# Modeling and Analyzing Collective Behavior Captured by Many-to-Many Networks

Carlos Henrique G. Ferreira
chgferreira@dcc.ufmg.br
Department of Computer Science - Universidade Federal de Minas Gerais, Brazil

Marco Mellia
mellia@polito.it
Department of Control and Computer Engineering - Politecnico di Torino, Italy

Jussara M. Almeida
jussara@dcc.ufmg.br
Department of Computer Science - Universidade Federal de Minas Gerais, Brazil

## Abstract

Several significant studies in the existing literature have relied on network models to gain insights into various collective behavior phenomena. Nevertheless, a facet that has been critically overlooked is the presence of numerous irrelevant edges that may obscure a more meaningful underlying topology, representing the targeted phenomenon. In fact, the literature provides ample evidence that overlooking these noisy edges may result in inaccurate and misleading interpretations. Nonetheless, employing these solutions presents various challenges, prominently the absence of foundational formalization regarding the appropriate application and expected outcomes. In this context, our focus centers on extracting salient edges, exploring backbone extraction methods, for the purpose of modeling and analyzing collective behavior. To address the gaps in the current literature regarding the use of such methods for modeling collective behavior, we undertake a comprehensive series of efforts. These include formalizing, analyzing, discussing, applying, and validating existing methods, many of which are drawn from parallel fields of study to computer science, and finally introducing novel methods to advance the state-of-the-art. We also demonstrate the effectiveness of these methods as fundamental tools for uncovering relevant patterns, applying them across diverse phenomena each with distinct requirements. Our contributions are multifaceted, including innovative methods, case studies yielding specific insights, and a comprehensive methodology for the selection, application, and validation of these methods. Moreover, our outcomes wielded a substantial impact on both the scientific community and society. They not only unveiled numerous opportunities for fellow researchers but also catalyzed the initiation of new and impactful research.

***Keywords:*** Network science, collective behavior, network backbone extraction

## 1 Introduction

Network science has emerged as a valuable field for modeling and studying the *collective behavior* of individuals in complex systems [31, 36]. It provides a range of theoretical tools to describe and analyze phenomena that are of great interest to our society. For instance, network science helps us understand how users spread ideas and information through content sharing on social media platforms [11, 23, 24, 26, 32], how voting behavior of House of Representatives members forms ideological groups that better represent a country's political scenario beyond traditional political parties [13], co-authorship patterns in publications [5], people's mobility between places using social web data [21, 38], among others.

However, many complex systems are structured by interactions that occur simultaneously among multiple individuals (or even components), which we refer to as *co-interactions* or *many-to-many* interactions, such as multiple users sharing the same piece of information in an information dissemination network, or a set of co-authors in a co-authorship network. Recent studies have highlighted the impact of these co-interactions on the topological structure of the network, especially when projected into undirected and weighted networks, as they exhibit rich and diverse patterns at different levels, including sequentiality, periodicity, and sporadicity [3, 20]. As a result, such networks tend to contain a large number of edges representing random, sporadic, and spurious interactions that are only weakly related, if at all, to the phenomenon under study. The large number of these *noisy* edges adds even more complexity to the analysis of network models, including the study of collective behavior, and requires the identification of co-interactions that are truly relevant to the target phenomenon. Moreover, most metrics and algorithms used for network analysis (e.g., community detection and recommendation systems) assume that the network structure derived from the interactions accurately represents the phenomenon under study [4, 29, 36]. Consequently, these algorithms take into account all available edges, including noisy and sporadic ones, potentially leading to misinterpretations and misleading conclusions. Hence, the presence of noisy edges highlights the importance of identifying the significant edges that are essential to understanding the phenomenon under study. Surprisingly, this critical step has been largely overlooked in the network models used for studying such phenomena.

The selection and extraction of salient edges, also known as *backbone*, from complex networks is tackled by extraction methods. These methods aim to filter out noisy edges and provide a reduced version of the network that captures the most important edges for the target phenomenon [18]. However,

In: V Concurso de Teses e Dissertações (CTD 2023), Ribeirão Preto, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2023.

ISSN 2596-1683

despite the importance of these methods, they are still underused. One of the challenges in using backbone extraction methods is to select the most appropriate method for a particular study, considering a predefined definition of edge salience. Each method relies on assumptions, revealing a unique underlying structure, requiring careful examination of properties and statistical models [6, 9, 10, 18, 25, 28, 35, 37]. Although this task is challenging, it provides an opportunity to explore different methods and improve results for many interesting phenomena. It also provides the opportunity to propose new approaches for backbone extraction, especially in the study of novel phenomena. Another major challenge is the evaluation and validation of backbone extraction methods. Since the actual structure of the network is by definition unknown, it is difficult to assess the quality of the extracted backbone using conventional assessment methods [30]. Therefore, unsupervised assessment strategies should be considered to overcome this limitation. In summary, backbone extraction methods provide a way to identify salient edges in complex networks. However, their application, selection, proposal, and evaluation have gaps in the literature, which we aimed to fill with our dissertation. The full dissertation is accessible in [12].

### 1.1 Goals

In this context, our dissertation was 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?*

The challenges associated with our guiding question have led to the definition of the following research goals:

**RG1: Uncovering topological and contextual properties of communities in many-to-many networks**: Our first goal was to identify communities that are representative of collective behavior in a target system and to characterize their structural and contextual properties that are fundamentally related to the phenomenon under study. As mentioned earlier, one of the main challenges is to identify the salient edges that form the backbone of the network for a given target system. Therefore, it is important to explore different methods for extracting the backbone, both existing and new, that fit the characteristics of the system and the phenomenon under study.

**RG2: Modeling the temporal dynamics of communities in many-to-many networks**:We were interested in analyzing the temporal dynamics of the identified communities by examining how the structural and contextual properties of the backbone evolve over time. From the structural perspective, we were interested in understanding and quantifying the dynamics at the individual member and community levels. With the contextual perspective that we tackled in our RG1, we can also examine the contextual properties of the phenomenon behind these communities (e.g., the topics of discussion, the patterns of co-interactions) 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:** We found that some methods of extracting the backbone extraction 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 there is often no ground truth. This largely depends on a comprehensive knowledge of the assumptions of both methods and phenomena. Our ultimate research goal, therefore, was to survey the key properties of such methods and potential phenomena to guide the selection, use, and evaluation of methods for the study of a particular phenomenon.

## 2 Results and Contributions

**RG1:** To model collective behavior in many-to-many networks, we relied on some phenomena focused on groups of users representative of communities. Through a range of case studies, we quantified the presence and impact of noise in these networks, uncovering structural and contextual (related to phenomena) patterns not previously observed in the literature. Given the need to deal with noise in the phenomena, we applied and proposed new methods for extracting backbones. First, we combined two threshold-based and neighborhood-based approaches and showed the importance of contextual information to identify salient edges in modeling and analyzing political ideologies in a network modeled from roll call votes [13, 17]. Our study not only demonstrated the importance of using contextual information derived from the phenomenon to assist the backbone extraction process, but also, unlike previous ideological analyses in the political context, compared the characteristics of collective behavior in fragmented and non-fragmented party systems over a long period of time.

In studying another phenomenon, notably, online discussions in social media applications, we found that social media applications have characteristics that challenge the modeling and analysis of collective behavior. Most notably, these include the heavy tail nature of content popularity and user activity, leading to many edges that are not necessarily relevant to the study. We then proposed to study this phenomenon on Instagram by modeling a tripartite network. Nevertheless, we did not find any backbone extraction techniques in the literature that *explicitly* exploit this type of structure to identify salient edges. Thus, we proposed TriBE, a probabilistic backbone extraction method that takes into account this type of structure and can be explored for any domain modeled by a tripartite network [14, 16]. Some important examples include co-authorship networks, co-developers networks in the context of software engineering, among others.

It is also important to highlight that this research goal was explored in other research projects we collaborated. Among the most important are: i) an analysis in the first case study conducted as part of a research project by an undergraduate student [27]; ii) a project conducted by a Master's student on the spread of misinformation on WhatsApp [33, 34], whose project was awarded by the *2022 Google Latin America Research Awards*; iii) an analysis of Twitter discussions about allegations of fraud in the United States elections conducted in a bachelor's thesis supervised by the dissertation author [7]; iv) and finally, a study of coordinated actions on Twitter around the election and attack on the United States Capitol, conducted in a bachelor's thesis also supervised by the dissertation author [22]. Notably, the latter study also proposed a new method for backbone extraction that combines two other strategies from the literature examined in our dissertation, therefore, advancing the state of the art in backbone extraction. These results validate our concern that it is important to consider network noise when modeling and analyzing collective behavior, which is mostly neglected in the literature. Moreover, all of these studies explored different network structures in the context of complex networks that are common in several other domains, demonstrating their applicability and importance.

**RG2:** We applied several metrics to the temporal analysis on the phenomena studied to capture individual-level dynamics and noted some important limitations with the method most commonly used in the literature. We then explored the use of machine learning methods, specifically network embedding techniques, to obtain latent representations of networks. However, we encountered an issue known as the *alignment problem* when applying static embedding techniques to successive networks representing different time windows. This problem arises because the resulting embeddings are not mapped to the same latent space, making it difficult to track node-level dynamics over time.

To overcome this challenge, we then proposed a technique for mapping network sequences into a time-aligned latent space. By combining state-of-the-art approaches like node2vec and dynamicWord2vec [19, 39], we jointly learned temporal node embeddings for successive networks, enabling consistent tracking of individual nodes over time. This advancement in modeling dynamics in temporal networks based solely on structural information is a significant contribution. Our method offers advantages over temporal dynamic analysis. We demonstrated its effectiveness and generalizability through two case studies, including the analysis of political and mobility networks. Thus, our approach builds upon previous studies that have demonstrated the efficiency of temporal embedding techniques across various domains.

**RG3:** In our last research goal, we addressed the challenge of selecting and evaluating backbone extraction methods in the absence of ground truth in a studied phenomenon. We reviewed ten state-of-the-art methods from reputable publications in computer science fields, such as [10, 18, 28, 35, 37], including prestigious venues like *Proceedings of the National Academy of Sciences* and *Nature Communications*. We provided a detailed description of their assumptions, advantages, and disadvantages, considering their statistical and structural properties for practical applicability. Furthermore, we identified network properties utilized by these models to capture collective behavior across different domains.

Our methodology explicitly considers both method and phenomenon properties for effective selection. It incorporates metrics that capture structural and contextual aspects, often overlooked in the literature, to evaluate the resulting backbone's emergent structure and its relevance to the studied phenomenon [15]. Our investigation surpasses the existing literature in terms of comprehensiveness and thoroughness when compared to similar attempts [8, 28], as it encompasses a larger number of methods and evaluation metrics. The results highlight the substantial variation in backbones obtained using different methods, underscoring the importance of method selection for gaining meaningful insights into the phenomenon under investigation. Finally, we have assembled all applied and developed backbone extraction methods, datasets, and details in a single data repository to ensure the reproducibility of our studies and enable their use in future research. The repository is accessible at: https://github.com/chgferreira/backbone_extraction.

### 2.1 Scientific and Social Impacts

The results of our dissertation have been disseminated through prestigious journals and conferences [13–17], including the *International Conference on Social Informatics*, *The Journal of Web Science*, *ACM Conference on Web Science*, *Elsevier Online Social Networks and Media Journal*, and *PlosOne*.

Furthermore, our involvement in parallel endeavors closely aligned with our dissertation has yielded significant contributions and consequential publications [2, 7, 22, 27, 33, 34], such as the *Brazilian Workshop on Social Network Analysis and Mining*, *International Conference on Social Informatics*, *Elsevier Information Processing & Management*, *ACM SIGMETRICS Performance Evaluation Review*, *International Conference on Advances in Social Network Analysis and Mining*, and once again, the *International Conference on Social Informatics*. These works demonstrate that our dissertation has not only contributed to research but has also inspired new avenues for investigation. Most importantly, it has catalyzed numerous projects among undergraduate and graduate students, fostering the author's personal and professional development during and after the PhD journey.

Moreover, our work extends beyond the academic realm to have a tangible social impact. This is notably evident through our collaborative research initiative with the *Ministério Público do Estado de Minas Gerais* [1]. By applying our dissertation findings, we have effectively identified electoral irregularities on social media platforms, transcending traditional academic boundaries.

## 3 Final Considerations

For many years, network-based models have been widely employed to study various phenomena in numerous domains by representing the interactions between agents or components through an undirected and weighted graph. However, despite their wide applicability, network models also bring some challenges that have received little attention, even though they may provide more accurate results. Our dissertation sheds light on this aspect by focusing on the modeling and analysis of collective behavior in many-to-many networks, drawing attention to the impact of noise on the network and its effects, which are often overlooked in network-based modeling efforts. Our results demonstrate the importance of considering noise in network-based models to improve their effectiveness. We hope that extracting the network backbone will be a crucial step for many-to-many network modeling and analysis.

## References


[1] M Araujo, C. Ferreira, J. Reis, A. Silva, and J. Almeida. 2023. Identificação e Caracterização de Campanhas de Propagandas Eleitorais Antecipadas Brasileiras no Twitter. In *Brazilian Workshop on Social Network Analysis and Mining*.

[2] M. F. C. Barros, C. H. G Ferreira, L. A .P. Junior, M. Mellia, J.M. Almeida, and B.P. Santos. 2022. Understanding mobility in networks: A node embedding approach. *ACM SIGMETRICS Performance Evaluation Review* (2022).

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

[4] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008 (2008).

[5] M. A. Brandão, P. O. S. Vaz de Melo, and M. M. Moro. 2017. Tie Strength Dynamics over Temporal Co-authorship Social Networks. In *International Conference on Web Intelligence*.

[6] Michele Coscia and Frank MH Neffke. 2017. Network backboning with noisy data. In *International Conference on Data Engineering*.

[7] Jose da Rosa Jr, Renan Linhares, Carlos Henrique Gomes Ferreira, Gabriel P Nobre, Fabricio Murai, and Jussara M Almeida. 2022. Uncovering Discussion Groups on Claims of Election Fraud from Twitter. In *Social Informatics: 13th International Conference, SocInfo*.

[8] Liang Dai, Ben Derudder, and Xingjian Liu. 2018. Transport network backbone extraction: A comparison of techniques. *Journal of Transport Geography* 69 (2018), 271–281.

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

[10] N. Dianati. 2016. Unwinding the hairball graph: Pruning algorithms for weighted complex networks. *Physical Review E* (2016).

[11] M. F. Feitosa, G. D. Gonçalves, C. H. G. Ferreira, and J. M. Almeida. 2022. Sentiment Analysis on Twitter Repercussion of Police Operations. In *Simpósio Brasileiro de Sistemas Multimídia e Web*.

[12] Carlos Henrique Gomes Ferreira. 2022. *Modeling and analyzing collective behavior captured by many-to-many networks*. PhD thesis. Universidade Federal de Minas Gerais. Available at http://hdl.handle.net/1843/43014.

[13] C. H. G. Ferreira, B de Souza M., and J. M. Almeira. 2018. Analyzing Dynamic Ideological Communities in Congressional Voting Networks. In *International Conference on Social Informatics*.

[14] C. H. G. Ferreira, F. Murai, A. P. C. da Silva, J. M. de Almeida, M. Trevisan, L. Vassio, I. Drago, and M. Mellia. 2020. Unveiling Community Dynamics on Instagram Political Network. In *ACM Conference on Web Science*.

[15] C. H. G. Ferreira, F. Murai, A. P.C. Silva, M. Trevisan, L. Vassio, I. Drago, M. Mellia, and J. M Almeida. 2022. On network backbone extraction for modeling online collective behavior. *Plos one* 17, 9 (2022), e0274218.

[16] C. H. G. Ferreira, F. Murai, Ana P. C. Silva, J. M. Almeida, M. Trevisan, Luca V., M. Mellia, and I. Drago. 2021. On the dynamics of political discussions on Instagram: A network perspective. *Online Social Networks and Media* 25 (2021), 100155.

[17] C. H. G. Ferreira, F. Murai, B Souza M., and J. M. de Almeida. 2019. Modeling Dynamic Ideological Behavior in Political Networks. *The Journal of Web Science* 7 (2019).

[18] D. Grady, C. Thiemann, and D. Brockmann. 2012. Robust classification of salient links in complex networks. *Nature communications* (2012).

[19] A. Grover and J. Leskovec. 2016. node2vec: Scalable feature learning for networks. In *International Conference on Knowledge Discovery and Data Mining*.

[20] R. Guidotti, G. Rossetti, L. Pappalardo, F. Giannotti, and D. Pedreschi. 2019. Personalized Market Basket Prediction with Temporal Annotated Recurring Sequences. *IEEE Transactions on Knowledge and Data Engineering* 11 (2019).

[21] Roger N. Kamoi, L. A. Pereira Jr, F. A. Verri, C. A. Marcondes, C. H. G. Ferreira, R. I. Meneguette, and A. M. Da Cunha. 2021. Platoon grouping network offloading mechanism for vanets. *IEEE Access* (2021).

[22] R. S. Linhares, Jose da Rosa Jr, C. H. G. Ferreira, G. P. Nobre, F. Murai, and J. M. Almeida. 2022. Uncovering Coordinated Communities on Twitter During the 2020 US Election. In *International Conference on Advances in Social Networks Analysis and Mining*.

[23] L. Malagoli, J. Stancioli, Carlos HG Ferreira, M. Vasconcelos, A. P. C. da Silva, and J. Almeida. 2021. Caracterizaçao do debate no twitter sobre a vacinaçao contra a covid-19 no brasil. In *Brazilian Workshop on Social Network Analysis and Mining*.

[24] L. G. Malagoli, J. Stancioli, C. H. G. Ferreira, M. Vasconcelos, A. P. C. da Silva, and J. M. Almeida. 2021. A look into covid-19 vaccination debate on twitter. In *ACM Web Science Conference 2021*.

[25] R. Marcaccioli and G. Liivan. 2019. A Pólya urn approach to information filtering in complex networks. *Nature communications* (2019).

[26] S. Martin-Gutierrez, Juan C Losada, and R. M Benito. [n. d.]. Impact of individual actions on the collective response of social systems. *Scientific reports* ([n. d.]).

[27] B. Matos, C. H. G. Ferreira, and J. M. Almeida. 2018. 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*.

[28] Z. P. Neal, R. Domagalski, and B. Sagan. 2021. Comparing alternatives to the fixed degree sequence model for extracting the backbone of bipartite projections. *Scientific reports* (2021).

[29] M.E.J. Newman. 2006. Finding community structure in networks using the eigenvectors of matrices. *Physical review E* (2006).

[30] MEJ Newman. 2018. Network structure from rich but noisy data. *Nature Physics* 14 (2018), 542–546.

[31] Mark EJ Newman. 2003. The structure and function of complex networks. *SIAM review* 45, 2 (2003), 167–256.

[32] G. P. Nobre, J. M. Almeida, and C.H.G. Ferreira. 2019. Caracterização de bots no Twitter durante as Eleições Presidenciais no Brasil em 2018. In *Brazilian Workshop on Social Network Analysis and Mining*.

[33] G. P. Nobre, C. H. G. Ferreira, and J. M. Almeida. 2020. Beyond Groups: Uncovering Dynamic Communities on the WhatsApp Network of Information Dissemination. In *International Conference on Social Informatics*.

[34] G. P. Nobre, C. H. G. Ferreira, and J. M. Almeida. 2022. A hierarchical network-oriented analysis of user participation in misinformation spread on WhatsApp. *Information Processing & Management* (2022).

[35] F. Radicchi, J. J Ramasco, and S. Fortunato. 2011. Information filtering in complex weighted networks. *Physical Review E* (2011).

[36] G. Rossetti and R. Cazabet. 2018. Community discovery in dynamic networks: A survey. *Comput. Surveys* 51 (2018), 35.

[37] M. Serrano, Marián Boguná, and A. Vespignani. 2009. Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences* (2009).

[38] T. H. Silva, P. O. S. V. de Melo, J. M. Almeida, J. Salles, and A. F. Loureiro. 2014. Revealing the City That We Cannot See. *ACM Transactions on Internet Technology* (2014).

[39] Z. Yao, Y. Sun, W. Ding, N. Rao, and H. Xiong. 2018. Dynamic Word Embeddings for Evolving Semantic Discovery. In *International Conference on Web Search and Data Mining*.