take a global perspective (the structure of the whole network or a target threshold) as reference to identify salient edges. RECAST, in turn, characterizes edges into four classes, allowing for different definitions of edge salience (see Section 2.3). Yet, by exploring two such classes, namely *Friends* and *Bridges*, one may produce backbones that also favor edges with heavier weights. Considering that the network model we build encodes user interactions, such methods favor keeping edges in the backbone based on repetitive and consistent patterns of interactions.

6.3.3 Steps 3 and 4 – Backbone Extraction and Evaluation

Having identified a set of backbone extraction methods that could be employed in a particular study, step 3 consists of applying the selected methods to the original network to extract the corresponding backbone. Specifically, each candidate method m in a set of methods M identified in step 2 is applied to extract a backbone $B^m = (V_b^m, E_b^m)$, such that $E_b^m \subseteq E$ consists of only edges considered salient by m and $V_b^m \subseteq V$ is the set of nodes with at least one edge in the backbone extracted by method m.[2] Step 4 consists of evaluating the quality of the produced backbones. In case multiple methods were selected in step 2, the best alternative should be chosen according to a trade-off between the metrics discussed next. The backbone produced by the best method would then be used to carry out the study in course.

Building on prior work [Newman and Girvan, 2004; Newman, 2006; Coscia and Neffke, 2017; Leão et al., 2018; Marcaccioli and Livan, 2019; Coscia, 2021], we consider

[2]In other words, after backbone extraction, all isolated nodes are disregarded.

metrics of backbone quality in two categories: topological, which are closely related to network and community structure, and contextual, which refers to phenomenon-specific attributes.

6.3.3.1 Topological Metrics

The topology-related metrics aim at quantifying the extent to which the network structure emerging from the backbone provides a clear view of how users are organized. Metrics such as node degree, density, clustering coefficient, number of connected components, modularity (see discussion below) characterize the structural properties of interactions considered as salient by the backbone extraction process. For the sake of brevity, we refrain from formally presenting all such metrics here and refer the reader to [Barabási et al., 2016] for formal definitions.

Recall that our main focus is on phenomena related to collective user behavior. Examples in the social media domain include efforts to promote particular ideas, brands, or ideologies. The graph concept that can be directly applied to this notion of collective behavior is *community*. Thus, the emergence of clearly defined (i.e., strongly structured) communities in the backbone offer potential evidence of groups of users actively engaging in common behavior. Identifying such communities is an important step to uncover relevant knowledge about the phenomenon being studied [Ferreira et al., 2019; Leão et al., 2018; Ferreira et al., 2020, 2021; Brito et al., 2020; Soro et al., 2020; Nobre et al., 2020, 2022].

The literature on community detection is quite extensive, with approaches focusing on specific concepts of communities defined via various network models [Labatut and Balasque, 2012; Yang et al., 2016a; Rossetti et al., 2019]. Recall, however, that we have adopted here the definition of a community which naturally implies groups of users who are more *similar* in terms of shared interactions and other behavioral patterns. To capture this definition, which implies that users in a given community are more strongly connected to each other than to the rest of the network, we again chose to apply Louvain's algorithm [Blondel et al., 2008; Newman and Girvan, 2004] to identify communities in the backbones. However, as stated in Chapter 3, this is an important step that could be modified from our methodology to employ alternative methods. In addition, we recommend reading and exploring the various strategies and metrics for detecting and evaluating communities in networks which are comprehensively summarized in [Rossetti et al., 2019].

6.3.3.2 Contextual Metrics

In addition to topological metrics, the quality of a backbone should be assessed with respect to how well it represents properties of the phenomenon under study. For example, by focusing on communities and, in particular, by exploring the contextual properties associated with them – i.e., characteristics of the communities that are not explicitly captured by the network topology, but are intrinsically related to the phenomenon under study – we may uncover properties that can help explain the emergence of different collective behavior patterns. In this way, we can gain insights into factors driving the phenomenon [Yang et al., 2016a; Gao and Liu, 2016; Liu et al., 2018b; Lu et al., 2020].

Unlike topological attributes, contextual criteria of backbone quality require additional information about the phenomenon. For example, in the case of social media applications, contextual information can be obtained through metadata that is usually collected when studying these applications. We thus also propose to assess how well the backbone captures phenomenon-specific properties by means of regression models. Specifically, we build upon prior work [Coscia and Neffke, 2017; Marcaccioli, 2020; Coscia, 2021], where contextual (phenomenon-specific) properties are used as explanatory variables to build linear regression models with edge weights as the response (dependent) variable. Although only linear regression models were used in these previous studies, nonlinear models can also be considered and are particularly appropriate when the chosen covariates are known or expected to have a nonlinear relationship with the edge weights. Since this is not the case for the studies presented in Sections 6.4 and 6.5, we use linear models to estimate the edge weights in the backbone (or in the entire network). Specifically, we consider the following regression model:

$$\gamma(i_1, i_2) = \beta_0 + \beta_1 \mathrm{X}_1 + \beta_2 \mathrm{X}_2 + ... + \beta_n \mathrm{X}_n + \epsilon. \tag{6.1}$$

where $\gamma_{(i_1,i_2)}$ is the weight of e_{i_1,i_2}, $X_1...X_n$ is a set of covariates related to the phenomenon, $\beta_0...\beta_n$ are the model coefficients and ϵ is an error factor.

The quality of the fitting of the model to the data captures how well the covariates (contextual properties) can be used to explain the edge weights (topological property). The better the fitting of the regression model, the more representative the considered edges (and corresponding weights) are of the underlying network structure driving the target phenomenon. In particular, we expect that the fitting of the regression model produced for the edges in a backbone (i.e., only edges in E_b) to be better than the fitting of the model produced using the entire (noisy) network (i.e., all edges in E). Similarly, we can compare the quality of different backbones by comparing the fitting of the models produced for them.

Although this approach has been used in previous studies [Coscia and Neffke, 2017; Marcaccioli, 2020; Coscia, 2021], we point out some possible limitations. First, prior work only considered the coefficient of determination R^2 (or its relative improvement for the backbone over the original network) as a quality measure. However, R^2 values may be misleading as they do not account for error estimates [Jain, 1991]. Therefore, we propose to assess the quality of the fitting by using both R^2 and the root mean square error (RMSE), which is the square root of the mean squared difference between estimated and observed values [James et al., 2013]. That is, given n edges, and the observed and estimated weights of these edges, γ_{i_1,i_2} and $\hat{\gamma}_{i_1,i_2}$ for each edge e_{i_1,i_2} respectively, the RMSE is defined as:

$$RMSE = \sqrt{\frac{\sum_{(i_1,i_2)\in E}\left(\gamma_{i_1,i_2} - \hat{\gamma}_{i_1,i_2}\right)^2}{|E|}}, \tag{6.2}$$

Smaller RMSE values suggest better (i.e., more accurate) fittings of the model. In order to be able to compare RMSE values for different networks/backbones, we use a normalized version of RMSE [Hastie and Tibshirani, 2017], where edge weights are normalized by the average value defined as:

$$NRMSE = \frac{RMSE}{\left(\frac{\sum_{(i_1,i_2)\in E}(\gamma_{i_1,i_2})}{|E|}\right)} \tag{6.3}$$

Another issue that deserves special attention is that backbones extracted by different methods may be quite different in terms of both the number of salient edges and the ranges of weight values, as the methods may favor very different edges during selection of the salient ones. On the one hand, one would like to assess the quality of each backbone using all (or most of) its edges. On the other hand, it may be interesting to compare different methods over the same set of (salient) edges. As a trade-off between these two scenarios, we propose to split the data into a training and a test set, whereas the latter consists of a smaller set of edges common to all backbones, and evaluate backbone quality in both sets.

Specifically, we first identify the largest common set of edges present in all extracted backbones $E_b^{\cap} = \bigcap_{m\in M} E_b^m$, where M is the set of alternative backbone extraction methods to be evaluated. We then propose to randomly select a sample T of $E_b^{\cap}$ as *test edges*. We choose to select 20% of $E_b^{\cap}$ as test edges here, but other sample sizes could be adopted [Gholamy et al., 2018]. Next, for each method $m \in M$, we build the regression model using all edges in $E_b^m \setminus T$, that is, all edges in the extracted backbone except those in the testing set are used as *training edges*.[3]

We first evaluate the quality of each regression model using both R^2 and NRMSE over the *training edges* to assess how well the model fits the training data. Note that the training data captures the majority of the edges in the original backbones. As such, by analyzing the model fitting to this data we are able to assess the extent to which each backbone is indeed capturing relevant information for the phenomenon under study.

We then assess the quality of each model in the common set of test edges T. That is, we use the trained models to estimate the weights of edges in T, and evaluate the quality of the fitting using NRMSE.[4] In a sense, T captures the consensus in terms of edge saliency among all methods. As such, we note that similar NRMSE values in the training and test sets for a given method suggests that this consensus is representative of the entire backbone extracted by the method. In contrast, larger NRMSE values in the test set suggests that the backbone extracted by the method deviates significantly from the other backbones (that is, the test set is not representative of the training data).

Note that we chose to use a sample of $E_b^\cap$ as test edges, instead of the complete set, to avoid favoring particular methods. For example, backbones with larger relative intersections with $E_b^\cap$ (i.e., the smaller backbones) could be favored in the quality assessment as edges in $E_b^\cap$ are more representative of the training data. Indeed, we did observe this effect to happen in a preliminary set of experiments when the complete set $E_b^\cap$ was used as test edges in our case studies. This effect was reduced as we adopted the strategy of using a sample of $E_b^\cap$ as test edges instead. This strategy has also the side effect of leaving more edges to build the model, which may lead to more accurate models.

We also note that although the above discussion focuses on quantitative assessment of backbone quality, one could also resort to visualization, especially of the denser components of each backbone, to identify possible differences between backbones extracted by different methods.

6.4 Case Study 1: *Online Discussions* on Instagram

We apply here our methodology for selecting and evaluating backbone extraction for our first case study examined in this chapter, which was also examined in Chapter 5. We then start with a brief review of the phenomenon under consideration.

[3]We do the same for the entire network, using set $E - T$ as training edges.

[4]We only use the NRMSE because it is better suited for checking how far the points of the common test set are from the regression line [James et al., 2013]

6.4.1 Characterization of the Phenomenon

Our first case study focuses on *online discussions on Instagram* as the phenomenon of interest. Different factors may drive users to comment on specific posts. For example, the common interest in a specific topic related to the original post, may drive users to elaborate their opinions and exchange ideas in their comments, thus generating and feeding on-going discussions (as we presented in Chapter 5). However, users may also comment on posts due to other factors, e.g., reaction to advertising campaigns or user personalization mechanisms [Perra and Rocha, 2019]. Not all these factors are truly related to online discussions. Therefore, a network model built from such commenting interactions may contain a number of irrelevant (i.e., non-salient) edges to the study of online discussions.

In the last chapter, we examined online discussions on Instagram by applying a specific backbone extraction, TriBE, to reveal the truly relevant (i.e., salient) edges. Our main focus was on characterizing the structural and dynamic properties of the communities that compose the extracted backbone, as representations of different groups of users actively engaged in online discussions. Despite the interest in the same phenomenon, our goal here is completely different. We are here interested in evaluating alternative methods of backbone extraction. In a sense, our present study should precede an investigation of the phenomenon, as we aim here to identify the most appropriate backbone extraction method for such an investigation. Specifically, we here analyze how different backbone extractions methods uncover the user interactions driving online discussions in posts and forming communities that can foster the spread of information (e.g., ideas or opinions). Our methodology can also be applied to other social media and other broader contexts. With respect to the former, we note that user interactions when commenting on various topics have already been studied in the context of other social media applications such as Twitter [Malagoli et al., 2021] and Facebook [Smith and Graham, 2019; Trevisan et al., 2020]. As for the latter, one can cite mobility networks [Coscia et al., 2020; Bonaventura et al., 2021] and biological networks [Putra et al., 2021], where some of the methods we use here have been applied without careful investigation.

We then sampled from the dataset described in Section5.3 focusing on content posted by political influencers in Brazil in the week surrounding the first round of the 2018 Brazilian general elections (i.e., from September 30th to October 6th). In particular, we gather posts from the eight main candidate runners.[5] We use all the

[5]We target the profiles: @jairmessiasbolsonaro, @fernandohaddadoficial, @lulaoficial, @cirogomes, @_marinasilva_, @guilhermeboulos.oficial, @cabodaciolo, @ad.alvarodias).

posts that these profiles made during the election week, as well as the comments they received from other Instagram users. Again, we chose not include users who commented on fewer than two posts, as these clearly reflect sporadic behavior. In total, we analyze here 41 099 who made 376 779 comments on 540 posts. These data cover a specific time window where user activity is high in terms of posts and comments, but user interaction patterns are representative of other time windows analyzed in these studies.

The following sections describe how we applied our proposed methodology to the study of online discussions on Instagram and how we performed each of the four steps.

6.4.2 Step 1 - Building the Network Model

We model the interactions among users commenting on the same Instagram posts with the same system properties presented in Section 5.2. That is, the *commenters* are the *individuals* (i.e., users commenting on posts made by influencers analyzed), *opportunities* are the posts created and the co-interactions are co-commenters commenting on the same post. Thus, the network model that has been shown to reveal communities of users participating in online discussions is defined as a weighted undirected graph $G_{Instagram} = (V, E)$, where:

- V is the set nodes representing the users who commented on posts; and
- E is the set of edges, where each edge $(i_1, i_2) \in E$ connects nodes i_1 and i_2, representing two users who commented on the same post. The edge weight $\gamma(i_1, i_2)$ is the number of posts that received comments by the same two users.

As mentioned earlier, many interactions captured by user comments on Instagram, and represented by edges in our network model, may not reflect *actual discussions*. For example, a very popular post naturally attracts many users, who often comment on them, often in an independent manner, without actually engaging in a discussion about the topic. Also, some users are more active than others. Thus, it is very likely that the most active users actually comment on many posts, acting completely independent of others who commented on the same posts, without however engaging in discussions. Such cases result in a set of edges in our network model that are simply random. Since these edges do not reflect discussions among users/commenters, they represent a *noise* for studying the target phenomenon. This observation requires the use of backbone extraction methods to identify and extract the salient edges for the study, filtering out those that are most likely just random.

6.4.3 Step 2 - Selection of Candidate Backbone Extraction Methods

As argued in the last paragraph, the network model may contain a number of edges that occur simply by chance, as a side effect of independent user behavior in the presence of very popular posts and/or very active users (i.e., commenters). This understanding of how inherent properties of the phenomenon may impact the network model helps us focus on backbone extraction methods that do take such properties into consideration to identify and extract only the salient edges, that is, those with greater evidence of reflecting online discussions. Specifically, we are looking for methods that consider the effects of *user activity level* and *post popularity* on the emergence of random and spurious (i.e., irrelevant) edges. As such, methods that are based on the assumption that edge salience is necessarily related to edge weight (e.g., methods that assume that edges with larger weights are more likely to be salient), either from a local or a global perspective, are not adequate as these methods tend to retain in the backbone only edges representing repetitive patterns of the most active users, disregarding interactions reflecting discussions carried out by less active (though still important) users.

Given these considerations, we select the following set of candidate methods for further evaluation: $M = \{\text{MLF, NC, Gloss Filter, TriBE}\}$. As discussed in Section 6.3, these methods are fundamentally local and statistical, and factor both *user activity level* and *content popularity*. TriBE explicitly builds a reference model based on both characteristics. MLF, NC and GloSS Filter evaluate the salience of an edge taking these factors into account indirectly, by exploring structural aspects of the network, notably node strength and degree, that are affected by user activity level and content popularity as explained in Section 6.3.2.4.

6.4.4 Step 3 - Backbone Extraction

After selecting the set M of candidate backbone extraction methods, we apply each method in M to the network model built in step 1 and extract the backbones. Recall that all the selected methods require a p-value as parameter. As mentioned in Section 6.3.2.3, we tested a range of possible values, varying from 0.001 to 0.1. Table 6.2 summarizes the topological characteristics of the original network and backbones extracted by the considered methods. Notably, the backbone results reported were produced with p-value equal to 0.05 for all methods. This value was selected as representative of all results tested. Results for other p-values are consistent, as reported in Table A.1 (A). We note that, for the sake of analyzing collective behavior, we chose to disregard from

Table 6.2. Online discussions on Instagram: Topological metrics of the network and backbones extracted by each candidate method (Columns 2-3 contain total numbers for the original network and corresponding percentages for backbones).

Network Model	Nodes	Edges	Avg. Deg.	Density	Avg. Clust.	# C.C.	# Comm.	Gini Index	Mod.
Original network	41 099	1 193 201	9236	0.2248	0.68	1	4	0.10	0.25
TriBE's backbone	70.06%	0.91%	119	0.0042	0.38	1	9	0.23	0.58
GloSS's backbone	65.45%	0.73%	103	0.0039	0.33	1	6	0.18	0.39
NC's backbone	100.00%	39.58%	3655	0.0890	0.46	1	5	0.28	0.52
MLF's backbone	94.78%	10.27%	1000	0.0257	0.40	1	5	0.53	0.49

the backbones and from the original network, prior to performing any analysis of them, any connected component with up to 3 nodes, focusing instead on larger components with possibly more impact and relevance to the phenomenon being studied.

Columns 2-7 of the table show the results of topological metrics, notably: nodes and edges (total numbers for the original network and corresponding percentages that remained in the extracted backbones), average degree (Avg. Deg.), density, average clustering coefficient (Avg. Clust.), and the number of connected components (# C.C.). We can see large differences across the backbone extraction methods for most topological metrics. Overall the backbones are sparser than the original graph once a fraction of the nodes (up to 34%) and most of the edges (up to 99%) have been removed. Moreover, the average clustering coefficient shows a moderate number of connected triangles in the network for all methods. Interestingly, the resulting backbones have only one connected component as is also the case for the original graph, which suggests the presence of key users promoting online discussions across different Instagram profiles who tie salient edges into a single component.

The three rightmost columns of Table 6.2 show results of community-related metrics. We note that the number of communities (# Comm.) is somewhat larger in all extracted backbones. This result is expected since backbones are sparser than the original graph. We analyze the community size distributions by employing the Gini index [Lerman and Yitzhaki, 1984], which measures a deviation of the given distribution to perfect equality (i.e., uniform distribution). The larger the Gini index, the greater inequality across community sizes. As the table shows, all backbones exhibit greater inequality of community sizes than the original network, notably the backbone extracted by MLF (largest Gini index). Considering only the backbones, the one extracted by GloSS Filter has communities with more evenly distributed sizes (lowest Gini index), though still with greater inequality than the original network. Finally, the rightmost column of Table 6.2 shows the modularity results as a measure of community quality. Clearly, all backbones exhibit more strongly connected communities (i.e., higher modularity) than the original network.

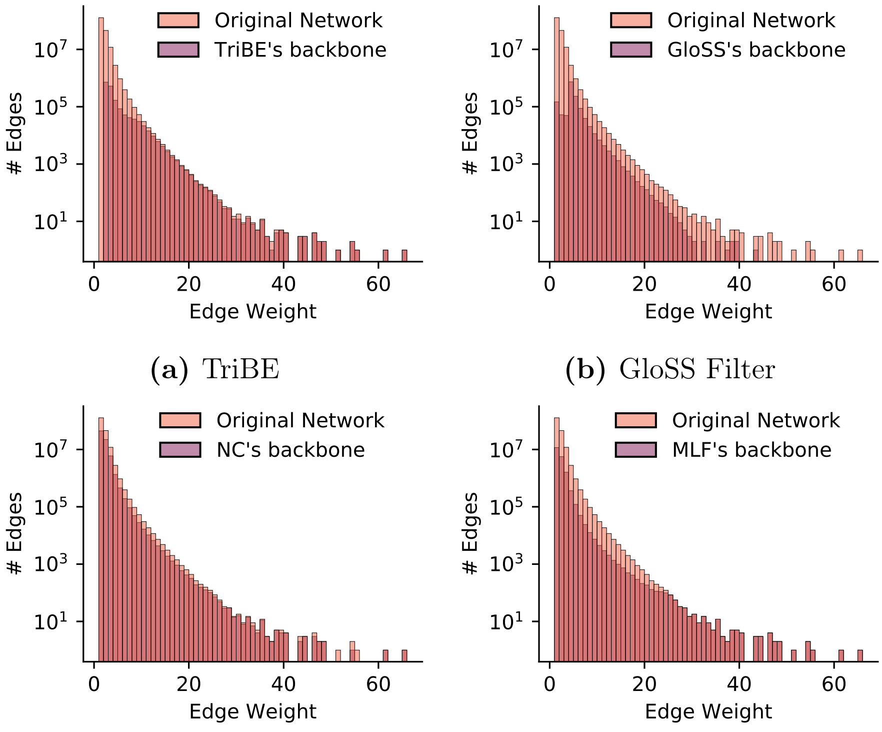
Since the different methods extract backbones with different topological structures, we further analyze how each method deals with noisy/irrelevant edges with respect to edge weights. Specifically, Figure 6.3 shows the distributions of weights of the edges retained in the backbone (i.e., salient edges) by each method. For comparison purposes, each graph in the figure also shows the distribution of edge weight in the original (complete) network. Recall that, as discussed in the previous section, the salience of an edge for the particular study of online discussions is not necessarily related to edge weight, thus our goal was not to retain edges with a particular range of weight values. Indeed, the figure shows some diversity in the weights of the edges that are removed by each method. TriBE (Figure 6.3a) removes many edges with small weights (notably all edges with weight equal to 1 are removed), as well as some edges with larger weights (around 38). GloSS Filter (Figure 6.3b), in turn, is more aggressive towards heavier edges, removing a fraction of edges through the whole range of values. In particular, note that all edges with the heaviest weights are filtered out by this method. MLF (Figure 6.3d) follows a similar pattern, though it is much less aggressive towards the heavier edges. Indeed, unlike GloSS Filter, MLF retains all the heaviest edges in the backbone. Finally, NC (Figure 6.3c) is the most conservative method, removing a smaller fraction of the edges through the whole range of weight values. In a nutshell, the results show that even though all four methods do share similarities (e.g., they all capture the effects of user activity level and post popularity), they also have their particularities when identifying an edge as salient, and may produce quite different backbones.

6.4.5 Step 4 - Backbone Evaluation

Having extracted the backbones by the selected methods, we turn to the final step of our methodology. We evaluate the quality of the extracted backbones from a topological and contextual perspectives, aiming at identifying the most adequate backbone extraction method, out of those selected as candidates, to the study of online discussions on Instagram.

6.4.5.1 Topological Evaluation

The topological properties of each backbone are presented in Table 6.2 and were discussed in the previous section. We here delve further into those results, aiming at comparing the backbones produced by the methods. Specifically, we compare the methods with respect to the structural quality of the communities identified in the corresponding backbones, as captured by the modularity metric, since communities are



(a) TriBE **(b)** GloSS Filter **(c)** Noise Corrected **(d)** MLF

Figure 6.3. Online discussions on Instagram: Weight distribution for edges retained in the backbone by each method (distribution for original/complete network shown for comparison purposes).

graph structures representing the different collective patterns driving the phenomenon under consideration.

As already mentioned, all four methods allow for better discrimination between communities (higher modularity), which should lead to a better understanding of the role of each community in promoting online discussions. The improvement in community structure (with respect to the original network) varies across methods. GloSS Filter, for instance, produces the smallest improvement (modularity increases from 0.25 to 0.39), while TriBE shows the largest improvement (from 0.25 to 0.58).

We also evaluate how similar the communities in the backbones and in the original network are, with respect to node membership. To that end, we employ the Normalized Mutual Information (NMI) metric which allows for comparison of two partitions even with different numbers of communities [Vinh et al., 2010]. Figure 6.4 shows a heatmap of the NMI values computed as a pairwise comparison between all four backbones as well as the original network. We note that the NMI for two backbones (or a backbone and the original network) is computed considering only the common subset of nodes in both graphs. In general, the results range from 0.49 to 0.65, suggesting low to medium similarity. The similarity is greater among the backbones produced by NC and MLF and the original network, which is consistent with other topological metrics shown in Table 6.2. As these two methods retain a much larger fraction of the original edges, the resulting backbones tend to have more structural similarities to the original network, including the number and composition of identified communities. And even for the methods that retain a smaller and similar percentage of edges, such as TriBE and GloSS, there is a difference in terms of the communities that these edges form.

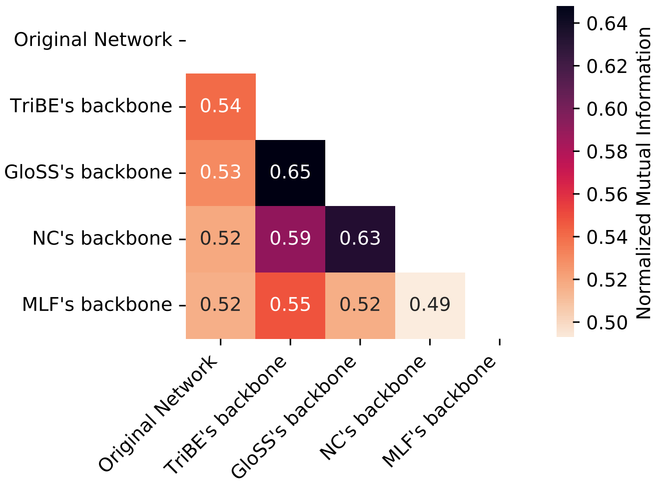


Figure 6.4. Online discussions on Instagram: Similarity of communities, estimated by Normalized Mutual Information (NMI), present in different backbones and original network.

6.4.5.2 Contextual Evaluation

We now shift our analysis to the quality of the backbones from a contextual point of view, that is, we assess how well the edges present in each backbone are able to capture phenomenon-specific properties. As described in Section 6.3, we do so by building a linear regression model that aims at explaining the weights of the edges retained in a backbone using a number of phenomenon-specific covariates. For comparison purposes, we do the same for the edges in the original network.

We build the regression models considering the following key assumption: *If two given users i_1 and i_2 do indeed engage in the same discussions by commenting on the same posts, the activities (comments) performed by each user individually are strongly correlated with the joint activities performed by both users (i.e., comments on the same post).* If, however, comments posted by one user (or both) are mostly reactions to popular content or to some automatic tool (e.g., advertising or personalization mechanisms), or simply sporadic and random behavior, the activities of the user, taken individually, are only weakly correlated (at best) with the joint behavior of both users.

Based on this assumption, we build a regression model for each backbone and the original network. Given the edge weight $\gamma(i_1, i_2)$ as dependent variable, the explanatory covariates are: (i) number of posts that user i_1 commented on, (ii) number of posts that user i_2 commented on, (iii) number of influencers that user i_1 commented on, and (iv) number of influencers that user i_2 commented on. We capture user activities by considering both the number of influencers and the number of posts each user commented on because, as reported in Section 5, it is often the case that the same influencer has multiple posts on different topics, each one attracting a different group of users (community). Thus, we expect that edges representing user interactions driven by joint engagement of both users in the same discussions to be reasonably well explained by these covariates, whereas edges reflecting random or sporadic behavior to be only poorly explained by these covariates. These edges should be removed from the backbone. Thus, the better the fitting of the model to the edge weights in a backbone, the better the quality of this backbone, from a contextual perspective.

Before discussing the results, we note that we did examine whether the covariates considered are indeed linearly related to the dependent variable, key assumption to use a linear regression model. We found that such linear relationship exists if a log transformation is applied to all covariates and to the dependent variable. Such transformation is often employed in variables with very skewed distributions, which is the case of edge weights (see Figure 6.3) and measures of user activity in social media applications [Schwartz et al., 2013; Malik et al., 2015].

Table 6.3 shows the results of model fitting for the four backbones as well as for the original network. As discussed in Section 6.3, we assess the quality of model fitting using the coefficient of determination R^2 and the NMRSE for the training edges and the NMRSE for the test edges.

Table 6.3. Online discussions on Instagram: Contextual evaluation of backbones by regression analysis.

Network Model	Training edges		Test edges (20% of common edges)
	R^2	NRMSE	NRMSE
Original network	0.25	0.48	0.72
TriBE's backbone	0.87	0.20	0.23
GloSS's backbone	0.65	0.28	0.36
NC's backbone	0.26	0.48	0.70
MLF's backbone	0.28	0.49	0.72

Focusing first on the results for the training edges, we see that both TriBE and

GloSS Filter produce large improvements over the original network, with respect to both R^2 and NMRSE. These two methods indeed are able to filter out a lot of noisy edges, retaining those that are more closely related to the phenomenon being studied, as reflected in the covariates used to build the regression models. TriBE, in particular, by explicitly considering both user activity level and post popularity to build its reference model, leads to the best results, with a very high R^2 of 0.87 and a small NRMSE of only 0.20. The other two methods, NC and MLF, in turn, despite filtering out many edges (as shown in Table 6.2), lead to only marginal improvements (if any) over the original network. This result is consistent with our observation in the last section that the backbones produced by these two methods exhibit greater topological similarity with the original network.

The same conclusion holds for the test edges. Compared to the original network, the model fittings for both TriBE and GloSS Filter present notable reductions in NRMSE, whereas the results for both NC and MLF remain very similar, with no improvement. This suggests that the edges considered as salient by NC and MLF deviate the most from the common set of edges considered salient by all methods. This observation suggests, in turn, that these two methods retain a large fraction of possibly non-salient edges, which ultimately impacts the fitting of the regression model.

In conclusion, when considering both topological and contextual metrics, we make two final observations. First, evaluating backbone quality based solely on one perspective may be misleading. For example, as shown in Table 6.2, the backbone extracted by NC has one of the largest values of modularity. Yet, the regression analysis reveals that, though well organized into strongly connected communities, the edges identified as salient by NC do not offer a more clear understanding of user behavior patterns driving the online discussions than the (poorly structured) original network. In contrast, the backbone extracted by GloSS Filter has the lowest modularity of all four backbones, even though the edges considered salient by this method are more strongly related to our study than the edges retained by NC and MLF (as captured by the regression model). Though keeping a large fraction of edges closely related to online discussions, Gloss Filter leaves out some important edges for the identification of the communities leading such discussions. Second, TriBE, in turn, stands out as the best approach in terms of both topological quality of the communities that emerge in the extracted backbone (modularity) and how well the selected edges are able to capture the user interactions that indeed are driving the online discussions.

6.5 Case Study 2: Coordinated Behavior on WhatsApp

We now turn to our second case study, which we examine in this chapter. Nevertheless, since this is the first time it is considered in this dissertation, we give a more detailed description in the next section.

6.5.1 Characterization of the Phenomenon

Our second case study concerns coordinated actions to disseminate information in WhatsApp groups. WhatsApp is a free messaging app widely used in many countries, with over 2 billion active users worldwide[6]. The platform connects users in end-to-end as well as group conversations. Despite being limited to only 256 simultaneous members, WhatsApp groups have been shown to be effective channels for the large dissemination of information [Resende et al., 2019a; Reis et al., 2020; Nobre et al., 2020, 2022], notably misinformation [Nobre et al., 2022], in spite of WhatsApp's recent efforts to mitigate such phenomenon (e.g., by limiting message forwarding [de Freitas Melo et al., 2019]).

We here adopt the following widely used definition of *coordination* [Cao et al., 2014; Yu et al., 2015; Pacheco et al., 2020, 2021]:

Coordinated behavior: *coordinated users typically exhibit a repetitive and synchronized pattern of activity.*

With such definition as background, we have recently unveiled the presence of communities of WhatsApp users for whom there is strong evidence of coordinated efforts to spread information at large by repeatedly sharing the same content in different groups [Nobre et al., 2020, 2022]. Such work, though not a central part of this dissertation, is worth mentioning as it represents collaboration in the application of our general approach. We find that these communities often transcend the formal group boundaries enforced by WhatsApp. They have members in multiple groups building an underlying network capable of sharing the same piece of content at very large scale on the platform. To reveal these communities, we applied a specific backbone extraction method to the network that connects users sharing similar content. Specifically, we applied the disparity filter (DF) as a strategy to filter out edges that do not offer strong evidence of coordination. For example, a highly popular piece of content is

[6] https://backlinko.com/whatsapp-users#whatsapp-statistics

most likely shared by multiple users independently, resulting in multiple edges in the network, even without any kind of user orchestration. However, the presence of a set of users who repeatedly spread similar content (more frequently than would be expected by pure chance) is a clearer indication that some coordination is taking place.

In contrast to these previous studies, our present goal is *not* to identify and study coordination (i.e., communities) of information spread on WhatsApp but rather to compare the use of alternative backbone extraction methods to such study. We aim at evaluating how well different strategies are able to extract the aforementioned underlying network of information spread over WhatsApp groups, uncovering communities for which greater evidence of coordination exists. Thus, once again, the present effort aims at identifying the most adequate backbone extraction method to be applied to a study of the given phenomenon (as those reported in [Nobre et al., 2020, 2022]).

Our present investigation relies on a dataset of anonymized messages shared in publicly accessible political-oriented WhatsApp groups in Brazil [Nobre et al., 2020, 2022], originally collected by the WhatsApp Monitor [Resende et al., 2019a]. We focus our analysis on the month of the general presidential election in Brazil (October 2018), a time of great political mobilization and strong evidence of message coordination and orchestration in WhatsApp [Resende et al., 2019b,a; Maros et al., 2020; Nobre et al., 2020, 2022]. In summary, we analyze 4 341 users who participated in 155 groups and shared 91 417 unique pieces of information, in the form of text messages, images, audios and videos.

6.5.2 Step 1 - Building the Network Model

Before describing the network model, we should explain the components of such a system in terms of our general approach. Fundamentally, these user interactions,just like those discussed in Chapters 4 and 5, can be modeled as a many-to-many network as follows. The *individuals* are WhatsApp *users*, i.e., users who have shared at least one piece of content during the observed time period and thus form the set of *individuals* $I_{WhatsApp} = \{i_1, i_2, ..., i_j\}$. Here, *opportunities* are taken as the pieces of content. In this way, we define the set of opportunities containing all the pieces of content posted as $O_{WhatsApp} = o^1, o^2, ...o^m$. We consider all users who shared the (near-)duplicate piece piece of content as a co-interaction. Thus, each opportunity (post) o^k leads to a single co-interaction in the set $C(o^k)$. Collectively, all opportunities during the observed period lead to a set $C_{WhatsApp}$. Having defined these elements, we use the same network model adopted in [Nobre et al., 2020, 2022], referred to as *media-centric* network, which is defined as an undirected and weighted graph $G_{WhatsApp}$=(V, E) such that:

- V is the set of nodes representing users who shared at least one message in one of the monitored groups during the period of analysis;
- E represents the set of edges, where each edge connects two users if they share similar content in the same or different groups. The similarity between message content was estimated using a set of heuristics for filtering and identifying (near-)duplicate content. We refer the reader to [Nobre et al., 2022] for more details on these heuristics. The edge weight represents the number of times the two users shared similar content, regardless of the group in which the sharing was made.

In light of the definition of coordination adopted, presented in the previous section, salient edges are those whose weights are unusually high, considering individual patterns or not. However, as defined, the network may contain a number of noisy edges, that emerge due to sporadic or weak interactions. For example, endogenous factors (e.g., temporary common interest or even large popularity of some particular content) may cause different users to share similar content, which may overshadow the actions of coordinated users who regularly and repeatedly share the same content. Therefore, the interest is to separate users who persistently engage in such common sharing from users who only sporadically exhibit such behavior. This separation implies favoring as salient those edges with heavier weights, either taking a local perspective (e.g., other edges incident to the same two nodes) or a global perspective (i.e., all edges in the network). This principle is used as guideline in the selection of candidate backbone extraction methods, as discussed next.

6.5.3 Step 2 - Selection of Candidate Backbone Extraction Methods

As argued, we aim at selecting methods that explore the heterogeneity of the network by identifying as salient the edges with unusually heavier weights, based on individual (local) or network (global) patterns, as representative of persistent and repetitive interactions. As argued in Section 6.3.2.1, Threshold, HSS and RECAST are global methods that explore the heterogeneity of the edge weight distribution, favoring as salient the edges with heavier weights in the whole network. From a local perspective, Polya Urn Filter and Disparity Filter (DF) select as salient those edges whose weights are heavier compared to the weights associated with a subset of the edges (e.g., edges incident to the same pair of nodes). Thus, we define the set M of candidate methods as

$M = \{$Threshold, HSS, RECAST, Poly Urn and DF$\}$. Recall that both Threshold and HSS explore structural properties, whereas the other three methods rely on statistical reference models to identify the salient edges.

6.5.4 Step 3 - Backbone Extraction

Table 6.4 summarizes the topological characteristics of the original network and the backbones extracted by each candidate method. As performed in the first case study, we here also disregard from the backbones and from the original network any connected component with up to 3 nodes. We note that the candidate methods have different parameters, which were set as follows. Disparity Filter (DF), Polya Urn Filter (Polya) and RECAST require a p-value. We here report results for p-value equal to 0.05, which are representative of those found for other values, as presented in Table A.2 (A). The Polya Urn method also requires a second parameter α, which is related to network heterogeneity. This parameter was set to 0.25, following a fine tuning process, as briefly mentioned in Section 2.3. The High Salient Skeleton (HSS) and Threshold approaches, on the other hand, take an arbitrary threshold value τ as input parameter. In both cases, we select τ so as to retain the top-k% most important edges, given the definition of salience adopted by each method. That is, we consider the distribution of the network feature exploited each approach (i.s., number of shortest paths in which an edge participates for HSS, and edge weight for Threshold), and select τ such that the top k% edges in the network, according to the given feature, are considered salient. Table 6.4 shows results for values of τ corresponding to the top 5% edges,[7] but we also tested for other values of k (thus of τ), as reported in Table A.2.

To further support the following discussions, we show visualizations of the original network and each backbone in Figure 6.5, focusing on their largest connected component[8]. In each graph, nodes belonging to the same community are represented by the same color, and edge weights are represented by both edge thickness and color (heavier/lighter edges are colored in red/blue).

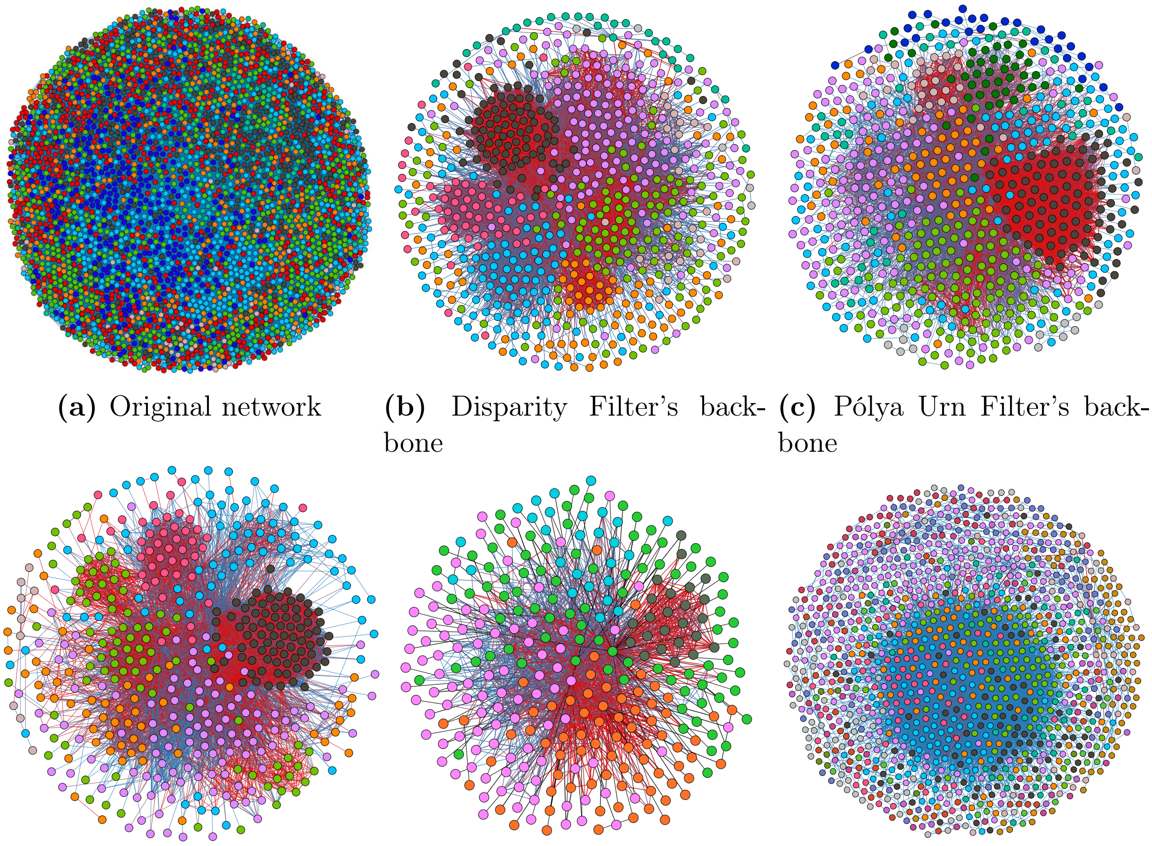
As shown in the table and in the graphs, DF, Polya and Threshold retain somewhat similar fractions of nodes (11-17%) and edges (3-4%) in the extracted backbones. Despite such large removal of nodes and edges, all three backbones have a number of connected components similar to the original network. Interestingly, we also find that

[7] We note that, according to Table 6.4, the fractions of edges retained by both HSS snd Threshold are slightly below 5%. This deviation is due to the removal of very small components (up to 3 nodes) of the backbone and of the original network.

[8] Note that the number of communities shown in these figures does not match the number in the Table6.4, as we are only showing the giant component for illustrative purposes.

Table 6.4. Coordinated behavior on WhatsApp: Topological metrics of the network and backbones extracted by each candidate method (Columns 2-3 contain total numbers for the original network and corresponding percentages for backbones).

Network Model	Nodes	Edges	Avg. Deg.	Density	Avg. Clust.	# C.C.	# Comm.	Gini Index	Mod.
Original network	4341	221002	103	0.0241	0.62	4	14	0.41	0.24
DF's backbone	16.51%	3.51%	21	0.0310	0.53	5	15	0.42	0.52
Polya's backbone	17.15%	4.76%	28	0.0391	0.59	5	15	0.47	0.48
Threshold's backbone	11.56%	4.29%	38	0.0776	0.73	3	8	0.35	0.45
RECAST's backbone	7.31%	0.85%	11	0.0384	0.49	2	8	0.49	0.37
HSS's backbone	43.42%	4.86%	11	0.0062	0.21	135	147	0.66	0.30



(a) Original network **(b)** Disparity Filter's backbone **(c)** Pólya Urn Filter's backbone

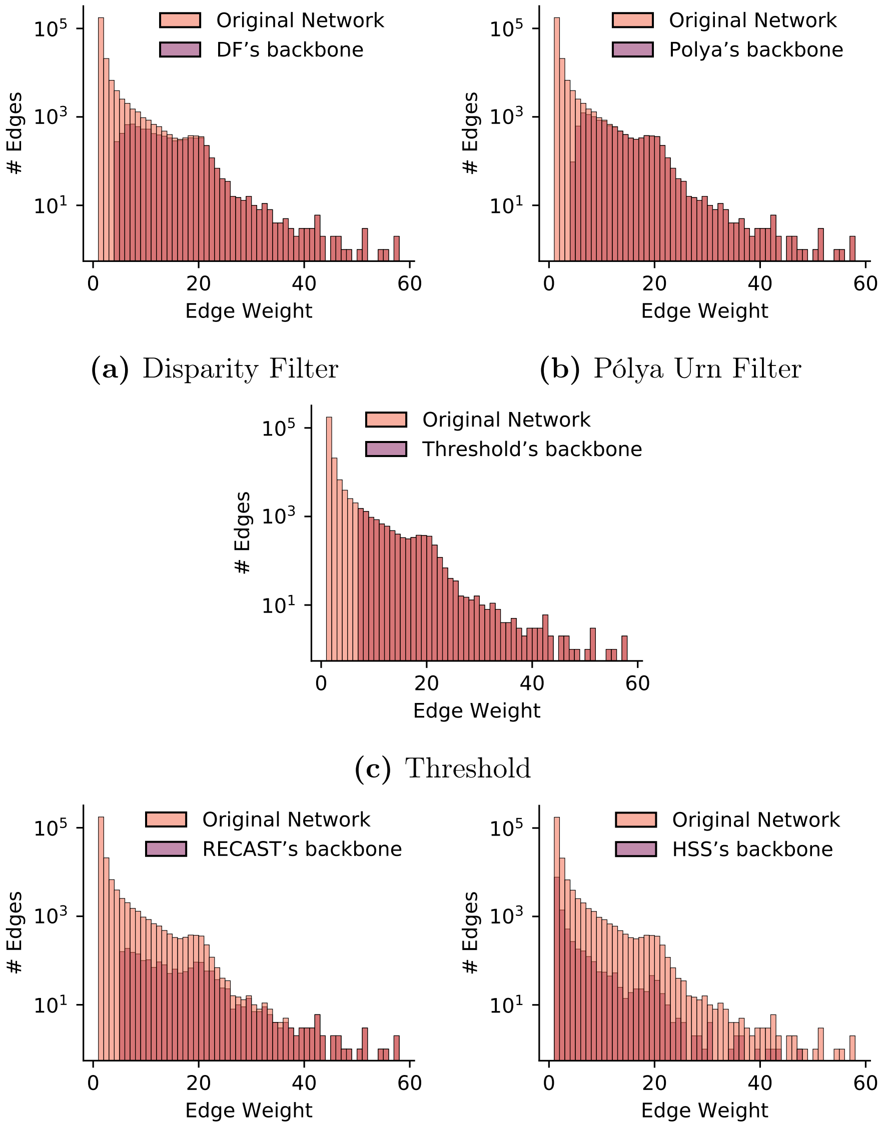
(d) Threshold's backbone **(e)** RECAST's backbone **(f)** HSS's backbone

Figure 6.5. Coordinated behavior on WhatsApp: Largest connected component of the original network and extracted backbone (node color indicates community membership, edge thickness and color indicates edge weight – heavy/light edges colored in red/blue).

these backbones have greater density (especially the backbone extracted by Threshold) and comparable (if not higher) average clustering coefficient to the original network. RECAST, in turn, is the most aggressive method, retaining only around 7% of the original nodes and fewer than 1% of the original edges. This leads to smaller number of components and average clustering coefficient, though the backbone's density is still comparable to that of the DF's and Polya's backbones. Moreover, it becomes much more clear, from the network visualizations in Figures 6.5-b) and 6.5-c), the large presence of heavier (red) edges in the backbones, especially if compared to the quite noisy original network. In the other extreme, HSS extracts a quite large backbone, with over 40% of the original nodes. This backbone has a very different topological structure compared to the others, being much sparser (lower density and average clustering) and much more fragmented (much larger number of connected components), as can be seen in Figure 6.5-f).

Turning to the analysis of the communities (3 rightmost columns of the table), we observe that the number of communities varies greatly across the backbones and the original network, being smaller for the backbones extracted by Threshold and RECAST. The former is possibly a result of the much denser backbone, whereas the latter possibly relates to the fewer nodes in the backbone. Nevertheless, the inequality in the distributions of community sizes is moderately similar, as captured by the Gini index. This can be observed also in the visualizations of Figures 6.5. The exception is, once again, HSS, whose backbone has a very large number of communities (147) with sizes more unevenly distributed, which most probably is a side effect of the more fragmented backbone. Indeed, as Figure 6.5-f) shows, it is much harder to distinguish these communities than those in the other backbones. Finally, once again we observe that all backbones have more strongly connected communities, as captured by the modularity metric, than the original network, including the backbone extracted by HSS (though to a lesser extent).

To further understand how the selected methods work, we analyze the distribution of edge weights. Figure 6.6 shows the distributions for each backbone and for the original network. The figure illustrates mainly how HSS differs from the other methods. All other four methods remove mostly edges with small weights, though DF and especially RECAST also remove small fractions of edges with more moderate weights. In particular, all edges with weights between 1 and 3 are removed by all four methods. HSS, in contrast, removes large fractions of edges across the whole range of weight values. In particular, it removes most of the heaviest edges, while still retaining a large fraction of edges with very small weights (between 1 and 3). Recall that, as presented in Section 2.3, HSS aims at retaining as salient those edges that are both heavy and are fundamental to keeping nodes connected. Our results suggest that, for the WhatsApp media-centric network studied here, it achieves neither of these goals, possibly due to a mismatch of them in the network structure. Indeed, we find that the fraction of shortest paths associated with each edge (main metric explored by HSS) is mostly uncorrelated with edge weight (Spearman coefficient of correlation of -0.026). As we will see in the next section, HSS's backbone is inferior to the others both from



(a) Disparity Filter **(b)** Pólya Urn Filter

(c) Threshold

(d) RECAST **(e)** HSS

Figure 6.6. Coordinated behavior on WhatsApp: Weight distribution for edges retained in the backbone by each method (distribution for original/complete network shown for comparison purposes).

a topological perspective and with respect to how well it captures edges that most probably are related to the phenomenon under investigation.

6.5.5 Step 4 - Backbone Evaluation

As our final step, we compare the five selected methods with respect to topological and contextual properties of the extracted backbones.

6.5.5.1 Topological Evaluation

As in our first case study, the results presented in Table 6.4 and illustrated in Figure 6.5 show that all backbones are composed of more strongly connected and more clearly discriminated communities than the original network. The improvements in community structure, as captured by the modularity metric, are particularly large for the DF, Poly Urn and Threshold approaches. The backbones extracted by these methods are mostly composed of well structured communities of users who repeatedly share the same content, which favors the information spread at large. RECAST and HSS, in turn, produce backbones with weaker community structures, for the reasons discussed in the previous section.

Delving deeper into the membership of the identified communities, we again compute the pairwise NMI value for all five backbones and the original network. Results are shown in Figure 6.7. Once again, we observe greater similarity in community membership for the backbones extracted by DF, Polya Urn and Threshold, while HSS has the most dissimilar community composition. Interestingly, the backbone extracted by RECAST has perfect NMI value in the comparison with all other backbones and with the original network. This suggests that the communities identified in RECAST's backbone are also present in all other graphs. However, being very aggressive in node and edge removal, RECAST's backbone offers only a partial and very limited view of the existing communities and their members, which is limited to the nodes with the heaviest edges and also the most connected to their respective neighbors.

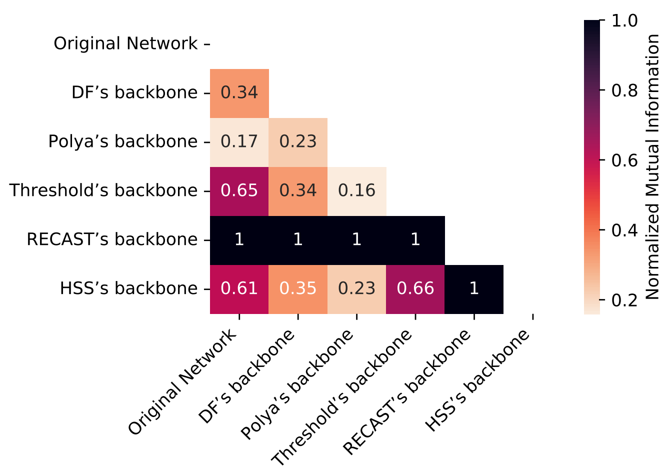


Figure 6.7. Coordinated behavior on WhatsApp: Similarity of communities, estimated by Normalized Mutual Information (NMI), present in different backbones and original network.

Table 6.5. Coordinated behavior on WhatsApp: Contextual evaluation of backbones by regression analysis.

Network Model	Training edges		Test edges (20% common edges)
	R^2	NMRSE	NMRSE
Original network	0.21	1.33	0.89
DF's backbone	0.40	0.39	0.43
Polya Urn's backbone	0.30	0.41	0.45
Threshold's backbone	0.22	0.37	0.43
RECAST's backbone	0.23	0.51	0.56
HSS's backbone	0.25	1.31	0.86

6.5.5.2 Contextual Evaluation

Our contextual evaluation of the backbone extraction methods is guided by the following key assumption: *If two users i_1 and i_2 are acting in coordination to share the same pieces of content repeatedly in one or multiple groups, such coordination should be reflected in their content sharing patterns.* Guided by this assumption, we build a regression model where the dependent variable $\gamma(i_1, i_2)$ is correlated with the following 11 explanatory variables: (i) total number of messages shared by i_1 (i_2); (ii) number of distinct messages shared by i_1 (i_2); (iii) number of messages with new content introduced (i.e., shared first) by i_1 (i_2); (iv) number of groups i_1 (i_2) participates in (inferred by the groups he/she shared content at least once); (v) Gini index of the number of messages shared by i_1 (i_2) across different groups; and (vi) number of common groups both i_1 and i_2 participate in. Note that all variables, but the last one, are computed separately for i_1 and i_2, thus contributing as two covariates to the model. They capture different facets of user activity, notably content sharing frequency, content diversity, introduction of new content, participation in different groups, and user preference towards particular groups or not (as captured by the Gini index). The only variable related to the joint behavior of both users is the number of common groups, which indirectly captures whether or not the two users act in the same subset of observed groups. For example, users may act in a coordinated manner by frequently sharing the same content (i.e., heavy edge), even though the number of groups in which both participate is small. This could indicate, for example, that each of them forwards content to a particular subset of groups. As in our first case study, we also tested for the assumption of linearity, finding that it holds reasonably well after a square root transformation is applied to all covariates and to the dependent variable.

Table 6.5 summarizes the results of the model fitting for each backbone as well as for the original network, for both training and test edges. Compared to our first case study, the fittings are in general poorer (note, for instance, the much lower R^2 values in the training edges). We emphasize that it is much more challenging to perform a contextual evaluation of the network and backbone structures in this case study because, as described, the dataset used consists of a sample of messages shared in only 155 groups. All monitored groups belong to the same context (political domain) and are strongly interconnected as many users belong to multiple groups. However, these groups offer only a very partial view of the phenomenon, as the same users most probably participate in other (unobserved) groups, where they share and forward content, contributing to the information spread at large (see [Resende et al., 2019a]). Thus, our analysis is inevitably constrained by the lack of an unknown number of edges that most probably exist in the real underlying network connecting these users.

Under this constraint, the contextual evaluation leads to results that mostly agree with those of the topological evaluation. The backbones extracted by DF, Polya Urn and Threshold have the lowest errors (NMRSE) in both training and test sets. In terms of R^2, the results are in general poor, as discussed, but better for DF and Polya Urn. Threshold performs comparable to the original network in this metric, and clearly worse than DF and Polya, suggesting that the global approach may leave out some important edges for the investigation. Indeed, as shown in Figure 6.6, both DF and Polya do retain a more diversified set of edges, in terms of edge weights. In the other extreme, RECAST and especially HSS perform quite poorly, with the latter offering mostly no improvement over the original network.

In sum, we find that both DF and Polya Urn are the best methods for revealing evidence of user coordination to sharing similar content in WhatsApp, with very similar results in terms of both topological and contextual properties. If one method is to be picked, DF is possibly the best choice as it extracts a backbone with slightly stronger community structures.

6.6 Discussion

We have proposed a principled methodology to select and evaluate the best methods for extracting the network backbone that more accurately represents the collective behavior driving a given phenomenon of interest. This methodology includes 4 steps to apply, evaluate and select the best method(s) for a given target phenomenon. We used two case studies with different requirements: Online discussions on Instagram and coordinated behavior in WhatsApp groups to validate it. Our characterization and application confirm that these are very different scenarios that require different solutions and illustrate the complexity of selecting appropriate methods for backbone extraction.

Moreover, we found that some methods extract rather useless backbones, while others are particularly suitable to capture and describe the phenomenon under study, taking into account a trade-off between topological and contextual measures. In this sense, we hope that our methodology helps to highlight and demonstrate the risks of using inappropriate and inadequate backbone extraction techniques.

6.7 Summary

In this chapter, we formalized and proposed a methodology for backbone extraction. This effort contributes, fundamentally and systematically, to several recent network-oriented studies relying on backbone extraction strategies to unveil useful knowledge about various phenomena of interest. In particular, our work fills a gap in the literature by emphasizing the need to: (1) carefully match assumptions and properties of the method with characteristics of the given phenomenon, showing that different methods may, in fact, extract quite different backbones, some which offer little (if any) useful knowledge to the study, and (2) consider different criteria to evaluate the quality of alternative backbones, especially when there is no ground truth (which is often the case).

In addition, we offered a reasoned characterization of nine state-of-the-art methods, including TriBE we proposed here, for extracting backbones and discussed their assumptions, properties, and issues to consider when applying them in practice. This characterization, developed based on our experience with the methods, advances existing knowledge available in the literature and is intended to aid in selecting candidate methods for a particular study. We also combine alternatives for validating the extracted backbones, both structurally (based on topological measures extracted from the network) and contextually (based on phenomenon-specific attributes). In this way, we filled the gap of a systematic procedure for comparing and selecting backbones by proposing a principled methodology for selecting the most appropriate backbone extraction method for a given phenomenon.

Chapter 7

Conclusions and Future Work

This chapter concludes this dissertation and is organized as follows: Section 7.1 presents the conclusions, followed by Section 7.2, which summarizes our main publications. Next, Section 7.3 discusses some limitations we observed during the development of this work. Finally, Section 7.4 describes possible future work.

7.1 Conclusions

For several years, researchers have used network-based models, particularly projections that we refer to here as many-to-many networks, to study various phenomena in the physical and online worlds. These models have indeed made it possible to extract an immeasurable range of knowledge in different knowledge domains. However, little attention has been paid to the various challenges naturally faced by these models. This dissertation set out to model and analyze collective behavior captured by many-to-many networks. We addressed several neglected challenges, most notably the presence of noise in the network and its drastic impact on the analysis. To achieve this, we defined three research goals:

- RG1: Uncovering topological and contextual properties of communities in *many-to-many* networks.
- RG2: Modeling the temporal dynamics of communities in *many-to-many* networks.
- RG3: Establishing a methodology for selecting and evaluating network backbone extraction methods in the face of a phenomenon modeled in *many-to-many* networks.

We addressed the first two research goals with results in the context of two very different case studies. Specifically, we were interested in uncovering and modeling structural, contextual, and temporal properties of communities in the context of a phenomenon under study. To this end, we proposed a general approach consisting of a sequence of steps that primarily address several challenges that have been mostly neglected in previous efforts to model collective behavior, especially when considered together. For some of these challenges, we proposed alternative solutions to those available in the literature. We then applied our general approach to these two very different case studies, bringing to light fundamental insights about the phenomena under study that would not have been observed without the challenges addressed here. We then extend this general approach by proposing a methodology that allows us to select and evaluate the backbone strategy that better captures the structural and contextual properties of the phenomenon. Finally, we analyzed this methodology in the face of two phenomena, one of those studied in the previous research goals and a new one. Thus, our main contributions to the modeling and analysis of collective behavior in many-to-many networks for RG1, RG2, and RG3 in the context of these three case studies can be summarized as follows:

RG1: Uncovering topological and contextual properties of communities in *many-to-many* networks

In the first case study, presented in Chapter 4, we were interested in examining the behavior of politicians, particularly members of the House of Representatives, during the legislative session in order to infer their political ideologies. To this end, we modeled roll call votes as a many-to-many network. In such system, the latent relationships that exist among congressmen who vote alike on roll call votes correspond to our notion of *co-interaction*. Thus, the collective behavior of interest in this case- the emergence of groups of congressmen with similar political ideologies- over time through co-interactions. As a natural consequence of such structure, some of these co-interactions are noisy and, when projected into a network, form a complete network (i.e., density equals one) and hide the actual ideological similarities that exist among congressmen. We have taken up the challenge of such modeling and analysis for two very different party systems (Brazil and the United States) covering a large 15-year period.

To identify the salient edges and remove the noisy ones from the network, we deliberately explored two strategies through contextual and structural information and uncovered the backbones and their topological and contextual properties. Specifically, we first applied a global threshold-based approach where the choice of the value to judge whether an edge is salient or not in each scenario is guided by contextual information (i.e., partisan discipline). In this way, we highlight how contextual information, which is often overlooked, can be used to identify salient edges. Yet, in characterizing such ideological communities and their structural and contextual properties, we found that they were nebulous to some degree in some periods, which is a particular feature of this phenomenon in Brazil. For this reason, we went a step further and proposed to apply a second local neighborhood-based aimed at extracting a backbone composed mainly of ideological communities composed only of the more polarized members. This approach allowed us to refine the originally extracted backbone structure as needed and to uncover new properties about the phenomenon that had not been observed before: more structured and more polarized ideological communities.

Most importantly, such strategy sheds light on the debate about the possibility of combining different methods to extract the backbone. So far, we have found that the above questions have been fundamental to gaining insights into the analyzed phenomenon, mainly considering that the original networks modeled are complete and do not favor the extraction of the patterns observed according to our strategies.

We then moved on to a second case study, presented in Chapter 5, which focuses on the study of online discussions, particularly in social media applications. Here, we focused on user comments on Instagram and proposed studying the co-interactions between users commenting on the same post by an influencer. We re-applied our general approach to four different scenarios (two political and two non-political) from two very different countries to investigate this phenomenon. Our novelty for modeling and analyzing collective behavior (online discussions) starts with networks of co-commenters, capturing the concept of co-interactions when groups of users collectively co-comment on the same post. In doing so, we found that the network model we used here is unprecedented in the literature because it captures co-interactions from the many-to-many perspective that can occur on topics of interest.

However, social media applications have characteristics that make hard the modeling and analysis of user behavior. Most notably, these include the heavy tail nature of content popularity and user activity, resulting in many edges that are not necessarily relevant to this study. Despite this, we have not found any backbone extraction techniques in the literature that explicitly exploit this information to identify salient edges. We then proposed TriBE, an alternative to backbone extraction that works by building a reference model under the null assumption *independent behavior* and considers these two main factors into account. Accordingly, we found that TriBE can capture co-interactions represented by edges arising from triggers that lead users to interact with certain content, thus reflecting real interactions driven by potentially interesting topics and not suffering from side effects such as those mentioned above.

Analyzing the structural and contextual properties of the backbones study extracted from TriBE, we found that there are stronger and well-structured communities representing groups of users who frequently participate in online discussions that were not seen in the original modeled networks without the application of TriBE. Moreover, under the assumption of independent behavior, we observed a much higher volume of interactions and thus intense discussions in the political context that, when characterized, reveal the presence of multiple triggers (e.g., topics) and patterns of communication and interest. Overall, TriBE has demonstrated its ability to capture how these interactions occur. When applied in four different contexts, both interesting similarities and topological and contextual differences between them became apparent.

RG2: Modeling the temporal dynamics of communities in *many-to-many* networks

We addressed the temporal properties of our two case studies discussed in RG1. Beginning with the first case study, i.e., a study of the behavior of House of Representatives politicians, we extracted patterns about the permanence and community affiliation of congressmen in the backbones over the observed period. Yet, we found that these metrics and others in the literature capture such community-level dynamics but do not track individual members over time. In addition, while strategies based on low-dimensional latent representations, such as embeddings, were beginning to be explored over time, we found that there were few and very limited efforts in this direction at the time. We then proposed a new method to jointly learn temporal node embeddings from a sequence of networks modeled by discrete-time windows - a particular contribution of this dissertation that builds on this case study. In summary, our method is based on two state-of-the-art approaches and allows us to obtain a temporal representation of the target system in the direction of temporal patterns. Furthermore, the results show that it is possible to capture the dynamics at the member level, which is an important contribution to the field as it allows a wide range of analyses.

In our second case, examining online discussions led by commenters on Instagram, we again quantified patterns related to the persistence and organization of individuals in communities extracted from the backbone over the analyzed period. We have shown that interest in particular debates changes dramatically over the weeks. This reflects the shifts in interest that occur in real-world society, particularly during election periods such as the one analyzed here. In addition, we also highlighted the dynamics of the discussion topics in which communities participate. In this regard, we observed a great diversity in dynamics, while some topics attract attention only for a short period, while others remain active for a longer time.

RG3: Establishing a methodology for selecting and evaluating network backbone extraction methods in the face of a phenomenon modeled in *many-to-many* networks

Finally, we dove into the major step of our general approach and addressed how to select and evaluate potential backbone methods in a studied phenomenon when ground truth is constantly lacking. To this end, we proposed a methodology in which we reviewed nine backbone extraction methods, characterize their assumptions and requirements, and discussed aspects to consider for their applicability in practice. Next, we identified the network characteristics that these approaches exploit by showing how they can be used to study different phenomena. Thus, our methodology explicitly considers the characteristics and correspondence of both methods and phenomenon as a step in method selection. In particular, our methodology brings together metrics for backbone quality that capture both structural and contextual (i.e., phenomenon-specific) aspects. Thus, it evaluates the resulting backbone from the perspective of the emerging structure and the extent to which it captures the phenomenon under study.

We applied the proposed methodology to two large-scale case studies related to phenomena with different requirements. Compared to other similar previous studies [Dai et al., 2018; Neal et al., 2021; Mukerjee et al., 2022], we offered a thorough and reasoned investigation that includes a larger number of state-of-the-art methods. Our results confirmed our interpretation that different methods can lead to very different backbones and that choosing the most appropriate method is paramount to gaining insight into the phenomenon under investigation. In particular, we found that our proposed method, TriBE, yielded the best results when analyzing online discussions on Instagram. In summary, we filled the gap of a systematic procedure for selecting and evaluating backbones by proposing a principled methodology for selecting the most appropriate backbone extraction method for a given phenomenon.

7.2 Publications

The results obtained according to our research goals were summarized in the following publications:

- Ferreira, C. H. G.; Souza, B. M.; Almeida, J. M.. Analyzing Dynamic Ideological Communities in Congressional Voting Networks. In: 10th International Conference on Social Informatics, 2018. (**RG1** and **RG2**).

- Ferreira, C. H. G.; Ferreira, F. M. ; Souza, B. M.; Almeida, J. M.. Modeling Dynamic Ideological Behavior in Political Networks. In: The Journal of Web Science, v. 1, p. 1-14, 2019. (**RG1** and **RG2**).

- Ferreira, C. H. G.; Murai, F.; Da Silva, A. P. C.; Almeida, J. M.; Trevisan, M.; Vassio, L.; Drago, I.; Mellia, M.. Unveiling Community Dynamics on Instagram Political Network. In: 12th ACM Conference on Web Science, 2020. (**RG1** and **RG2**).

- Ferreira, C. H. G.; Murai, F.; Da Silva, A. P. C.; Almeida, J. M.; Trevisan, M.; Vassio, L.; Drago, I.; Mellia, M.. On the Dynamics of Political Discussions on Instagram: A Network Perspective. In: Elsevier Online Social Networks and Media Journal, 2021. (**RG1** and **RG2**).

- Ferreira, C. H. G.; Murai, F.; Da Silva, A. P. C.; Trevisan, M.; Vassio, L.; Drago, I.; Mellia, M., Almeida, J. M.. On network backbone extraction for modeling online collective behavior. *Submitted to Plos One*, 2022. (**RG3**).

Our general approach and particular efforts developed here have been used in some collaborations. Therefore, we also note our involvement in those studies, which, although not part of this dissertation, should be mentioned because they are closely related to the efforts pursued here. We list the publications as results of these collaborations and explain our contribution in each case:

- Souza, B. M.; Ferreira, C. H. G.; Almeida, J. M.. *Analisando a governabilidade presidencial a partir de padrões de homofilia na Câmara dos Deputados: Estudos de Casos no Brasil e nos EUA*. In: VII Brazilian Workshop on Social Network Analysis and Mining, 2018.

- Nobre, G. P.; Ferreira, C. H. G.; Almeida, J. M.. Beyond Groups: Uncovering Dynamic Communities on the WhatsApp Network of Information Dissemination. In: 12th International Conference on Social Informatics, 2020.

- Nobre, G. P.; Ferreira, C. H. G.; Almeida, J. M.. Misinformation Dissemination on WhatsApp: A Hierarchical Network-Oriented Analysis. In: Information Processing & Management, 2021.

- Malagoli, L. G., Stancioli, J., Ferreira, C. H., Vasconcelos, M., Couto da Silva, A. P., Almeida, J. M.. A look into COVID-19 vaccination debate on Twitter. In: 13th ACM Web Science Conference, 2021.

- Barros, M. F., Ferreira, C. H., Santos, B. P. D., Júnior, L. A., Mellia, M., Almeida, J. M.. Understanding mobility in networks: A node embedding approach. In: Performance Evaluation Review, 2022.

In summary, the first work is a parallel effort to understand more aspects related to political governance issues on the phenomenon studied in our first case study (Chapter 4.1). The next two works focused on the spreading (mis-)information on WhatsApp by using our overall solution for modeling and analyzing many-to-many networks. However, only one method was used to extract the backbone, and the focus was on analyzing the content properties and propagation dynamics of users involved in misinformation, considering three perspectives: Individuals, WhatsApp groups, and communities of users. Next, the work of Malagoli *et al.* had our contribution in terms of the use of contextual characterization strategies, which we used in Chapter 5, for the study of a similar phenomenon but in a different case study (specifically Twitter). Finally, the work of Barros *et al.* is an extension and application of the temporal embedding method proposed here for the context of computer networks and focuses on the extraction of mobility patterns of groups of people.

7.3 Limitations

This dissertation investigated and addressed several challenges in modeling collective behavior through network modeling. However, the results should be considered in light of the existing limitations that are listed below.

Community Detection:

All of our case studies were based on a single community definition and a single algorithm was used to capture that definition. However, the definition of community in networks is very broad in the literature. As mentioned in Section 3.2, other algorithms consider different definitions to uncover different patterns of collective behavior. In addition, there is the possibility of exploring algorithms that consider both the structure and attributes of the network [Chunaev, 2020]. The growing interest in approaches of this type would allow exploration of contextual attributes to enrich the knowledge of the phenomenon and favor identifying communities that best capture the patterns of interest.

Temporal dynamics:

We have analyzed the dynamics of users, whether at the community or individual level, at various points in this dissertation. In particular, at the community level, we measured the extent to which individuals switch communities using normalized mutual information. Although this provides us with insights into community dynamics, this measure is limited from a community evolution perspective. In addition, other metrics available in the literature could be used to assess community dynamics [Rossetti et al., 2019]. A second limitation of the temporal context relates to the size of the fixed duration time windows. No studies were conducted to quantify the impact of this parameter on network modeling.

Sampled dataset:

In the case study of information dissemination via WhatsApp explored in Chapter 6, we used datasets that represent samples of the existing political WhatsApp groups. Thus, the media co-sharing network we built represents the potential channels through which information can be disseminated, but we cannot state if or when these channels were actually used as we only see a subset of the entire network. This is a limitation inherent in many social media applications due to crawling or API limitations, privacy issues, or even difficulties in managing and processing the often large amounts of data involved [Campan et al., 2018; Kumar and Carley, 2019].

Scalability:

We have not evaluated the computational cost and, consequently, the scalability of the different backbone extraction methods we used in this dissertation. However, given the modeling of larger-scale phenomena, this may be an important and limiting factor. Moreover, evaluating this aspect would allow for a more rational decision on which method is most appropriate for a given phenomenon, taking into account not only performance in capturing structural and contextual properties, but also execution time.

7.4 Future Work

We address here the main future work that offers a range of possibilities and, most importantly, explores most of the limitations mentioned above.

Social Phenomena Evolution:

Society changes, and as a direct consequence, the properties of phenomena change as well. In this way, co-interactions can take place in different ways, which poses some challenges to the choice of the backbone extraction method. In addition to the immediate application of the entire methodology developed here in other phenomena, the proprieties of both phenomena and the methods certainly need to be updated over time. Another point is that different strategies can be used to explore the notion of co-interaction based on contextual features that open up new immediate possibilities. For example, in the context of coordinated behavior, this possibility has recently been explored in various ways in [Pacheco et al., 2021].

Temporal dynamics:

One way to analyze the evolution of these communities in more detail is to characterize them in terms of events, particularly births, deaths, mergers, and splits of communities. An alternative for modeling such events that trigger such dynamics is the heuristic *event graph formalism* [Cazabet and Rossetti, 2019]. By adopting such a model, it would be possible to analyze in more detail how the phenomenon evolves in terms of community membership, how it differs in the original network and the backbone, and ultimately provide another measure to quantify the presence of noise in the original network.

Another possibility to be explored is the proposal of backbone extraction techniques that consider temporal properties over a sequence of time windows. The idea is to identify the best backbone consisting of users that remain essentially stable in their interactions across different time windows. Furthermore, considering such dimension allows us to study phenomena where consistent user interactions may be of central importance, e.g., co-authorship, coordinated behavior, mobility, and others. Finally, we highlighted that this issue of community stability could be quantified using a latent space obtained through our temporal node embedding technique proposed in Chapter 5.

As for temporal dynamics, another possibility is to analyze the effects of time windows of different sizes. Indeed, this has already been discussed in the literature [Krings et al., 2012; Ribeiro et al., 2013; Valdano et al., 2015], but only recently we observe an advance in approaches that are quantitatively interpretable and allow us to determine the impact of the size of the chosen time window [Orman et al., 2021; Chiappori and Cazabet, 2021]. Thus, there is the possibility of dynamically exploring the identification of the time window when the co-interactions become more extensive than generally observed.

Sampled dataset:

When working with a sampled dataset, one question that arises is how to achieve a reliable extraction of the backbone since the sampling procedure may directly impact the topological structure of the network and, consequently, on the salience analysis of an edge. Some works have addressed the challenges of working with samples from the perspective of potential bias [González-Bailón et al., 2014; Miranda Filho et al., 2015] and sampling strategies [Gjoka et al., 2010; Liakos et al., 2017]. Another related work has proposed a strategy based on the noise-corrected backbone extraction method to select edges in networks based on sampling processes [Coscia, 2021]. The goal is to select edges to achieve the largest information gain over the entire network structure. This first effort opens new possibilities for other backbone extraction techniques that can be extended to address this problem.

In addition, an immediate opportunity is to quantify how sensitive backbone extraction techniques are to sampled networks by using full networks as a ground truth. This can be done using the Instagram dataset that we used in Chapter 5. Given the crawling date, our dataset is complete because it contains all users who commented on the posts of the target influencers. In this way, it is possible to apply several sampling alternatives described in the literature, simulate the criteria of APIs such as Twitter that provide samples of only 1% of tweets, and compare the backbones extracted from the networks with and without sampling methods to quantify the extent to which this affects the study of a particular phenomenon.

Scalability:

It is not new that the volume of stored data is growing every year, as it is the interest in various phenomena that can be studied with those data. Therefore, scalability is an issue that arises in the methods used here. In terms of network modeling and community detection, there are several approaches that explore parallel and distributed execution, such as Apache Sparth Graphx[1], Neo4j[2], and Distributed Graph Analytics (DGA)[3]. However, for backbone extraction, a few studies evaluate the scalability of the approaches reviewed here [Radicchi et al., 2011; Coscia and Neffke, 2017], but even those do not take advantage of parallel and distributed computing. Thus, there is an opportunity to make a rather technical but very valuable contribution to the performance analysis and parallelization of backbone extraction methods, which hopefully, after this dissertation, will be considered a crucial step for the modeling and analysis of many-to-many networks.

[1] https://spark.apache.org/docs/latest/graphx-programming-guide.html

[2] https://neo4j.com/

[3] https://sotera.github.io/distributed-graph-analytics/

Bibliography

Abbar, S., Zanouda, T., Berti-Equille, L., and Borge-Holthoefer, J. (2016). Using twitter to understand public interest in climate change: The case of qatar. In *Tenth International AAAI Conference on Web and Social Media*.

Abraham, A. and Hassanien, A.-E. (2012). *Computational social networks: Tools, perspectives and applications*. Springer Science & Business Media.

Agathangelou, P., Katakis, I., Rori, L., Gunopulos, D., and Richards, B. (2017). Understanding online political networks: The case of the far-right and far-left in greece. In *International Conference on Social Informatics*.

Ahn, Y.-Y., Han, S., Kwak, H., Moon, S., and Jeong, H. (2007). Analysis of topological characteristics of huge online social networking services. In *International Conference on World Wide Web*.

Al-Garadi, M. A., Varathan, K. D., Ravana, S. D., Ahmed, E., Mujtaba, G., Khan, M. U. S., and Khan, S. U. (2018). Analysis of online social network connections for identification of influential users: Survey and open research issues. *ACM Computing Surveys*, 51(1):1--37.

Ames, B. (2001). *The Deadlock of Democracy in Brazil*. University of Michigan Press.

Andris, C., Lee, D., Hamilton, M. J., Martino, M., Gunning, C. E., and Selden, J. A. (2015). The rise of partisanship and super-cooperators in the u.s. house of representatives. *PLOS ONE*, 10.

Arinik, N., Figueiredo, R., and Labatut, V. (2017). Signed graph analysis for the interpretation of voting behavior. In *International Workshop on Social Network Analysis and Digital Humanities*.

Arslan, P., Corazza, M., Cabrio, E., and Villata, S. (2019). Overwhelmed by negative emotions? maybe you are being cyber-bullied! In *Symposium on Applied Computing*.

Awal, G. K. and Bharadwaj, K. K. (2019). Leveraging collective intelligence for behavioral prediction in signed social networks through evolutionary approach. *Information Systems Frontiers*, 21(2):417--439.

Badawy, A., Ferrara, E., and Lerman, K. (2018). Analyzing the digital traces of political manipulation: The 2016 russian interference twitter campaign. In *International Conference on Advances in Social Networks Analysis and Mining*.

Baeza-Yates, R., Ribeiro-Neto, B., et al. (1999). *Modern information retrieval*, volume 463. ACM press New York, New York.

Bafumi, J., Gelman, A., Park, D. K., and Kaplan, N. (2005). Practical issues in implementing and understanding bayesian ideal point estimation. *Political Analysis*, 13:171--187.

Baldassarri, D. and Gelman, A. (2008). Partisans without constraint: Political polarization and trends in american public opinion. *American Journal of Sociology*, 114(2):408--446.

Bamler, R. and Mandt, S. (2017). Dynamic word embeddings. In *International Conference on Machine Learning*, Proceedings of Machine Learning Research.

Banerjee, A., Dhillon, I. S., Ghosh, J., and Sra, S. (2005). Clustering on the unit hypersphere using von mises-fisher distributions. *Journal of Machine Learning Research*, 6:1345--1382.

Barabási, A.-L. et al. (2016). *Network science*. Cambridge university press.

Barros, C. D., Mendonça, M. R., Vieira, A. B., and Ziviani, A. (2021a). A survey on embedding dynamic graphs. *ACM Computing Surveys*, 55(1).

Barros, M. F., Ferreira, C. H., Santos, B. P. d., Júnior, L. A., Mellia, M., and Almeida, J. M. (2021b). Understanding mobility in networks: A node embedding approach. *Performance Evaluation Review*.

Bateman, D. A. and Lapinski, J. (2016). Ideal points and american political development: Beyond dw-nominate. *Studies in American Political Development*, 30(2):147--171.

Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., Young, J.-G., and Petri, G. (2020). Networks beyond pairwise interactions: structure and dynamics. *Physics Reports*.

Belhadi, A., Djenouri, Y., Srivastava, G., Djenouri, D., Lin, J. C.-W., and Fortino, G. (2021). Deep learning for pedestrian collective behavior analysis in smart cities: A model of group trajectory outlier detection. *Information Fusion*, 65.

Benson, A. R., Abebe, R., Schaub, M. T., Jadbabaie, A., and Kleinberg, J. (2018a). Simplicial closure and higher-order link prediction. *Proceedings of the National Academy of Sciences*, pages 11221--11230.

Benson, A. R., Gleich, D. F., and Leskovec, J. (2016). Higher-order organization of complex networks. *Science*, 353(6295).

Benson, A. R., Kumar, R., and Tomkins, A. (2018b). Sequences of sets. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.

Bhattacharyya, S. and Chatterjee, S. (2017). Spectral clustering for multiple sparse networks. *Biometrika*, 103(1):1--28.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008:1--13.

Bogdanov, P., Busch, M., Moehlis, J., Singh, A. K., and Szymanski, B. K. (2014). Modeling individual topic-specific behavior and influence backbone networks in social media. *Social Network Analysis and Mining*, 4(1):204.

Bonaventura, M., Aiello, L. M., Quercia, D., and Latora, V. (2021). Predicting urban innovation from the us workforce mobility network. *Humanities and Social Sciences Communications*, 8(1):1--9.

Bordier, C., Nicolini, C., and Bifone, A. (2017). Graph analysis and modularity of brain functional connectivity networks: searching for the optimal threshold. *Frontiers in neuroscience*, 11:1--9.

Brady, W. J., Wills, J. A., Burkart, D., Jost, J. T., and Van Bavel, J. J. (2019). An ideological asymmetry in the diffusion of moralized content on social media among political leaders. *Journal of Experimental Psychology: General*, 148.

Brito, A. C. M., Silva, F. N., and Amancio, D. R. (2020). A complex network approach to political analysis: Application to the brazilian chamber of deputies. *PLOS ONE*, 15(3):1–21.

Budge, I. and Laver, M. J. (2016). *Party Policy and Government Coalitions*. Springer.

Caetano, J. A., Almeida, J., and Marques-Neto, H. T. (2018). Characterizing politically engaged users' behavior during the 2016 us presidential campaign. In *International Conference on Advances in Social Networks Analysis and Mining*.

Caetano, J. A., Magno, G., Gonçalves, M. A., Almeida, J. M., Marques-Neto, H. T., and Almeida, V. A. F. (2019). Characterizing attention cascades in whatsapp groups. In *ACM Conference on Web Science*.

Camacho, J., Stouffer, D. B., and Amaral, L. A. N. (2007). Quantitative analysis of the local structure of food webs. *Journal of theoretical biology*, 246(2):260--268.

Campan, A., Atnafu, T., Truta, T. M., and Nolan, J. (2018). Is data collection through twitter streaming api useful for academic research? In *IEEE international conference on big data (Big Data)*.

Cann, T. J., Weaver, I. S., and Williams, H. T. (2021). Ideological biases in social sharing of online information about climate change. *Plos one*, 16(4).

Cao, H., Chen, Z., Xu, F., Li, Y., and Kostakos, V. (2018). Revisitation in urban space vs. online: A comparison across pois, websites, and smartphone apps. *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):1--24.

Cao, J., Ding, C., and Shi, B. (2019). Motif-based functional backbone extraction of complex networks. *Physica A: Statistical Mechanics and its Applications*, 526.

Cao, Q., Yang, X., Yu, J., and Palow, C. (2014). Uncovering large groups of active malicious accounts in online social networks. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*.

Cazabet, R. and Rossetti, G. (2019). *Challenges in Community Discovery on Temporal Networks*, pages 181--197. Springer International Publishing, Cham.

Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., and Leskovec, J. (2017). Anyone can become a troll: Causes of trolling behavior in online discussions. In *ACM conference on computer supported cooperative work and social computing*.

Cherepnalkoski, D., Karpf, A., Mozetič, I., and Grčar, M. (2016). Cohesion and coalition formation in the european parliament: Roll-call votes and twitter activities. *PLOS ONE*, 11(11).

Chi, F. and Yang, N. (2010). Twitter in congress: Outreach vs transparency. Mpra paper, University Library of Munich, Germany.

Chiappori, A. and Cazabet, R. (2021). Quantitative evaluation of snapshot graphs for the analysis of temporal networks. In *International Conference on Complex Networks and Their Applications*. Springer.

Chunaev, P. (2020). Community detection in node-attributed social networks: a survey. *Computer Science Review*, 37:100286.

Clinton, J., Jackman, S., and Rivers, D. (2004). The statistical analysis of roll call data. *American Political Science Review*, 98(2):355--370.

Cohen, R. and Havlin, S. (2010). *Complex networks: structure, robustness and function*. Cambridge university press.

Coscia, M. (2021). Noise corrected sampling of online social networks. *ACM Transactions on Knowledge Discovery from Data*, 15(2):1--21.

Coscia, M. and Neffke, F. M. (2017). Network backboning with noisy data. In *International Conference on Data Engineering*.

Coscia, M., Neffke, F. M., and Hausmann, R. (2020). Knowledge diffusion in the network of international business travel. *Nature Human Behaviour*, 4(10):1011--1020.

Coscia, M. and Rossi, L. (2019). The impact of projection and backboning on network topologies. In *International Conference on Advances in Social Networks Analysis and Mining*.

Crovella, M. E., Taqqu, M. S., and Bestavros, A. (1998). Heavy-tailed probability distributions in the world wide web. *A practical guide to heavy tails*, 1:3--26.

Cruickshank, I. J. and Carley, K. M. (2020). Characterizing communities of hashtag usage on twitter during the 2020 covid-19 pandemic by multi-view clustering. *Applied Network Science*, 5(1):1--40.

Csermely, P., London, A., Wu, L.-Y., and Uzzi, B. (2013). Structure and dynamics of core/periphery networks. *Journal of Complex Networks*, 1(2):93--123.

Cui, P., Wang, X., Pei, J., and Zhu, W. (2018). A survey on network embedding. *IEEE Transactions on Knowledge and Data Engineering*.

Dai, L., Derudder, B., and Liu, X. (2018). Transport network backbone extraction: A comparison of techniques. *Journal of Transport Geography*, 69:271--281.

Dal Maso, C., Pompa, G., Puliga, M., Riotta, G., and Chessa, A. (2015). Voting behavior, coalitions and government strength through a complex network analysis. *PLOS ONE*, 9.

Darwish, K., Magdy, W., and Zanouda, T. (2017). Trump vs. hillary: What went viral during the 2016 us presidential election. In *International Conference on Social Informatics*.

Davenport, M. A. and Romberg, J. (2016). An overview of low-rank matrix recovery from incomplete observations. *Journal of Selected Topics in Signal Processing*, 10(4):608–622.

David, E. and Jon, K. (2010). *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, USA. ISBN 0521195330.

De Domenico, M. and Altmann, E. G. (2020). Unraveling the origin of social bursts in collective attention. *Scientific reports*, 10(1):1--9.

de Freitas Melo, P., Vieira, C. C., Garimella, K., de Melo, P. O. V., and Benevenuto, F. (2019). Can whatsapp counter misinformation by limiting message forwarding? In *International conference on complex networks and their applications*. Springer.

de Melo, P. O. V. (2015). How many political parties should brazil have? a data-driven method to assess and reduce fragmentation in multi-party political systems. *PloS one*, 10.

De Melo, P. O. V., Viana, A. C., Fiore, M., Jaffrès-Runser, K., Le Mouël, F., Loureiro, A. A., Addepalli, L., and Guangshuo, C. (2015). Recast: Telling apart social and random relationships in dynamic networks. *Performance Evaluation*, 87:19--36.

Del Re, E. C. (2013). *Collective Behavior*, pages 422--429. Springer Netherlands, Dordrecht.

Desposato, S. W. (2006). Parties for rent? ambition, ideology, and party switching in brazil's chamber of deputies. *American Journal of Political Science*, 50(1):62--80.

Dianati, N. (2016). Unwinding the hairball graph: Pruning algorithms for weighted complex networks. *Physical Review E*, 93(1):012304.

Ehm, W. (1991). Binomial approximation to the poisson binomial distribution. *Statistics & Probability Letters*, 11:7--16.

Eidelman, V., Kornilova, A., and Argyle, D. (2018). How predictable is your state? leveraging lexical and contextual information for predicting legislative floor action at the state level. In *International Conference on Computational Linguistics*.

ElSherief, M., Nilizadeh, S., Nguyen, D., Vigna, G., and Belding, E. (2018). Peer to peer hate: Hate speech instigators and their targets. In *International Conference on Web and Social Media*.

Ferreira, C. H., Murai, F., Silva, A. P., Almeida, J. M., Trevisan, M., Vassio, L., Mellia, M., and Drago, I. (2021). On the dynamics of political discussions on instagram: A network perspective. *Online Social Networks and Media*, 25:100155.

Ferreira, C. H. G., Murai, F., Couto da Silva, A. P., de Almeida, J. M., Trevisan, M., Vassio, L., Drago, I., and Mellia, M. (2020). Unveiling community dynamics on instagram political network. In *ACM Conference on Web Science*.

Ferreira, C. H. G., Murai, F., de Souza Matos, B., and de Almeida, J. M. (2019). Modeling dynamic ideological behavior in political networks. *The Journal of Web Science*, 7.

Fiorina, M. P. and Abrams, S. J. (2008). Political polarization in the american public. *Annual Review of Political Science*, 11(1):563–588.

Fortunato, S. (2010). Community detection in graphs. *Physics reports*, 486(3-5):75--174.

Fortunato, S. and Barthelemy, M. (2007). Resolution limit in community detection. *Proceedings of the national academy of sciences*, 104(1):36--41.

Fortunato, S. and Hric, D. (2016). Community detection in networks: A user guide. *Physics reports*, 659:1--44.

Founta, A. M., Chatzakou, D., Kourtellis, N., Blackburn, J., Vakali, A., and Leontiadis, I. (2019). A unified deep learning architecture for abuse detection. In *ACM Conference on Web Science*.

Fournier-Viger, P., He, G., Cheng, C., Li, J., Zhou, M., Lin, J. C.-W., and Yun, U. (2020). A survey of pattern mining in dynamic graphs. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*.

Fraga, B. S., da Silva, A. P. C., and Murai, F. (2018). Online social networks in health care: A study of mental disorders on reddit. In *International Conference on Web Intelligence*.

Fu, X., Yu, S., and Benson, A. R. (2019). Modelling and analysis of tagging networks in Stack Exchange communities. *Journal of Complex Networks*.

Galuppo Azevedo, F. and Murai, F. (2021). Evaluating the state-of-the-art in mapping research spaces: A brazilian case study. *PloS one*, 16(3):e0248724.

Gao, C. and Liu, J. (2016). Network-based modeling for characterizing human collective behaviors during extreme events. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(1):171--183.

Garretón, M., Rihm, A., and Parra, D. (2019). #default #interactiveart #audiencexperience. In *The World Wide Web Conference*.

Gehman, S., Gururangan, S., Sap, M., Choi, Y., and Smith, N. A. (2020). RealToxicityPrompts: Evaluating neural toxic degeneration in language models. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3356--3369, Online. Association for Computational Linguistics.

Ghalmane, Z., Cherifi, C., Cherifi, H., and El Hassouni, M. (2020). Extracting backbones in weighted modular complex networks. *Scientific Reports*, 10(1):1--18.

Gholamy, A., Kreinovich, V., and Kosheleva, O. (2018). Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation. *Technical Report*.

Gjoka, M., Kurant, M., Butts, C. T., and Markopoulou, A. (2010). Walking in facebook: A case study of unbiased sampling of osns. In *2010 Proceedings IEEE INFOCOM*.

Gleeson, J. P., Cellai, D., Onnela, J.-P., Porter, M. A., and Reed-Tsochas, F. (2014). A simple generative model of collective online behavior. *Proceedings of the National Academy of Sciences*, 111(29):10411--10415.

González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J., and Moreno, Y. (2014). Assessing the bias in samples of large online networks. *Social Networks*, 38:16--27.

Gorkovenko, K. and Taylor, N. (2017). Understanding how people use twitter during election debates. In *International BCS Human Computer Interaction Conference*.

Goyal, P., Chhetri, S. R., and Canedo, A. (2018a). dyngraph2vec: Capturing network dynamics using dynamic graph representation learning.

Goyal, P., Kamra, N., He, X., and Liu, Y. (2018b). Dyngem: Deep embedding method for dynamic graphs. In *International Workshop on Representation Learning for Graphs*.

Grabowicz, P. A., Ramasco, J. J., Gonçalves, B., and Eguíluz, V. M. (2014). Entangling mobility and interactions in social media. *PloS one*, 9(3):e92196.

Grady, D., Thiemann, C., and Brockmann, D. (2012). Robust classification of salient links in complex networks. *Nature communications*, 3(1):1--10.

Granovetter, M. S. (1973). The Strength of Weak Ties. *The American Journal of Sociology*, 78(6):1360--1380.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., and Lazer, D. (2019). Fake news on twitter during the 2016 us presidential election. *Science*, 363(6425):374--378.

Grover, A. and Leskovec, J. (2016). node2vec: Scalable feature learning for networks. In *International Conference on Knowledge Discovery and Data Mining*.

Gruzd, A. and Roy, J. (2014). Investigating political polarization on twitter: A canadian perspective. *Policy & internet*, 6(1):28--45.

Guidotti, R., Rossetti, G., Pappalardo, L., Giannotti, F., and Pedreschi, D. (2019). Personalized market basket prediction with temporal annotated recurring sequences. *IEEE Transactions on Knowledge and Data Engineering*, 11.

Gupta, A., Yang, W., Sivakumar, D., Silva, Y., Hall, D., and Nardini Barioni, M. (2020). Temporal properties of cyberbullying on instagram. In *The Web Conference*.

Hamedmoghadam, H., Ramezani, M., and Saberi, M. (2019). Revealing latent characteristics of mobility networks with coarse-graining. *Scientific reports*, 9(1):1--10.

Han, J., Li, W., and Deng, W. (2016). Multi-resolution community detection in massive networks. *Scientific reports*, 6(1):1--12.

Hastie, T. J. and Tibshirani, R. J. (2017). *Generalized additive models*. Routledge.

Hinds, P. J., Carley, K. M., Krackhardt, D., and Wholey, D. (2000). Choosing work group members: Balancing similarity, competence, and familiarity. *Organizational Behavior and Human Decision Processes*, 81(2).

Hix, S., Noury, A., and Roland, G. (2005). Power to the parties: Cohesion and competition in the european parliament, 1979–2001. *British Journal of Political Science*, 35(2):209--234.

Hong, Y. (2013). On computing the distribution function for the poisson binomial distribution. *Computational Statistics & Data Analysis*, 59:41--51.

Hoppe, F. M. (1984). Pólya-like urns and the ewens' sampling formula. *Journal of Mathematical Biology*, 20(1):91--94.

Hu, H. and He, X. (2019). Sets2sets: Learning from sequential sets with neural networks. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.

Huang, L., Wang, C.-D., and Chao, H.-Y. (2019). Higher-order multi-layer community detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Huang, X., Li, J., and Hu, X. (2017). Accelerated attributed network embedding. In *International Conference on Data Mining*.

Jaakonmäki, R., Müller, O., and Vom Brocke, J. (2017). The impact of content, context, and creator on user engagement in social media marketing. In *The Hawaii International Conference on System Sciences*.

Jacobs, A. (2015). *A unified view of generative models for networks: models, methods, opportunities, and challenges*. PhD dissertation, University of British Columbia.

Jaffe, K., Ter Horst, E., Gunn, L. H., Zambrano, J. D., and Molina, G. (2020). A network analysis of research productivity by country, discipline, and wealth. *Plos one*, 15(5):e0232458.

Jain, R. (1991). *The art of computer systems performance analysis - techniques for experimental design, measurement, simulation, and modeling*. Wiley professional computing. Wiley.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.

Jia, C., Li, Y., Carson, M. B., Wang, X., and Yu, J. (2017). Node attribute-enhanced community detection in complex networks. *Scientific Reports*, 7(1).

Jiang, J., Chen, E., Yan, S., Lerman, K., and Ferrara, E. (2020). Political polarization drives online conversations about covid-19 in the united states. *Human Behavior and Emerging Technologies*, 2(3):200--211.

Jin, R., McCallen, S., and Almaas, E. (2007). Trend motif: A graph mining approach for analysis of dynamic complex networks. In *International Conference on Data Mining*.

Jones, J. J., Settle, J. E., Bond, R. M., Fariss, C. J., Marlow, C., and Fowler, J. H. (2013). Inferring tie strength from online directed behavior. *PLOS ONE*, 8.

Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*.

Kaalia, R. and Rajapakse, J. C. (2019). Refining modules to determine functionally significant clusters in molecular networks. *BMC genomics*, 20(9):1--14.

Kang, J., Yoon, J., and Han, J. (2020). Why do instagram users tag friends in comments? In *The Web Conference*.

Kao, H.-T., Yan, S., Huang, D., Bartley, N., Hosseinmardi, H., and Ferrara, E. (2019). Understanding cyberbullying on instagram and ask.fm via social role detection. In *The Web Conference*.

Kim, S., Jiang, J.-Y., Nakada, M., Han, J., and Wang, W. (2020). Multimodal post attentive profiling for influencer marketing. In *The Web Conference*.

Kobayashi, T., Takaguchi, T., and Barrat, A. (2019). The structured backbone of temporal social ties. *Nature communications*, 10(1):1--11.

Kornilova, A., Argyle, D., and Eidelman, V. (2018). Party matters: Enhancing legislative embeddings with author attributes for vote prediction. In *Annual Meeting of the Association for Computational Linguistics*.

Krings, G., Karsai, M., Bernhardsson, S., Blondel, V. D., and Saramäki, J. (2012). Effects of time window size and placement on the structure of an aggregated communication network. *EPJ Data Science*, 1(1).

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583--621.

Krzakala, F., Moore, C., Mossel, E., Neeman, J., Sly, A., Zdeborová, L., and Zhang, P. (2013). Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935--20940.

Kumar, S. and Carley, K. M. (2019). What to track on the twitter streaming api? a knapsack bandits approach to dynamically update the search terms. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*.

Kumar, S., Hamilton, W. L., Leskovec, J., and Jurafsky, D. (2018). Community interaction and conflict on the web. In *The World Wide Web Conference*.

Kumar, S., Honnungar, S., and Leskovec, J. (2019). Representation learning in continuous entity-set associations. In *arxiv.org*.

Kumar, T., Darwin, K., Parthasarathy, S., and Ravindran, B. (2020). Hpra: Hyperedge prediction using resource allocation. In *ACM Conference on Web Science*.

Kusner, M., Sun, Y., Kolkin, N., and Weinberger, K. (2015). From word embeddings to document distances. In *International Conference on Machine Learning*.

Labatut, V. and Balasque, J.-M. (2012). Detection and interpretation of communities in complex networks: Practical methods and application. In *Computational Social Networks*, pages 81--113. Springer.

Lambiotte, R., Delvenne, J.-C., and Barahona, M. (2008). Laplacian dynamics and multiscale modular structure in networks. *arXiv preprint arXiv:0812.1770*.

Lauderdale, B. E. and Herzog, A. (2016). Measuring political positions from legislative speech. *Political Analysis*, 24(3):374--394.

Leão, J. C., Brandão, M. A., de Melo, P. O. V., and Laender, A. H. (2018). Who is really in my social circle? *Journal of Internet Services and Applications*, 9.

Lee, C. and Wilkinson, D. J. (2019). A review of stochastic block models and extensions for graph clustering. *Applied Network Science*, 4(1):1--50.

Lee, J., Lee, Y., Kim, J., Kosiorek, A., Choi, S., and Teh, Y. W. (2019). Set transformer: A framework for attention-based permutation-invariant neural networks. In *International Conference on Machine Learning*.

Lee, S. H., Magallanes, J. M., and Porter, M. A. (2017). Time-dependent community structure in legislation cosponsorship networks in the congress of the republic of peru. *Journal of Complex Networks*, 5(1):127--144.

Lehmann, J., Gonçalves, B., Ramasco, J. J., and Cattuto, C. (2012). Dynamical classes of collective attention in twitter. In *International Conference on World Wide Web*.

Lerman, R. (2020). Whatsapp, instagram surge in popularity as people search for coronavirus news. `https://www.washingtonpost.com/technology/2020/06/16/whatsapp-instagram-covid-news/`. Accessed: 2020-08-03.

Lerman, R. I. and Yitzhaki, S. (1984). A note on the calculation and interpretation of the gini index. *Economics Letters*, 15(3).

Lerner, J., Tranmer, M., Mowbray, J., and Hancean, M.-G. (2019). Rem beyond dyads: relational hyperevent models for multi-actor interaction networks. *arXiv preprint arXiv:1912.07403*.

Leskovec, J., Lang, K. J., and Mahoney, M. (2010). Empirical comparison of algorithms for network community detection. In *International conference on World wide web*.

Levorato, M. and Frota, Y. (2017). Brazilian Congress structural balance analysis. *Journal of Interdisciplinary Methodologies and Issues in Science*.

Levy, O. and Goldberg, Y. (2014). Neural word embedding as implicit matrix factorization. In *International Conference on Neural Information Processing Systems*, volume 2.

Li, J., Dani, H., Hu, X., Tang, J., Chang, Y., and Liu, H. (2017). Attributed network embedding for learning in a dynamic environment. In *The Conference on Information and Knowledge Management*.

Liakos, P., Ntoulas, A., and Delis, A. (2017). Rhea: Adaptively sampling authoritative content from social activity streams. In *IEEE International Conference on Big Data (Big Data)*.

Liebig, J. and Rao, A. (2016). Fast extraction of the backbone of projected bipartite networks to aid community detection. *EPL (Europhysics Letters)*, 113(2):28003.

Liu, J.-G., Guo, Q., Zhang, Y.-C., et al. (2018a). Collective iteration behavior for online social networks. *Physica A: Statistical Mechanics and its Applications*, 499.

Liu, P., Benson, A. R., and Charikar, M. (2019). Sampling methods for counting temporal motifs. In *International Conference on Web Search and Data Mining*.

Liu, Q., Liu, Q., Yang, L., and Wang, G. (2018b). A multi-granularity collective behavior analysis approach for online social networks. *Granular Computing*, 3(4):333--343.

Lord, L.-D., Expert, P., Fernandes, H. M., Petri, G., Van Hartevelt, T. J., Vaccarino, F., Deco, G., Turkheimer, F., and Kringelbach, M. L. (2016). Insights into brain architectures from the homological scaffolds of functional connectivity networks. *Frontiers in Systems Neuroscience*, 10:85.

Lu, Y., Wang, X., Shi, C., Yu, P. S., and Ye, Y. (2019a). Temporal network embedding with micro-and macro-dynamics. In *International Conference on Information and Knowledge Management*.

Lu, Y., Yu, L., Cui, P., Zang, C., Xu, R., Liu, Y., Li, L., and Zhu, W. (2019b). Uncovering the co-driven mechanism of social and content links in user churn phenomena. In *International Conference on Knowledge Discovery & Data Mining*.

Lu, Y., Yu, L., Zhang, T., Zang, C., Cui, P., Song, C., and Zhu, W. (2020). Exploring the collective human behavior in cascading systems: a comprehensive framework. *Knowledge and Information Systems*, 62(12):4599--4623.

Maaten, L. v. d. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9:2579--2605.

Mackay, C. (2012). *Extraordinary popular delusions and the madness of crowds*. Simon and Schuster.

Mahdavi, S., Khoshraftar, S., and An, A. (2018). dynnode2vec: Scalable dynamic network embedding. In *Workshop on Advances in High Dimensional Big Data*.

Mainwaring, S. and Shugart, M. S. (1997). *Presidentialism and Democracy in Latin America*. Cambridge University Press.

Malagoli, L. G., Stancioli, J., Ferreira, C. H., Vasconcelos, M., Couto da Silva, A. P., and Almeida, J. M. (2021). A look into covid-19 vaccination debate on twitter. In *13th ACM Web Science Conference 2021*.

Malang, K., Wang, S., Phaphuangwittayakul, A., Lv, Y., Yuan, H., and Zhang, X. (2020). Identifying influential nodes of global terrorism network: A comparison for skeleton network extraction. *Physica A: Statistical Mechanics and its Applications*, 545:123769.

Malik, M. M., Lamba, H., Nakos, C., and Pfeffer, J. (2015). Population bias in geo-tagged tweets. In *Ninth international AAAI conference on web and social media.*

Mann, T. E. and Ornstein, N. J. (2016). *It's Even Worse Than it Looks: How the American Constitutional System Collided with the New Politics of Extremism.* Basic Books.

Marcaccioli, R. (2020). *Null models and complexity science: disentangling signal from noise in complex interacting systems.* PhD dissertation, UCL (University College London).

Marcaccioli, R. and Livan, G. (2019). A pólya urn approach to information filtering in complex networks. *Nature communications*, 10(1):1--10.

Maros, A., Almeida, J., Benevenuto, F., and Vasconcelos, M. (2020). Analyzing the use of audio messages in whatsapp groups. In *The Web Conference.*

Martin-Gutierrez, S., Losada, J. C., and Benito, R. M. (2020). Impact of individual actions on the collective response of social systems. *Scientific reports*, 10.

Mateo, D., Horsevad, N., Hassani, V., Chamanbaz, M., and Bouffanais, R. (2019). Optimal network topology for responsive collective behavior. *Science advances*, 5(4).

Mattsson, C. (2020). *Financial Transaction Networks to Describe and Model Economic Systems.* PhD dissertation, Northeastern University.

Mattsson, C. E. and Stuart, G. (2020). Tracking movement of resources in a financial transaction network. US Patent App. 16/796,332.

McClellan, C., Ali, M. M., Mutter, R., Kroutil, L., and Landwehr, J. (2017). Using social media to monitor mental health discussions- evidence from twitter. *Journal of the American Medical Informatics Association*, 24:496--502.

McGee, J., Caverlee, J., and Cheng, Z. (2013). Location prediction in social media based on tie strength. In *International Conference on Information & Knowledge Management.*

Menczer, F., Fortunato, S., and Davis, C. A. (2020). *A first course in network science.* Cambridge University Press.

Mendonça, I., Trouve, A., and Fukuda, A. (2017). Exploring the importance of negative links through the european parliament social graph. In *International Conference on E-Society, E-Education and E-Technology.*

Meng, C., Mouli, S. C., Ribeiro, B., and Neville, J. (2018). Subgraph pattern neural networks for high-order graph evolution prediction. In *AAAI Conference on Artificial Intelligence*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26:3111--3119.

Milenković, T. and Pržulj, N. (2008). Uncovering biological network function via graphlet degree signatures. *Cancer Informatics*, 6.

Miranda Filho, R., Almeida, J. M., and Pappa, G. L. (2015). Twitter population sample bias and its impact on predictive outcomes: a case study on elections. In *IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*.

Mitchell, M. and Newman, M. (2002). Complex systems theory and evolution. *Encyclopedia of Evolution*, pages 1--5.

Mitra, T., Wright, G., and Gilbert, E. (2017). Credibility and the dynamics of collective attention. *ACM on Human-Computer Interaction*, 1:1--17.

Mondal, M., Silva, L. A., and Benevenuto, F. (2017). A measurement study of hate speech in social media. In *ACM conference on hypertext and social media*.

Moody, J. and Mucha, P. J. (2013). Portrait of political party polarization. *Network Science*, 1(1).

Muñoz, C. L. and Towner, T. L. (2017). The image is the message: Instagram marketing and the 2016 presidential primary season. *Journal of Political Marketing*, 16(3-4):290--318.

Mueller, W., Silva, T. H., Almeida, J. M., and Loureiro, A. A. (2017). Gender matters! analyzing global cultural gender preferences for venues using social sensing. *EPJ Data Science*, 6(1):5.

Mukerjee, S., Yang, T., Stadler, G., and González-Bailón, S. (2022). What counts as a weak tie? a comparison of filtering techniques to analyze co-exposure networks. *Social Networks*, 68:386--393.

Mukhopadhyay, D., Mishra, K., Mishra, K., and Tiwari, L. (2020). Cyber bullying detection based on twitter dataset. In *Machine Learning for Predictive Analysis*. Springer.

Namaki, A., Shirazi, A., Raei, R., and Jafari, G. (2011). Network analysis of a financial market based on genuine correlation and threshold method. *Physica A: Statistical Mechanics and its Applications*, 390(21-22).

Neal, Z. P. (2020). A sign of the times? weak and strong polarization in the us congress, 1973–2016. *Social Networks*, 60:103--112.

Neal, Z. P., Domagalski, R., and Sahan, B. (2021). Comparing models for extracting the backbone of bipartite projections.

Newman, M. (2018). Network structure from rich but noisy data. *Nature Physics*, 14:542--546.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2):167--256.

Newman, M. E. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3):036104.

Newman, M. E. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69:026113.

Newman, N., Fletcher, Schulz, A., Andi, S., and Nielse, R. K. (2020). Da, & nielsen, r.(2020). reuters institute digital news report 2020. *Reuters Institute for the Study of Journalism. Retrieved from `https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf`.*

Newman, N., Fletcher, R., Kalogeropoulos, A., and Nielsen, R. (2019). Reuters institute digital news report 2019. Technical report, Reuters Institute for the Study of Journalism.

Nguyen, J. (2018). Politics and the twitter revolution: A brief literature review and implications for future research. *Social Networking*, 7(4):243--251.

Nguyen, V.-A., Boyd-Graber, J., Resnik, P., and Miler, K. (2015). Tea party in the house: A hierarchical ideal point topic model and its application to republican legislators in the 112th congress. In *International Joint Conference on Natural Language Processing*.

Ni, K., Ramanathan, N., Chehade, M. N. H., Balzano, L., Nair, S., Zahedi, S., Kohler, E., Pottie, G., Hansen, M., and Srivastava, M. (2009). Sensor network data fault types. *ACM Transactions on Sensor Networks*, 5(3):1--29.

Nobre, G. P., Ferreira, C. H., and Almeida, J. M. (2022). A hierarchical network-oriented analysis of user participation in misinformation spread on whatsapp. *Information Processing & Management*, 59(1).

Nobre, G. P., Ferreira, C. H. G., and Almeida, J. M. (2020). Beyond groups: Uncovering dynamic communities on the whatsapp network of information dissemination. In *Social Informatics*. Springer International Publishing.

Oliveira, L., de Melo, P. V., Amaral, M., and Pinho, J. A. (2018). When politicians talk about politics: Identifying political tweets of brazilian congressmen. In *AAAI Conference on Web and Social Media*.

Oliveira, L. S. D., Vaz-de Melo, P. O., Amaral, M. S., and Pinho, J. A. G. (2020). Do politicians talk about politics? assessing online communication patterns of brazilian politicians. *ACM Transactions on Social Computing*, 3:1--28.

Olson, R. S. and Neal, Z. P. (2015). Navigating the massive world of reddit: Using backbone networks to map user interests in social media. *PeerJ Computer Science*, 1.

Orman, G. K., Türe, N., Balcisoy, S., and Boz, H. A. (2021). Finding proper time intervals for dynamic network extraction. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(3):033414.

Pacheco, D., Flammini, A., and Menczer, F. (2020). Unveiling coordinated groups behind white helmets disinformation. In *Companion Proceedings of the Web Conference 2020*, pages 611--616.

Pacheco, D., Hui, P.-M., Torres-Lugo, C., Truong, B. T., Flammini, A., and Menczer, F. (2021). Uncovering coordinated networks on social media: Methods and case studies. In *Proceedings of the International AAAI Conference on Web and Social Media*.

Papadimitriou, T., Gogas, P., and Gkatzoglou, F. (2020). The evolution of the cryptocurrencies market: A complex networks approach. *Journal of Computational and Applied Mathematics*, page 112831.

Paranjape, A., Benson, A. R., and Leskovec, J. (2017). Motifs in temporal networks. In *International Conference on Web Search and Data Mining*.

Paudel, P., Nguyen, T. T., Hatua, A., and Sung, A. H. (2019). How the tables have turned: studying the new wave of social bots on twitter using complex network analysis techniques. In *International Conference on Advances in Social Networks Analysis and Mining*.

Perkins, A. D. and Langston, M. A. (2009). Threshold selection in gene co-expression networks using spectral graph theory techniques. In *BMC bioinformatics*, volume 10.

Peron, T. D. and Rodrigues, F. A. (2011). Collective behavior in financial markets. *EPL (Europhysics Letters)*, 96.

Perra, N. and Rocha, L. E. (2019). Modelling opinion dynamics in the age of algorithmic personalisation. *Scientific reports*, 9(1).

Pierri, F., Artoni, A., and Ceri, S. (2020). Investigating italian disinformation spreading on twitter in the context of 2019 european elections. *PLOS ONE*, 15:1–23.

Pizzuti, C. and Socievole, A. (2017). An evolutionary motifs-based algorithm for community detection. In *International Conference on Information, Intelligence, Systems & Applications*, pages 1--6.

Pons, P. and Latapy, M. (2005). Computing communities in large networks using random walks. In *International symposium on computer and information sciences*. Springer.

Poole, K. T. and Rosenthal, H. (1985). A spatial model for legislative roll call analysis. *American Journal of Political Science*, 29(2):357--384.

Poole, K. T. and Rosenthal, H. (2000). *Congress: A Political-economic History of Roll Call Voting*, volume 58. Oxford University Press on Demand.

Poole, K. T. and Rosenthal, H. (2001). D-nominate after 10 years: A comparative update to congress: A political-economic history of roll-call voting. *Legislative Studies Quarterly*, pages 5--29.

Poole, K. T. and Rosenthal, H. L. (2011). *Ideology and congress*, volume 1. Transaction Publishers.

Porter, M. A., Mucha, P. J., Newman, M. E. J., and Warmbrand, C. M. (2005). A network analysis of committees in the u.s. house of representatives. *Proceedings of the National Academy of Sciences*, 102(20).

Putra, P., Thompson, T. B., Chaggar, P., and Goriely, A. (2021). Braiding braak and braak: Staging patterns and model selection in network neurodegeneration. *Network Neuroscience*, 5(4).

Qian, L., Bu, Z., Lu, M., Cao, J., and Wu, Z. (2015). Extracting backbones from weighted complex networks with incomplete information. In *Abstract and Applied Analysis*, volume 2015. Hindawi.

Radicchi, F., Ramasco, J. J., and Fortunato, S. (2011). Information filtering in complex weighted networks. *Physical Review E*, 83(4):046101.

Rahimi, A., Cohn, T., and Baldwin, T. (2015). Twitter user geolocation using a unified text and network prediction model. In *International Joint Conference on Natural Language Processing*.

Raisi, E. and Huang, B. (2017). Cyberbullying detection with weakly supervised machine learning. In *International Conference on Advances in Social Networks Analysis and Mining 2017*.

Rajadesingan, A., Resnick, P., and Budak, C. (2020). Quick, community-specific learning: How distinctive toxicity norms are maintained in political subreddits. In *Internacional Conference on Web and Social Media*, volume 14.

Ratkiewicz, J., Fortunato, S., Flammini, A., Menczer, F., and Vespignani, A. (2010). Characterizing and modeling the dynamics of online popularity. *Physical review letters*, 105(15).

Reece, A. G. and Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, 6(1):1--12.

Reis, J. C., Melo, P., Garimella, K., Almeida, J. M., Eckles, D., and Benevenuto, F. (2020). A dataset of fact-checked images shared on whatsapp during the brazilian and indian elections. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14.

Resende, G., Melo, P., CS Reis, J., Vasconcelos, M., Almeida, J. M., and Benevenuto, F. (2019a). Analyzing textual (mis)information shared in whatsapp groups. In *ACM Conference on Web Science*.

Resende, G., Melo, P., Sousa, H., Messias, J., Vasconcelos, M., Almeida, J., and Benevenuto, F. (2019b). (mis) information dissemination in whatsapp: Gathering, analyzing and countermeasures. In *The World Wide Web Conference*.

Ribeiro, B., Perra, N., and Baronchelli, A. (2013). Quantifying the effect of temporal resolution on time-varying networks. *Scientific reports*, 3(1).

Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., and Benevenuto, F. (2016). Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, 5(1):1--29.

Rice, S. A. (1925). The behavior of legislative groups: A method of measurement. *Political Science Quarterly*, 40(1):60--72.

Rohall, D. E., Milkie, M. A., Lucas, J., and Lucas, J. W. (2013). *Social psychology: Sociological perspectives*. Pearson Higher Ed.

Rosenthal, H. (2018). Introduction to the issue in honor of keith t. poole. *Public Choice*, 176(1):1--5.

Rossetti, G. and Cazabet, R. (2018a). Community discovery in dynamic networks: a survey. *ACM Computing Surveys*, 51:35.

Rossetti, G. and Cazabet, R. (2018b). Community discovery in dynamic networks: A survey. *ACM Computing Surveys*, 51.

Rossetti, G., Milli, L., and Cazabet, R. (2019). Cdlib: a python library to extract, compare and evaluate communities from complex networks. *Applied Network Science*, 4.

Rossi, R. A., Ahmed, N. K., and Koh, E. (2018). Higher-order network representation learning. In *The Web Conference*.

Rosvall, M., Delvenne, J.-C., Schaub, M. T., and Lambiotte, R. (2019). Different approaches to community detection. *Advances in network clustering and blockmodeling*, pages 105--119.

Rubinov, M. and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059--1069.

Saeedian, M., Jamali, T., Kamali, M., Bayani, H., Yasseri, T., and Jafari, G. (2019). Emergence of world-stock-market network. *Physica A: Statistical Mechanics and its Applications*, 526.

Saha, K., Chandrasekharan, E., and De Choudhury, M. (2019). Prevalence and psychological effects of hateful speech in online college communities. In *ACM Conference on Web Science*.

Sakamoto, T. and Takikawa, H. (2017). Cross-national measurement of polarization in political discourse: Analyzing floor debate in the us the japanese legislatures. In *International Conference on Big Data*.

Sartori, G. (2005). *Parties and Party Systems: A framework for analysis*, volume 1. ECPR press.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Lucas, R. E., Agrawal, M., Park, G. J., Lakshmikanth, S. K., Jha, S., Seligman, M. E., et al. (2013). Characterizing geographic variation in well-being using tweets. In *Seventh International AAAI Conference on Weblogs and Social Media*.

Serrano, M. Á., Boguná, M., and Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences*, 106:6483--6488.

Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, 5(1).

Shearer, E. (2018). Social media outpaces print newspapers in the us as a news source. *Pew research center*, 10.

Shiga, M., Takigawa, I., and Mamitsuka, H. (2007). A spectral clustering approach to optimally combining numericalvectors with a modular network. In *International Conference on Knowledge Discovery and Data Mining*.

Shuo, L. and Chai, B. (2016). Discussion of the community detection algorithm based on statistical inference. *Perspectives in Science*, 7:122--125.

Silva, M., de Oliveira, L. S., Andreou, A., de Melo, P. O. S. V., Goga, O., and Benevenuto, F. (2020). Facebook ads monitor: An independent auditing system for political ads on facebook.

Silva, T., De Melo, P. V., Almeida, J., Musolesi, M., and Loureiro, A. (2014a). You are what you eat (and drink): Identifying cultural boundaries by analyzing food and drink habits in foursquare. In *International AAAI Conference on Web and Social Media*.

Silva, T. H., de Melo, P. O. V., Almeida, J. M., Musolesi, M., and Loureiro, A. A. (2017). A large-scale study of cultural differences using urban data about eating and drinking preferences. *Information Systems*, 72:95--116.

Silva, T. H., Vaz de Melo, P. O. S., Almeida, J. M., Salles, J., and Loureiro, A. A. F. (2014b). Revealing the city that we cannot see. *ACM Transactions on Internet Technology*, 14(4).

Silveira, B., Murai, F., and da Silva, A. P. C. (2020). Predicting user emotional tone in mental disorder online communities. *arXiv preprint arXiv:2005.07473*.

Slater, P. B. (2009). A two-stage algorithm for extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences*, 106(26):66--66.

Smelser, N. J. (2011). *Theory of collective behavior*. Quid Pro Books.

Smith, N. and Graham, T. (2019). Mapping the anti-vaccination movement on facebook. *Information, Communication & Society*, 22(9).

Soro, F., Allegretta, M., Mellia, M., Drago, I., and Bertholdo, L. M. (2020). Sensing the noise: Uncovering communities in darknet traffic. In *Mediterranean Communication and Computer Networking Conference (MedComNet)*, pages 1–8.

Sporns, O. and Kötter, R. (2004). Motifs in brain networks. *PLoS Biol*, 2(11):e369.

Stosic, D., Stosic, D., Ludermir, T. B., and Stosic, T. (2018). Collective behavior of cryptocurrency price changes. *Physica A: Statistical Mechanics and its Applications*, 507.

Sun, L., Bai, Y., Du, B., Liu, C., Xiong, H., and Lv, W. (2020). Dual sequential network for temporal sets prediction. In *International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Tang, J. (2017). Computational models for social network analysis: A brief survey. In *International Conference on World Wide Web Companion*, page 921–925.

Tasente, T. (2020). The #Brexit on the Facebook pages of the European institutions. *Technium Social Sciences Journal*, 3(1):63–75.

Tausczik, Y. R. and Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24--54.

Thelwall, M. (2017). The heart and soul of the web? sentiment strength detection in the social web with sentistrength. In *Cyberemotions*, pages 119--134. Springer, London.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American society for information science and technology*, 61(12):2544--2558.

Tipping, M. E. and Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611--622.

Traag, V. A., Van Dooren, P., and Nesterov, Y. (2011). Narrow scope for resolution-limit-free community detection. *Physical Review E*, 84(1):016114.

Traag, V. A., Waltman, L., and Van Eck, N. J. (2019). From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):1--12.

Trevisan, M., Giordano, D., Drago, I., Munafò, M. M., and Mellia, M. (2020). Five years at the edge: Watching internet from the isp network. *IEEE/ACM Transactions on Networking*, 1:1–14.

Trevisan, M., Vassio, L., Drago, I., Mellia, M., Murai, F., Figueiredo, F., Couto da Silva, A. P., and Almeida, J. M. (2019). Towards understanding political interactions on instagram. In *ACM Conference on Hypertext and Social Media*.

Tsourakakis, C. E., Pachocki, J., and Mitzenmacher, M. (2017). Scalable motif-aware graph clustering. In *International Conference on World Wide Web*.

Tsur, O. and Lazer, D. (2017). On the interpretability of thresholded social networks. In *International Conference on Web and Social Media*.

Turner, R. H., Smelser, N. J., and Killian, L. M. (2020). Collective behaviour.

Uyheng, J. and Carley, K. M. (2021). Characterizing network dynamics of online hate communities around the covid-19 pandemic. *Applied Network Science*, 6(1):1--21.

Valdano, E., Ferreri, L., Poletto, C., and Colizza, V. (2015). Analytical computation of the epidemic threshold on temporal networks. *Physical Review X*, 5(2):021005.

Valverde-Rebaza, J. C., Roche, M., Poncelet, P., and de Andrade Lopes, A. (2018). The role of location and social strength for friendship prediction in location-based social networks. *Information Processing & Management*, 54(4):475--489.

Vergeer, M. (2015). Twitter and political campaigning. *Sociology compass*, 9(9):745--760.

Vinh, N. X., Epps, J., and Bailey, J. (2010). Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, 11:2837--2854.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146--1151.

Wang, T., Peng, J., Peng, Q., Wang, Y., and Chen, J. (2020). Fsm: Fast and scalable network motif discovery for exploring higher-order network organizations. *Methods*, 173.

Wang, Y. H. (1993). On the number of successes in independent trials. *Statistica Sinica*, pages 295--312.

Waseem, Z., Davidson, T., Warmsley, D., and Weber, I. (2017). Understanding abuse: A typology of abusive language detection subtasks. In *First Workshop on Abusive Language Online*.

Watson, A. (2020). Share of adults who trust news media most of the time in selected countries worldwide as of february 2020. `https://www.statista.com/statistics/718019/social-media-news-source/`. Accessed: 2020-08-03.

Waugh, A. S., Pei, L., Fowler, J. H., Mucha, P. J., and Porter, M. A. (2009). Party polarization in congress: A network science approach. *Forthcoming*, 1.

Weaver, I. S., Williams, H., Cioroianu, I., Jasney, L., Coan, T., and Banducci, S. (2019). Communities of online news exposure during the uk general election 2015. *Online Social Networks and Media*, 10.

Weber, D. and Neumann, F. (2020). Who's in the gang? revealing coordinating communities in social media. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*.

Weerasinghe, J., Flanigan, B., Stein, A., McCoy, D., and Greenstadt, R. (2020). The pod people: Understanding manipulation of social media popularity via reciprocity abuse. In *The Web Conference*.

Wehmuth, K., Fleury, É., and Ziviani, A. (2016). On multiaspect graphs. *Theoretical Computer Science*, 651:50--61.

Wehmuth, K., Fleury, É., and Ziviani, A. (2017). Multiaspect graphs: Algebraic representation and algorithms. *Algorithms*, 10(1):1.

Wei, W. and Carley, K. M. (2015). Measuring temporal patterns in dynamic social networks. *ACM Transactions on Knowledge Discovery from Data*, 10(1):9.

Wiese, J., Min, J.-K., Hong, J. I., and Zimmerman, J. (2015). "you never call, you never write": Call and sms logs do not always indicate tie strength. In *ACM Conference on Computer Supported Cooperative Work and Social Computing*.

Wu, Z., Braunstein, L. A., Havlin, S., and Stanley, H. E. (2006). Transport in weighted networks: partition into superhighways and roads. *Physical review letters*, 96(14):148702.

Xie, J., Kelley, S., and Szymanski, B. K. (2013). Overlapping community detection in networks: The state-of-the-art and comparative study. *Acm computing surveys*, 45(4).

Xie, Y., Li, C., Yu, B., Zhang, C., and Tang, Z. (2020). A survey on dynamic network embedding. *arXiv preprint arXiv:2006.08093*.

Xu, R., Wang, M., Xie, Y., and Yang, Y. (2018). Distinct collective inference learning in social networks. In *International Conference on Computer Supported Cooperative Work in Design*.

Xu, X., Yuruk, N., Feng, Z., and Schweiger, T. A. (2007). Scan: a structural clustering algorithm for networks. In *International Conference on Knowledge Discovery and Data Mining*.

Yan, X., Jeub, L. G., Flammini, A., Radicchi, F., and Fortunato, S. (2018). Weight thresholding on complex networks. *Physical Review E*, 98.

Yang, C., Liu, C., Zhang, X., Nepal, S., and Chen, J. (2015). A time efficient approach for detecting errors in big sensor data on cloud. *IEEE Transactions on Parallel and Distributed Systems*, 26(2):329--339.

Yang, D., Zhang, D., and Qu, B. (2016a). Participatory cultural mapping based on collective behavior data in location-based social networks. *ACM Transactions on Intelligent Systems and Technology*, 7:1--23.

Yang, X., Kim, S., and Sun, Y. (2019). How do influencers mention brands in social media? sponsorship prediction of instagram posts. In *International Conference on Advances in Social Networks Analysis and Mining*, New York, NY, USA. Association for Computing Machinery.

Yang, Z., Algesheimer, R., and Tessone, C. J. (2016b). A comparative analysis of community detection algorithms on artificial networks. *Scientific reports*, 6(1):1--18.

Yao, M., Chelmis, C., and Zois, D. S. (2019). Cyberbullying ends here: Towards robust detection of cyberbullying in social media. In *The World Wide Web Conference*.

Yao, Z., Sun, Y., Ding, W., Rao, N., and Xiong, H. (2018). Dynamic word embeddings for evolving semantic discovery. In *International Conference on Web Search and Data Mining*.

Yin, H., Benson, A. R., Leskovec, J., and Gleich, D. F. (2017). Local higher-order graph clustering. In *International Conference on Knowledge Discovery and Data Mining*.

Yitzhaki, S. (1979). Relative deprivation and the gini coefficient. *The quarterly journal of economics*, pages 321--324.

Yoon, S.-e., Song, H., Shin, K., and Yi, Y. (2020). How much and when do we need higher-order informationin hypergraphs? a case study on hyperedge prediction. In *The Web Conference*.

Yu, F., Liu, Q., Wu, S., Wang, L., and Tan, T. (2016). A dynamic recurrent model for next basket recommendation. In *International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery.

Yu, H., Hsieh, C., Si, S., and Dhillon, I. (2012). Scalable coordinate descent approaches to parallel matrix factorization for recommender systems. In *International Conference on Data Mining*.

Yu, L., Sun, L., Du, B., Liu, C., Xiong, H., and Lv, W. (2020). Predicting temporal sets with deep neural networks. In *International Conference on Knowledge Discovery & Data Mining*.

Yu, R., He, X., and Liu, Y. (2015). Glad: group anomaly detection in social media analysis. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 10(2).

Zampieri, M., Malmasi, S., Nakov, P., Rosenthal, S., Farra, N., and Kumar, R. (2019). Predicting the type and target of offensive posts in social media. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*.

Zarei, K., Farahbakhsh, R., and Crespi, N. (2019). Typification of impersonated accounts on instagram. In *International Performance Computing and Communications Conference*.

Zeno, G., Fond, T. L., and Neville, J. (2020). Dynamic network modeling from motif-activity. In *The Web Conference 2020*.

Zhan, M., Tu, R., and Yu, Q. (2018). Understanding readers: Conducting sentiment analysis of instagram captions. In *International Conference on Computer Science and Artificial Intelligence*, New York, NY, USA. Association for Computing Machinery.

Zhu, L., Guo, D., Yin, J., Ver Steeg, G., and Galstyan, A. (2016). Scalable temporal latent space inference for link prediction in dynamic social networks. *IEEE Transactions on Knowledge and Data Engineering*, 28(10).

Appendix A

Parameter Sensitivity Analysis

We report in this appendix additional results from a sensitivity analysis we performed to parameterize the backbone extraction methods. Table A.1 shows results of the impact of the p-value (input parameter) on the methods selected as candidates for the study of online discussions on Instagram (case study 1 reported in Section 6.4). Table A.2 shows corresponding results for the methods selected as candidates for the study of coordinated behavior on WhatsApp (case study 2, discussed in Section 6.5). For the latter, we note that, for some of the parameter values tested (e.g., p-values lower than 0.05), empty set of edges in the intersection of all backbones is empty. Such empty intersection renders our method comparisons unfeasible, from the contextual perspective, since we rely on training and test sets selected among edges common to all backbones to perform the regression analysis (see Section 6.3). Therefore, we disregard such cases, marked as '-' in the table. Also, as discussed in Section 6.5, both HSS and Threshold methods are parameterized by setting a threshold equivalent to retaining at most top-k% of the edges. Table A.2 presents results as we vary k.

Table A.1. Online discussions on Instagram: Impact of method parameters on topological and contextual metrics.

Strategy	% N	% E	# Comm.	Mod.	Parameter		R^2	NMRSE (Test)
TriBE	28.32	0.03	44	0.74	p-value	0.001	0.95	0.16
TriBE	52.83	0.48	19	0.69		0.010	0.89	0.22
TriBE	70.06	0.91	10	0.58		0.005	0.87	0.22
TriBE	99.86	9.43	6	0.43		0.100	0.80	0.23
GloSS	68.29	2.58	7	0.32	p-value	0.100	0.35	0.28
GloSS	65.45	0.73	6	0.39		0.050	0.65	0.36
GloSS	58.59	0.27	7	0.58		0.010	0.71	0.40
GloSS	51.72	0.11	9	0.73		0.001	0.66	0.91
NC	100.00	24.63	5	0.58	p-value	0.001	0.21	1.46
NC	100.00	30.63	5	0.57		0.010	0.23	0.88
NC	100.00	39.58	5	0.52		0.050	0.26	0.70
NC	100.00	46.62	5	0.48		0.100	0.27	0.57
MLF	98.79	16.16	7	0.51	p-value	0.100	0.17	0.60
MLF	94.78	10.27	19	0.49		0.050	0.28	0.72
MLF	74.61	3.95	8	0.55		0.010	0.47	0.79
MLF	49.78	1.16	12	0.62		0.001	0.49	1.28

Table A.2. Coordinated behavior on WhatsApp: Impact of method parameters on topological and contextual metrics.

Strategy	% N	% E	# Comm.	Mod.	Parameter		R^2	NRMSE (Test)
DF	22.96	5.74	14	0.49	p-value	0.100	0.41	0.50
DF	16.51	3.51	15	0.52		0.050	0.40	0.43
DF	9.48	1.21	14	0.55		0.010	-	-
DF	4.18	0.26	12	0.59		0.001	-	-
Polya	19.36	6.20	12	0.46	p-value	0.100	0.30	0.52
Polya	17.15	4.76	15	0.48		0.050	0.30	0.45
Polya	11.82	2.66	12	0.48		0.010	-	-
Polya	7.62	1.21	11	0.50		0.001	-	-
Threshold	19.55	8.13	12	0.42	Threshold / k	10%	0.24	0.57
Threshold	11.56	4.29	10	0.45		5%	0.22	0.43
Threshold	4.06	0.94	6	0.41		1%	-	-
Threshold	1.54	0.10	4	0.32		0.1%	-	-
RECAST	7.31	0.85	7	0.37	p-value	0.100	0.22	0.49
RECAST	7.31	0.85	8	0.37		0.050	0.23	0.56
RECAST	2.80	0.14	8	0.36		0.010	-	-
RECAST	2.80	0.14	8	0.38		0.001	-	-
HSS	56.90	9.93	135	0.25	Threshold / k	10%	0.24	1.36
HSS	43.42	4.86	148	0.30		5%	0.25	0.86
HSS	23.15	0.88	141	0.56		1%	-	-
HSS	6.26	0.09	70	0.98		0.1%	-	-