PhD Project
Null models of complex networks
to include temporal and higher order constraints

supervision: Mehdi Naima and Lionel Tabourier

contact: `mehdi.naima@lip6.fr`, `lionel.tabourier@lip6.fr`

1 Introduction

In complex network analysis a *Null Model* (or *neutral model*) is a random graph that matches one specific graph in some of its features, or more generally satisfies a collection of constraints, but which is otherwise an unbiased random structure. Seminal works on this topic include null models preserving the number of nodes and edges, known as Erdős-Rényi graphs [BB98], preserving the degree distribution [Bol80, BD11, FLNU18], or more complex graph features [VKBK21, TK24].

The development of null models have proven essential in many domains since they serve as baselines for assessing goodness of fit, facilitate data analysis and allow the construction of families of graphs of interest [MIK$^+$04, BLM$^+$06]. They have been used in seminal works for community detection in network science [NG04], brain networks in neurosciences [VM22], in food-web relationships and protein-protein interactions [HRP08] in biology, and in social network analysis [Was94, JGN01].

This project aims to advance the field of complex network analysis by developing and generalizing random graph models taking into consideration not only standard graphs but also extensions to temporal graphs and higher-order graphs.

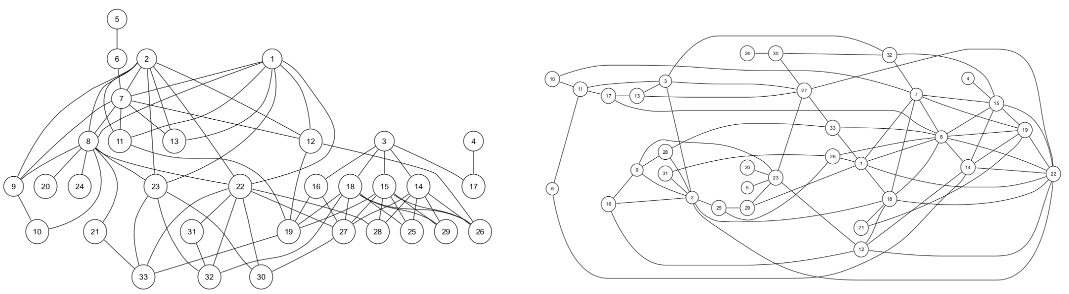

Figure 1: Real structure of the food web for the Chesapeake bay ecosystem (left) vs. a realization of a null model preserving its degree sequence (right). Extracted from [BD11].

2 Objectives

In this PhD project, there are two main methodological aspects to investigate. The first one concerns **mining patterns in the data that allow to characterise their structure and dynamics**. Indeed, while there is a rich toolbox to identify relevant structural features on graphs, there is still a need to develop their equivalent to characterise synthetically the properties of interaction data through time and higher order interactions (interactions involving more than two entities). Our team acquired an expertise in the design and development of tools to analyse temporal data from real-world interaction networks [LVM18, BTM23, Nai23, SNS25a]. We plan to use this knowledge for the identification and measurement of the features which are specific to the data collected.

The second aspect concerns generation processes. Some simple graph models were proposed around 25 years ago to help describe some coarse-grain properties observed in real world, such as the heavy-tailed degree distribution or high clustering, but they proved limited when trying to account for more elaborate properties. Consequently, useful techniques were proposed to generate graph models with stronger properties [RPKL07, VKBK21, TK24]. Yet, improved access to data uncovered the important role of the temporality of the interactions, as well as the number of agents that they involve. There is now a **need for realistic models that also account for the dynamics of the interactions as well as their order**.

3 Theory and methods

In this PhD project, we present the theoretical contributions into three parts corresponding to the three types of objects that we aim to model, namely static, temporal and higher-order graphs.

3.1 Static graphs

The goal of any null model is to reproduce some essential aspects of the complex structure observed using a few simple characteristics of the graph. In this way, we identify the fundamental parts of its structure. However, generating null models presents intertwined technical and interpretative challenges. The main obstacle is that it is difficult to generate unbiased neutral models—that is, models that are truly random concerning all aspects except those specified by the model. Each proposed model often requires a complex adaptation of existing methods. As a result, the literature contains a significant number of studies attempting to establish the randomness of specific complex models [KTV99, RJ00, TRC11]. In practice, it is common to lack an appropriate method when proposing a neutral model tailored to a particular family of graphs.

To address this challenge, we prioritize two approaches. The first one is the use of Exponential Random Graph Models (ERGM) [S^+02], which have become a standard for generating neutral models, in particular in the context of social network analysis. Although they do not strictly guarantee an unbiased element within the considered set, the versatility of these models —both in terms of selecting properties for generation and the ability to interpret ERGM regression parameters— has contributed to their widespread popularity over the past thirty years [LKR13]. The second approach is the use of a recently proposed method known as Probabilistic K-Swap (PKS), which enables the generation of a wide range of null models while ensuring the unbiased nature of the process [TK24].

While these methods share the same objective, they are based on different principles and have distinct advantages and drawbacks. Indeed, whereas PKS is not subject to the overfitting and degeneracy issues that can limit the use of ERGM, it is not a regression-based method. Its interpretation is guided solely by the model's ingredients, making PKS less flexible in practice. We aim to determine which of these methods is best suited for each practical case study.

3.2 Temporal graphs

Temporal graphs extend classical graphs to the case of interactions that happen at specific moments in time. Examples of networks appropriately represented by temporal graphs include communication networks, social interactions or financial transactions.

To deepen our understanding of the underlying dynamical systems, it is crucial to analyze temporal networks in ways that respect their core temporal structure. In particular, randomizing these graphs while maintaining time-respecting walks is essential to preserve the integrity of temporal causality and information flow, which are fundamental to the system dynamics. As highlighted in the literature [HS12, KKK00, LRS19], time-respecting walks capture the ordered sequence of interactions that underpin real-world processes, i.e., a potential cause must temporally precede its result. Failing to preserve this causality structure would distort the temporal coherence, undermining the study of phenomena such as diffusion, communication, or influence within evolving systems.

There exists different models of temporal graph randomization, see [GGK$^+$22, HS12] for a review. However, these models usually randomize by fixing either a topological pattern (graph pattern) or the temporal pattern such that fixing the number of interactions per time step, while spatio-temporal patterns play a fundamental role and we want to develop model that fix them in this PhD. There is some recent work [SNS25b] in that direction, but the spatio-temporal patterns are fixed approximately and not exactly. Extending the PKS method [TK24] to the context of temporal graphs is an interesting avenue.

3.3 Higher-order graphs

Unlike classical graph models, which simply consider the relations between pairs of nodes via edges, higher-order graphs allow for interactions between groups of nodes, i.e., groups of entities can interact simultaneously [BGHS23, Bia21]. There are different models of higher-order graphs: Hypergraphs generalize from edges of two nodes to hyperedges, each consisting of an arbitrary number of nodes. Indeed, it has been found that in many real-world cases one needs to consider interactions that are more than pairwise dyadic interactions, such as various kinds of social networks (collaboration, affiliation, discussion) but also biological networks (protein interactions, gene-protein interactions. Higher-order graphs are increasingly popular to model both the structure and the dynamics, using additional structure for extensions of Graph Neural Networks, and in Graph Signal Processing (GSP). In Graph Signal Processing, SCs (Simplicial Complexes) and CCs (Cell Complexes) are particularly popular for processing (directed) flow signals on the edges of a graph [BS20, SBT21] due to their ability to capture direction on edges and relationships between edges and higher-order structure.

The growing field of higher-order graph analysis requires random models for analogous reasons as these models are required in the analysis of classical (pairwise) graphs. For instance, graph randomization corresponds to a controlled modification of an observed (real-world) higher-order network that outputs a new network with similar properties as the original. This is particularly useful for data anonymization, where we aim to alter a higher-order graph such that the privacy of the individual nodes is protected, while maintaining the overall composition of the network such that useful information can be extracted. Graph randomization can also be seen as a tool to generate new graphs that can be used for data augmentation in learning processes.

There are some works in this direction have been done on SCs and CCs in the literature [CB16, YPVP17, HS24] but there are still many gaps on the controlled generation of these models, that is we want to fix some high order structure and randomise the graph otherwise.

4 Applications

The applications that we develop in the context of this PhD project focus on online social network analysis, which is the main application field that our research team has developed during the last 15 years. Several ongoing projects would benefit of improved generation models.

In the context of the CNRS MITI project MoNReSo (2025-2026), we are especially interested in the description and classification of social ego-networks across several case studies corresponding to different data collection methods. These networks are centered around an individual (*ego*) and the set of people with whom they maintain relationships (*alters*), as well as the relationships between these alters. They provide a rich description of an individual's social environment and, as such, offer a precise ego-centered depiction of the concept of social ties [FG21]. In particular, Christophe Prieur (LISIS, UGE), who is a sociologist participating to the MITI project, has gathered around 10,000 Facebook ego-networks enriched with survey information. We will explore several aspects: first, is there a reduced set of structural characteristics that account for the overall structure of these networks, or are they, on the contrary, irreducible in the sense that no such characteristics can be found? And in the latter case, do neutral models allow for the establishment of new typologies of personal networks depending on the contexts considered? Christophe Prieur would help provide and analyse data from a sociological perspective during the PhD.

As part of the PostGenAI@Paris project supported by Sorbonne University, our team proposes to contribute to the CAP dedicated to augmented deliberation by investigating the structure and dynamics of online collaborative projects. Our aim here is to to evaluate if there are structural and dynamical elements of a collaboration that mark a productive or unproductive deliberation process. In that purpose being able to classify these dynamic collaboration networks in relation to standard null models is a favoured research path. Data such as github programming projects makes an interesting playground to investigate such methods.

Finally, collaboration networks as well as other co-occurrence networks can be viewed as high-order networks where authors of a papers form a hyperlink or a simplex. Therefore, having more refined neutral models would allow a better understanding of their inherent structure to characterise them.

5 Profile required and environment

The PhD would take place at Sorbonne University (Jussieu Campus, Paris 5), under the supervision of Mehdi Naima and Lionel Tabourier. Mehdi Naima is an associate professor at LIP6 (Sorbonne University/CNRS). His research focuses on algorithm analysis as well as temporal graph algorithms (more information at `https://busyweaver.github.io/`). Lionel Tabourier is a Professor at LIP6. During his previous works, he contributed to realistic graph generation methods, as well as the analysis of information spreading on social platforms (more information at `https://lioneltabourier.github.io/`).

We are looking for students whose primary interests are algorithmics and programming for data mining. A taste for interdisciplinary questions is also important.

References

[BB98] Béla Bollobás and Béla Bollobás. *Random graphs.* Springer, 1998.

[BD11] Joseph Blitzstein and Persi Diaconis. A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet mathematics*, 6(4):489–522, 2011.

[BGHS23] Christian Bick, Elizabeth Gross, Heather A Harrington, and Michael T Schaub. What are higher-order networks? *SIAM review*, 65(3):686–731, 2023.

[Bia21] Ginestra Bianconi. *Higher-order networks.* Cambridge University Press, 2021.

[BLM+06] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4-5):175–308, 2006.

[Bol80] Béla Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1(4):311–316, 1980.

[BS20] Sergio Barbarossa and Stefania Sardellitti. Topological signal processing over simplicial complexes. *IEEE Transactions on Signal Processing*, 68:2992–3007, 2020.

[BTM23] Alexis Baudin, Lionel Tabourier, and Clémence Magnien. Lscpm: Communities in massive real-world link streams by clique percolation method. In *30th International Symposium on Temporal Representation and Reasoning*, 2023.

[CB16] Owen T Courtney and Ginestra Bianconi. Generalized network structures: The configuration model and the canonical ensemble of simplicial complexes. *Physical Review E*, 93(6):062311, 2016.

[FG21] Guillaume Favre and Michel Grossetti. Les réseaux personnels en france ont-ils changé? une comparaison entre 2001 et 2017. *Revue française de sociologie*, (2):167–208, 2021.

[FLNU18] Bailey K Fosdick, Daniel B Larremore, Joel Nishimura, and Johan Ugander. Configuring random graph models with fixed degree sequences. *Siam Review*, 60(2):315–355, 2018.

[GGK+22] Laetitia Gauvin, Mathieu Génois, Márton Karsai, Mikko Kivela, Taro Takaguchi, Eugenio Valdano, and Christian L Vestergaard. Randomized reference models for temporal networks. *Siam Review*, 64(4):763–830, 2022.

[HRP08] Desmond J Higham, Marija Rašajski, and Nataša Pržulj. Fitting a geometric graph to a protein–protein interaction network. *Bioinformatics*, 24(8):1093–1099, 2008.

[HS12] Petter Holme and Jari Saramäki. Temporal networks. *Physics reports*, 519(3):97–125, 2012.

[HS24] Josef Hoppe and Michael T Schaub. Random abstract cell complexes. *arXiv preprint arXiv:2406.01999*, 2024.

[JGN01] Emily M Jin, Michelle Girvan, and Mark EJ Newman. Structure of growing social networks. *Physical review E*, 64(4):046132, 2001.

[KKK00] David Kempe, Jon Kleinberg, and Amit Kumar. Connectivity and inference problems for temporal networks. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 504–513, 2000.

[KTV99] Ravi Kannan, Prasad Tetali, and Santosh Vempala. Simple markov-chain algorithms for generating bipartite graphs and tournaments. *Random Structures & Algorithms*, 14(4):293–308, 1999.

[LKR13] Dean Lusher, Johan Koskinen, and Garry Robins. *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge University Press, 2013.

[LRS19] Renaud Lambiotte, Martin Rosvall, and Ingo Scholtes. From networks to optimal higher-order models of complex systems. *Nature physics*, 15(4):313–320, 2019.

[LVM18] Matthieu Latapy, Tiphaine Viard, and Clémence Magnien. Stream graphs and link streams for the modeling of interactions over time. *Social Network Analysis and Mining*, 8:1–29, 2018.

[MIK+04] Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenshtat, Michal Sheffer, and Uri Alon. Superfamilies of evolved and designed networks. *Science*, 303(5663):1538–1542, 2004.

[Nai23] Mehdi Naima. Temporal betweenness centrality on shortest walks variants. *(accepted for publication at Applied Network Science), arXiv preprint arXiv:2305.01080*, 2023.

[NG04] Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

[RJ00] John M Roberts Jr. Simple methods for simulating sociomatrices with given marginal totals. *Social Networks*, 22(3):273–283, 2000.

[RPKL07] Garry Robins, Pip Pattison, Yuval Kalish, and Dean Lusher. An introduction to exponential random graph (p*) models for social networks. *Social networks*, 29(2):173–191, 2007.

[S+02] Tom AB Snijders et al. Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2):1–40, 2002.

[SBT21] Stefania Sardellitti, Sergio Barbarossa, and Lucia Testa. Topological signal processing over cell complexes. In *2021 55th Asilomar Conference on Signals, Systems, and Computers*, pages 1558–1562. IEEE, 2021.

[SNS25a] Felix I Stamm, Mehdi Naima, and Michael T Schaub. Efficient sampling of temporal networks with preserved causality structure. *(accepted for publication at SIAM International Conference on Data Mining (SDM25)), arXiv preprint arXiv:2501.09856*, 2025.

[SNS25b] Felix I Stamm, Mehdi Naima, and Michael T Schaub. Efficient sampling of temporal networks with preserved causality structure. *arXiv preprint arXiv:2501.09856*, 2025.

[TK24] Lionel Tabourier and Julien Karadayi. Probabilistic k-swap method for uniform graph generation beyond the configuration model. *Journal of Complex Networks*, 12(1):cnae002, 2024.

[TRC11] Lionel Tabourier, Camille Roth, and Jean-Philippe Cointet. Generating constrained random graphs using multiple edge switches. *Journal of Experimental Algorithmics (JEA)*, 16:1–1, 2011.

[VKBK21] Katherine Van Koevering, Austin Benson, and Jon Kleinberg. Random graphs with prescribed k-core sequences: A new null model for network analysis. In *Proceedings of the Web Conference 2021*, pages 367–378, 2021.

[VM22] František Váša and Bratislav Mišić. Null models in network neuroscience. *Nature Reviews Neuroscience*, 23(8):493–504, 2022.

[Was94] Stanley Wasserman. Social network analysis: Methods and applications. *The Press Syndicate of the University of Cambridge*, 1994.

[YPVP17] Jean-Gabriel Young, Giovanni Petri, Francesco Vaccarino, and Alice Patania. Construction of and efficient sampling from the simplicial configuration model. *Physical Review E*, 96(3):032312, 2017.