A tale of controllability: evolving network motifs in random Complex Networks

Ph.D. student: Bogdan-Eduard-Mădălin Mursa

Scientific supervisor: Prof. Univ. Dr. Anca M. Andreica

Department of Computer Science
Babeş-Bolyai University, Cluj Napoca, România

Ph.D. thesis summary

Keywords: Complex networks, Dynamic networks, Network generator, Network motifs, Topology

2023

Abstract

This thesis explores the common characteristics shared by complex systems, such as galaxies, natural food chains, and social communities, that are nurtured by the immeasurable amount of interactions between their atomic entities. While scientific frameworks have traditionally focused on analyzing smaller components of systems to understand the whole, this approach may result in overlooking important patterns and losing the big picture. To address this gap, our study makes use of the Complex Networks Theory as a framework able to examine the correlations between patterns of interconnections, or network motifs, and other components of varying scales present in the topology of the complex networks modeled from real-world systems. Our study suggests that network motifs could be a key variable for understanding the evolution and dynamics of complex systems. Based on our empirical evidence, we developed a new null-model that can generate random networks with motifs with a high accuracy and precision. This model is expected to have a significant impact on network motif analysis, providing researchers with a better understanding of how patterns emerge in real-world complex systems.

Contents

List of publications

The rankings have been chosen in accordance with the classification of journals, as per the 2018 ranking published by UEFISCDI. Similarly, the classification of conferences is based on the CORE database, and the rank employed is the one that was active during the publication period. For instance, the 2018 rank (CORE2018) was used for publications made in 2019.

Publications in Web of Science - Journals Citation Index Expanded

[28] **Mursa, B.E.M.**, Andreica, A. (2023): Generating Random Complex Networks with Network Motifs using Evolutionary Algorithm-based Null Model.
(**SUBMITTED**) **(IF=8.139 on publication year 2023)**

[33] **Mursa, B.E.M.**, Andreica, A., Dioşan, L. (2021): Network motifs: A key variable in the equation of dynamic flow between macro and micro layers in complex networks. Knowledge-Based Systems, 213:106648
doi: https://doi.org/10.1016/j.knosys.2020.10664 **(IF=8.66 on publication year 2021)**
Rank A, 8 points

[25] Mester, A., Pop, A., **Mursa, B.E.M.**, Greblă, H., Dioşan, L., Chira, C. (2021). Network analysis based on important node selection and community detection. Mathematics, 9(18):2294
doi: https://doi.org/10.3390/math9182294 **(IF=2.592 on publication year 2021)**
Rank A, 2 points

Publications in Web of Science - Conference Proceedings Citation Index

[29] **Mursa, B.E.M.**, Andreica, A., and Dioşan, L. (2018): Parallel acceleration of subgraph enumeration in the process of network motif detection. In 2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), pages 191–198. Institute of Electrical and Electronics Engineers
doi: 10.1109/SYNASC.2018.0003 **(Index used = CORE2018)**
Rank C, 2 points

[32] **Mursa, B.E.M.**, Andreica, A., Dioşan, L. (2019): Study of connection between articulation points and network motifs in complex networks. In Proceedings of the 27th European Conference on Information Systems (ECIS), 127. AIS eLibrary
https://aisel.aisnet.org/ecis2019_rp/127 **(Index used = CORE2018)**
Rank A, 8 points

[30] **Mursa, B.E.M.**, Andreica, A., Dioşan, L. (2019): An empirical analysis of the correlation between the motifs frequency and the topological properties of complex networks. In Procedia Computer Science, 23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES), volume 159, pages 333–341. Elsevier BV
doi: 10.1016/j.procs.2019.09.188 **(Index used = CORE2018)**
Rank B, 4 points

[31] **Mursa, B.E.M.**, Andreica, A., Dioşan, L. (2019): Mining network motif discovery by learning techniques. In Proceedings of the Hybrid Artificial Intelligent Systems 16th International Conference (HAIS), pages 73–84, Cham. Springer International Publishing
doi: 10.1007/978-3-030-29859-3_7 **(Index used = CORE2018)**
Rank C, 2 points

Publications in journals and conference proceedings indexed by international databases (BDI)

[27] **Mursa, B.E.M.** (2023): Examining the social behavior of ant colonies using Complex Networks. In Proceedings of the Studia Universitatis Babes-Bolyai Informatica,

volume 67, pages 49–64.

Rank D, 1 point

Publications scores: 27 points.

Chapter 1

Introduction

What do a galaxy, a natural food chain, a social community, or other complex systems have in common? Although they are different, these structures share common characteristics by an evaluative dynamic nursed by the immeasurable amount of interactions existing between the entities they contain. Knowing that for each action, there is an equal and opposite reaction, the interactions are what give life to systems apparent static at first glance.

Being surrounded by an immense quantity of stimuli, our brain observed that filtering the information or paying awareness only to specific interactions can be more than enough to ensure the effectiveness of its daily activities [2, 24]. Moreover, by making use of abstraction, our brain can reuse information associated with patterns observed in specific systems in new and unknown systems later encountered [41].

Although unnatural to how our brain was used to perceive the world, our race evolved a scientific framework based on a common analytical approach of splitting the problem into smaller conundrums - divide et impera [11]. It is quite common now, even more since the popularity gain by statistics, to associate a global characteristic to a system by analyzing the most atomic components of its own. Although such an approach is simpler to follow and more comfortable to generalize proofs, it is not necessarily accurate as analyzing individually smaller components of a system one is in danger of losing the big picture.

Without any willingness to neglect the benefits brought by such analytical methodology, which significantly boosted the process of understanding the world, literature presents the need for more robust frameworks that are capable of understanding the apparent chaos as it is and not making order in it. Understanding the importance of patterns and how our brain perceives the world around us, an intriguing question arises: can the patterns be a missing variable in understanding the evolution of complex systems and their dynamics? By dynamic, we understand the non-linear mathematical behavior of a complex system over a certain period [14]. Our work is willing to represent an empirical study of correlations between the properties of these patterns and other components of different scales [1]. We intend to use this empirical evidence as a foundation for a new null-model that can generate random networks with a specific concentration of motifs.

1.1 Motivation

We found network theory as a robust enough framework to serve as a tool for the research we intend to conduct. The mechanism approached by this field is to encode the entities and the interactions between them as topological structures represented as graphs, composed by nodes and edges. What makes these mathematical objects so powerful is that they encapsulate in their underlying graph the contextual information offered by the modeled real-world system. This means that each edge will have a meaning, and it is explainable by understanding a real-world dynamic [8]. Metaphorically, we can say that each complex network has a story.

The interdisciplinary applicability of network theory makes it such an elegant framework, emerging notions from physics, mathematics, computer science, biology, and many others. Statistics are used as a mechanism of quantifying properties of the topological spaces, as it can offer the analytical approach to study the essential components of the networks: the nodes. The study of these fundamental components has formulated a large number of microscale properties from which we can remind a few of them: nodes degree, clustering coefficient [44], assortativity degree [36] or power-law exponent [6]. Fundamental properties of Stability Theory of dynamical systems [9, 18] were used as the argument to sustain the focus of the study on microscale properties in complex networks. This argument states that a dynamic triggered in a complex system by a given perturbation at the macroscale will eventually converge to its smallest structural components, encoding in these microscale components the results of the macro dynamic [4, 15]. Hence, knowing that action at a global scale is persisted at the local scale was a robust argument for continuing a microscale-based research.

Applying this convergence principle in a reversed engineering manner, the studies were extended to a macroscale analysis. In this type of examination, the microscale properties are correlated with macroscale network dynamics. The literature recalls the real-world networks that are characterized by the famous theory of six degrees of separation [5], namely small-world networks. These networks tend to form clusters with short paths between every two

[1] *scale (micro, meso, macro)* found in complex systems - a telescopic perspective of the analyzed structures

nodes of the network, which makes them robust structures without any trade-off in the performance of information flow between its members [38]. Other remarkable networks studied in the literature are sparse networks, which are most of the real-world networks governed by a power-law degree distribution [40] or random networks - structures generated using probability theory to help the study of real-world networks [34]. All these properties are quantifying the relationship of the nodes with their neighbors, as it can be observed in the analytical principle of *divide et impera*.

Although the accepted analytical framework [19, 20, 23, 42] can be considered reliable and came with answers to many questions proposed by researchers, something is missing in the big picture. We consider that the convergence from macro to microlayers [2] and its impact in the in-between layer (meso) are topics not studied sufficient, which require a better emphasis in the current analytical framework of the complex networks. Motivated by the existence of applications that study the importance of mesoscale structures in domain-oriented contexts (biology, ecology or sociology) [1, 26, 39, 47] in respect to the other layers of the network, macro, respectively micro, we consider the mesoscale structures as bridges between the macro and microscale analysis, the missing key in solving the equation of the dynamical convergence from macro to microlayers. Such an approach could lead to a universal framework that could imply better precision of the dynamics of the complex networks surrounding us.

Our focus is on improving existing approaches of understanding complex networks by not following a simple analytical method but trying to observe these topological spaces in their entire complexity and beauty.

The state-of-the-art presents various mesoscale structures such as communities - groups of nodes which are more densely connected than the rest of the network [10], block models - decomposed functional classes of entities within a complex network [17], or modules - community-like components with specific contextual roles [16]. These structures are widely studied in applications addressing problems from fields such as Biology, Physics, or Social Sciences [3, 12, 13, 21, 35, 35, 43, 43], most of the times presented as stand-alone components and not correlated with properties of components found on different scales.

Wharrie et al. [46] studied how the presence of triangles is an essential factor in the evolution of communities - raising awareness on the existence of a relationship between the structures of a given network. Starting from their valuable results, our work is willing to follow a similar principle by presenting an empirical study proposing more detailed research of the correlations between all layers of a complex network - macro, meso, and micro.

[2] *layer (micro, meso, macro)* - structural segmentation level of the networks

As a first step in the pursuit of this goal, in our proposal we are willing to have a research focused on the study of mesoscale properties named network motifs - patterns of interconnections which can occur in significantly higher numbers in complex networks than in randomized networks which hold the same degree of distribution as the first network. Considering their nature of existence, network motifs will always have behind them contextual reasoning brought by the dynamics of the network.

Learning from nature's wisdom and understanding strong network motifs contributing to the dynamics of the network, we consider them a perfect candidate as a first step in the direction of our cross-layer analysis. This thesis proposes an experimental analysis of the correlation between the properties of microscale components and the existence of network motifs, respectively, the existence of network motifs and the properties of macroscale structures. The results obtained will emphasize how the existence of these mesoscale structures can help to discretize types of networks in ones with a more complex source of information. This experiment is also willing to empirically validate the convergence of the dynamic flows between all layers of the complex networks, having network motifs as a bridge in the dynamic convergence from macro to microlayers.

By introducing the study of network motifs as possible components facilitating the convergence mechanism, we intend to strengthen existing conclusions about the micro, meso and macro-layers dynamic previously analyzed through the perspective of modular structures such as triangles [46].

Applied to fields where connections between structures or components play an essential role, this study can help further understand molecular, self-assembly, protein-based cellular adaptation, or social interactions.

Using the results of the experiments focused on describing the convergence mechanism, as this thesis's final resolution, we propose a network generator model to generate random networks that contain motifs. Such a model would be an original contribution to the models existing in the literature that currently do not explicitly consider the evolution of network motifs in the random networks they generate. Considering the uncommonness of real-world networks evolving network motifs, such a model would be a valuable tool for generating synthetic networks for experiments that need large network samples with motifs, improving the quality of the studies and their applications.

1.2 Thesis structure

The thesis is structured in 9 chapters as following:

- **Chapter I - Introduction** presents a general picture of the theory of complex networks and its state-of-the-art applications, with a particular focus on network motifs. The Motivation section outlines the main objective of the thesis which is to propose a new null-model able to generate synthetic networks that contain network motifs and the necessary steps to achieve it, highlighting the potential significance and impact in complex networks research, leading to a deeper understand of their intricate dynamics.

- **Chapter II - Theoretical insights** provides a comprehensive examination of the historical background of Complex Networks Theory (CNT), elucidating the scale viewpoint of network topology through Stability Theory and detailing the distinct properties of each layer at the micro, meso, and macro levels. This chapter serves as a robust theoretical foundation to aid in the comprehension of the subsequent chapters' studies and experiments.

- **Chapter III - Parallel acceleration of subgraph enumeration in the process of network motif detection** addresses the challenge of performance in identifying network motifs, known to be an NP problem. Despite the existence of several algorithms attempting to solve this problem, no perfect approach exists in terms of time performance. The chapter proposes a redesign of the ESU algorithm, which utilizes parallel programming with process-driven and thread-driven models.

- **Chapter IV - Articulation Points and Isolated Communities: Catalysts for Network Motif Evolution** reports on a research study aimed at investigating the potential relationship between articulation points and network motifs in complex networks. The analysis was conducted on a variety of real-world networks to determine the co-existence of these two properties and shed light on the topological features that may contribute to the development and evolution of network motifs.

- **Chapter V - Unraveling the relationship between micro-level topological properties and the propagation of motifs in network topologies** builds upon previous research presented in Chapter IV which established a strong correlation between articulation points and meso-level network motifs. The current study extends this by investigating the relationship between micro-level topological properties and the appearance of motifs. By using network properties, the study aims to gain insight into the mechanisms underlying the appearance of motifs and how network dynamics propagate through its structures.

- **Chapter VI - Prediction of network motifs existence in Complex Networks** describes a novel approach to predict the existence of network motifs in complex networks using Artificial Intelligence (AI) models. This method uses the topological properties identified in Chapters IV and V as facilitators for the evolution of network motifs and trains AI models to determine if a given complex network contains motifs. The aim of this approach is to serve as a pre-processing step to save time in motif detection, which is known to be an NP-complete problem, by using the proposed AI model for prediction.

- **Chapter VII - The study of the empirical evidence of dynamic convergence existing between complex networks' layers** presents an analytical framework that unifies the findings of previous chapters by demonstrating the existence of dynamic convergence between layers of complex networks. The framework shows that the existence of network motifs is directly related to a variety of microscale and macroscale properties, providing evidence of a macro-meso-micro convergence. The ultimate goal of this work is to propose a null-model capable of generating synthetic networks with network motifs in a controllable manner, and this framework aims to facilitate the achievement of this goal.

- **Chapter VIII - Generating Random Complex Networks with network motifs using Evolutionary Algorithm-based null model** proposes a new null-model that generates random networks containing motifs based on a series of micro-scale properties. The chapter builds on previous research that established a correlation between network motifs and various micro- and macro-scale properties of network topology. The proposed null-model, and the end goal of this thesis, aims to facilitate further study of network motifs by providing a way to generate synthetic networks for experimental purposes.

- **Chapter IX - Conclusions** summarizes the findings and contributions of the previous chapters, highlighting their originality and potential benefits for the research community. It also identifies areas for future work to enhance the proposed null-model.

1.3 Original contributions

Throughout the course of this thesis, which aims to propose a null model capable of generating random networks with network motifs, several theoretical and experimental studies were conducted. These studies not only validated research questions and hypotheses but also led to a number of original contributions to the current state-of-the-art through a methodology focused on gathering empirical evidence via statistical significance:

- A detailed comparison between the state-of-the-art algorithms for extracting network motifs in terms of performance and scalability [29].
- Proposing a new parallel model that makes use of parallel and distributed programming to significantly improve the running time of an existing algorithm used to extract network motifs [29]. This parallel model will be used as a primary tool for motif detection in the experiments described in Chapters IV, V, VI, VII, and VIII.
- Proving empirically that articulation points will occur with a higher degree of probability in real-world networks than in randomized networks [32].
- Proving that articulation points through their isolated communities that they link are prone to evolve network motifs [32].
- Proving the altruistic behavior of the ants in the social networks, they are part of centrality measures and other network properties [27].
- Proving the performance of information flow in ant colonies, as optimal short paths between any two ants, and validating the small-world characteristic of the networks modeled from their social interactions [27].
- Revealing a structural overlap between the top nodes detected by centrality measures at network level respectively community level [25].
- Performing a comprehensive study over a large set of micro respectively macro scale properties that revealed a significant correlation between Local Clustering Coefficient and Assortativity Degree and the existence of network motifs, leading to the conclusion that network topologies evolving high values for these two properties will be prone to contain network motifs [30]. The results obtained served as a foundation for the experiments described in Chapters VI, VII, and VII.
- Proposing of classification model that can predict if a network contains or not network motifs based on its topological properties. This approach serves as a time-saving preprocessing step for the state-of-the-art solutions used to detect motifs [31].

- Formalizing an analytical framework that can prove that the convergence of a specific dynamic between the network's layers leads to the materialization of network motifs. This study was accompanied by an experiment that validated the possibility of controlling the evolution network motifs in random networks [33] and serves as empirical proof for the null model proposed in Chapter VIII.

- Proposing a null model architecture based on Evolutionary Algorithms that generate random networks with motifs based on a series of parameters which includes the two most impacting topological properties identified as factors of triggering the convergence between network's layers that stabilize in network motifs: Clustering Coefficient and Assortativity Degree [28].

Chapter 2

Theoretical insights

This chapter covers all the theoretical concepts that will be used in the following chapters, going through the history of Complex Network Theory as it emerged from Graph Theory.

The examination of a network topology is conducted across diverse scales—micro, meso, and macro. Microscale properties quantify measurements representing characteristics of the most fundamental components of the network, nodes, or edges. In graph and network theory, these properties vary in complexity from metrics that reveal a simple count to more complicated metrics that compute different distributions [7, 8, 22, 36, 37, 44].(E.g., node degree, local clustering coefficient, assortativity degree, etc.). On the other hand, mesoscale properties are focused on studying complex network samples of nodes with the interactions between them. The studies revealed that specific groups of nodes interact in distinctive ways that lead to remarkable properties. Mesoscale structures like communities or network motifs are used in applications from various fields to extrapolate theories about real-world complex network dynamics. Especially in social sciences, communities are represented in the literature as the fundamental and important structures in the evolution of a social system [35, 43]. Finally, macroscale properties describe the behavior of a network as a whole trying to identify and classify the network's topologies under categories with common microscale properties.

The scale perspective of Complex Networks analysis will be used as a theoretical foundation for all experiments that will be presented in the following chapters.

Chapter 3

Parallel acceleration of subgraph enumeration in the process of network motif detection

Network motifs bring a great interest to many fields, because they are a perfect candidate to speed up the applied research in the understanding of complex networks dynamics. One of the biggest problems raised in the process of finding these motifs is the performance. Starting from splitting the initial network into subgraphs to the subgraphs clusterization, each algorithm used addresses an *NP* problem. Hence there are many algorithms that are trying to solve network motifs discovery, however there is no perfect approach talking from a time performance perspective, but only solutions that are faster than others.

In this chapter we describe a redesign of one of the most competitive algorithms for subgraphs enumeration found in the literature called ESU [45]. In the proposed approach a given network can be processed by using parallel programming, either with a process-driven model, or a hybrid one (process-driven and thread-driven). Finally, the existing and proposed models are compared using benchmark graphs, revealing competitive results.

The results were published in paper **[29]** and represent an original contribution to the existing state-of-the-art algorithms used to extract network motifs by significantly improving the running waiting time and not affecting the accuracy of the results.

Moreover, we consider the results proposed in this chapter as being complementary with all other experiments we led as it represented an indispensable tool we used to extract motifs from the networks we studied in the main research track.

Chapter 4

Articulation Points and Isolated Communities: Catalysts for Network Motif Evolution

Complex networks are powerful mechanisms one can use to model real-world networks as topological spaces. The beauty of these structures is provided by the infinite degree of analysis one is allowed to do using them. Biologically it is almost impossible for the human mind to comprehend the behaviour of these systems, but when modelled as complex networks different properties of the network topology can reveal precious information. Starting from the two key properties of the participants in a complex network and the relations between them, one can derive further properties that reflect specific behaviours for entities or groups of entities.

Examples of these further remarkable properties include entities which create unique bridges between two or more communities (known as articulation points) or the appearance of patterns of interconnections between entities (known as Network Motifs).

In this chapter it is presented a study on the co-existence of these two properties, articulation points and network motifs, and how their appearance is correlated, by using results obtained in analysing a variety of real-world networks.

The results of this experiment were published in paper **[32]** and represent a first step in our research project that is willing to prove that meso-structures such as network motifs are key components facilitating the dynamic propagation between network layers.

Chapter 5

Unraveling the relationship between micro-level topological properties and the propagation of motifs in network topologies

Based on the evidence found in our previous studies presented in Chapter 4, which revealed a significant correlation between the existence of articulation points and meso-level components such as network motifs, in this chapter we will extend this study by presenting analytical research between a consistent set of micro-level topological properties from Graph and Complex Networks Theory and the appearance of the motifs. The purpose of this study is to use network properties to provide a better understanding of how and why network motifs appear, learning how the dynamic of a network is propagated through its structures.

The results presented in this chapter were published in paper **[30]**. They represent an original contribution in understanding what micro-level properties are indicators to the existence of network motifs. With these results, we can unify the co-existence of various micro and meso-level structures, namely network motifs, and continue studying the co-existence of network motifs with macro-level properties that will bring us to the final step of presenting a complete picture of the convergence phenomena happening in network topologies.

Chapter 6

Prediction of network motifs existence in Complex Networks

Properties of complex networks represent a powerful set of tools that can be used to study the complex behaviour of these systems of interconnections. They can vary from properties represented as simplistic metrics (number of edges and nodes) to properties that reflect complex information of the connection between entities part of the network (assortativity degree, density or clustering coefficient). Such a topological property that has valuable implications on the study of the networks dynamics are network motifs - patterns of interconnections found in real-world networks. Knowing that one of the biggest issue with network motifs discovery is its algorithmic NP-complete nature, in this chapter we present a method to detect if a network is prone or not to generate motifs by making use of its topological properties while training various classification models. This approach wants to serve as a time saving pre-processing step for the state-of-the-art solutions used to detect motifs in Complex networks.

The results and the model proposed were published in paper **[31]** and they represent an original contribution to the optimization of the processes used to extract network motifs. Similar to the improvements presented in Chapter 3, the results described in this chapter represent a complementary work to our main research track as it facilitates the waiting time implied in our experiments for motifs extraction.

Chapter 7

The study of the empirical evidence of dynamic convergence existing between complex networks' layers

As often described in the narative of this thesis, Complex Networks theory represents a powerful tool to model real-world systems as graphs with non-trivial topological features. Static by their definition, complex networks are limited to be the reflection or the snapshot of the dynamical systems they encode in a given moment. Frankly, studies show that the network preserves the characteristics of the dynamic flow between various structures evolved at different scales: micro, meso, and macro, respectively.

Considering that micro components will end by encapsulating any stability state initiated by a perturbation of the macroscale, the most popular framework used to analyze complex networks associates in a reverse-engineering manner properties of the micro-scale components with global network characteristics. Although the principle of convergence is widely agreed upon, we consider the transition from micro to macro one not studied enough, and the highlighting of an in-between layer being necessary exemplifies the convergence mechanism better. Our work is willing to improve existing approaches of understanding the intricate beauty behind network dynamics by embracing the chaos they evolve and not order it using inflexible analytical mechanisms. We formulate a series of research questions that are willing to demonstrate that the existence of a mesoscale structure, namely network motifs, is connected with a variety of microscale, respectively, macroscale properties - hence strengthening the proof of macro-meso-micro convergence.

Our approach could provide a substantial improvement in the models used to understand molecular self-assembly, protein-based cellular adaptation, social interactions, and many other fields, where connections between structures or components play an essential role. The research presented in this chapter was published in the paper **[33]** and it outlines a comprehensive an original analysis framework that ultimately aims the proposal of a null-model for the evolution of network motifs in networks. This analytical framework is based on the principle of causalization of topological properties, which foster the existence of motifs. The results of this analysis represent a significant contribution to the field of network science and provide a valuable resource for future research.

Chapter 8

Generating Random Complex Networks with network motifs using Evolutionary Algorithm-based null model

As discussed in the previous chapter, the existence of network motifs is mainly attributed to the apparent chaotical behavior of the atoms in the systems that the networks are modeling, behavior that emerges in patterns that will be encoded in the network. Motifs were not formalized as structures due to their rarity but only explained through their contextual appearance in the networks. The series of experiments described in the previous chapters are part of the goal of explaining the dynamics of network motifs and they have demonstrated a correlation between the presence of network motifs and various micro- and macro-scale properties of the network topology. This chapter introduces a novel null model architecture that employs Evolutionary Algorithms (EA) to generate random networks containing motifs. The proposed architecture is characterized by a set of parameters, including the Clustering Coefficient and Assortativity Degree, which have been identified as key factors in triggering the convergence between network layers leading to the emergence of network motifs. Such a model has the potential of significantly improve the study of network motifs by generating any number of synthetic networks that can serve as experimental material for the community of researchers.

This chapter presents the results of our research submitted in **[28]**, which conclude with the proposal of a null-model that controls the evolution of synthetic networks with network motifs using Evolutionary algorithms. This proposal is the crowning achievement of our extended research efforts described throughout this thesis.

Chapter 9

Conclusions and future work

The pursuit of finding solutions to challenges in life requires curiosity and the willingness to explore uncharted territories. Such an endeavor can be both beautiful and demanding, and discovering solutions can bring a sense of unparalleled satisfaction. The roadmap of this thesis reflects such an attitude and the author considers himself fortunate to have been able to engage in the pursuit of new knowledge - a dream he had cherished since childhood. Although the primary objective was well-defined, unexpected findings along the way contributed to the author's personal and professional growth. As the saying goes, "the wiser you become, the more questions you have," and staying focused on the roadmap presented a logistical challenge that was overcome with the guidance of the author's supervisor and mentors - to whom he is grateful.

In conclusion, this thesis marks the beginning of a new phase of research. It has been a fulfilling journey, and we are delighted to present the results in the form of a captivating story. As a network enthusiast, we firmly believe that every network has a story to tell, and this thesis is a testament to that belief. Through the review of thousands of networks, this work validated research questions, supported hypotheses, and demonstrated experiments. The findings of this thesis have contributed to the understanding of complex networks, and we are eager to continue exploring new questions and discovering more stories.

The objective of our research was to investigate the evolution of network motifs and explore the feasibility of controlling their formation. Network motifs are infrequent in real-world systems due to their specific dynamics, making them valuable indicators for understanding complex networks. However, their rarity has impeded the development of universal laws governing their existence, thereby presenting significant challenges to their study. Our research therefore required us to explore unknown territory, acknowledging that there would be times when we would not have access to established solutions or answers.

The initial step of this research project involved an investigation of tools for identifying network motifs. Specifically, we explored various algorithms commonly used for determining if a group of nodes and edges can be classified as a motif. Our study revealed that the algorithms had poor performance due to the NP-hard nature of the motif extraction problem. This posed a significant challenge for our research project, as we would need to run the algorithms numerous times, resulting in unacceptably long waiting times for the results needed for our analysis. To address the performance issue of motif extraction algorithms, a reliable and efficient process was developed prior to conducting the desired analysis. A parallelized version of the ESU algorithm was introduced and presented as a substantially improved solution in comparison to other approaches (Chapter 3). The development of this algorithm was a significant accomplishment that not only served the goal of the research project but also benefited the larger research community in the field of complex networks.

The research continued with the collection of a diverse dataset of networks from various domains and sizes to ensure a balanced dataset for analysis. Alongside the extraction of motifs from these networks, we also collected their topological properties (Chapter 2) to investigate potential correlations with the presence of motifs. The parallelized ESU algorithm proved to be a significant time-saver, allowing us to efficiently gather a comprehensive dataset in a surprisingly short period.

Following the principles of dynamical convergence from the Stability Theory, we developed a set of research questions to guide our experiments. These questions were centered around the atoms or nodes of the system, as they encode the dynamics that converge to them and eventually settle into a new state. We used this approach to formulate a hypothesis explaining why and how nodes and edges appear or disappear in complex networks, providing an overarching explanation for the evolution of motifs. According to our hypothesis, when a specific event occurs in a system, the atoms behave in a way that creates structures guided by the event's dynamics. From the perspective of a modeled system as a network, this is encoded as nodes having specific edges, which are detected as network motifs.

Building upon the theoretical insights obtained from our analysis, we conducted a series of statistical tests to investigate the relationship between the extracted topological properties of the networks and the presence of motifs. The resulting complex statistical analysis provided answers to all of the research questions we had formulated, confirming our convergence hypothesis. Our findings revealed that a set of topological properties exhibit a strong correlation with the existence of motifs in the network.

We started by investigating the correlation between the presence of Articulation Points (*AP*) in a network and the existence of network motifs. *AP* nodes are critical bridges between two or more communities and play a significant role in facilitating information flow between them. Through a series of experiments presented in Chapter 4, we demonstrated that *AP*s have a controllability property, which enables them to propagate specific dynamics to the communities they control while isolating them from others. This leads to a rapid evolution of motifs in those networks. Our findings have important implications, particularly in social networks, where *AP*s are often influential people who can induce ideologies in the communities they lead. This has a dangerous application, especially in social networks, where *AP*s are usually influential people and can induce ideologies in the communities that they lead.

The other two topological properties highly associated with the presence of network motifs were assortativity degree (ρ) and local clustering coefficient (CC_l), owing to their ability to encode specific network behaviors. ρ is a measure that indicates whether nodes in a network are more likely to be connected to other nodes with the same degree, which is comparable to people in a community being more likely to have friends who are equally popular. This particular behavior is highly specific and explains the emergence of certain motifs during the process of forming links between nodes of the same degree. Local clustering coefficient (CC_l) is another topological property that is highly correlated with the existence of network motifs. It indicates the tendency of nodes in a network to form tightly interconnected groups or clusters. This behavior provides a favorable environment for the evolution of motifs among the nodes within the same cluster, particularly if they tend to link with nodes of similar degrees following the ρ principles of preferential attachment.

After obtaining the results of the analysis, we proposed a comprehensive framework that confirmed the convergence properties from Stability Theory in complex networks, thereby achieving our initial objective of comprehending the evolution of network motifs in such networks (Chapter 7). Furthermore, based on these outcomes, we suggested a new tool that simplifies the extraction of motifs as a preprocessing step. The task of extracting topological properties from complex networks is significantly faster than extracting motifs, even with the use of parallelization. Since specific topological properties and their values have been shown to indicate the presence of motifs in networks, we propose training a series of AI-based models that can predict the presence or absence of motifs in a network based on its topological properties. In this way, it is possible to determine from an early stage whether the network is worth processing using the trained AI-based model as a recommender system. Of note, one of the trained models achieved outstanding performance, detecting the presence or absence of motifs with over 95% accuracy (Chapter 5).

Despite the promising results obtained from our previous analyses, gave birth to a new more supreme goal. We identified a research gap in the current state-of-the-art literature, specifically the absence of a null-model that can generate networks containing motifs. While existing null-models aim to generate specific network types using stochastic methods to add links between nodes until a desired property is met, none address the challenge of creating networks that contain motifs. We believed that the algorithmic approach used by most null-models in the literature, which aims to simulate a synthetic behavior, contradicts the natural environment required for motifs to exist. As a result, we developed a new null-model architecture based on an Evolutionary Algorithm (EA) that can evolve networks to replicate the continuous dynamics of a real-world system and able to generate complex networks containing motifs. The EA component aimed to optimize ρ and CC_l properties simultaneously and to devise a more realistic mechanism to add edges to the network, a crucial factor for motif existence. We controlled the proposed null-model using a set of parameters based on a novel fitness function, which we introduced for the first time in the literature. The experiment described in Chapter 8 showed that the proposed architecture was reliable, as it generated networks with strong motifs. This contribution represents an important addition to the literature, with potential implications for future research in this field.

In future work, we consider the study of coexistence dynamics between network motifs and other meso-scale topological structures, such as communities or chains. This exploration holds the potential to unveil nuanced behaviors in the evolution of network motifs' controllability. Leveraging these structures could enhance the precision of the existing null-model introduced in Chapter 8, enabling the design of a mechanism for regulating motif size and not only their existence. Additionally, we plan to explore diverse null-model architectures, such as stochastic block models or deep generative models, and conduct comparative analyses. This approach promises to unlock further avenues for controlling motif evolution in generated networks.

In conclusion, while we acknowledge that no work can ever be considered completely finished, we are confident that our research has contributed significantly to the field of complex network analysis. Throughout this journey, we have proposed a series of results innovative to the state-of-the-art and opened up new research tracks and considered suggestions for improvement at the end of each chapter, with the hope of inspiring young researchers to continue pushing the boundaries of knowledge. Looking back on our achievements, we take pride in the completeness of the thesis and we are excited to see how the result will impact future research in this area.

References

[1] Albert, I. and Albert, R. (2005). Conserved network motifs allow protein-protein interaction prediction. *Bioinformatics (Oxford, England)*, 20:3346–3352.

[2] Alho, K. (1992). Selective attention in auditory processing as reflected by event-related brain potentials. *Psychophysiology*, 29(3):247–263.

[3] Altay, Y., Cao, S., Che, H., Abdelmohsen, L., and Hest, J. (2019). Adaptive polymeric assemblies for applications in biomimicry and nanomedicine. *Biomacromolecules*, 20(11):4053–4064.

[4] Angulo, M., Liu, Y.-Y., and Slotine, J.-J. (2014). Network motifs emerge from interconnections that favor stability. *Nature Physics*, 11:848–852.

[5] Bakhshandeh, R., Samadi, M., Azimifar, Z., and Schaeffer, J. (2011). Degrees of separation in social networks. In *Proceedings of the 4th Annual Symposium on Combinatorial Search, SoCS 2011*, pages 18–23. Association for Computing Machinery.

[6] Barabási, A.-L. (2009). Scale-free networks: A decade and beyond. *Science*, 325(5939):412–413.

[7] Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in networks. *Science*, 286(5439):509–512.

[8] Barabási, A.-L. and Pósfai, M. (2016). *Network science (pages 21–29)*. Cambridge University Press, Cambridge.

[9] Barbu, V. (2016). *Stability Theory (pages 123–159)*. Springer, Cham.

[10] Chakraborty, T., Srinivasan, S., Ganguly, N., Bhowmick, S., and Mukherjee, A. (2013). Constant communities in complex networks. *Scientific reports*, 3:1825.

[11] Cormen, T., Leiserson, C., Rivest, R., and Stein, C. (2001). *Introduction To Algorithms (pages 65–97)*. The MIT Press.

[12] Deritei, D., Rozum, J., Regan, E., and Albert, R. (2019). A feedback loop of conditionally stable circuits drives the cell cycle from checkpoint to checkpoint. *Scientific Reports*, 9:16430.

[13] Findlay, M., Freitas, D., Mobed-Miremadi, M., and Wheeler, K. (2017). Machine learning provides predictive analysis into silver nanoparticle protein corona formation from physicochemical properties. *Environmental Science: Nano*, 5(1):64–71.

[14] Forrester, J. (1989). The beginning of system dynamics. *McKinsey Quarterly*, 4.

[15] Ghys, E. (2013). The lorenz attractor, a paradigm for chaos. *Progress in Mathematical Physics*, 66:1–54.

[16] Guimerà, R. and Amaral, L. A. N. (2005). Cartography of complex networks: modules and universal roles. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(2):P02001.

[17] Harrison, C. W., Scott, A. B., and Ronald, B. (1976). Social structure from multiple networks. i. blockmodels of roles and positions. *American Journal of Sociology*, 81(4):730–780.

[18] Hirsch, M., Smale, S., and L Devaney, R. (2004). *Differential Equations, Dynamical Systems and an Introduction to Chaos (pages 140–142)*. Academic Press.

[19] Hoel, E. (2018). *Agent Above, Atom Below: How Agents Causally Emerge from Their Underlying Microphysics (pages 63–76)*. Springer, Cham.

[20] Hoel, E., Albantakis, L., and Tononi, G. (2013). Quantifying causal emergence shows that macro can beat micro. *Proceedings of the National Academy of Sciences of the United States of America*, 110(49):19790–19795.

[21] Itzkovitz, S., Levitt, R., Kashtan, N., Milo, R., Itzkovitz, M., and Alon, U. (2005). Coarse-graining and self-dissimilarity of complex networks. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 71(1):016127.

[22] Kunegis, J., Blattner, M., and Moser, C. (2013). Preferential attachment in online networks: Measurement and explanations. In *Proceedings of the 5th Annual ACM Web Science Conference*, WebSci '13, pages 205–214. Association for Computing Machinery.

[23] Marchiori, M. and Possamai, L. (2015). Micro-macro analysis of complex networks. *PLOS ONE*, 10(1):1–27.

[24] Mayer, A. R., Hanlon, F., Franco, A., Teshiba, T. M., Thoma, R., Clark, V., and Cañive, J. (2008). The neural networks underlying sensory gating. *NeuroImage*, 44(1):182–189.

[25] Mester, A., Pop, A., Mursa, B.-E.-M., Greblă, H., Dioşan, L., and Chira, C. (2021). Network analysis based on important node selection and community detection. *Mathematics*, 9(18):2294.

[26] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827.

[27] Mursa, B.-E.-M. (2023). Examining the social behavior of ant colonies using complex networks. In *Proceedings of the Studia Universitatis Babes-Bolyai Informatica*, volume 67, pages 49–64.

[28] Mursa, B.-E.-M. and Andreica, A. (2023). Generating random complex networks with network motifs using evolutionary algorithm-based null model. *Swarm and Evolutionary Computation*. **SUBMITTED**.

[29] Mursa, B.-E.-M., Andreica, A., and Dioşan, L. (2018). Parallel acceleration of subgraph enumeration in the process of network motif detection. In *2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, pages 191–198. Institute of Electrical and Electronics Engineers.

[30] Mursa, B.-E.-M., Andreica, A., and Dioşan, L. (2019a). An empirical analysis of the correlation between the motifs frequency and the topological properties of complex networks. In *Procedia Computer Science*, volume 159, pages 333–341. Elsevier BV.

[31] Mursa, B.-E.-M., Andreica, A., and Dioşan, L. (2019b). Mining network motif discovery by learning techniques. In *Hybrid Artificial Intelligent Systems*, pages 73–84, Cham. Springer International Publishing.

[32] Mursa, B.-E.-M., Andreica, A., and Dioşan, L. (2019c). Study of connection between articulation points and network motifs in complex networks. In *In Proceedings of the 27th European Conference on Information Systems (ECIS)*, page 127. AIS eLibrary.

[33] Mursa, B.-E.-M., Dioşan, L., and Andreica, A. (2021). Network motifs: A key variable in the equation of dynamic flow between macro and micro layers in complex networks. *Knowledge-Based Systems*, 213:106648.

[34] Nasini, S., Martínez-de Albéniz, V., and Dehdarirad, T. (2017). Conditionally exponential random models for individual properties and network structures: Method and application. *Social Networks*, 48:202–212.

[35] Nguyen, N. P., Dinh, T. N., Shen, Y., and Thai, M. T. (2014). Dynamic social community detection and its applications. *PLOS ONE*, 9(4):1–18.

[36] Noldus, R. and Van Mieghem, P. (2015). Assortativity in complex networks. *Journal of Complex Networks 2015*, 3(4):507–542.

[37] Raj P. M., K., Mohan, A., and Srinivasa, K. G. (2018). Basics of graph theory. In *Practical Social Network Analysis with Python*, pages 1–23. Springer International Publishing.

[38] Ravid, G. and Rafaeli, S. (2004). Asynchronous discussion groups as small world and scale free networks. *First Monday*, 9(9).

[39] Ricketts, S., Francis, M., Farhadi, L., Rust, M., Das, M., Ross, J., and Robertson-Anderson, R. (2019). Varying crosslinking motifs drive the mesoscale mechanics of actin-microtubule composites. *Scientific Reports*, 9:12831.

[40] Scholz, M. (2015). Node similarity as a basic principle behind connectivity in complex networks. *Journal of Data Mining & Digital Humanities*, 2015:1–7.

[41] Shugen, W. (2002). Framework of pattern recognition model based on the cognitive psychology. *Geo-spatial Information Science*, 5:74–78.

[42] Sun, P. and Ma, X. (2017). Understanding controllability of complex networks from microcosmic to macrocosmic. *New Journal of Physics*, 19:013022.

[43] Viswanath, B., Mislove, A., Cha, M., and Gummadi, K. P. (2009). On the evolution of user interaction in facebook. In *Proceedings of the 2nd ACM Workshop on Online Social Networks*, WOSN '09, pages 37–42. Association for Computing Machinery.

[44] Watts, D. and Strogatz, S. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393:440–442.

[45] Wernicke, S. (2006). Efficient detection of network motifs. *Transactions on Computational Biology and Bioinformatics*, 3:347–359.

[46] Wharrie, S., Azizi, L., and Altmann, E. G. (2019). Micro-, meso-, macroscales: The effect of triangles on communities in networks. *Physical Review E*, 100(2):022315.

[47] Yeger-Lotem, E., Sattath, S., Kashtan, N., Itzkovitz, S., Milo, R., Pinter, R. Y., Alon, U., and Margalit, H. (2004). Network motifs in integrated cellular networks of transcription–regulation and protein–protein interaction. *Proceedings of the National Academy of Sciences*, 101(16):5934–5939.