Computational Analysis of Emergent Behaviour in Collaboration Networks

PhD Thesis – Abstract

in order to obtain the degree of PhD from

Polytechnic University of Timisoara

in the field of Computers and Information Technology

author: ing. Gabriel BARINA

supervisor: prof.univ.dr.ing. Mircea VLĂDUȚIU

The expanding domain of Network Science facilitates the understanding of existing patterns of connection in nature and our own society, both physical and social. Social Network Analysis, the application of the broader field of Network Science, has received an increase of interest from the scientific community, due to its relevance in analyzing the intricate nature of social dynamics, emergent human behaviour, collaboration and influence.

The goal of this thesis is to use Computer Science as an underlying tool for simulating complex models in a dynamic fashion, as well as to uncover crucial aspects regarding social and economic collaboration and emergent behaviour.

The main objectives of this thesis are:

1. Create a state-of-the-art emergent and collaboration network - by using the analytical power of computers (*e.g.* data-mining, machine-learning, *etc.*) -, based on real-world data, in order to analyze and compare its fundamental properties to other, similar networks.
2. Propose a new metric capable of quantifying the sociability of a node in regard to the social features.
3. Create a simulator capable of simulating emergent relationships between (economic) agents and releasing it as a tool.
4. Using the created simulator as a tool, simulate and analyze the share of total payoff, the distribution of payoff, as well as the ergodicity of economic networks.

In order to achieve the first two objectives, in the first part of this thesis I analyze the emergent network formed by musicians. By employing both traditional methods of analyzing complex networks, as well as original elements exposed with this thesis, I analyze the network of musicians not only from a collaborative point of view (*i.e.* between musicians), but from an economic point of view (*i.e.* their activity to produce musical content) as well. As such, I bring the following contributions to this thesis:

- Big Data mining: the nature of information needed for this study meant that the data itself was not readily available and needed to be gathered from several online repositories. Furthermore, inspired by the Jazz musicians network, MuSeNet is not limited to one genre, but instead takes into account the bands and musicians from all musical genre.
- Centrality analysis: I analyze MuSeNet from the perspective of important centralities.
- Machine learning (unsupervised): by using the analytical power of computer analysis and simulation, I segregate fundamental communities with the help of a 2D force-directed (community detection) layout algorithm. Furthermore, I identify the

overlapping of genres, detect influential agents, as well determine the "Kevin Bacon" of the music industry.
- Motif analysis: I apply a novel approach of analyzing and differentiating networks by identifying existing motifs within these networks.
- *S*-metric: by introducing a state-of-the-art metric into literature, I determine the sociability of several networks, and by comparing them to MuSeNet, I discuss their real-world effects.

In the second part of the thesis I present a state-of-the-art socio-economic simulator, based on both empirical observations, as well as innovative economic models. Having precise simulation capabilities, it is used to analyze the processes behind collaborations and interactions, as well as the emergent distribution of payoff at the macroscopic scale, with specific, well-defined rules at a microscopic scale. As such, the contributions brought in the second part of this thesis refer to the creation of a state-of-the-art trade and economic simulation application, with the following features:

- Driven by heuristics: with additional improvement to the mechanism to model the behaviour of economic agents, by using the tolerance-based interaction model as foundation.
- Tailored according to main schools of economic thought: with high flexibility in terms of economic theories, agent models, and interaction assumptions.
- Real-life features: complex network topologies, evolution of economic agent roles, dynamic creation of new economic agents, diversity in product types, dynamic evolution of product prices, and investment decisions at agent-level.
- Valuable simulation application: by using TrEcSim, I analyze the following attributes of (economic) exchange networks:
  - Static and dynamic distributions of payoff: I analyze inter-agent dynamic and emergent behaviour.
  - Ergodicity: I employ computer simulation in order to obtain accurate insight regarding the intrinsic fairness of economic systems, based on network topology and producer/consumer placement.

Thesis structure includes an introductory chapter, a chapter pertaining theoretical foundations, and one describing the state-of-the-art of the chosen field. In the next two chapters, I present the two emergent collaborative networks analyzed (MuSeNet and TrEcSim), each with its own dedicated chapter. The thesis ends with a chapter dedicated to the conclusions, contributions, as well as to the research direction, references (219 titles consulted and cited) and an annex. This thesis extends itself over 122 pages and contains 62 figures and 15 tables.

In the first chapter I present a brief introduction to the field of Network Science. Containing elements from exact sciences – *e.g.* computer science, physics, mathematics, *etc.* – Network Science facilitates the study of our society, its behaviors and relationships by using the computer as a tool for modeling and simulating Big Data. Social Network Analysis – one of the main branches of Network Science – has caught the attention of the scientific community, due to its applicability in analyzing and understanding of real-world networks, both from a topological level of a given network (*i.e.* how each nodes are connected to each other), and from a behavioural level (*i.e.* how each nodes interact with each other). Since both of these are analyzed using empirical studies (*e.g.* statistical analysis, direct measurements, indirect measurements, *etc.*), Social Network Analysis can lead to the creation of valid models for the observed real-world networks.

The motivation behind the research presented in this thesis is to observe and understand the professional relationships of agents (both musical and economic), how they form new links based on their common attributes (*e.g.* role, profession, location, preference, *etc.*), and watching this collaboration network evolve with each new node, all the while staying within the framework of Computer Science.

Ever since my Master of Science studies, after being exposed to the concept of complex networks, my research involved more and more often the usage of Network Science in Computer Science. As a result, I opted to use computer analysis and simulation as research methodology for this thesis as well, especially due to the fact that a purely mathematical approach would not offer the possibility of gathering, analyzing and modeling the huge amount of intricate data required to model complex networks. That being said, in order to make use of a mathematical approach, researchers are limited to a purely statistical analysis on networks with either a regular, or a random graph topology. Therefore, by using the analytical power of computer analysis and simulation, I show that the generated inter-agent relationships are indeed realistic and dynamic in nature, and as a result, they can be used in real-world applications.

This research, along with its results allows us to elucidate the emergence and mechanisms of various social phenomenon and whether they share dynamical and structural features or not with other natural, social processes. Closely observing social phenomena like influential agents, collaborations between two or more agents, or even the formation of a new agent (or link) will constitute an excellent opportunity to understand network formation processes and influence dynamics. Indeed, Network Science brings a better understanding for the structure and behaviour of social and economic networks, thus proving that human interaction is not only important in Social Science, but it is also essential for many other fields such as technology and engineering.

In the second chapter I present the theoretical foundations necessary for understanding this thesis. Thus, I present a classification of complex networks into four basic categories (*i.e.* biological, social, technological and semantic), basic topologies (*i.e.* regular mesh, random, small-world and scale-free), as well as metrics specific to complex networks (*i.e.* centrality, degree distribution, average path length, clustering coefficient, modularity) used in this thesis. Related to these metrics, I also present two novel approaches of analyzing networks, namely: network motifs and metric fidelity.

One important common property of all networks is that they can be represented as graphs, as well as sub-graphs called (network) motifs. Motifs are defined as being recurrent and statistically significant sub-graphs or patterns of complex networks. Since each and every one of these sub-graphs, defined by a particular interaction-pattern between graph nodes, reflects a specific function in the network, as a whole, they can also be used to compare various networks. However, their detection is still computationally challenging. This is due to the large amount of combinations which need to be detected and compared. To this end, the smaller the size of the motif, the easier is to detect; as such, I rely only on motifs of size when analyzing MuSeNet and comparing it to other networks. Even though there are a few approaches by various authors studying network functionality using motifs of up to 6, I found that using smaller motifs not only do I obtain far less distinct patterns, but they are also much more numerous to be found in graphs, thus yielding far more relevant results.

A new and alternative method of quantitatively comparing networks – one that is also used in this thesis – is to compute each network's metric fidelity and to compare them among each other based on individual metric measurements. Tailored to express the similarity between any two generic vectors, it can offer insight on network model resemblance or synthetic model realism compared to a real world network.

In the third chapter I present the state-of-the-art regarding the process of collaboration in complex networks. In general terms, complex networks are formed by a set of social actors connected together based on certain rules. These nodes, though mostly autonomous, geographically distributed, and heterogeneous in terms of their operating environment, culture, social capital and goals, all share the same basic property: they collaborate with each other in order to achieve common or compatible goals. Thus, the more significant this outcome, the higher the participation and commitment level will be among the collaborators. From a structure-point-of-view, this process can be represented by a physical connection between the collaborating nodes. Studies performed over a variety of complex networks have resulted in mapping distinctive types of relationships, yet featuring similar properties, namely the desire to collaborate, in one way or another. As such, a handful of such examples are listed, most of which constitute prerequisites to MuSeNet:

- **Co-authorship Network:** a contemporary landmark in the academic research of the collaboration process represents the study of the authors of different scientific publications. These authors are no longer isolated agents, but are part of a multidisciplinary collaborative network. Among other things, the analysis of these networks offers new perspectives regarding the number of published articles, their quality, the location of the authors, but also the evolution of the network over time. As a structure, this network is small-world, with clearly delimited communities. This is due to the fact that the collaborative action is adapted from one node (author) to another, based on the scientific area of the respective group.
- **Marvel Universe:** Marvel Entertainment has been in business for over 70 years, continuously developing characters, plots and media (*e.g.* movies, television shows, games, *etc.*), only to realize that for a newcomer, jumping into this plethora of information would be an intimidating process of manual and time-consuming research. Aiming to simplify this process, and to overcome its disadvantages, the community behind the Marvel universe resorted to the power of graphs. Thus, a connection was created between superheroes who appeared, even for a short time in the same movie, cartoons, series, game, *etc.* The network obtained - consisting of ~ 6,500 nodes (superheroes) and ~ 10,000 edges (relationships) -, although a synthetic one, is very similar to other naturally formed networks however, the clustering coefficient differs drastically from that of real-life collaboration networks, due to the way how characters are distributed throughout the media. This completely contradicts the way how real-life scientists collaborate in writing scientific papers, and is due to the networks 'artificial origins.
- **IMDB Actor's Network:** Derived from a famous statement made by Kevin Bacon himself, a whole science was dedicated to this, sparking an interesting concept in the domain of social networks: the Bacon number; it is defined as the number of degrees of separation any given Hollywood actor has from Kevin Bacon. The higher the Bacon number, the farther away from Kevin Bacon that particular actor is. The computation of a Bacon number for any given actor is based on the shortest path algorithm, applied to the co-stardom network.
- **Jazz musicians network:** Similar to the previously presented studies, an analysis involving the collaboration network of Jazz musicians represents yet another prerequisite of MuSeNet. In this particular study, the authors present both the collaboration network formed between two individuals, where two musicians are connected if they have played together in the same band, as well as the collaboration band network. This network is formed by creating connections between bands who feature at least one common member. One of the most interesting results was the

segregation of musicians in communities, either because of racial discrimination or because of their geographical location.

In the fourth chapter I create and analyze MuSeNet, the collaborative network or musicians. Based on the mentioned publications regarding collaboration in complex networks, we can notice the emergence of certain communities based on existing ties within the network. Inspired by these studies, I considered it paramount to address the existing relation-ships – both collaborative and economic – between musicians. As such, by staying within the framework of Computer Science, MuSeNet, a novel approach of mapping and analyzing the community formed by musical artists – without limiting it to just one genre –, was introduced into literature. The necessary data were collected from several sources, obtaining information on ~ 20,000 musicians and ~ 5,000 bands.

Network Analysis

In Figure 4.2 we can see the relevant emerging communities that form over MuSeNet based on genre, namely Pop/Rock 24.56%, Jazz 16.72%, Blues 15.8%, Classical 8% and Country 5.35%. Even though the proportion of music styles is already a known fact, what network analysis unveils are the existing spatial distributions as well as their overlapping properties. As such, the most popular genres are also the ones clustered together, as there are more collaborating artists. The topologically marginal genres are also the ones less popular, confirming the fact that there is a correlation between the communities ’center of gravity and their real-world popularity. As a general rule of thumb, the further a genre-community is from the absolute centre of MuSeNet, the less popular it is. This holds true for the opposite also.

As the most dominant music style, the community formed by Pop/Rock artists is very central and also tightly clustered, meaning that artists in this industry prefer to work together with others alike. On the opposite side lies the community formed by Jazz musicians. This community tends to dissipate and overlap multiple styles. This is due to the very collaborative nature of Jazz musicians together with musicians of various other genres. The same conclusion can be drawn for Classical music which, in today’s world, implies composing contributions for movie scores, commercials, and melodic lines for other genres. Finally, Country music shows a similarity to the pop/rock community, namely that all artists are linked more with each other rather than with musicians from other genres. However, the community has a more eccentric position which I correlate with its popularity.

In Figures 4.3 - 4.6 we can observe the distributions of centralities in MuSeNet, specific to other networks as well, namely a distribution of power-law, degree, eigenvector and pagerank. Noteworthy is the dominant cluster in Figure 4.5, consisting of nodes with a very high eigenvector. Upon close inspection, I determined that this community is made up of mature musicians, such as Alphonso Johnson, with a recording studio; the fact that most published music goes through their studio makes them,as a whole, the central community in MuSeNet. Referring to the previously mentioned idea of meritocracy vs. topocracy presented in a recent study by authors Borondo *et al.*, this community is the one that thrives mostly in the topocratic environment of the music industry, making the most out of its influence in the music industry. This also holds true from an economic point of view, as content creation is a form of economic activity. Moreover, this real-world influence is replicated in the graph.

Finally, similar to the IMDB study which denotes Kevin Bacon as the most influential node in the Hollywood actor network, I identify Dave Grohl as the ”Kevin Bacon” of the music industry. This aspect is clearly visible in Figure 4.6, where I show the betweenness distribution, a classical method of computing influence.

Graph Metric Analysis

In order to obtain other relevant results, I compared MuSeNet with other distinct networks: Jazz and IMDB – due to their similar approach –, but also social networking models, such as Faccebook, Twitter or Google+, in order to put in perspective the particular features that artists have as opposed to everyday Internet users. Thus, I used the topological metrics which are specific for every complex network, namely: average degree, average path length, average clustering coefficient, modularity, graph edge density and graph diameter.

Interestingly, the Facebook model is at an average level in terms of sociability, while the IMDB actor's network is more sociable and MuSeNet on the contrary, less sociable. This difference can be explained as follows: Facebook users (*i.e.* everyday users) interact and create new friendships at what we call a normal rate. Actor's everyday job, however, relies on co-starring with other actors, in a different movie every time, due to the fact that casts for movies are very broad. This makes their network very clustered and thus seems more sociable, in our terms. Musicians, however, do not usually create art (work) with many others. They mainly rely on their own band (of approximately five members on average), and not more then on the other artists from their own genre. This makes links in MuSeNet less dense, clustering very high and the community structure powerful. By applying the sociability term on MuSeNet, it can easily be considered as a "non-sociable" network.

Motif Distribution

A popular approach in Network Science to analyze the functional abilities of a given network is by uncover structural design patterns – *i.e.* repetitive sub-graphs –, consisting of a well-defined number of nodes, which are specific only to that given network. For this particular study, I considered only size-3 motifs, due to the simplicity of detection, their large number of appearances in the network, as well as the relevance they offer. Therefore, I determined the distribution of patterns for each empirical network using the FANMOD algorithm, one of the fastest detection algorithms. As a result, I obtained Figure 4.7, in which we can see that the Jazz musicians network behaves more like a normal social network – having a uniform distribution of motifs – while the IMDB and MuSeNet networks have a predominant motif characterizing them. As the second and last step, I apply the fidelity metric to compare the motif distribution vectors with one another, where a value of 1 means complete similarity, while a value of 0 means complete dissimilarity.

In the fifth chapter I present TrEcSim, a novel socio-economic simulator, as well the simulation results obtained with it. Switching from the musical industry to the domain of Ecomomics, it is very important to understand the conditions in which certain economic agents fare better than others at individual-level. Also, it is important to discern the types of social and economic networks that are associated with the best outcomes at system-level; however, due to the fact that economic networks are non-linear, un-predictable complex systems, it is very difficult to analyze them based only on real-time quantitative observations. Another approach for the analysis of economic systems is to simplify them using mathematical models, or to simulate them with simulators capable of simulating complex agent behaviour.

As such, I extend the existing economic models, simulators and empirical observations by creating TrEcSim. The Trade and Economic Simulator is a state-of-the-art economic network simulator, where agent decision is driven by certain heuristics, that were tailored according to main economic theories and is designed to support the following real-life features: complex network topologies, evolution of economic agent roles, dynamic creation of new economic agents, diversity in product types, dynamic evolution of product prices, and investment decisions at agent-level. Here, my scope is to gain a better understanding of economic networks and to analyze inter-agent dynamic behaviour by means of computer modeling and simulation. As such, by employing computer simulation, I want to address the following objectives:

- Simulate economic networks based on four underlying network topologies – mesh, small-world, random and scale-free – in order to get a better insight on static and

dynamic distributions of payoff.
- Analyze the influence of topological features (*i.e.* network topology and placement of agents according to their roles) on the distribution of payoff.
- Implement a new mechanism for modeling the behaviour of economic agents, inspired by the tolerance-based interaction model.

In short, TrEcSim allows the computational analysis of collaborative and emerging economic networks on a macroscopic scale, with parameters set at the micro level.

One of the most interesting studies that has attracted the attention of the scientific community is the model described by the authors Borondo *et al.* regarding integrated economic systems, a model on which TrEcSim is based on. The analysis of these systems is particularly important because the creation of new economic ties is costly and often influenced by the social network that co-exists within them. Therefore, in a real economic system, revenues can be classified according to their source, either from the producer or from the middleman, the intermediary in a transaction. Thus, the authors Borondo *et al.* argue that once the number of connections in the network decreases, the network goes from a meritocratic state (fair, where the income is determined by the talent of the individual agent) to a topocratic state (unfair, where the topological position of the economic agent determines the obtained income). However, the model described by the authors Borondo is a simplified one, due to the mathematical and statistical approaches used:

- Assumes only the random topology as the underlying network (*i.e.* random).
- The number of economic agents and their roles as producer/rock-star or intermediary are predefined and fixed.
- The agents only produce one type of product, with a fixed (pricing) value.
- All analysis is done in a single iteration.

All these aspects impact the realism of the model, as well as the study as a whole. Conclusively, I addressed these issues within TrEcSim: it uses any complex system as a basis for simulations, either created or imported. Moreover, the evolution of agents and roles is variable, and such an example can be seen in Figure 5.11: the producers of the product *Pr1* on the left side can also be consumers of the product *Pr2* on the right, in another simulation cycle. Similarly, the evolution of products, their quality and quantity, as well as their price, are also variable, and the income obtained by agents can be invested in various actions.

Model Description

The framework itself can be split into three main components: initialization phase, transactional phase and decisional phase; an overview of TrEcSim's framework is presented in Figure5.10, corresponding to the implementation of the extended model. In the initialization phase four main processes take place, in order to create the simulation interface, namely: *createNetwork*, which creates the network of economic agents, based on the settings made or the imported data; *createProducts* which defines products and their attributes; *createProductions* which defines producers and their production attributes; and *createNeeds* which defines the sub-set of needs for each economic agent.

In the transactional phase, economic agents identify current needs based on the importance factor, quality, quantity and price of the product. Thus, each new iteration of the transactional phase starts with the *getBestProduction* process, to compute the best option from where to procure the necessary products, while the *getAffordableQuantity* process determines the maximum quantity that can be obtained depending on the quality and quantity of the products. The transactional phase ends with the *finalizeTransaction* process, which completes the transactions according to the attributes mentioned above, and to determine the individual income, based on Equation 5.4.

The transactional phase, while important from an economical point of view, does not entail by itself the network's dynamicity. On the other hand, it is in the decisional phase where the dynamics of the network and its topology are determined. It is in this phase, where the economic agents probabilistically decide which of the following actions to adopt:

- Action 1 (determined by the *decideCreateLink* process): creating new links between two economic agents and thus circumventing a given number of middlemen. Based on information gathered so far, the algorithm computes which economic agent would be best suited to link to, in order to improve the current agent's economical stance.
- Action 2 (determined by the *decideCreateProduction* process): investing in the creation of a new product by allowing the current economic agent to start producing a specific *Pr* product, based on demands
- Action 3 (determined by the *decideImproveProduction* process): invest in improving current production quality or quantity by looking through all of economic agent's current products and deciding upon improving either quality or quantity for a given product *Pr*.
- Action 4 (determined by the *decideExpand* process): expanding the network by creating a new economic agent. This option requires that the algorithm analyzes several attributes before computing its outcome: the percentage of the current funds which will be transferred to the new agent; the advantages and disadvantages of creating different products; the advantages and disadvantages of being linked to the new agent; how much of the current agent's debt it should inherit; *etc.*

To compute the best investment action for the current economic agent (*determineCurrentDecision* process), the *getPastDecisionScores* and the *getPastDecisionScoresFromNeighbours* processes determine which of the past decisions were most profitable for the current agent and an arbitrary number of neighboring economic agents. Once an affordable decision has been computed, the *makeDecision* process – the last from the decisional phase – implements the decision.

Simulation Results

To get conclusive results, I used the same initial settings in all my simulations. As such, we generated 10 networks of 100,000 agents each and with different densities for each basic topology: mesh, small-world, random and scale-free. Thus, the networks obtained are presented in Table 5.3, and are created to simulate two cases for each topology: one where the producers are assigned randomly to the existing economic agents, and one where the algorithm within TrEcSim assigns the producer roles based on a probability that is proportional to the agent's degree; as a result, the higher the agent's degree, the higher its probability of becoming a producer.

Simulation Results for the Extended Model

By plotting the amassed payoff of both producers and middlemen for all of the 10 distinct densities (in each of the considered topologies) based on the random allocation of producers and averaging the values, I uncover the difficulty that producers encounter when trying to surpass the payoffs gained by the middlemen. In the 2D mesh (Figure 5.14), small-world(Figure 5.15) and random (Figure 5.16) networks the producers surpass the 50% threshold of the total (network-wide) payoff only after a significant number of iteration cycles (≈230). However, in the scale-free network (Figure 5.17) the total payoff earned by the producers does not surpass that of the middlemen. Therefore, when randomly assigning the producers throughout the network, increasing only the network density does not guarantee that the producers will get the larger portion of the total payoff. Such a result can be attributed to the presence of many highly connected middlemen which act as exchange hubs.

On the other hand, when preferentially assigning the roles of producers to the economic agents with the highest degrees, and plotting the share of total payoffs resulted from simulations, I obtain an evident transition from a topocratic network layout (*i.e.* where the middleman obtain most of the payoff) to a meritocratic one (*i.e.* where the producers obtain the largest share of the payoff), as presented in Figure 5.18 (mesh network), Figure 5.19 (small-world network), Figure 5.20 (random network), and Figure 5.21 (scale-free network). This scenario, however, is not completely accurate with respect to the real-world, as the layout of economic networks are not (necessarily) governed by such prerequisites. Furthermore, the simulation results contradict the conclusion that an increased network density alone automatically leads towards a meritocratic economic network.

By analyzing the evolution of payoff for both cases, I can clearly identify the unfair advantages of a topocratic environment over a meritocratic one (regardless of the underlying topology), as well as the presence of emergent behaviour among the economic agents. As such, each agent is adapting to its current environment and is investing in viable actions accordingly. This is most prominent in the scale-free network topology, where the presence of hubs limits the payoff of each agent.

Simulation Results for the Distribution Payoff

To obtain the distribution of income for each economic role, I used the previously obtained data and carefully analyzed the distributions between the two categories: producers and middlemen, obtaining the charts presented in Figures 5.22 - 5.25.

When producers are assigned randomly throughout the network, the averaged income for the producers is represented by positively skewed distribution; in other words, only a handful of agents benefit from an increased income compared to the rest of the producing agents. Conversely, the normalized share of total payoff for the middlemen closely resembles a normal (*i.e.* Gaussian) distribution, meaning that there are a lot more economic agents that gain (percentage-wise) the maximum payoff when comparing their payoff to both their producing counterparts, and the rest of the middlemen. When using the scale-free topology and assigning the producers randomly throughout the network I obtain a log-normal distribution. This is not only due to the presence of hubs in the network, but also because some hubs become producers.

Upon assigning the producers preferentially to high-degree agents in the mesh, small-world and random networks, I obtain a Gaussian distribution of payoff for the producers, as presented in Figures 5.26 - 5.28, while for the middlemen I obtain a positively skewed distribution. The fact that in both cases the distribution patterns alternate indicates that the physical location of economic agents plays an important role regarding the payoff of the agents, not only for the middlemen, but also for the producers acting as intermediaries. When the producers are assigned preferentially in scale free networks (Figure 5.29), the obtained charts clearly depict a fat-tailed, power-law distribution of payoff for producers, where only a handful of economic agents earn a lot (*i.e.* those in a favorable topological location), while the rest of them benefit from minimal payoff. Additionally, I obtain a log-normal distribution of payoff among the middlemen. Similarly, as in previous cases, the obtained results are a clear indication of emergent behaviour among the economic agents, adapting to the way they were assigned throughout the network.

Simulation Results Pertaining the Ergodicity in Economic Networks

In economic networks, it is important to analyze both static (*i.e.* population-level) payoff distribution, as well as dynamic (*i.e.* time distribution at individual-level) payoff distribution. Furthermore, it is also important to compare and correlate the two distributions according to the ergodic theory: if the two distributions are similar, then the economic system is ergodic and may be considered as being fair: the individual agent has good chances of improving its payoff if it takes the right decisions, but it can also be punished if it takes the wrong decisions. If the two distributions are substantially different, then the system is non-erogodic (or path-dependent) and considered as unfair. I investigate ergodicity further by analyzing the distribution of wealth based on the number of iterations (time) spent in a particular payoff category, as well as on the number of economic agents (space) in each payoff category and comparing their payoff distributions, in order to determine the fairness of economic exchange networks

After identifying each distributions, I obtained the data from the Tables 5.5 - 5.8. This fitting process was done by employing EasyFit, a software system for data fitting in dynamical systems. The numerical analysis shows that there is variable similarity between the payoff distributions in space (*i.e.* number of economic agents) and time (*i.e.* time interval), depending on the underlying topology. In general, the scale-free topology (Figures 5.36 and 5.37) is the only one with significant differences in distributions (between 22% and 80%), while the mesh (Figures 5.30 and 5.31), small-world (Figures 5.32 and 5.33) and random (Figures 5.34 and 5.35) networks have small to very small variation (between 2% and 27%), proving that our economic (extended) model is ergodic. An investigation of all simulation scenarios makes us conclude that assigning the producers preferentially rather than randomly does not change the ergodicity of the tested models, with the notable exception of the scale-free topology.

In order to obtain more, in-depth information regarding the ergodicity of economic networks, I also investigate if there are any economic agents, who based on the outcome of the chosen actions of investments, have reached a point where they can no longer afford to undertake – *i.e.* invest in – any additional action(s). As a result, the percentages of bankrupt economic agents are indeed within the margins of empirical observations for all the fundamental network topologies – albeit more pronounced in the case of the scale-free topology –, both for random and preferential agent assignment; this leads us to the conclusion that the obtained results are indeed good ergodicity indicators for economic exchange networks.

In the last chapter I reiterate over the relevant conclusions from this thesis. Complex networks are comprehensively studied due to their important applications in various fields, from Medicine and Sociology, to Architecture, Music, Engineering and Economy, as well as an amalgam of these fields. They can also be considered collaboration networks, because they represent actors (indirectly) connected through their common collaboration entity, be it movie acting or economic activity.

In the first half of this thesis, I presented a state-of-the-art analysis of MuSeNet, an emergent network formed solely by musical artists. Very similar to other complex networks, MuSeNet presents all of the usual properties: it is scale-free – meaning that artists 'connectivity distributions are in a power-law form – and has a high degree of centrality. With this study, the sociability of several networks were also highlighted via graph metrics: MuSeNet is a more closed network than the IMDB actors network – and other usual friendship networks –, due to the fact that music artists do not usually work with many others, since they rely on their on band and associated acts; additionally, links are also formed at a much slower rate, compared to the Facebook model. Motif-based analysis was also used to numerically express the characteristic aspects of collaboration networks, a technique that has recently been adopted from Systems Biology. In light of the study nominating Kevin Bacon as the most influential node in the IMDB actor network, I found Dave Grohl to be the "Kevin Bacon" of the music industry. Moreover, by analyzing MuSeNet from the perspective of important centralities, I reached the conclusion that similar to both the IMDB actors network and the Jazz musicians network, certain artists have higher centrality indices than the rest. As such, I found artists like Greg Errico to have the highest degree and Pagerank, and Alphonso Johnsonto have the highest eigenvector centrality. A second important empirical observation is the existence of a small single dominant community of nodes with very high eigenvector centrality. This is the community formed by artists who currently own a record studio. It is through their studios that most music is recorded and produced and it is because of this topocratic environment they managed to secure a thriving, central role in MuSeNet. With the broader perspective of social networks analysis in mind – namely to better understand and model complex networks –, the obtained results pave the way for a better understanding of the particular concepts of social collaboration, our society as a whole and the role we play in it, especially from a socio-economic point of view. For instance, we would often identify individuals who would benefit from a topological opportunity, though without any creative contribution to the network itself. Hence, we can not fully understand a meritocratic network without factoring in topocracy.

Conclusively, in the second half of this thesis, I presented a state-of-the-art economic simulator; TrEcSim was specifically created to simulate economic activities with high flexibility in terms of economic theories, agent models, and interaction assumptions. One such simulated economic model concludes that an increased economic interconnectivity fosters meritocracy, as opposed to topocracy, which is promoted in a poorly connected network. At first glance, the findings presented by Borondo *et al.* were also confirmed after simulating multiple network topologies. However, by analyzing the payoff distribution in a meritocratic environment based on agent roles, I showed that the topological placement of the economic agents directly influences the payoff distribution within the separate categories of producers and middleman. Indeed, the payoff distribution within the same economic agent category is strongly non-uniform, often following a fat tailed, power-law distribution. This observation holds true for both middlemen and producers acting as intermediaries. Nevertheless, I found that the distribution inside each agent role is not influenced by the network's topology, but instead by the placement strategy of agents within the net-work. Indeed, when producers are assigned randomly to topological positions, the payoff distribution within the producers category is fat-tailed (only a handful of producers benefit from an increased payoff), while the payoff of the middlemen category closely resembles a Gaussian distribution. Conversely, when the topological positions of producers are assigned preferentially, the payoff distributions of the two role categories reverse. Taken together, these results also highlight the emergent behaviour economic agents exhibit on a macroscopic scale, in order to further themselves in a specific economic community. By applying a new, state-of-the-art approach, I gained even more valuable insight regarding the distribution of the income for each agent-role in various economic exchange networks. In all cases, the evolution of the total payoffs closely followed the overall results already obtained by other means, and offered yet another argument in reference to the un-fair advantages of a topocratic economic network over a meritocratic one, regardless of the network's topology, as well as the presence of emergent behaviour among the economic agents. By analyzing both time and space payoff distribution fitting, I concluded that the payoff distribution generated with TrEcSim is indeed ergodic – *i.e.* fair – for all topologies except the scale-free topology; moreover, the ergodicity seems to be determined by the topology type alone, as agent-role assignment does not play a role in this case. A good ergodicity indicator is also the presence of a (limited) number of financially bankrupt economic agents, otherwise non-existent in the rockstar model.

Admittedly, the contributions brought with this thesis to the field of Network Science are significant. The tools and results presented leave room to further the research and experimentation I started many years ago; moreover, the work started also promotes new approaches and research in the field of Social Network Analysis.

Future Research Directions

Obtaining relevant results is the driving force for any researcher, even more so when the domain one is working in is still in its relative infancy. As such, in light of the recent advancements in the field of Social Network Analysis and the direction my studies have brought me in this field during my studies, I foresee the following contributions to have immediate effect on the research I started:

- Improved heuristic algorithm: additional effort will be put into the enhancement of TrEcSim's heuristic algorithm. Currently, the algorithm analyzes past decisions made by the economic agents and computes its outcome, however it will be improved as to allow the heuristic algorithm to use this information and create a buffer simulation; in other words, the algorithm will create a side-simulation based on several steps ahead and (probabilistically) analyze its results, greatly improving the accuracy of choices in the process.
- Genetic algorithm: converting from the existing heuristic algorithm to a genetic algorithm will improve TrEcSim in more than just a couple of ways. The implementation of said algorithm will allow users to find fit solutions in a short computational time, while the random mutation guarantees a wider range of solutions.
- Economic theories: improved implementation of the main schools of economic thought will greatly increase TrEcSim's applicability and usability in the field of social network analysis, herein including the economic domain as well; two such theories are ”the theory of marginality” and ”the labor theory of values”.
- Realism: by implementing new mechanisms into TrEcSim, it will undoubtedly improve the realism of the simulator further by taking into consideration several real-world factors like information asymmetry – which often occurs in transactions – as well the role of government involvement and regulatory red-taping. Adding cost (or other form of burden on the economic agent) in maintaining certain actions in place (*e.g.* links, new products, improved production, *etc.*) will also contribute to said realism.
- Interface: an even more customizable interface is necessary in order to interface the mentioned improvements with the user, as well as to allow flexibility during and after simulations.
- Extensive simulations: the simulations in this thesis represent just a few possible scenarios that we can realistically analyze using TrEcSim. Consequently, continuing the research and simulating real-world economical systems by using other possible configuration settings within TrEcSim can yield significant results, for instance pertaining product saturation and product shortage.

Selective Bibliography

1. Acemoğlu, D., Como, G., Fagnani, F., and Ozdaglar, A. (2013). Opinion fluctuations and disagreement in social networks.Mathematics of Operations Research, 38(1):1–27
2. Alberich, R., Miro-Julia, J., and Rosselló, F. (2002). Marvel universe looks almost like areal social network.arXiv preprint cond-mat/0202174
3. Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks.Re-views of modern physics, 74(1):47
4. Barina, G., Sicoe, C., Udrescu, M., and Vladutiu, M. (2017). Simulating trade in eco-nomic networks with trecsim. InEuropean Network Intelligence Conference, pages 169–185. Springer
5. Barina, G., Topirceanu, A., and Udrescu, M. (2014). Musenet: Natural patterns in the music artists industry. In 2014 IEEE 9th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI), pages 317–322. IEEE.
6. Barina, G., Udrescu, M., Barina, A., Topirceanu, A., and Vladutiu, M. (2019). Agent based simulations of payoff distribution in economic networks. Social Network Analysis and Mining, 9(1):63
7. Barina, G., Udrescu, M., Topirceanu, A., and Vladutiu, M. (2018). Simulating payoff distribution in networks of economic agents. In2018 IEEE/ACM International

Conference on Advances in Social Networks Analysis and Mining (ASONAM), pages 467–470. IEEE
8. Blume, L. E. and Durlauf, S. N. (2005).The economy as an evolving complex system, III: Current perspectives and future directions. Oxford University Press
9. Borgatti, S. P., Mehra, A., Brass, D. J., and Labianca, G. (2009). Network analysis in the social sciences.science, 323(5916):892–895
10. Borondo, J., Borondo, F., Rodriguez-Sickert, C., and Hidalgo, C. A. (2014). To each according to its degree: The meritocracy and topocracy of embedded markets. Scientific reports, 4:3784
11. Chen, D., Lü, L., Shang, M.-S., Zhang, Y.-C., and Zhou, T. (2012). Identifying influential nodes in complex networks. Physica a: Statistical mechanics and its applications,391(4):1777–1787
12. Dawid, H., Gemkow, S., Harting, P., van der Hoog, S., and Neugart, M. (2014). Agent-based macroeconomic modeling and policy analysis: the eurace@ unibi model
13. Duma, A. and Topirceanu, A. (2014). A network motif based approach for classifying online social networks. In Applied computational intelligence and informatics (SACI), 2014 IEEE 9th international symposium on, pages 311–315. IEEE
14. e Fonseca, B. d. P. F., Sampaio, R. B., de Araújo Fonseca, M. V., and Zicker, F. (2016). Co-authorship network analysis in health research: method and potential use. HealthResearch Policy and Systems, 14(1):34
15. Fass, C., Ginelli, M., and Turtle, B. (1996). Six degrees of kevin bacon. Plume Books
16. Franceschet, M. (2011). Collaboration in computer science: A network science approach. Journal of the American Society for Information Science and Technology,62(10):1992–2012
17. Freeman, L. (2004). The development of social network analysis. A Study in theSociology of Science, 1
18. Giaquinto, G., Bledsoe, C., and McGuirk, B. (2007). Influence and similarity between contemporary jazz artists, plus six degrees of kind of blue. Ph. D. Dissertation, Citeseer.
19. Kantarci, B. and Labatut, V. (2013). Classification of complex networks based on topological properties. In Cloud and Green Computing (CGC), 2013 Third InternationalConference on, pages 297–304. IEEE
20. Matthews, L. and Richard, P. (2017). Who is central to a social network? it depends on your centrality measure
21. Peters, O. (2019). The ergodicity problem in economics. Nature Physics, 15(12):1216–1221
22. Staudt, C., Schumm, A., Meyerhenke, H., Görke, R., and Wagner, D. (2012). Static and dynamic aspects of scientific collaboration networks. In Advances in Social NetworksAnalysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on, pages 522–526. IEEE
23. Topirceanu, A., Barina, G., and Udrescu, M. (2014a). Musenet: Collaboration in themusic artists industry. In2014 European Network Intelligence Conference, pages 89–94. IEEE
24. Topirceanu, A., Udrescu, M., and Vladutiu, M. (2013). Network fidelity: A metric toquantify the similarity and realism of complex networks. In Cloud and Green Computing(CGC), 2013 Third International Conference on, pages 289–296. IEEE
25. Topirceanu, A., Udrescu, M., Vladutiu, M., andMarculescu, R.(2016). Tolerance-based interaction: a new model targeting opinion formation and diffusion in social networks.PeerJ Computer Science, 2:e42
26. Wernicke, S. (2006). Efficient detection of network motifs.IEEE/ACM Transactions on Computational Biology and Bioinformatics, 3(4)