Computational Analysis of Emergent Behaviour in Collaboration Networks

Teză destinată obţinerii
titlului ştiinţific de doctor inginer
la
Universitatea Politehnica Timişoara
în domeniul Calculatoare şi Tehnologia Informaţiei
de către

ing. Gabriel Barina

Conducător ştiinţific: prof.univ.em.dr.ing. Mircea Vlăduţiu
Referenţi ştiinţifici: acad. Mircea Petrescu — Univ. Politehnica din București
prof.univ.dr.ing. Liviu Miclea — Univ. Tehnică din Cluj-Napoca
prof.univ.dr. Nicolae Bibu — Univ. de Vest din Timișoara

Ziua susţinerii tezei: 06.11.2020

BUPT

Dear *Sandra*,

I love you so...
You are my special little princess — I hope you know!
Hand in hand, I will watch you grow,
Watch you dress in white, from head to toe.
Until then though, come what may,
I'll shadow your movements night and day;
So, if you'll ever feel scared, or in the need,
I will step out of the shadows, to protect n' heed!

BUPT

Seriile Teze de doctorat ale UPT sunt:

1. Automatică
2. Chimie
3. Energetică
4. Ingineria Chimică
5. Inginerie Civilă
6. Inginerie Electrică
7. Inginerie Electronică şi Telecomunicaţii
8. Inginerie Industrială
9. Inginerie Mecanică
10. Ştiinţa Calculatoarelor
11. Ştiinţa şi Ingineria Materialelor
12. Ingineria sistemelor
13. Inginerie energetică
14. Calculatoare şi tehnologia informaţiei
15. Ingineria materialelor
16. Inginerie şi Management

Universitatea Politehnica Timişoara a iniţiat seriile de mai sus în scopul diseminării expertizei, cunoştinţelor şi rezultatelor cercetărilor întreprinse în cadrul Şcolii doctorale a universităţii. Seriile conţin, potrivit H.B.Ex.S Nr. 14 / 14.07.2006, tezele de doctorat susţinute în universitate începând cu 1 octombrie 2006.

Copyright © Editura Politehnica – Timişoara, 2020

Această publicaţie este supusă prevederilor legii dreptului de autor. Multiplicarea acestei publicaţii, în mod integral sau în parte, traducerea, tipărirea, reutilizarea ilustraţiilor, expunerea, radiodifuzarea, reproducerea pe microfilme sau în orice altă formă este permisă numai cu respectarea prevederilor Legii române a dreptului de autor în vigoare şi permisiunea pentru utilizare obţinută în scris din partea Universităţii Politehnica Timişoara. Toate încălcările acestor drepturi vor fi penalizate potrivit Legii române a drepturilor de autor.

România, 300159 Timişoara, Bd. Republicii 9,
Tel./fax 0256 403823
e-mail: editura@edipol.upt.ro

BUPT

Cuvânt înainte

Teza de doctorat a fost elaborată pe parcursul activităţii mele în cadrul Departamentului de Calculatoare şi Tehnologia Informaţiei al Universităţii Politehnica Timişoara.

Mulţumiri deosebite se aduc conducătorului de doctorat, domnului prof.univ.em.dr.ing. Mircea Vlăduţiu, pentru că m-a acceptat ca și doctorand și m-a îndrumat să depășesc cu succes această etapă din viața mea, cât şi domnului prof.univ.dr.ing. Mihai Udrescu-Milosav pentru ideile, contribuţiile şi sprijinul neîntrerupt oferit în perioada studiilor mele doctorale. Au dat dovadă de nenumărate ori de răbdare, motivare cât şi cunostinţe vaste, calităţi ce m-au ajutat în materializarea acestei teze de doctorat.

Doresc să le mulţumesc tuturor membrilor colectivului Advanced Computing Systems and Architectures (ACSA) — prof.univ.em.dr.ing. Mircea Vlăduţiu, prof.univ.dr.ing. Mihai Udrescu-Milosav, conf.dr.ing, Lucian Prodan, conf.dr.ing. Alexandru Topîrceanu, lect.dr.ing. Flavius Opriţoiu, lect.dr.ing. Alexandru Iovanovici — din cadrul departamentului pentru sfaturile lor. M-au acceptat încă din prima zi în echipa lor, partajând nu doar resursele fizice existente al colectivului, dar şi cunostinţele lor. Un lucru la fel de important este şi darul prieteniei lor. De asemena, doresc să-mi exprim recunostinţa atât față de conf.dr.ing. Alexandru Topîrceanu, cu care am scris mai multe articole ştiintifice și am obținut rezultate notabile, cât şi față de ing. Călin Sicoe pentru contribuţia lui la crearea simulatorului TrEcSim.

În final, doresc să mulţumesc familiei mele — atât părinţilor mei dragi Dorina şi Gheorghe, cât şi bunicilor Iuliana şi Francisc Martin, Iuliana Barina — pentru toată susţinerea lor oferită încă din primul an de licentă, respectiv iubitei mele soţii Alexandra, care m-a ajutat şi susţinut moral în perioada redactării articolelor şi a tezei curente.

Vă mulţumesc din suflet pentru tot!

Timişoara, 2020 ing. Gabriel Barina

BUPT

Barina, Gabriel

Computational Analysis of Emergent Behaviour in Collaboration Networks

Teze de doctorat ale UPT, Seria X, Nr. XX, Editura Politehnica, 2020, 120 pagini, 62 figuri, 15 tabele.
ISSN: XXXXXX-XXXXXXX
ISSN-L: XXXXXX-XXXXXXX
ISBN: XXXXXX-XXXXXXX

Cuvinte cheie: complex networks, network comparison, music artists, economic agents, simulation, collaboration, sociability, influence, emergence, payoff distribution, fairness

Rezumat: Domeniul interdisciplinar *Network Science* se ocupă cu analiza reţelor sociale, facilitând o mai bună inţelegere a societăţii noastre, atât din punct de vedere fizic, cât şi social. Astfel, odată cu studiul comportamental al agenţilor (social, muzical, economic, *etc.*), putem obţine o mai bună inţelegere a proceselor de interacţiune, colaborare şi influenţare.

Prezenta teză se bazează pe cercetarea proceselor menţionate cu ajutorul sistemelor de calcul. Astfel, folosind calculatorul ca instrument de bază în analiza reţelor, propun simularea modelelor complexe emergente într-o manieră dinamică, pentru o mai bună înțelegere a dinamicităţii sociale înconjurătoare, a comportamentului uman și a modului in care acestea se influențează reciproc.

În prima parte a tezei analizez reţeaua emergentă formată din muzicieni. Prin aplicarea unor metode tradiţionale de analiză ale reţelelor complexe, respectiv a elementelor autentice expuse odată cu teza de faţă — aplicarea *network motif*-urilor pentru a extrage proprietăţile topologice ale reţelelor sociale, generarea metricii de *sociability* ($\mathcal{S}$-metric) pentru diferenţierea reţelelor similare —, analizez reţeaua muzicienilor din punct de vedere colaborativ (*i.e.* dintre muzicieni), dar şi din punct de vedere economic (*i.e.* activitatea lor de a produce conţinut muzical).

În cea de-a doua parte a tezei prezint un simulator socio-economic original, bazat atât pe observaţii empirice, cât și pe modele economice inovatoare. Având capabilităţi precise de simulare, este folosit la analiza proceselor de colaborare și interacţiune, cât și la distribuţia emergentă de venit la scară macroscopică, cu reguli specifice bine-definite în prealabil la scară microscopică. Elementele de originalitate constau atât în folosirea unei abordări euristice pentru determinarea acţiunilor agenţilor economici, cât și în implementarea teoremelor economice recunoscute in mediul economic.

Lucrarea de faţă se evidenţiază prin aplicarea conceptelor din domeniul Netwok Science atât pe date empirice, cât și pe scenarii simulate, pentru o mai bună înţelegere (economică și comportamentală) a reţelei noastre înconjurătoare și a proceselor ce stau la baza acesteia.

BUPT

Acknowledgements

I would like to express my sincere gratitude to my advisor, prof.em. Mircea Vlăduţiu, PhD for believing, encouraging and guiding me through this important phase of my life, as well as to prof. Mihai Udrescu-Milosav, PhD for the ideas, contributions and unremitting support during my quest in obtaining the title of *Philosophiae Doctor*. They proved — and shared — their patience, motivation and vast knowledge on more than one occasion, aiding me in my research and in the writing of this thesis.

Besides my advisors, I would also like to thank all members of the Advanced Computing Systems and Architectures (ACSA) research group — prof.em. Mircea Vlăduţiu, PhD; prof. Mihai Udrescu-Milosav, PhD; assoc.prof. Lucian Prodan, PhD; assoc.prof. Alexandru Topîrceanu, PhD; lect. Flavius Opriţoiu, PhD; lect. Alexandru Iovanovici, PhD — for their insightful suggestions; they accepted me in their team as an intern, giving me access not only to their physical resources, but also to their knowledge-base and their friendship.

I would also like to express my sincere recognition to both assoc.prof. Alexandru Topîrceanu, PhD, with whom I wrote several scientific articles and obtained notable results, as well as to Călin Sicoe, who played a pivotal part in designing and creating TrEcSim with its current interface and performance.

Last but not least, I would like to thank my family, my dearest parents Dorina and Gheorghe, as well as grandparents — Iuliana and Francisc Martin, together with Iuliana Barina — for supporting me and aiding me get to where I am now, ever since the first year of my bachelor's studies. Similarly, I would like to thank my beloved wife Alexandra for supporting me morally during this period and aiding me in the elaboration of this thesis.

BUPT

Abstract

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 (emergent) models in a dynamic fashion, as well as to uncover crucial aspects regarding social and economic collaboration and emergent behaviour. As such, in the first half of this thesis I analyze the emergent network of musicians, entitled *MuSeNet*. By employing both traditional network analysis methodologies to the resulting network — *e.g.* centrality analysis, unsupervised machine learning, *etc.* —, as well as state-of-the-art methodologies presented in this thesis — *i.e.* applying *network motifs* in order to extract important topological properties of the underlying graph, introducing the $\mathcal{S}$-metric into literature, in order to determine the sociability of several networks —, I analyze MuSeNet, the *Musical Society Network*, from the context of collaboration. Additionally, due to the musicians' underlying activity of creating content, MuSeNet also brings relevance from an economic point of view.

In the second part of this thesis, I present a novel socio-economic simulator, inspired by empirical observations and state-of-the-art economic models. Capable of simulating complex scenarios, the *Trade and Economic Simulator (TrEcSim)* is able to use any network topology, in order to accurately simulate collaboration and interaction between economic agents, as well as the emergent payoff distribution on a macroscopic scale, with rules of interaction defined at a microscopic scale. Introducing a novel heuristic approach to drive the behaviour of agents according to theories pertaining to main schools of economic thought, TrEcSim is indeed a valuable tool for simulating the dynamics of trade in economic networks.

Admittedly, the contributions brought with this thesis to the field of Network Science are significant. By applying state-of-the-art concepts on both empirical data, as well as on simulated scenarios, I obtain relevant information pertaining our own society — both behavioural and economical —, and the processes which take part in it. Additionally, the tools and results presented leave room to further the research I started many years ago, closely promoting new approaches in the field of Social Network Analysis as well.

Keywords: *complex networks, network comparison, music artists, economic agents, simulation, collaboration, sociability, influence, emergence, payoff distribution, fairness*

Table of Contents

BUPT

BUPT

List of Figures

BUPT

BUPT

BUPT

BUPT

BUPT

List of Tables

1. Introduction

A major trend in Network Science is the study and understanding of the underlying social models of our society, its dynamics, human behaviours, relationships and connections [2, 69, 85, 91, 97, 106, 116, 201]. Even though Network Science is considered as a stand-alone science, it contains elements from exact sciences: Computer Science, Mathematics and Physics [198], as presented in Figure 1.1. As such, combining graph theory with statistical mechanics, social structures, data mining and information visualization [162], the benefit of understanding these complex processes is of paramount importance for researchers in fields like Computer Science [77], Biology [166], Psychology, Criminology [89], Philosophy, Economics, Marketing [88], Finances and even Warfare [69, 91].

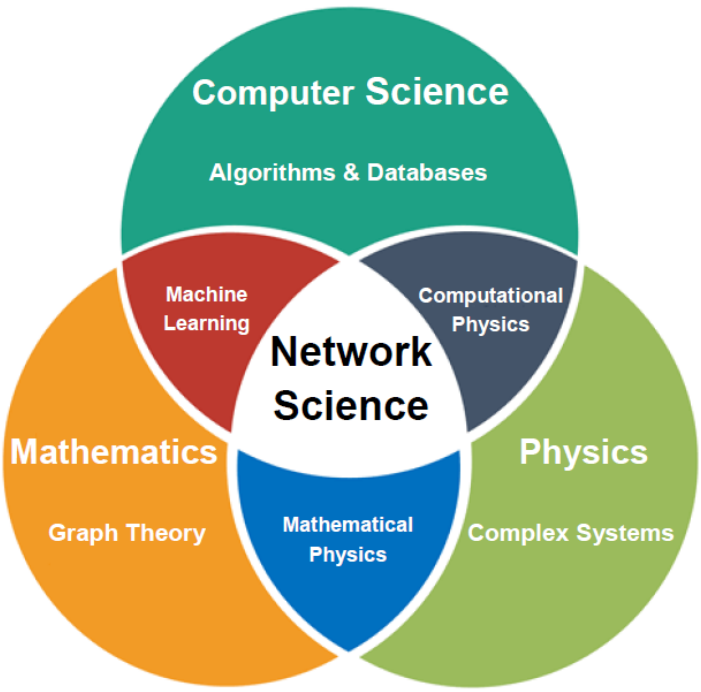


Figure 1.1 Overview of the overlapping fields of Computer Science (Algorithms and Databases), Mathematics (Graph Theory) and Physics (Complex Systems) in the resulting Network Science.

In the field of Computer Science, for instance, Network Science can be employed in applications which entail considerable amount of data and complexity [77, 150]. Indeed, the ever-increasing use of social networks, such as Facebook, Google Plus, Twitter, *etc.* is producing huge volume of data, which can only be analyzed by using the tools and methodologies of Network Science [9].

In the fields of Finance and Marketing, understanding the existing markets and consumers can greatly affect products and their properties (*e.g.* type, quantity, quality, price, *etc.*). By making use of available techniques and models in order to understand the strengths and weaknesses of individuals — as well as groups — [69], we can gain valuable information pertaining customer behaviour, interaction and influence [143].

The field of Politics uses techniques from Network Science in order to study the relations between political parties, as well as their projected influence on voters. Whether an agreement, a cooperative or collaborative consensus is longed-for, political parties generally focus on the overall public rather than on any given individual [100]; this behaviour is not unlike computers connected together within a network, in which the correct packet delivery — and overall throughput — is more important than the performance of an individual computer.

Warfare has always placed emphasis on counter-intelligence, in order to stop enemy propaganda, all the while greatly affecting the enemy's morale. Such a psychological warfare was successfully employed through several wars, by employing airborne leaflet propaganda. Military forces of various nations have used aircraft to drop propaganda leaflets, in order to influence the behaviour of both combatants and non-combatants alike, in enemy-controlled territories; similarly, certain humanitarian missions, with the aid of leaflet propaganda, can turn the populace against their heads of state, or prepare them for the arrival of enemy combatants [103, 144, 170].

The field of Social Science itself encapsulates a large number of other sciences stemming from the study of human behaviour. By using advanced analytics, we are able to identify different types of relationships between social actors; social leaders who influence the behaviour of others in the network; the mechanism of collaboration *etc.* Thus, the results obtained by using the tools and methodologies of Social Science are successfully applied in our everyday life.

BUPT

1.1 Thesis Domain

Network Science has received an increased interest during the last couple of decades, due to the tremendous development in Big Data [99, 177]. Regardless of the underlying topological model (*e.g.* friendship networks, protein interaction models, citation networks, musicians and actors networks, economic networks, political structures, recipe networks *etc.* [7, 69, 78, 108, 191, 206]), empirical observations pertaining our own surrounding world show the same overall properties, be it of natural or synthetic origin [73]. As such, 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 of5 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 [98].

By using the tools and methodologies available in Social Network Analysis, researchers can extract relevant patterns from said models, by analyzing the properties of their respective nodes and edges (*e.g.* type, direction, weight, *etc.*) [73, 78]. That being said, there is no single methodology of identifying similarities and dissimilarities within network, and instead the analysis is often based on numerical comparison, statistical analysis or empirical observation. As such, the analysis done in this thesis is based only on the topological properties of the underlying graph.

Computational Social Network Analysis is still in its infancy; even so, an ever-growing interest is shown from the scientific community. The need to better understand social processes has created new possibilities of collaboration between the domains of Computer Science and Social Science [91]. By employing computers at their full potential, researchers can analyze complex mathematical models more accurately and at a much faster rate, before validating the results with empirical data [95]. Furthermore, with the aid of computer simulation, researchers can identify and analyze emergence in complex systems. When describing collective behaviour, emergence occurs when a given entity is observed to have properties its individual parts do not have on their own, or when a certain behaviour, interaction, dependency or relationship arises at a macroscopic scale according to simple rules of interaction defined — or observed — on a microscopic scale [30, 102, 104, 119, 123, 140, 160]. In today's Computer Science, emergence represents one of the most important challenge for the engineering and analysis of complex systems.

1.2 Motivation

Following the presented context of Social Network Analysis and its application in a multitude of scientific fields, this thesis is based on using the tools and methodologies available in Computer Science in order to model and analyze complex networks, as well as the underlying behaviour of its agents.

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 [1], or a random graph topology [39]. 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.

1.3 Contributions

This thesis builds upon the emergent and collaborative behaviour of (economic) agents when grouped together in communities and exposed to certain conditions or restrictions. To this end, I analyze both a natural (real-world network), as well as synthetic networks created via a simulation application. As such, the first part of this thesis deals with MuSeNet [24, 193], a real-world collaboration network of musicians. 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 [86, 90], 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 [110, 159]. Furthermore, I identify the overlapping of genres, detect influential agents, as well determine the "Kevin Bacon" [75, 209] of the music industry.
- Motif analysis: I apply a novel approach of analyzing and differentiating networks by identifying existing motifs within these networks.
- $\mathcal{S}$-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.

Considering that creating content is a form of economic activity, the second half of this thesis deals with TrEcSim, a state-of-the-art trade and economic simulation application [23, 25, 26], with the following features:

- Driven by heuristics: with additional improvement to the mechanism to model the behaviour of economic agents [26], by using the tolerance-based interaction model [1, 196] 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, dynamic creation and 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.

Admittedly, the contributions — and associated results — brought with this thesis to the scientific community are recognized at an international level through the publications submitted, accepted and presented at international conferences or journals, relevant to the domain of Computer Science, as presented in section 6.1.

2. Theoretical Background

Underpinned by empirical studies in many real-world systems (*e.g.* social networks, economical networks, communication networks, *etc.*), complex networks have gained significant research interest, due to their applicability in many scientific and social fields [196, 65, 195], including Computer Science, Biology, Sociology [193], and Economy [69]. Over the last couple of decades however, complex networks have received an increased boost of interest due to the tremendous development in Big Data techniques and technologies, including Network Science [99]. Based on the context which they model, complex networks can be classified into four major types, namely:

- Biological networks: metabolic networks, transcription regulatory networks, protein-protein interaction networks, protein structure networks, neural networks, ecological networks, natural food chains, *etc.* [21, 69, 206].
- Social networks: friendship networks, citation networks, voter networks, world markets, political structures, actor's network, musician's network *etc.* [188, 206, 165].
- Technological networks: computer networks, electrical circuits, road networks, *etc.* [21].
- Semantic networks: word-net, recipe networks *etc.* [147, 191].

One of the fundamental properties of all networks is the presence of network-motifs. Initially introduced by Milo *et. al.* [150], they represent recurring and statistically significant sub-graphs or patterns. Each of these sub-graphs, defined by the interaction-model between its various nodes, may represent a framework in which particular functions are achieved in an efficient manner. In today's research motifs are considered to have a notable importance due to their underlying functional properties [139]. As such, taking into account their capacity to uncover structural design principles of complex networks, motifs became a popular approach in Network Science when analyzing the functional abilities of a given network. Even though detecting network motifs is surprisingly not an easy endeavor [65], any of the previously mentioned network-types can be analyzed using the mentioned motif-approach: in biological networks, motifs can be associated with functional roles of transcription regulation networks which control the expression of genes [13], while in transcription networks, motifs serve as basic building blocks. Yet another example is the understanding of how some cellular components are preserved across species but others evolve at a much faster rate [215]. A study was published fairly recently by Wang *et. al.* [205], in which the authors use a motif-based approach, instead of the traditional centrality measures, in order to detect important nodes in specific networks – not unlike the pursued method in MuSeNet. For this, the networks need to be grouped from a topological point of view, instead from a conceptual one. To this end, we obtain the following networks: regular networks, random networks, small-world networks and scale-free networks [206]. Regular and random networks represent the

BUPT

basics of complex networks [51, 71]. They can be observed and analyzed by applying by any of the fundamental properties of complex networks: average path length, clustering coefficient and degree distribution [188, 206]. A similar importance convey both the small-world and scale-free network topologies [210, 22]; since their introduction to literature, most of the new network-types fall into one of these categories, representing the object of intense studies.

BUPT

2.1 Social Networks

A social network is a construction formed by individuals (*e.g.* actors, musicians, economic agents *etc.*) with bidirectional connections (*i.e.* relationships, friendships) between them, effectively resembling a real-world structure of our own society. Social networks derive from complex networks, together with which they form part of the nascent field of Network Science [38, 69], ever since their first introduction to literature in the 1970's [151]. Based on graph theory and network theory, empirical observations of real-world networks and sociology, their main purpose is to model the various relationships from our own society [207].

Only recently has the field of social networks started gaining attention from researchers around the world. This is due to the fact that social networks (*e.g.* Facebook, Twitter, Google Plus *etc.*) gained attention not only in Social Science, but in Network Science as well. Furthermore, not only do these networks offer valuable data for researchers, but the results of these studies attract more and more people into this field of science.

Social networks provide valuable information on how various relationships evolve and how they interact among themselves. The main characteristics of social networks are their network topology and the agent interaction model. An interesting property is that specific patterns present at a small scale with a certain group of agents can also be found on a much larger scale. This is due to the fact that social networks are emergent and self-organizing. On the other hand, the larger the network size, the more difficult it is to analyze that given network. To this end, studies are done on relevant groups, with clearly defined properties, as their results can also be mapped on the larger, extended network.

BUPT

2.2 Network Metrics

When studying social networks, we make use of several metrics, specific to these networks. To this end, the ones used throughout the course of this thesis will be presented, in short, in the following subsections.

2.2.1 Centrality

In terms of graph theory and network analysis, the centrality of a vertex measures its relative importance within a graph. Some of the practical applications include how influential a person is within a social network, how important a room is within a building and how well-used a road is within an urban network [151]. There are four main measures of centrality, which I have extensively used in my thesis in order to analyze networks in terms of structural properties and similarity: degree, betweenness, closeness, and eigenvector.

Degree Centrality

The degree centrality is the simplest centrality property, which is defined as the number of edges a particular node has [151]. The degree distribution $P<k>$ is an important aspect when studying empirical networks as they usually possess a uniform, normal or power-law degree distribution.

Closeness Centrality

An important node centrality metric in networks is the closeness centrality, defined by the length of their shortest path L [80, 164, 207]. As such, the more central a node n_i is, the lower its total distance to all other nodes [172], as presented in Figure 2.1.

Figure 2.1 Individuals who are highly connected to others within their own cluster will have a high closeness centrality [141].

This metric can only be applied to networks with disconnected components, since the distance between nodes in disconnected components of a network is infinite [80, 155, 164, 207]. The closeness of a given node n_i can be defined with the following equation:

$$C(i) = \frac{1}{\sum_j d(i,j)} \tag{2.1}$$

Betweenness Centrality

In network theory, the betweenness is a centrality metric of a particular by quantifying the control of a node over the communication between other nodes [79, 87, 203]. It is equal to the number of instances a node acts as bridge along the shortest path from one node to another — as presented in Figure 2.2.

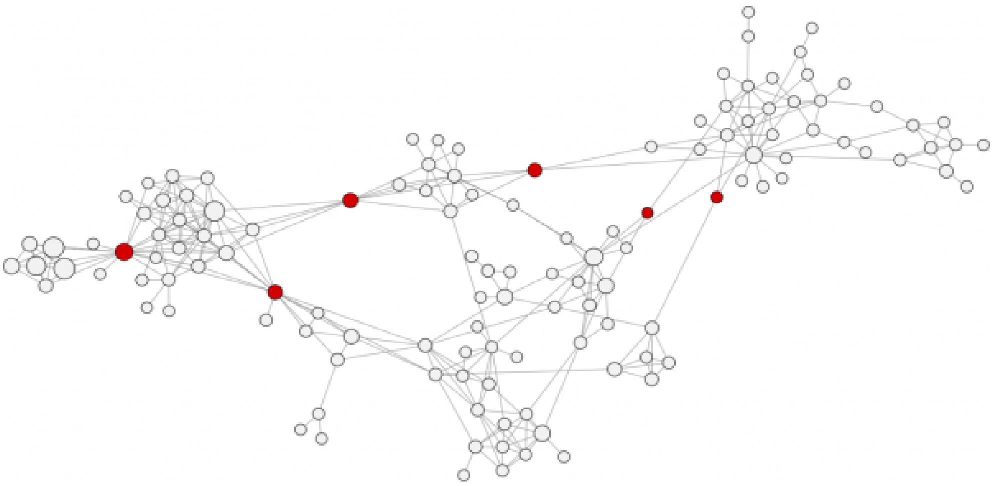


Figure 2.2 Individuals who act as a bridge between clusters in the network have a high betweenness centrality [141].

The betweenness centrality is a more useful measure of both the weight (global meaning) and importance (local meaning) of a node n_i; its betweenness and can be defined, for all pairs of nodes n_a and n_b, as follows:

$$Btw(i) = \sum_{a \neq i \neq j} \frac{\sigma_{ab}(i)}{\sigma_{ab}} \tag{2.2}$$

Eigenvector Centrality

The eigenvector centrality is a measure of the influence of a particular node in a network [18, 157]. It is calculated by evaluating how well an individual is connected compared to the parts of the network with the greatest connectivity. Individuals with a high eigenvector score have many connections, with the same being true for their connections, as seen in Figure 2.3. Google's PageRank is a variant of the Eigenvector centrality measure [18].

2.2.2 Degree Distribution

In the study of graphs and networks the simplest and perhaps the most important characteristic of any particular node is its degree. The degree of the node n_i is the number of connections it has to other nodes in the network and is denoted by $deg(n_i)$. Depending on the type of graph, there can be an in-degree $(deg-(n_i))$ and an out-degree $(deg+(n_i))$, for incoming respectively outgoing connections. Undirected graphs, like social networks, only have the degree characteristic [7].

The diameter of the network is the longest distance among all distances between any pair of nodes in the network [206]. Nodes with a (much) higher degree are called hubs, as they tend to facilitate communication for distant nodes and in the end, almost always become even more connected. The degree distribution $P(k)$ describes the probability that a

Figure 2.3 Highly connected individuals within a highly inter-connected cluster present a high eigenvector centrality [141].

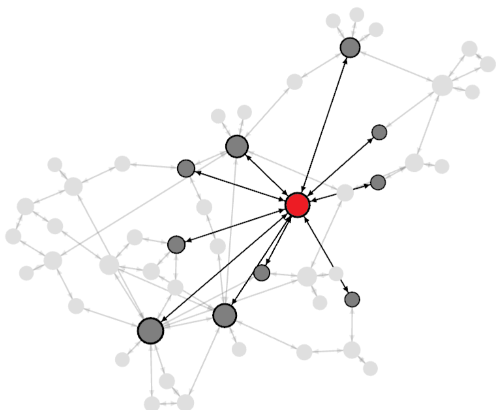


Figure 2.4 Visual representation of the degree distribution [10].

randomly selected node has the mentioned degree of k and is defined as the ratio between the total number of nodes with the degree of k (N_k) and the total number of nodes, N [206]:

$$P(k) = \frac{N_k}{N} \tag{2.3}$$

A regular network has a simple degree sequence due to the fact that most existing nodes have the exact same number of edges, and as such, a plot of the degree distribution contains a single sharp spike (*i.e.* delta distribution), much like the one Figure 2.5. As a result, increasing the randomness in the network will also increase the shape of this peak, while a fully random network would have a Poisson-like distribution of degrees [174]; in the past few years, however, significant empirical results showed that for most large-scale real networks the degree distribution deviates significantly from this Poisson distribution: individual nodes are more connected, much like in a scale-free network, thus following a power-law distribution [174, 175, 206]:

$$P(k) \approx k^{-\gamma} \tag{2.4}$$

where γ is an empirically observed value, $\gamma \in [2, 3]$ for a power-law specific to social networks, as presented in Figure 2.5. In statistics, a power law is a functional relationship between two

BUPT

quantities, where one quantity varies as a power of another. This power-law distribution falls off more gradually than an exponential one, allowing for a few nodes of very large degree to exist. As this form of distribution is not subject to network scale, it is characteristic for scale-free networks [174, 175, 206] and has attracted particular attention for their structural and dynamical properties [174].

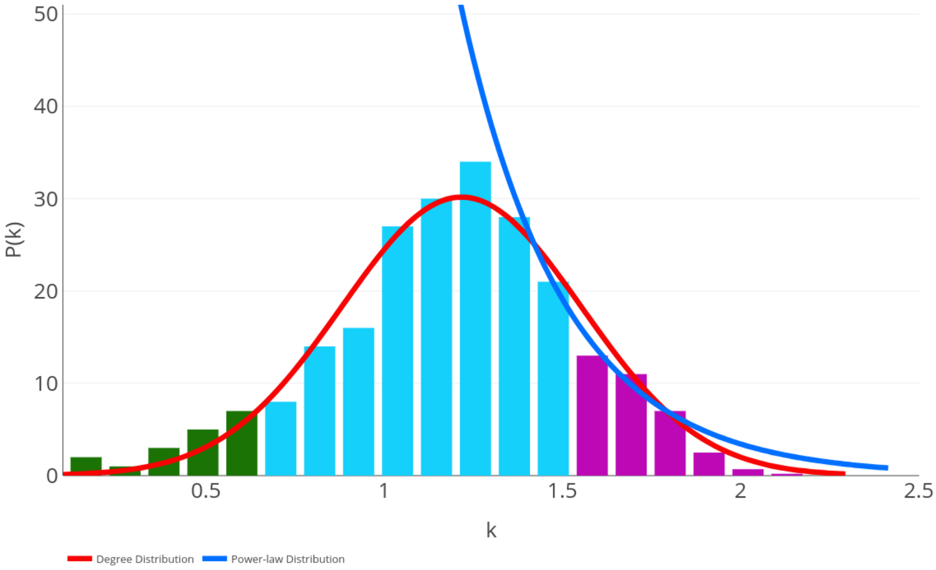


Figure 2.5 Degree distribution in a traditional random network (red) compared to a power-law distribution in a real-world network (blue). The lowly linked nodes are represented with green, the average linked with cyan while the highly linked nodes are represented with magenta.

2.2.3 Average Path Length

The average path length, L, is one of the three most robust measuring concept in network topology, that is defined being the distance d_{ij} between two nodes (in this case i and j), which represents the shortest path between these pairs of network nodes; hence the diameter D of a network is defined to be the maximal distance among all distances between any pair of nodes in the network. In this case the average path length of the network is defined as the mean distance between two nodes, averaged over all pairs of nodes [206], as expressed by equation 2.5:

$$L = \frac{2}{N(N-1)} \sum_{i \neq j} d(n_i, n_j) \tag{2.5}$$

where N is the size of the given graph and v are the graph's vertices. For instance, in a friendship network, L is the average number of friends existing in the shortest chain connecting two persons in the network [174, 206]; in a road network, L is the average number of roads a driver has to change in order to get from one city to any other city, *etc.*

A remarkable property is the fact that the average path length of most real complex networks is relatively small, even in those cases where these kinds of networks have a lot less edges than a typical globally coupled network with an equal number of nodes. This

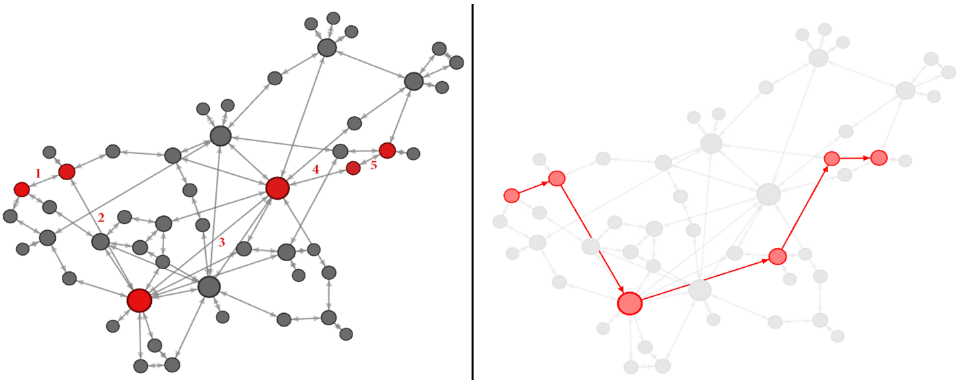


Figure 2.6 Visual representation of the average path length [10].

smallness led to the concept of the small-world effect [210, 188]; as a result, most models of real networks are created with this taken into consideration.

One of the first models which proved that the average path length depends mostly on the system size was the random network model. It was followed by the Watts and Strogatz model, while later there were the scale-free networks starting with the Barabási-Albert model. All these models had one thing in common: they all predicted that the average path length changes proportionally to $\log_N$ - where N is the number of nodes in the network [7]. The real-life applications of this concept are numerous. In a network like the World Wide Web, a short average path length facilitates the quick transfer of information and reduces costs. The effectiveness of mass transfer in a metabolic network can be evaluated by studying its average path length; the same goes for a power grid network.

2.2.4 Clustering Coefficient

In study of graph theory, the clustering coefficient, C, is a measure of the degree to which nodes in a graph tend to cluster together. More precisely, the clustering coefficient can be defined as the average fraction of pairs of neighbors of the node n_i, as seen in Figure 2.7a.

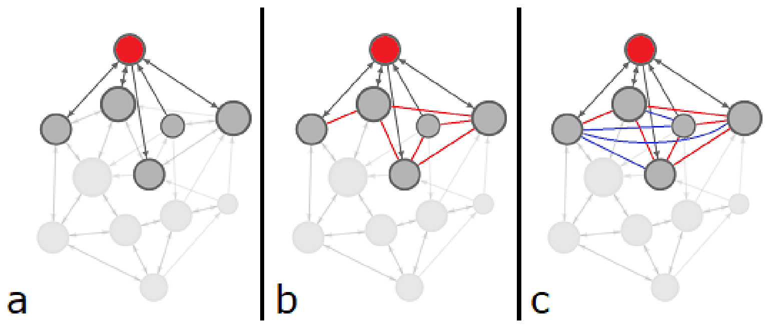


Figure 2.7 Visual representation of the clustering coefficient [10].

Now, assuming that a node n_i in the network has k edges all connecting to n other nodes, n_i is automatically the neighbor of all these nodes [174], case represented in Figure 2.7b. The clustering coefficient of node n_i is thus defined as the ratio between the number E edges that actually exist among these n nodes and the total possible number $(n_i(n_i - 1)/2)$, represented in Figure 2.7c and described using the following expression:

$$C = 2\frac{E}{n_i(n_i - 1)} \tag{2.6}$$

To this end, we can easily conclude that the maximum clustering coefficient of a network is 1. Such a network is a fully connected graph with point-to-point connections [174], while a completely random network has a clustering coefficient approximately equal to $1/N$, N being the total number of nodes from the network. This, however, is a small value compared to the clustering coefficient of other observable networks, of which can be defined using the following expression:

$$\frac{1}{N} \ll C < 1 \tag{2.7}$$

2.2.5 Modularity

Modularity (Q) is one of the measures which helps determines the structure of a network. It was intended as a means to measure the strength of division of a particular network into modules, also known as communities [157, 202], with substantial importance in understanding the dynamics of various networks [156].

If a network is characterized by a number of individual connected nodes which signify a certain degree of interaction between them, communities are defined as groups of densely organized nodes that are only sparsely connected with the rest of the network. To this end, it may prove crucial to identify distinct communities, as these may present particular properties, such as: node degree, clustering coefficient, betweenness, centrality, *etc.*, compared to that of the network as a whole [157].

There are several means for calculating the modularity of a network [156], but the most popular version of the concept is the randomization of the edges in such a way, that the degree of each node is left unchanged and applying the following expression:

$$Q = \sum_{\zeta_i \in \zeta_{set}} \left[\frac{E_{\zeta_{i_{in}}}}{E} - \left(\frac{2E_{\zeta_{i_{in}}} + E_{\zeta_{i_{out}}}}{2E} \right)^2 \right] \tag{2.8}$$

where ζ_{set} is the set of all the communities, ζ_i is a specific community in ζ_{set}, $E_{\zeta_{i_{in}}}$ is the number of edges between various nodes in the community ζ_i, $E_{\zeta_{i_{out}}}$ is the number links which exit the community ζ_i to the outside of it, and E is the total number of edges in the network. Applying equation (2.6) to a network, we obtain a value for its modularity equal to $Q \in [0, 1]$. Thus, the closer the Q is to 1, the stronger the community structure [49].

2.2.6 Network Motifs

One important common property of all these networks is that they can be represented as graphs, as well as sub-graphs called (network) motifs [150]. 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. As already mentioned in this thesis, network motifs may offer a deep insight into the network's functional abilities, yet their detection is still computationally challenging [65]. 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 3 — as illustrated in Figure 2.8 —, 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 [13].

Figure 2.8 All existing motifs of size 3 in a directed graph. The code of each motif corresponds to the decimal value of its serialized adjacency matrix [193].

2.2.7 Metric Fidelity

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 (φ) [194] and to compare them among each other based on individual metric measurements. Tailored to express the similarity between any two generic vectors — in a weighted or unweighted context and depending on the nature of the comparison —, it can offer insight on network model resemblance or synthetic model realism compared to a real world network. The metric fidelity can be expressed with the following equation:

$$\varphi^j = \begin{cases} \frac{1}{n}\sum_i \frac{m_i}{2m_i - m_i^j}, & \text{if } m_i^j < m_i \\ \frac{1}{n}\sum_i \frac{m_i}{m_i^j}, & \text{if } m_i^j \geq m_i \end{cases} \tag{2.9}$$

where j is the index of empirical distribution model being compared to the reference, $i = \{1, 2, \ldots, n\}$ is the index of the motif used to compare the respective models, while n represents the number of motifs which are common. The closer the φ metric is to 1, the more similar the models are. The measurements on the reference model are m_i respectively m_i^j on the model being compared [65].

2.2.8 S-metric

In order to better compare two or more networks with each other, I modeled the S-metric, with which I can express the so-called sociability of any given complex network [193]. As a state-of-the-art metric introduced into literature with the advent of this thesis, it takes into consideration the basic graph metrics — *i.e.* average degree (AD), average path length (L), average clustering coefficient (C), modularity (Mod), graph edge density (Dns) and graph diameter (Dmt) — when comparing a given network to a reference model. In order to obtain an optimal expression, I first normalize the offset from the reference value of each metric, after which I either add (direct proportional) or subtract (indirect proportional) the resulting normalized values. Thus, sociability is defined by the following equation:

$$S_i^j = \sum_{i=1}^{6} \frac{k_i(m_i - m_j)}{m_j} \tag{2.10}$$

where S_i^j, the sociability of network i towards reference model j, is the sum of the six normalized metrics: average degree ($k_1 = +1$), average path length ($k_2 = -1$), average clustering coefficient ($k_3 = +1$), modularity ($k_4 = -1$), density ($k_5 = +1$) and network diameter ($k_6 = -1$). The signs (+/-) of the metrics reflect if the particular metric is direct (AD, C, Dns) of indirect (L, Mod, Dmt) proportional to a more sociable network. Due to the fact that there are an equal number of such elements, equation 2.10 can be simplified as follows:

$$S_i^j = \sum_{i=1}^{6} \frac{k_i(m_i)}{m_j} \tag{2.11}$$

2.3 Topologies

Knowing the basic properties of any given complex network, such as the average path length L, clustering coefficient C and the degree distribution $P(k)$ is just the first step in comprehending its structure; the next step is to elaborate a mathematical model which allows for this particular network's analysis [206], taking into consideration this model's layout and connection patterns of various elements. This model is the network's topology and, as with any network-type, a social network too can be described by it. Network topologies can be grouped into two main classes [206], namely:

- Basic topologies: the most wide-spread of network topologies, herein including the mesh, star, bus and ring topologies.
- Social network-specific topologies: topologies based on one or more basic topology, but with a more complex structures capable of representing real-world relations, much like the random, small-world, scale-free, *etc.* network-topologies.

The analysis of these topologies is done with the help of graph theory, derived from mathematical theory and used to describe relations between objects (*i.e.* nodes, agents). In order to obtain relevant results for this thesis, a number of social network-specific networks have been studied.

2.3.1 Regular Mesh

One of the most common network topology, due to the its tightly interconnected layout. Although its model captures both the small-world and large-clustering properties of many real networks, it is easy to observe its limitations: a globally coupled network with N nodes has $N(N-1)/2$ edges, while most large-scale networks are not fully connected: their edge-count is usually of order N rather than $2N$ [206].

The mesh topology is known to have a uniform degree distribution, a high average path length and a low clustering coefficient, being used as a basis for other, more complex network topologies.

2.3.2 Random

Random networks (also known as the Erdős-Rényi model) consist of nodes with random connections among them [36, 66].

In mathematics, a random graph is the general term used to describe probability distributions over graphs or random processes which generates them. The theories which govern these networks lie at the intersection between graph theory and probability theory, with their main goal being the possibility to determine at what connection probability ρ a particular property of a graph will most likely arise [36, 66, 206]. From a mathematical point of view, random graphs are used to determine certain properties of other, ordinary (basic) graphs. Its practical applications can be found in all areas in which complex networks need to be modeled.

Random networks are formed by using any of the basic network topology — though usually a mesh topology is used — with a set of N vertices and randomly adding successive edges between them [71]. The aim of the studying random networks is to determine at what stage a particular property of the graph is likely to arise, since it can happen quite suddenly. Erdős-Rényi proved that, if the probability ρ is greater than the estimated value in equation 2.12, then almost every random graph is connected [71, 206]:

$$\rho \approx \frac{\ln N}{N} \tag{2.12}$$

where N is the total number of nodes in the network. Other important attributes of random networks have also been observed and reported in literature [56, 71, 82]: the average path length decreases dramatically when long-range links are inserted into the network. On the other hand, upon a close-up inspection of these networks, the clustering coefficient remains low, as there is no rule tying local nodes together in a cluster. Its practical applications can be found in all areas in which complex networks need to be modeled. Such a network can be seen in Figure 2.9.

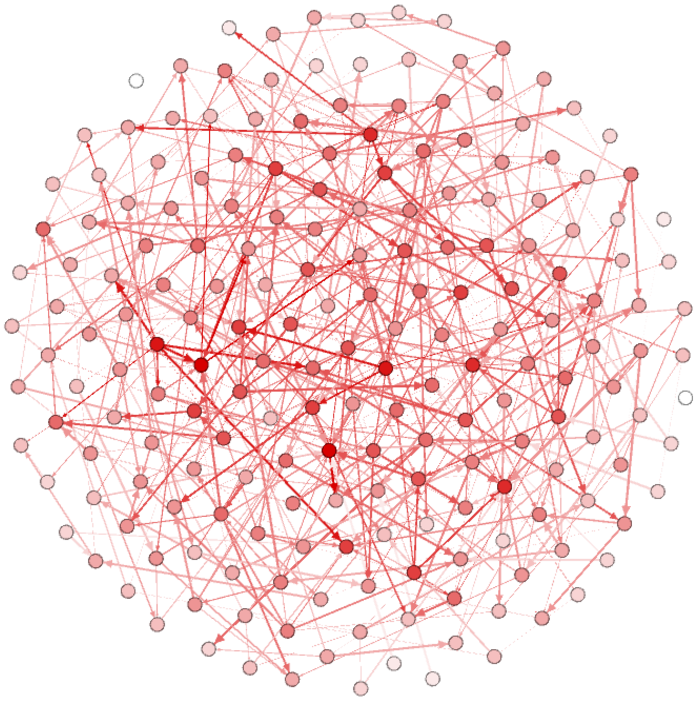


Figure 2.9 A basic example of an Erdős-Rényi model.

2.3.3 Small-world

Small-world networks are specific to social networks as they present attributes found in real-world societies. The topology is based on a graph with a generally low amount of interconnectivity, in which the majority of nodes are not neighbors, but in which the average path length between is small. More specifically, as the network grows, its average path length L only grows at a logarithmic rate, described using equation 2.13, whilst leaving the clustering coefficient unchanged [208]:

$$L \approx \log N \tag{2.13}$$

where N is the total number of vertices present in the network. This characteristic property is found in many of today's empirical networks such as the world-wide web, musicians' and actors' network, natural food-chains, gene networks, *etc.* [87, 151, 206].

The main metrics used to define and analyze small-world networks are the small average path length L and the high clustering coefficient C, creating a bridge between regular and random networks, as seen in Figure 2.10.

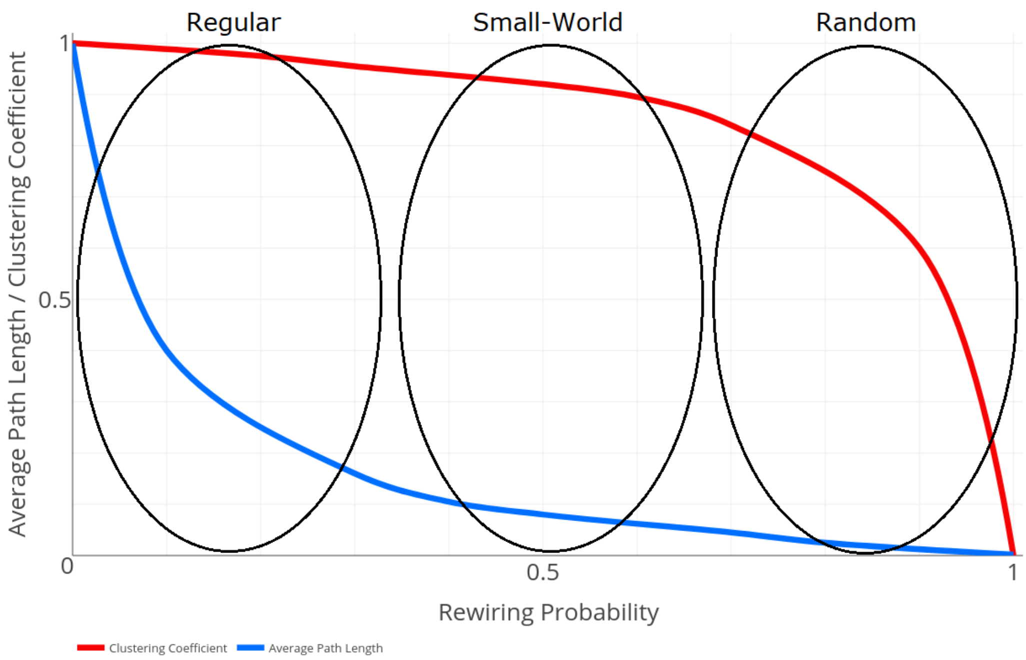


Figure 2.10 When taking into consideration the average path length L (blue) and the clustering coefficient C (red), the small-world network is positioned between the regular and random networks.

2.3.4 Scale-free

As with the small-world networks, scale-free networks are ideal in representing an amalgam of real-world (social) relationships, especially since connections within the network are formed based on preferential attachment: when creating new links, nodes with a high degree will be able to increase their degree even more, while nodes with small degrees will stagnate in the process [206].

Individual nodes of a scale-free network follow a power-law distribution [7, 148], meaning that the degree k of fraction of nodes from within the network — denoted with $P(k)$ — can be expressed using equation 2.3. This means that within the network most nodes have a very small degree, multiple nodes act as local hubs while very few nodes are hubs for most clusters formed in the network, as illustrated in Figure 2.11.

This kind of distribution is present in a vast majority of networks, all the way from social networks to biological networks, economic networks or networks representing comic-book lores [6, 21, 161, 163, 206].

A popular example demonstrating the occurrence of scale-free networks in real life is the collaboration of movie actors. Several studies [81, 169] have shown that all actors are linked by a relative small number of links and that many actors have been in direct contact with other actors. One actor in particular has an overwhelming number of connections, namely Kevin Back, effectively serving as a hub for connecting other actors together with each other. This led to the development of the well-known parlor game ”Six Degrees of Kevin Bacon”, focusing on the fact that no actor is more than 6 steps away from Kevin Bacon [21, 55, 75].

A similar study also exists in the musicians network; limited to jazz musicians only, the formed network closely resembled to the previous networks [90], and as such, I felt inspired

BUPT

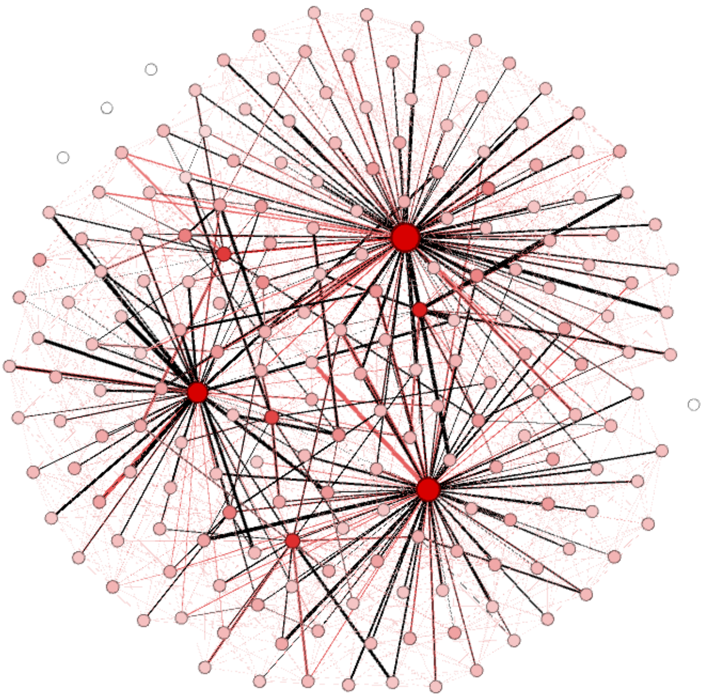


Figure 2.11 A scale-free network with preferential attachments: nodes with a high degree are able to increase their degree even further, forming so-called hubs.

(and compelled) to analyze their whole network, effectively creating the prerequisite of MuSeNet.

2.4 Social Network Analysis

Identifying certain aspects of any complex network and being able to use them to draw important conclusions are just few steps out of many any researcher needs to take in order to document noteworthy results. The prerequisites of such an endeavor involve collecting data and identifying or creating the tools and methods capable of processing raw data. Some of these, which are relevant to this thesis are listed in the following subsections.

2.4.1 Databases

- Stanford Large Network Dataset Collection: perhaps the most complex and well-known general-purpose network analysis dataset library curated by Jure Leskovec [124]. It contains collections of over one hundred networks of various network-types (*i.e.* directed, undirected, bipartite, temporal, labeled, as well as multigraphs) and relationship categories (*e.g.* social networks, communication networks, citation networks, collaboration networks, web graphs, amazon networks, internet networks, road networks, wikipedia networks, articles network, temporal networks, *etc.*).

- Complex Networks and Soft Matter Lab: personal repository of Herman Makse, with an impressive collection of networks of various categories (*e.g.* biological networks, social networks, transportation networks, collaboration networks, *etc.*) free to use by anyone [131].

- Awesome Public Datasets: a list of high-quality open datasets in public domains collected and tidied from blogs, answers, and user responses [52]; biological networks, weather networks, complex networks, computer networks, economic networks, government networks, software networks, transportation networks, language networks, *etc.* are just a few types of relationships that can be found on this repository curated by Xiaming Chen.

- AllMusic Guide: a comprehensive online, human-readable database referencing thousands of musical artists and bands, currently owned by All Media Network, LLC [127].

- IMDB 5000 Movie Dataset: over 5000 movie-data taken from the IMDB website, organized and listed publicly by Chuan Sun [189].

2.4.2 Tools

- Gephi: the leading tool in visualization and analysis of large networks [27], fully supported by an active community due to the fact that it is an open-source application. It features plug-in support, has a rich variety of tools for measuring and analyzing social networks and it provides a rich framework that helps developers extend the existing modules using the Java language. Its main features include the possibility of importing graph data in multiple formats, visualizing data using a multitude of intuitive layouts, measuring and displaying graph metrics, coloring nodes and communities based on custom criteria, filtering out nodes based on custom conditions, exporting data as images *etc.* One strong aspect is the amount of supported input formats: gml, gdf, gexf, graphml, csv, spreadsheet, Tulip TPL, Pajek NET, GraphViz DOT, NetDraw VNA, making Gephi a very useful network visualization tool.

- Cytoscape: an open source software platform, which, initially designed for visualizing molecular interaction networks and biological pathways, became a general platform for complex network analysis and visualization [179, 184] due to the fact that it too has a powerful community, responsible for improving or extending its modules.

- Centrifuge: a premium browser-based application with an integrated suite of capabilities that can help rapidly understand and glean insight from new data sources, visualize network-data by interacting with it and collaborate in real-time with others [190].
- Graphviz: an open source graph visualization software which represents structural information as diagrams of abstract graphs and networks, with many useful features for analyzing networks [70].
- SocNetV: a cross-platform, user-friendly tool for the analysis and visualization of social networks [112].
- Pajek: is a free windows applications for analyzing large network-types [28].
- iGraph: a collection of open-source network analysis tools with the emphasis on efficiency, portability and ease of use [60].

BUPT

3. Collaboration in Complex Networks

In general terms, complex networks are formed by a set of social actors connected together based on certain rules [69, 153], depending on the analyzed point of view: an actor might be a single person, a team, or an organization, while a connection might be a relationship among two people, a collaboration or common member of an organization.

Traditional studies of social networks have been carried out through different fields [74, 120, 178], and even though such studies have exposed much about the configuration of communities, they suffer from two deficiencies that make them poor sources of data for the kind of approach to network analysis. First and foremost, the collected data are not numerous; most data sets contain no more than few a hundreds or thousands of actors, with only a handful of studies exceeding 1,000,000 actors [83, 152, 217]. Also, gathering the data from these studies is a laborious process, making the statistical accuracy of many results poor [83, 152]. Secondly, they contain erroneous results due to the subjective nature of respondents' replies. As such, a more promising and reliable source of data are collaboration networks themselves.

A collaboration network consists of a multitude of nodes representing distinctive entities, like organizations, departments or people. 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 [44, 178, 186].

Collaboration is an activity which can be found within various social networks where a network is considered as a linkage between two nodes. Since this kind of ongoing analysis of the process of collaboration from a social networking point of view is still rather new, it is reasonable to assume that collaboration theory stands to expand in future. Collaboration is encouraged when there is an expected beneficial outcome common for the collaborators within the entity: the more significant this outcome, the higher the participation and commitment level will be among the collaborators. It also allows a mix of size and scale of businesses, organizations or communities to share experiences, capabilities, resources and ideas and become more competitive. Such processes find their "natural" environment on the Internet, where collaboration and social dissemination of information are made easier by current innovations and the proliferation of the web.

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 [178]. As such, a handful of such examples are listed, most of which constitute prerequisites to MuSeNet.

BUPT

3.1 Collaboration in Social Networks

3.1.1 Co-authorship Network

An interesting study regarding collaboration in social networks is the one pertaining the so-called co-authorship network [152, 154]. This network is a hallmark of contemporary academic research. Scientists and authors are no longer isolated, independent agents, but members of a much larger, multi-disciplinary groups [185, 186]. When these agents collaborate, share ideas and co-author papers, they inadvertently create a new connection among themselves [68]. This network — such as the one depicted in Figure 3.1 — once analyzed, can offer answers to a broad variety of questions related to collaboration patterns, such as number of published papers, their quality, author's physical location and how these patterns evolve over time [154], as well as a handful of other interesting properties [152].

In almost all studied cases, the formed communities constitute a small-world network topology, with an average path length between scientists and intermediate collaborators scaling logarithmically with the size of their community, having an average of five or six intermediaries between any two randomly chosen nodes. These networks are also highly clustered. This property is a tell-tale sign that scientists introduce their collaborators to one another, an important aspect for the development of the whole scientific community. The power-law property was also identified in the form of distribution values of both the number of collaborators of scientists and the numbers of papers they write.

The author Newman [154] also pinpointed important statistical differences between divergent scientific communities. Some differences, like the fact that experimental high-energy physics, which is famous for its diversified collaborations among authors, has a much higher average number of collaborators per author than any other examined field, be it Biology or Computer Science, were obvious. However, biomedical research, for example, presented the lowest degree of clustering out of the examined fields. This hints to rare collaborations among scientists in this field, even with the existence of common collaborators. Biomedicine is also the only field which is dominated by many people with few collaborators, rather than by few people with many collaborators, as opposed to other fields [152].

3.1.2 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 graph database. As such, they devised a model using graph theory, where two Marvel characters are considered linked if they jointly appear in the same Marvel comic book, show, movie or game. The purpose behind all this extensive research was to keep an accurate record of each Marvel character and its background, but in doing so, they inadvertently created their very own network. The result is essentially a dynamic database — as opposed to most databases, which are relational —, which was used in order to make recommendations for new fans who want to get into the mythos, but have no time to sort through what would seem to be a bottomless well of lore, spanning through decades.

The study pertaining the Marvel Universe was further analyzed by authors Alberich *et al.* [6]. The resulting network — which was formed of 6,486 nodes (corresponding to individual Marvel characters) and 96,662 edges (corresponding to individual relationships) —, presented most of the characteristics of a real-life collaboration networks [6], rather than

BUPT

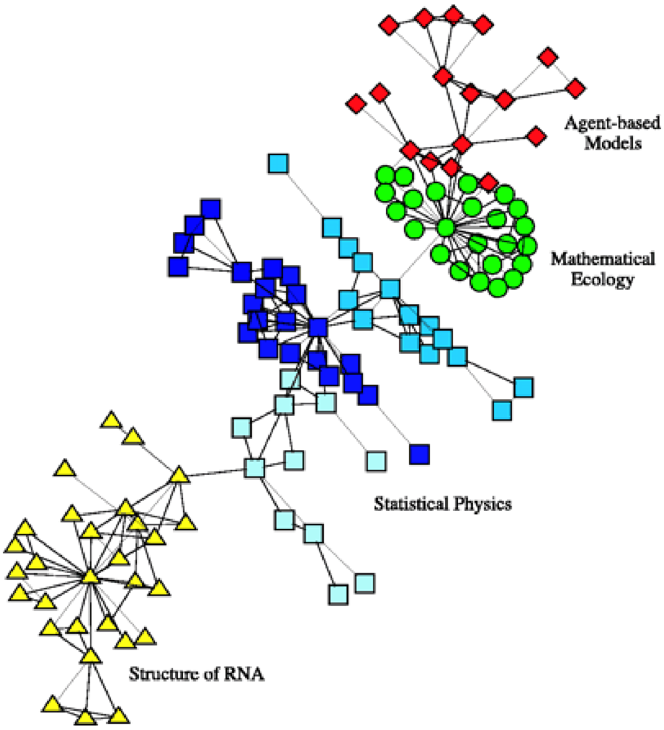


Figure 3.1 Visual representation of the co-authorship network, as presented in the original paper. The nodes represent scientists with links among them based on common works, further grouping them based on their research-topics [154].

that of a random collaboration networks. Interestingly however, the clustering coefficient differed 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' synthetic origins [6].

3.1.3 IMDB Actors' Network

Derived from a famous statement made by Kevin Bacon himself [75, 209], 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 [7, 133]. One of the most famous contributions in this aspect is the paper based on the winning entry of the Graph Drawing Competition [4], presenting the analysis of large and complex temporal multivariate networks derived from the Internet Movie Database, IMDB. IMDB has become, over time, a huge and very rich dataset with many attributes. It is currently used to create visualization graphs based on actor relationships for each particular year, starting from 1911 through recent years. Since the information found on movie database grows with each new movie release, using all of the data may result in networks that are not that transparent and hard to analyze. Therefore studying only a subset of the IMDB network, more specifically the adult collaboration network [81] brought many benefits, for instance by removing any nodes characterized by long-spanning careers and focusing the resulted network more onto the time evolution.

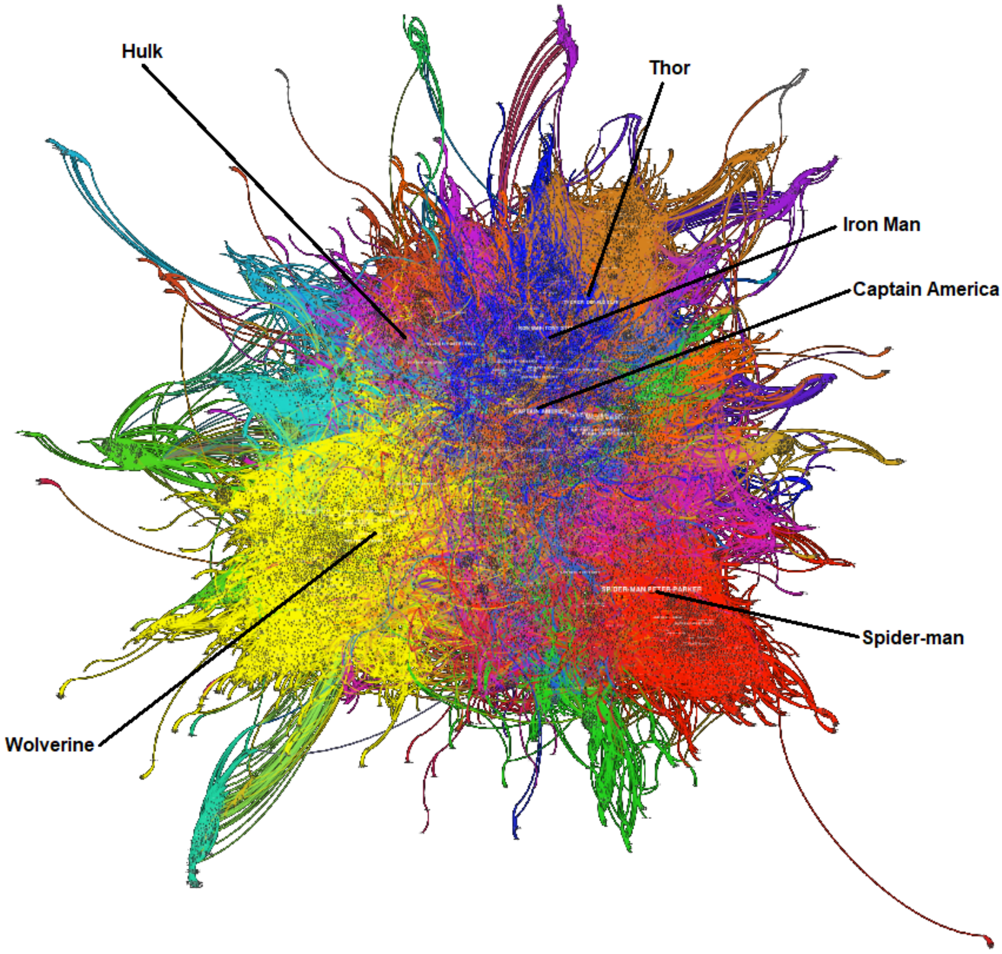


Figure 3.2 Visual representation of the connections between the Marvel Universe super-heroes split into three major clusters and a multitude of smaller ones. The large clusters are formed of the X-Men (yellow colored cluster on the left side), Spider-man (red color on the right side) and the Avengers (blue colored cluster in the middle). The smaller clusters, like the red colored nodes sprinkled throughout the network are formed of superheroes that belonged to small Marvel Universe factions that were interconnected with larger groups [46, 176].

3.1.4 Jazz Musicians' Network

Similar to the previously presented studies, an analysis involving the collaboration network of jazz musicians [45, 86, 90] represents yet another prerequisite of MuSeNet. In this particular study, the authors Gleiser *et al.* 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 [90]. This network is formed by creating connections between bands who feature at least one common member, as presented in Figure 3.4. The resulting network revealed the following information pertaining the community of jazz musicians of that period: going from the center towards the tips of the network, the separation and the splitting of the branches into two communities represents the manifestation of racial segregation present at that particular time. Although collaboration between races did exist, most bands were exclusively formed by people of a single race or, in some cases, within a specific geographical location. After a closer inspection of

BUPT

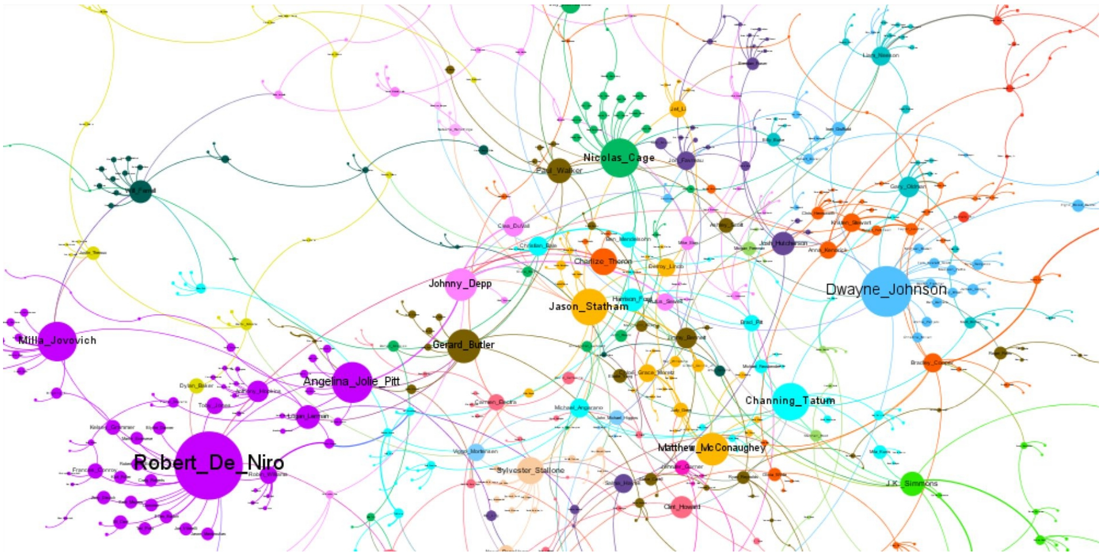


Figure 3.3 Preview of the collaborative IMDB actors' network [197].

the names of several artists, the authors Gleiser *et al.* concluded that the nodes located on the left part of the community represent African Americans, while the ones on the right represent Caucasians [90].

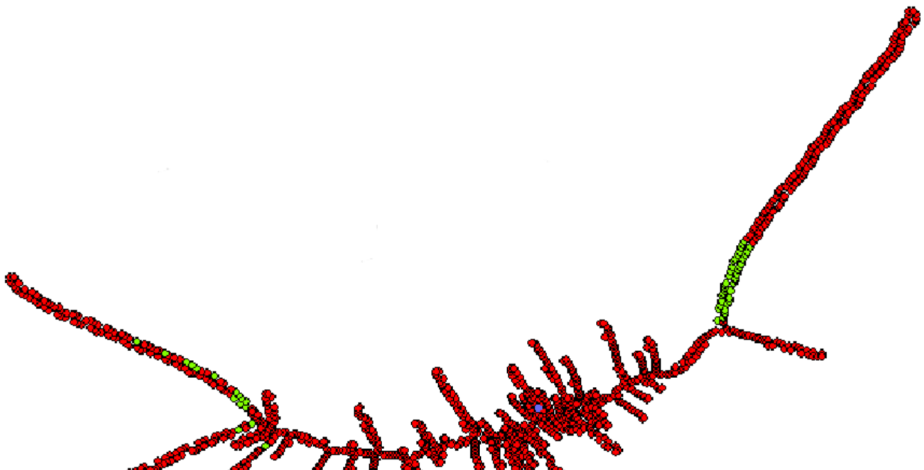


Figure 3.4 Graphical representation of the network formed by jazz musicians, as per original publication. The nodes highlighted with green nodes indicating musicians with a very high degree [90].

The collective empirical networks — both the network formed by jazz musicians, as well as the previously-mentioned IMDB actor's network — to which I compare MuseNet are presented in Figure 3.5 using a custom representation. Overall, the IMDB network has a clustered appearance with no visible community structure, while the jazz network has lower clustering and a very high interaction density. In both cases, the nodes represent artists (actors/jazz musicians), and are linked by at least one artistic collaboration.

BUPT

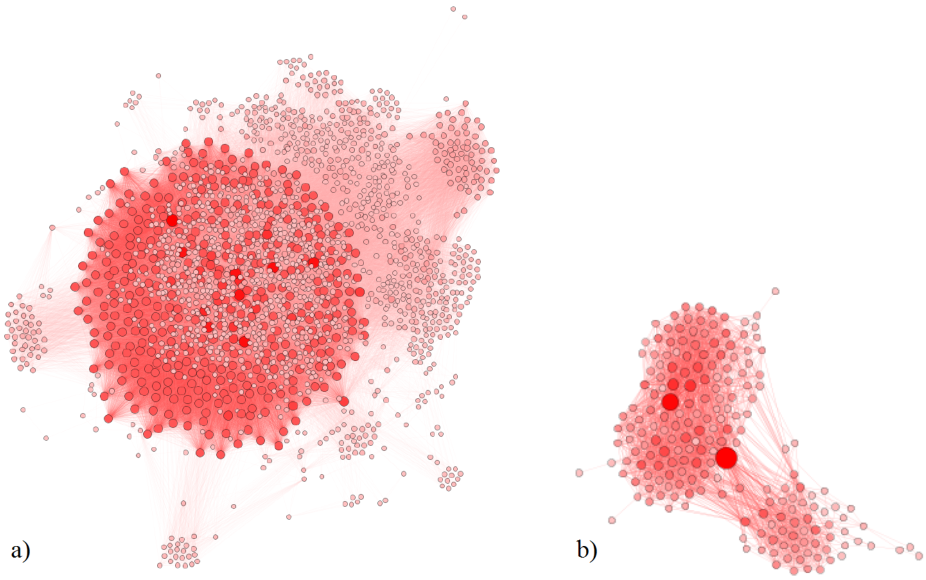


Figure 3.5 Representation of a) the IMDB movie actors co-appearance network and b) contemporary jazz musicians collaboration networks. Nodes depict individual artists, which are linked by an existing professional collaboration. Node-coloring and sizing are proportional to their degree [24].

BUPT

3.2 Collaboration in Economic Networks

Since economical improvements and financial stability are an important factor in our day-to-day life, researchers have started observing and analyzing the interest of world-wide companies in their effort of establishing industrial networks by means of studying aspects such as social interaction and contractual relationships [64]. Due to presence of a plethora of companies on the market, each with its own division, we can identify both global (*i.e.* outside), as well as local (*i.e.* within the company itself) connections. These connections act as visual representations of their dependencies generated by their activities, as well as their resources, both used and created [96]. Results obtained prove that certain companies rely and trade with each other, but in order to further themselves — and to obtain a profit —, they are ultimately forced to either cut down production costs or increase production quality. Also, an increased tendency towards specialization has aided to the introduction of advanced manufacturing technologies in order to supply niche markets [64].

BUPT

3.3 Emergent Properties of Complex Networks

A common characteristic of complex systems — including complex networks — is the presence of certain emergent properties at a macroscopic scale [119, 158, 160]. Driven by a rather small variety of a reasonably homogeneous agents collaborating together [123, 187], emergence is based on simple rules of interaction defined or observed for each agent at a microscopic scale [30, 102, 104, 140]. Colloquially, this can be expressed as "the whole is greater than the sum of its parts". However, despite the agents' homogeneity, the resulting emergent property — and implicitly, the entire system — is irreducible, due to the individual, self-organizing behaviour of each agent [123, 187].

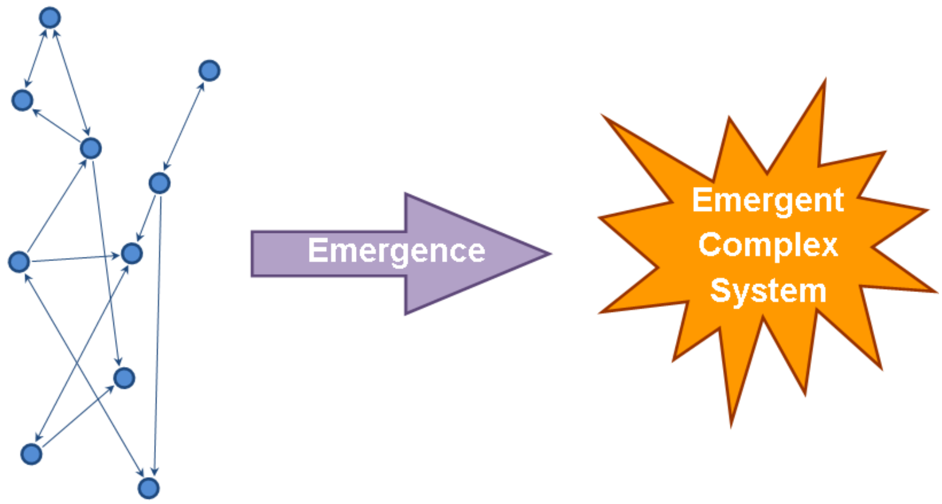


Figure 3.6 Abstract illustration of local agents interacting and connecting with each other in unpredictable ways. As a result, certain regularities emerge in their behaviour, resulting in a complex system.

As emergence generally describes the appearance of interactions, dependencies, or relationships between agents [119, 158, 160], it has important applications in fields like Computer Science, Physics, Mathematics, Biology or Social Science. While emergence is usually used as reference to the unexpected appearance of organized behaviour or collaboration within a given complex system, it actually refers to any phenomena which are challenging or even impossible to anticipate from the characteristics of individual agents that make up the system [187].

One important characteristic of emergence is novelty itself [123]. This stems from the fact that observations drawn both at a microscopic scale, as well at a macroscopic scale are distinct, and concomitantly, the existing (programmed) relationships between each agent (at microscopic scale) and the collective behaviour observed at macroscopic scale are often non-obvious [123]. As a result, emergence leads to the change of the initial network — as depicted in Figure 3.6 —, accounting for the appearance of an outcome which was not directly specified by the initial rules of interaction set at a microscopic scale [123, 187].

As the analysis of emergent properties represents a key interest in the fields of Computer Science, Mathematics, Physics *etc.*, I consider it relevant to make reference to it wherever applicable throughout this thesis.

3.4 Diffusion Models in Social Networks

A major trend in Network Science is the study and understanding of social diffusion models, namely how people influence each other and how they can be influenced [2, 69, 85, 91, 106, 116, 201].

The benefit of understanding these complex processes is of paramount importance for researchers in fields like Biology [166], Psychology, Criminology [89], Philosophy, Economics, Marketing [88], Finances and even Warfare [69, 91]. The type and distribution of specific diffusion models in a community at a certain time is a reflection of the distribution of influential people in that particular community [106, 117]. Social influence is the ability of certain individuals (*e.g.* musicians, economic agents, *etc.*) to influence others in their respective fields. Influential people leverage others to participate in certain activities, agree with their ideas, adopt their style and follow their lead [118]. At first, this is limited to a small portion of the networks' population; however, if this initial group is large enough, then the behaviour grows and spreads to a significant portion of the population, while otherwise the behaviour collapses so that no one in the population chooses to adopt the respective behaviour [109].

In today's world, the most efficient and notable tools of social influence propagation are the ubiquitous social platforms like Facebook, Twitter, LinkedIn or Google Plus [92, 183]. Using such platforms in order to influence the spread of rumors, political and religious opinions, interests, stories, epidemics, social media recommendations and analytics, behavioral targeting or viral marketing are just a few examples of means of shaping communities or entire networks [50]. Consumer groups use social influence to motivate others to act as a whole in order to obtain an envisioned economic, political or general consumer goal [142]. Present interest is focused on determining when, where, how and why a product or idea may be sold and/or bought by individuals in our society, and how the psychological factors behind this process can be influenced [143]. This is achieved by combining elements from Psychology, Sociology, and Economy [69, 206]. One such pertinent example is the so-called word-of-mouth marketing. The goal behind it is to trigger specific individuals with above-average influential abilities within the network, which, in turn, will activate other individuals via a viral (oral) propagation; more specifically, consider a network represented by a directed graph $G = (\nu, \varepsilon)$ consisting of $N = |\nu|$ nodes/agents ($|\nu|$ is the cardinality of node set ν) and a set of edges ε, representing social ties [23, 50]. Additionally, consider the function $p : \varepsilon \rightarrow [0, 1]$ that associates a given probability of $p(u, v)$ — representing the influence of u over v —, due to the existing (u, v) relationship. As a result, this leads to the reasoning that whenever u shares an action or opinion, v will do the same by following u, with a probability of $p(u, v)$, and represents the means of reaching a large target-audience driven by influence propagation from only a small set of users [50]. Other examples can be drawn from the field of Psychology, where a recent study [59] describes how an amalgam of emotional states can be transferred directly from one individual to another via mimicry or emotionally-relevant bodily actions like facial expressions — also known as emotional contagion —, greatly influencing the mood of the whole community they form [218]. Similarly, from a biological or military point of view specifically influencing certain nodes can lead beneficial results: vaccinating certain individuals or influencing them may lead to early disease eradication [14], effectively creating a blockade [72], or improving quality of life by reducing the cases of obesity [54]; detecting high-profile people of various criminal or military organizations can prevent tragedies [76]; *etc.*

In order to identify influential nodes in any network, certain network metrics can be applied [48]. The degree centrality without a doubt, is such a straightforward and efficient metric, but only under certain circumstances. This metric can become irrelevant in certain

BUPT

edge-cases, as a node linked to a few high influential neighbors may have much higher influence compared to a node having a larger number of less influential neighbors. Other, well-known global centrality metrics (*e.g.* betweenness, closeness, degree, PageRank, *etc.*) [40] can yield far more accurate results, but only applied on directed networks and at a high cost of computational complexity — as needed in case of certain large-scale networks [48]. Recent studies have brought forward more advanced measures meant to identify influential nodes. The first such noteworthy solution is a random-walk-based algorithm entitled LeaderRank [130], which, contingent on the results of its authors, significantly outperforms the PageRank metric. Yet another means of identifying influential nodes is by creating a semi-local centrality measure as a trade-off of low-degree centrality and time-consuming centrality measures [14, 48], but yielding similar results. The listed metrics are all important in order to establish both the influential nodes and their attributes, each yielding distinct results. The reason why there isn't a well-defined metric of establishing influential nodes comes from Borgatti and Everett's 2006 review article [37], and reconfirmed on several occasions [29, 61, 121]. The results presented show that the accuracy of identifying influential nodes — especially by centrality indices, which only rank nodes but do not quantify the difference between them — is highly dependent on network topology and since complex networks have a heterogeneous topology [29, 180], any given centrality measure which is appropriate for identifying highly influential nodes will most likely be inappropriate for the remainder of the network.

Conclusively, when taking into consideration relevant metrics for MuSeNet, I did so by applying the vast majority of the mentioned metrics, focusing on the betweenness and PageRank centralities for detecting influential nodes, and the eigenvector centrality for the communities themselves.

BUPT

4. MuSeNet - A Musician Collaboration Network

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. Most of these communities present typical properties of social networks, for instance (but without losing generality): the small-world property found the actors' network; a high degree of clustering coefficient present in the recipe network; the scale-free property in the Marvel universe, where the distribution degree follows a power-law [53, 206]; *etc.*

Inspired by these studies, I considered it paramount to address the existing relationships — 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 [24, 193]. In the following sections of this chapter, I describe the methodology used for obtaining the required data, its graphical representation and conclusions obtained by comparing it to other empirical networks.

4.1 Data Acquisition

The nature of information needed for such a study meant that the data itself was not readily available. As such, the first step was the gathering of the needed data. This would be done by using the website AllMusic [127] online database, currently owned by All Media Network, LLC. However, collecting the data was not an easy endeavor, mainly due to the fact that the website lacked a much-needed API (*i.e.* application programming interface) or means of downloading raw data. This meant that a script had to be written, in order parse each internal link automatically, and to retrieve the required information. After running for 24 hours, it accessed 781 pages, resulting in 19,881 artists (15,501 after filtering) with the following dataset saved into an SQL database:

- ID: the internal reference to be used to identify a specific artist in our scripts.
- Name: full name of the artist.
- URL: the URL of the artist, pointing to his/her profile on the AllMusic website.
- Genre: the conventional category that a particular artist identifies with.
- Style: (a list of) style(s) an artist identifies with.
- Member-of: a list of bands he/she was part of, if any.
- Active-period: the time-period in which the artist was reported as being active.

Taking into account future contributions to this thesis, information was also gathered from the bands' perspective, totaling in the identification and classification of 5,132 distinct bands. The only adaptation required was associating each node with a particular band (instead of an individual musician) and linking them based on presence of at least one common artist (*i.e.* a musician who has performed for both interconnected bands); as such, I am able to connect music artists based on their common bands, common music genre or music styles.

After collecting the data, I proceed to create the network of musicians, using an approach similar to the state-of-the-art methodology [90, 81]. Considering each artist as an individual node, I placed the links (between them) according to artist-compatibility. Particularly, compatibility is defined as the number of common bands two musicians have performed for. Thus, the more bands two nodes have in common, the greater the weight of the link between them. It is worth mentioning that this step greatly influences the structure of the resulting network: a different layout of ties — for instance, based not on common bands, but on overlapping activity years, gender, music genre, music style *etc.* — would offer different insights of the same dataset. This study only focuses on analyzing how common bands affect the clustering of artists in a complex network, and as such, I did not follow through with the additional insights mentioned; however, they do represent additional ways of furthering the research on this subject.

Having created the network, I followed through with computing the MuSeNet social network of musicians into a .gdf file, a valid input file I could load up in Gephi [27], the leading tool in visualization and analysis of large networks. For the purpose of this particular study I have truncated the weights of the resulting network's connections, obtaining an unweighted graph, where a link denotes one or more common bands between two musicians, while the absence of a given link denotes the absence of artistic interaction. The reasoning behind using an unweighted graph instead of a weighted one comes as a desire of optimizing and balancing out the interaction phenomena: it is shown to yield more accurate results for the study in terms of determining whether it is a community based meritocracy or topocracy

[39, 153]. To that end, I have left the attributes of genre, style and active year as parameters for doing the clustering of artists. An overview of the main steps taken in order to create MuSeNet is presented in Figure 4.1.

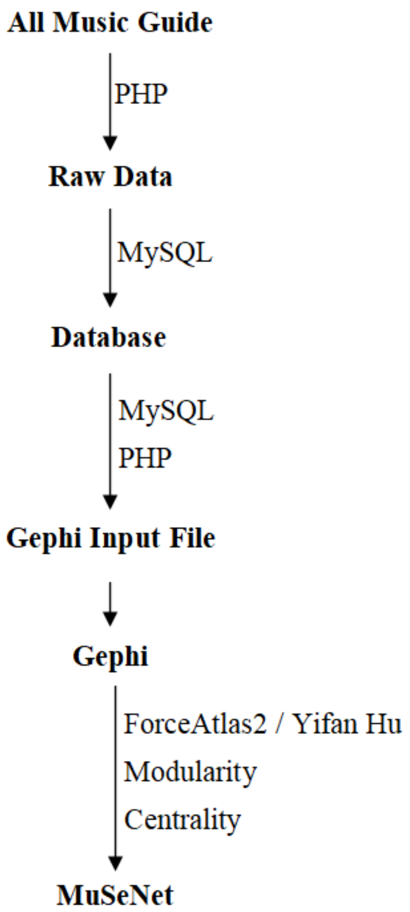


Figure 4.1 Visual representation of the steps taken in order to create MuSeNet [24].

4.2 Network Analysis

In this section I present the metrics and visualizations obtained by applying the techniques of Social Network Analysis.

In Figure 4.2 the relevant communities that form over the musical network are highlighted. Nodes are placed using ForceAtlas2 [110], a 2D force-directed layout algorithm [159] available in Gephi. Also, they are colored according to the community they belong to. The communities themselves are detected using the fast community detection algorithm implemented in Gephi [34]. This algorithm was chosen in light of the existing methodology to break down a social network into clusters, namely by using a modularity class clustering and extracting their representative features [156].

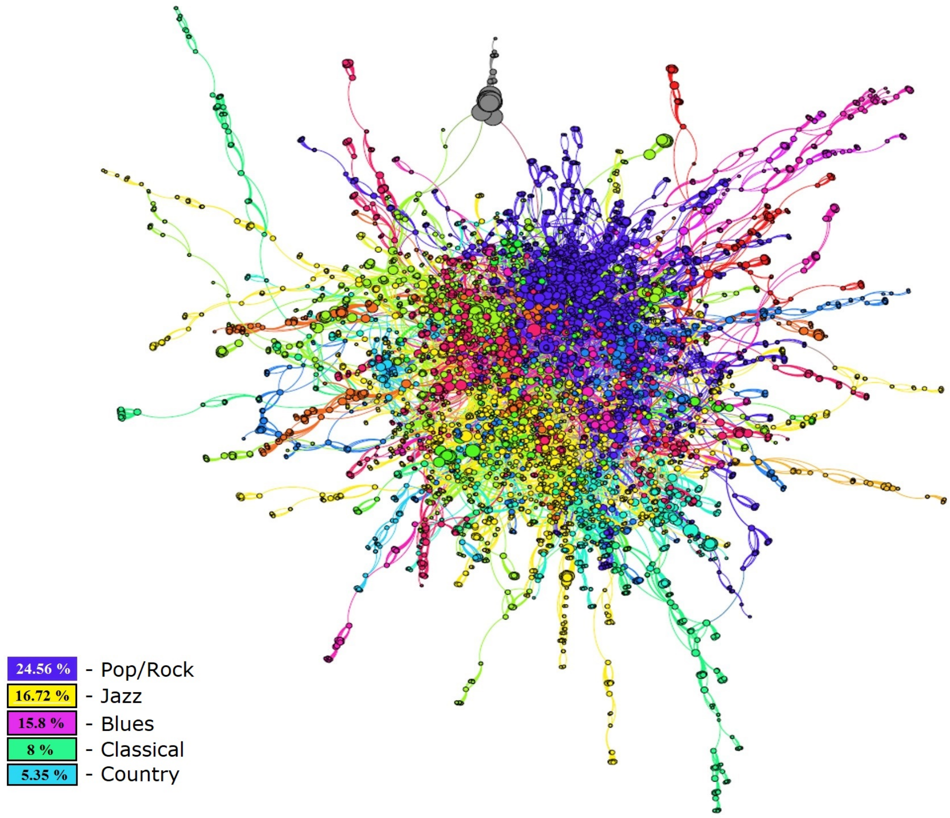


Figure 4.2 Graphical overview of MuSeNet generated in Gephi. Node-coloring highlights the distinct musical genres communities obtained by applying the community detection algorithm [24].

One of the analytical advantages of social networks analysis lies in the emergent community structure of the network it is applied to. Artists are grouped together by partially overlapping musical genres. As such, the relevant communities which emerge, based on genre, are: pop-rock (24.56%), jazz (16.72%), blues (15.8%), classical (8%), country (5.35%) and others.

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, like avant-garde, reggae, vocal, or religious, so I can confirm there is a correlation between the communities' center of gravity and their real-world popularity. As a general rule of thumb, the further a

BUPT

genre-community is from the absolute center 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 — represented with violet in Figure 4.2 — 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, highlighted with yellow. 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 (green) which, in today's world, implies composing contributions for movie scores, commercials, and melodic lines for other genres. Finally, country music (which is highlighted with cyan in Figure 4.2) 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.

By applying the main centrality metrics — power-law distribution of degree (Figure 4.3), betweenness (Figure 4.4), eigenvector centrality (Figure 4.5) and PageRank (Figure 4.6), which is specific for social networks, both empirical and synthetic [7, 56] — to MuseNet, it helped drawing the following aspects.

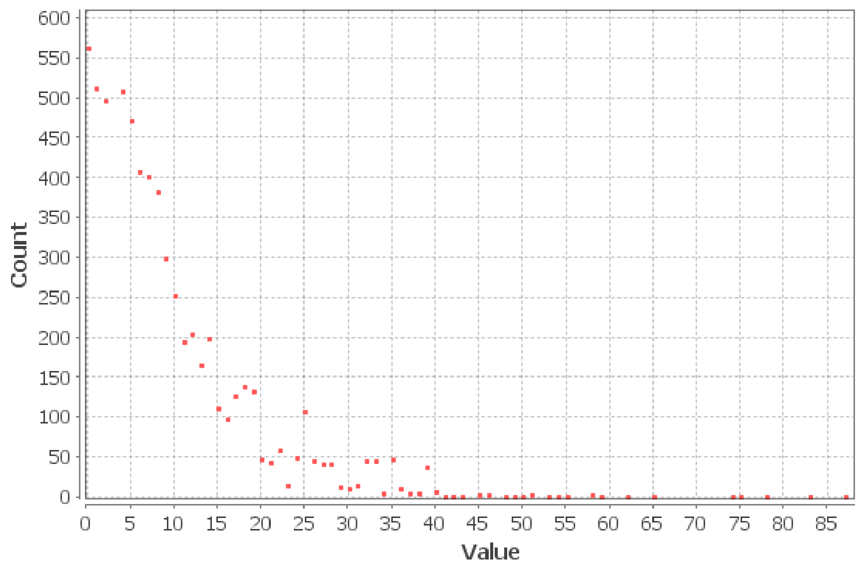


Figure 4.3 Power-law degree distribution in MuSeNet [24].

In Figure 4.5 the graphical results of applying the eigenvector centrality metric are presented. Noteworthy is the cluster visible highlighted in red, which proves that there is a small single dominant community of nodes with very high eigenvector centrality. On closer inspection, this community is formed by mature artists who currently own (or have owned) a record 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 versus topocracy presented in a recent study by Borondo *et al.* [39], 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.

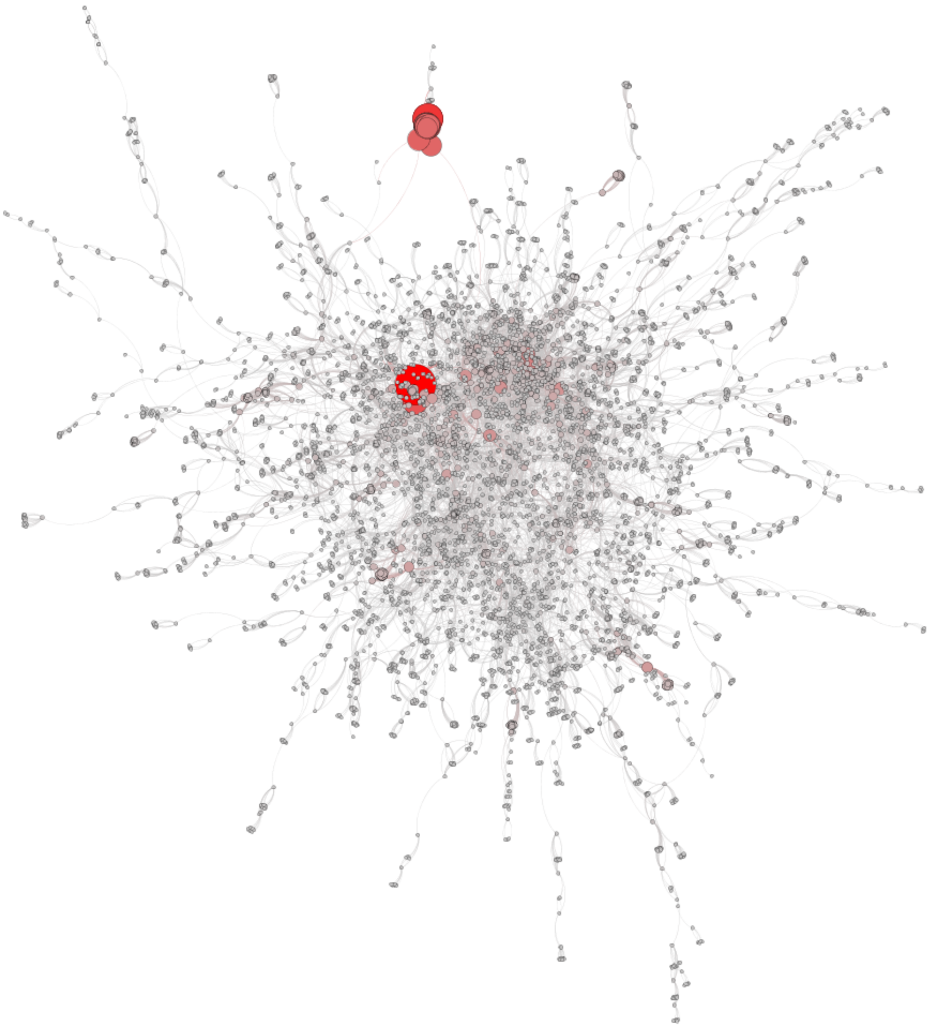


Figure 4.4 Graphical representation of the degree centrality measurement of MuSeNet [24].

Table 4.1 shows the top 5 artists with the highest centralities in the music industry. I have measured all four centralities since they highlight different aspects of importance in a network.

The musician with the highest degree is Greg Errico (Table 4.1a), an artist and producer who's resume spans across the most important musical genres, until today. He was member of the band named "Sly and the Family Stone" and performed with artists like David Bowie, Santana, Larry Graham, *etc.* from a multitude of genre-communities like rock, jazz, or fusion. On the other hand, betweenness (Table 4.1b) depicts importance in terms of interaction control. Dave Grohl, a member of "Foo Fighters" and "Nirvana", lies the crossroads of most collaboration paths between all other artists.

The eigenvector centrality highlights members of the mentioned community of producers (colored with red in Figure 4.5) alongside of with Greg Errico, Alphonso Johnson and many others, as seen in Table 4.1c.

In link analysis, where the PageRank centrality is mostly used, a web page will have a high PageRank if it has some combination of high in-links, low out-links, and specific in-

BUPT

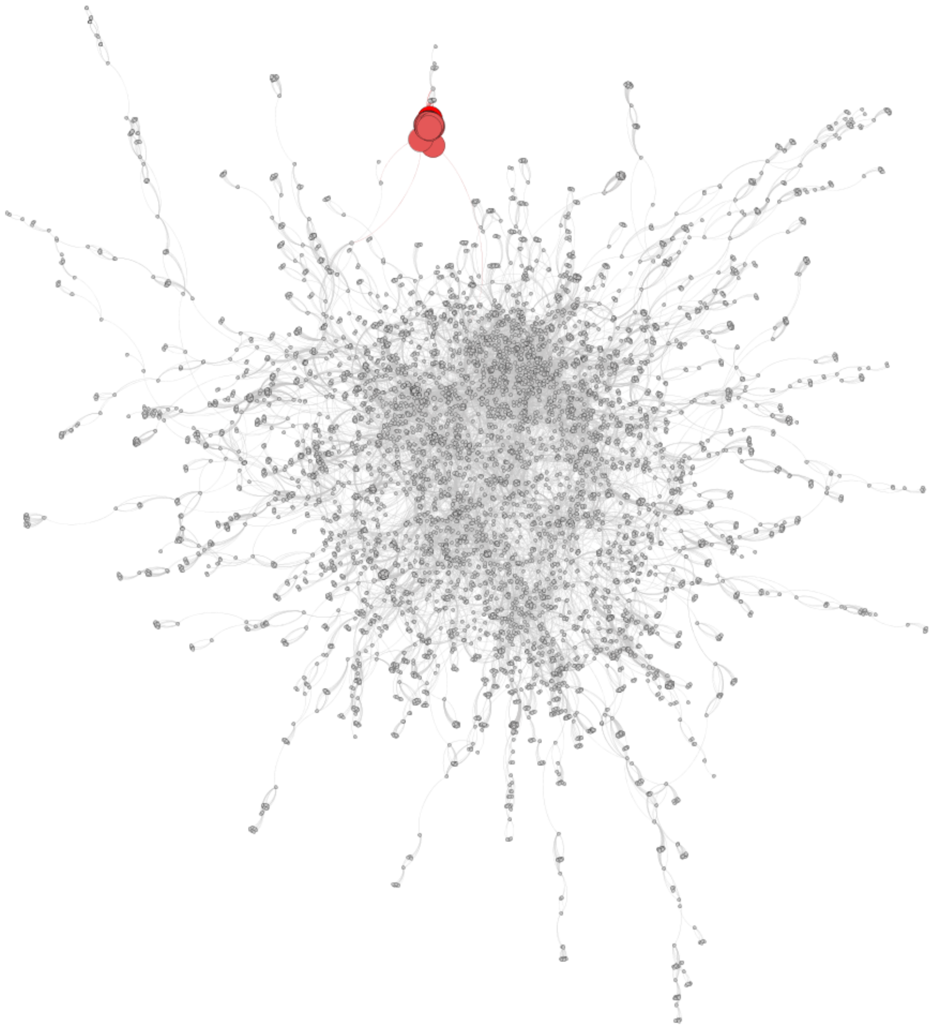


Figure 4.5 Graphical representation of eigenvector centrality measurement of MuSeNet. The community highlighted with red is formed by influential artists who currently own a record studio [24].

links from other high ranking pages. In the world of musicians, these artists with a high PageRank like Greg Errico, John Wetton, and Lu Edmonds — as depicted in Table 4.1d — have most likely been influenced by either a lot of people, a few very important people, or some combination of the two.

Finally, similar to the IMDB study which denotes Kevin Bacon as the most influential node in the Hollywood actor network [75], 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. Dave Grohl is an American rock-musician, multi-instrumentalist, singer, songwriter, producer and film director. He is best known for being the lead vocalist, guitarist, main songwriter and founder of the band ”Foo Fighters”, drummer and song-writer of ”Nirvana”, ”Them Crooked Vultures”, ”Queens of the Stone Age”, *etc.* He has also performed session work as a drummer for a variety of associated acts, like ”Garbage”, ”Nine Inch Nails”, ”David Bowie”, ”Paul McCartney”, ”The Prodigy”, ”Slash”, ”Iggy Pop”, ”Tenacious D”, ”Lemmy”, ”Stevie Nicks”, *etc.*

BUPT

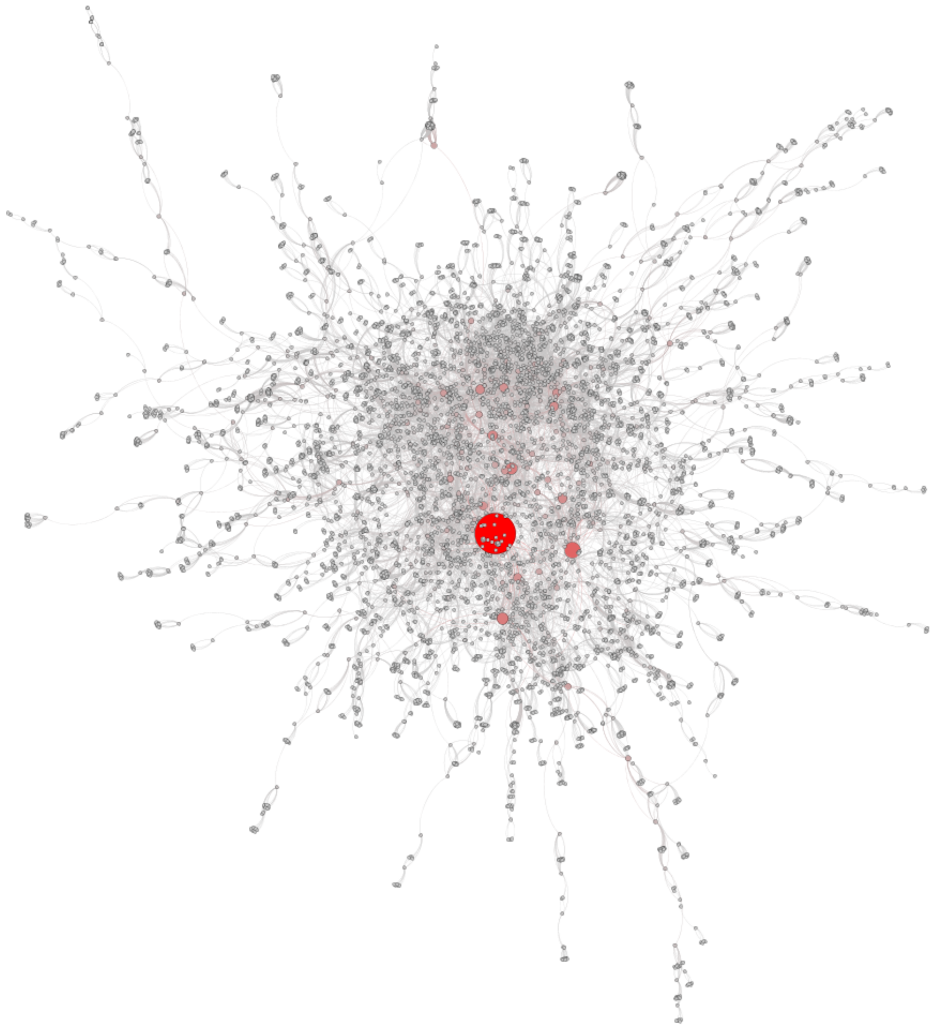


Figure 4.6 Dave Grohl is the ”Kevin Bacon” of the music industry, fact denoted applying the PageRank centrality on MuSeNet [24].

MuSeNet can further be analyzed from different perspectives as mentioned in a previous section, which I consider would yield yet additional and useful information.

Artist	Degree
Greg Errico	81
Alphonso Johnson	79
Dave Walker	67
Don Airey	65
John Wetton	62

(a) Musicians with the highest degree.

Artist	Betweenness
Dave Grohl	.0124
Josh Freese	.0091
Chris Shiflett	.0084
Lu Edmonds	.0075
John Wetton	.0073

(b) Musicians with the highest betweenness.

Artist	Eigenvector
Alphonso Johnson	.764
Greg Errico	.754
David Brown	.689
Graham Lear	.657
Neal Schon	.652

(c) Musicians with the highest eigenvector.

Artist	Pagerank
Greg Errico	2.925
John Wetton	2.777
Lu Edmonds	2.7
Jimmy DeGrasso	2.672
Alphonso Johnson	2.641

(d) Musicians with the highest pagerank.

Table 4.1 Numeral results of the centrality metrics applied to MuSeNet.

BUPT

4.3 Similarity Analysis

Along with the obtained MuSeNet network, I also use the IMDB network of actors [81, 169], the Jazz network of musicians [45, 86], as well as reference online social networks of Facebook, Twitter and Google Plus available from a previous publication by authors Topirceanu *et al.* [194], and the Stanford Large Network Dataset Collection [124]. The IMDB actors network and the Jazz musicians networks were chosen based on their similar approach, while the Facebook, Twitter, Google Plus models were chosen in order to put in perspective the particular features that artists have as opposed to everyday Internet users. The network analysis is done from two perspectives: a metrics-based comparison, by using the fidelity metric [194], and a motif-based comparison [149].

4.3.1 Metric Fidelity Comparison

The metric fidelity comparison is done using the topological metrics which are specific for every complex network [7, 151, 188, 206]: average degree (AD), average path length (L), average clustering coefficient (C), modularity (Mod), graph edge density (Dns) and graph diameter (Dmt). Conclusively, the mentioned metrics, as well as the obtained values are presented in Table 4.2, collectively representing the difference of *sociability* from the perspective of three types of collaboration networks. Interestingly, the Facebook model is situated at an average level of sociability, due to the fact that all of the metrics are centered on empirically representative values [194, 113].

As opposed to this, the IMDB actor network and MuSeNet each lie at an extreme. The former proves to be more sociable (*i.e.* significantly greater AD, shorter L, higher C, higher Dns, and shorter Dmt), while the latter the least sociable. From a social perspective the following conclusion can be drawn: 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.

Conclusively, the Twitter and Google Plus networks, much like Facebook, are moderately sociable networks, while jazz musicians — surprisingly — are comparable to actors when it comes to this metric. The explanation for this phenomenon can be seen in MuSeNet itself, as jazz musicians work with many artists, and foremost with the majority from their own genre.

In order to better convey the interpretations, I apply the S-metric described in equation 2.11 by taking into consideration the basic graph metrics (which are also present in Table 4.2) and compare them to reference models; to this end, I use the on-line social networks' model-distribution of metrics as the reference, and compare the metrics of all collaboration networks together. As such, the sociability of the collaboration networks using the Facebook model as a reference is given in Table 4.3. The Facebook model compared to itself will have a sociability $S = 0$. Any model that is considered as less sociable will have $S < 0$, while all models that are more sociable in terms of their graph metrics will have $S > 0$. We can also observe that MuSeNet is indeed on the "unsociable" side, while the Jazz and IMDB networks are more sociable. Changing the reference model (Facebook, Twitter, Google Plus) would

BUPT

Network	AD	L	C	Mod	Dns	Dmt
Facebook	22.23	2.34	0.256	0.577	0.005	7
Twitter	12.39	2.68	0.239	0.28	0.054	7
Google Plus	12.15	3.9	0.404	0.44	0.035	12
Jazz	27.7	2.23	0.633	0.441	0.141	6
IMDB	113.5	1.55	0.996	0.476	0.062	4
MuSeNet	13.18	7.64	0.884	0.844	0.002	23

Table 4.2 Relevant measures for each empiric network: average degree (AD), average path length (L), average clustering coefficient (C), modularity (Mod), graph edge density (Dns) and graph diameter (Dmt).

S	Facebook	Twitter	Google Plus
Jazz	29.34	4.23	5.80
IMDB	19.33	11.62	11.76
MuSeNet	-3.56	4.34	-2.46

Table 4.3 Sociability values for similar collaboration networks compared to Facebook, Twitter and Google Plus.

only change the values associated to each network's sociability, leaving the scale and signum the same.

Table 4.4 presents the fidelity values of each collaboration network when compared both to themselves and to the online social networks Facebook, Twitter and Google Plus. The results show a similarity of 45-65% between all collaboration networks and online networks. This is due to the sociability difference compared to the moderate one of the reference models. On the other hand, the metric comparison supports my sociability evaluation as it shows the IMDB and Jazz networks — both described a highly sociable — much more similar (67%) than compared to MuSeNet (<50%).

4.3.2 Motif Distribution Fidelity

Comparing the before-mentioned networks based on the existing motifs was done using a two-step approach.

First, I measure the distribution of size-3 motifs on each empirical network using FANMOD [212], which is based on RAND-ESU [211], one of the fastest detection algorithms available. As a result, I obtain the distribution depicted in Figure 4.7, offering a different perspective over the already presented conclusions. The Jazz musicians network behaves more like a normal social network — having a uniform distribution of motifs — while the IMDB and

φ	Facebook	Twitter	Google Plus	Jazz	IMDB	MuSeNet
Jazz	.647	.595	.615	-	.672	.517
IMDB	.472	.535	.537	.66	-	.472
MuSeNet	.486	.451	.574	.491	.479	-

Table 4.4 Network fidelity values of the three collaboration networks (rows) towards the six references (columns). A higher $0 \leq \varphi \leq 1$ value denotes a higher similarity.

BUPT

Network	Facebook	Jazz	IMDB	MuSeNet
Jazz	.818	-	.36	.662
IMDB	.16	.171	-	.231
MuSeNet	.501	.595	.433	-

Table 4.5 Network fidelity values of the collaboration networks (rows) towards the four used references (columns) in terms of motif distributions.

MuSeNet networks have a predominant motif characterizing them. The motif size used in this study is fixed to subgraphs with 3 nodes only, as they are much more numerous to be found in graphs, and thus substantially more relevant [13]. The motifs can be seen in the lower part of Figure 4.7.

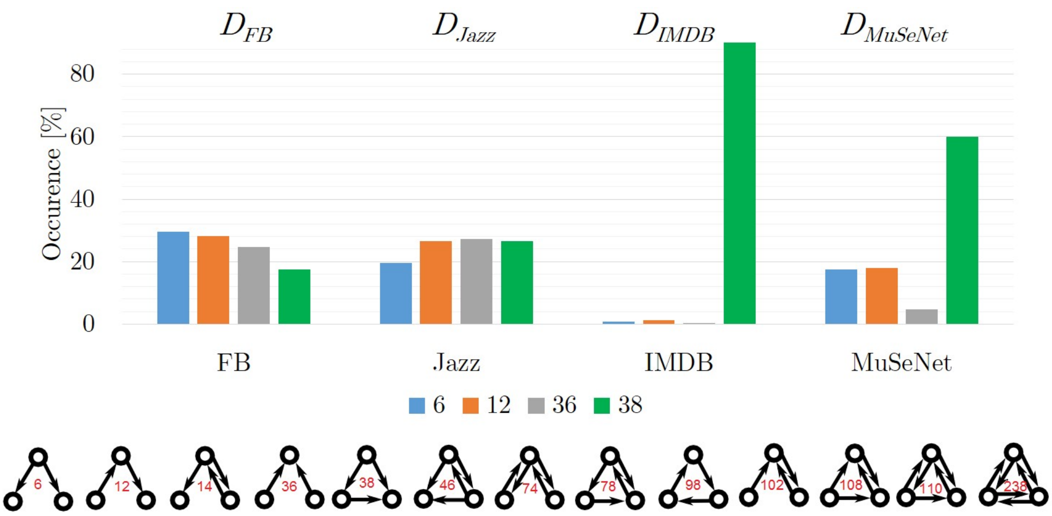


Figure 4.7 The resulting motif distribution of the chosen empirical network topologies. The occurrence of each motif is expressed in percentage in the central histogram. As can be seen, only distinct motifs (not all) characterize each network. All 13 motifs of size 3 are depicted at the bottom of the figure [24].

As the second and last step, I apply the fidelity metric to compare the motif distribution vectors with one another. The obtained values are given in Table 4.5. A value of 1 means complete similarity, while a value of 0 means complete dissimilarity. The data is interpreted as, for example: the Jazz network has a similarity of 81.8% towards the Facebook model *etc.*

BUPT

5. Agent-based Simulations of Collaboration and Payoff Distributions in Economic Networks Using TrEcSim

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. To that extent, economic and social networks facilitate understanding the dynamics of our society and the socio-economic roles we play. For instance, we can identify individuals who benefit from current topological features (*e.g.* hub nodes [65, 219], influential spreaders [196, 216], vital agents [8, 129], *etc.*). Indeed, the last couple of decades have witnessed notable developments in Big Data and machine learning, which have boosted various applications of Network Science [58].

In the field of Economics, 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 [107]; however, due to the fact that economic networks are non-linear, unpredictable complex systems, it is very difficult to analyze them based only on real-time quantitative observations [16, 20, 35]. As such, I extend the existing economic models, simulators and empirical observations by creating *TrEcSim*[1] [23, 25, 26], an extended, dynamic version of the embedded markets mathematical model from Borondo *et al.* [39], to which I will further refer to as the *rockstar model*.

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 [26, 204].

- 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 [1, 196].

[1] Short for Trade and Economic Simulator, TrEcSim is freely available at https://github.com/trecsim/trecsim

BUPT

- The analysis of ergodicity in economic systems — and with it the intrinsic fairness of the system — based on network topology and producer/consumer placement.

BUPT

5.1 State-of-the-Art

Visual representation for analyzing the dynamics of social and economic networks fosters data analysis and, at the same time, facilitates a convenient way of representing the results of network analysis. As such, many of the existing simulation tools offer the possibility of visualizing the simulated economic networks, as well as the obtained results. Data exploration is performed through displaying nodes and links using a variety of layouts and by attributing colors and size to nodes/agents, according to certain relevant network properties and centralities (*e.g.* modularity, degree, betweenness, *etc.*).

Visual representations of networks may be a powerful method for conveying complex information, but care should be taken in interpreting node and graph properties from visual displays alone, as they may misrepresent structural properties; indeed, such structural network properties are better captured through statistical, quantitative tools. Nonetheless, the typical application for any economic simulator is to visualize, analyze, evaluate or verify specific economic scenarios, or theoretical economic models.

5.1.1 The Rockstar Model

A recent attempt to explain how payoffs are distributed in complex economic networks, based on the nature of interactions between producers and consumers, was introduced by Borondo *et al.* [39]; to this end, the authors explore the way that the revenue (*i.e.* payoff) resulted from transactions is distributed among the types of economic agents[2] (producers, consumers, and middlemen) according to the density of economic ties (*i.e.* the network's links or edges).

When considering an ideal model such as Arrow-Debreu, every potential transaction that can create a surplus actually takes place [15]. However, such a scenario is unattainable in the real world. In a real-world economic network, most pairs of agents are not connected directly; instead, their connection is realized via chains of intermediaries — middlemen — who expect to benefit from their topological positions within the network by mediating transactions between other agents. This embeddedness of markets is especially important since forming economic links is expensive, thus further restricting markets to the structure of the social networks that co-exist with them [101]. Therefore, in a real-world market, the payoff can be classified based on the corresponding source: either from producing and selling content (*i.e.* goods or services), or from filling the role of a middleman [39], as presented in Figure 5.1.

Borondo *et al.* argue that as economic networks become more sparse a transition occurs, namely from the meritocratic (fair) distribution to a topocratic (unfair) one, where the topological placement of economic agents becomes the most important factor determining the compensation it receives [15]. Conclusively, they argue that as networks become more sparse, the topology transitions from the meritocratic regime of the Arrow-Debreu model — in which agents' income source is their individual talent — to a topocratic one, as illustrated in Figure 5.2. However, in [39] the authors assume random networks as underlying topologies, which do not represent realistic configurations of real-world markets [199].

Both topocratic and meritocratic topologies are, however, extreme cases of socio-economic networks (markets), none of which presenting an ideal situation in a real-world application [199], not to mention its redistributive consequences, would it be so [200]. This is due to the fact that possible transactions might not take place due to uncertain quality of goods on the market or search frictions [5, 33, 126, 171, 214]. As such, with the introduction of middleman, these aspects would be compensated for by their ability of reducing informa-

[2] Economic agents are represented as network nodes.

BUPT

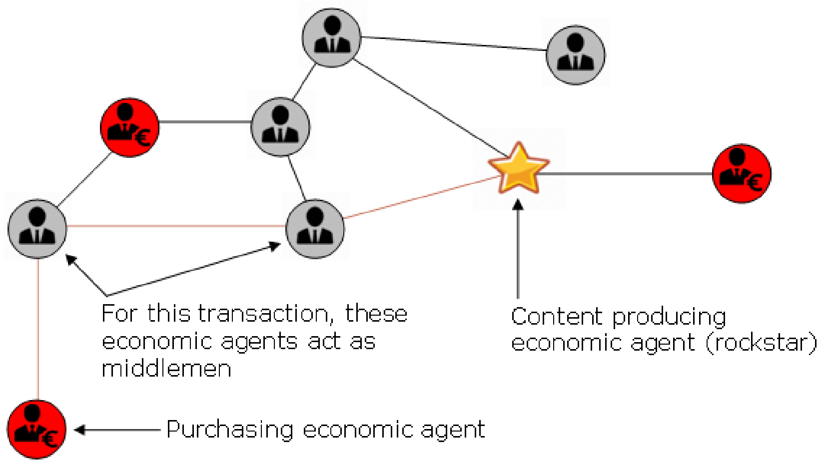


Figure 5.1 Graphical representation of a static transaction iteration presented by Borondo *et al.*: the economic agents (grey node) act as a middleman in intermediating the transaction between the content producing node (*i.e.* the *rockstar* in the rockstar model, represented by the yellow star) and the consumer (red node).

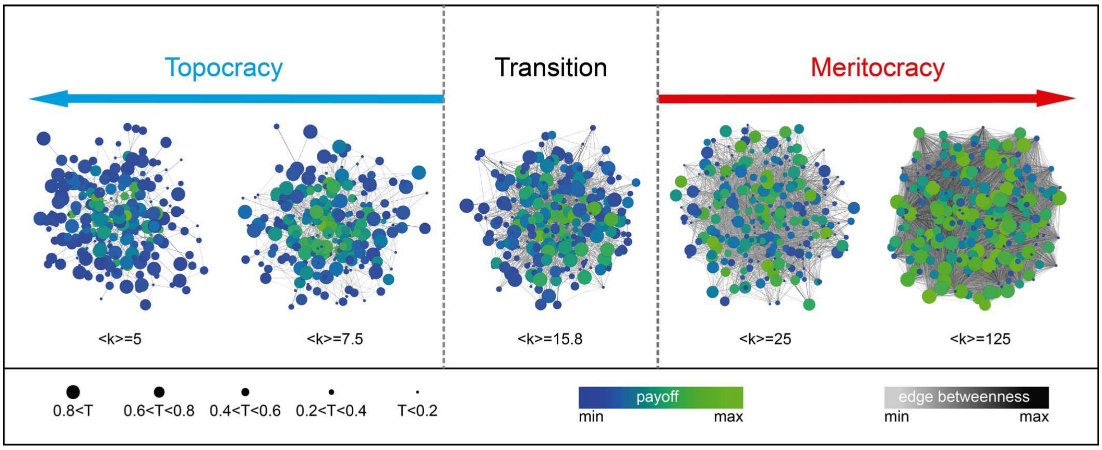


Figure 5.2 Visual representation of payoff distribution for different levels of average connectivity in a network of size $N = 250$ [39].

tion asymmetries and search frictions, or even controlling information flows in the network. Socialist Ronald Burt has made a similar remark regarding topocratic networks, pointing out that the position of a middleman in a network can be viewed as a source of advantage, as these agents are part of the ”social capital of structural holes” [42, 43]. Nonetheless, the rockstar model presents the specific issues of mathematical modeling, namely it omits a lot of real-world details. Indeed, the model is a distilled representation which is able to approximate some aspects in a real-world economic network, but cannot fully cover it [17]. Moreover, the rockstar model represents only a single iteration of economic interactions, and the economic agents (rockstars) produce only a single type of products (content) having a fixed, predefined price (value).

5.1.2 Economic Theories

Ever since the dawn of time, the conscious man realized that, starting from a certain point, he would not be able to obtain all of the goods he needed all by himself; as such,

BUPT

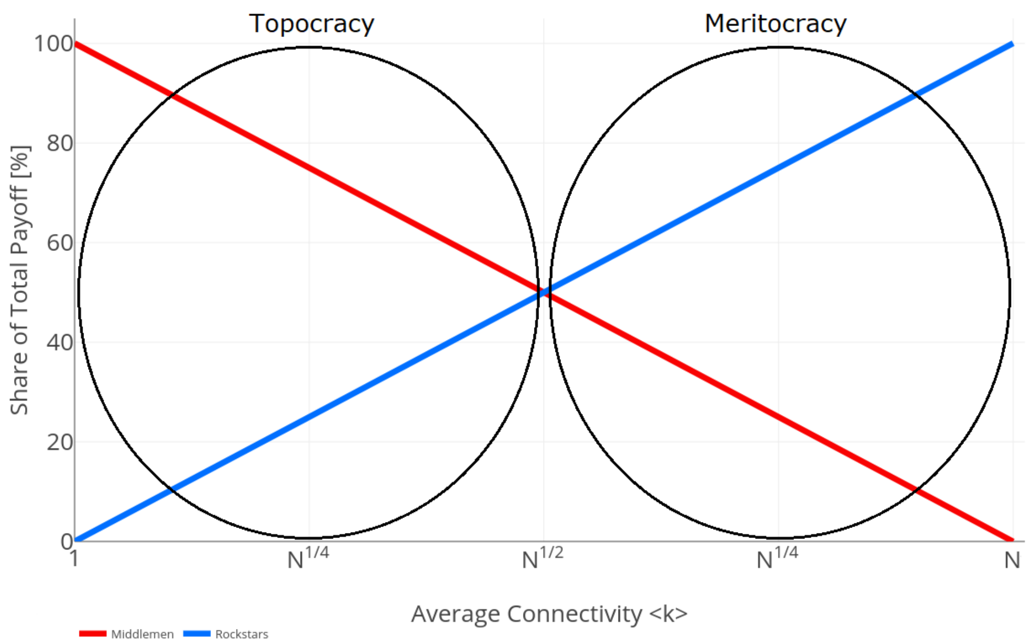


Figure 5.3 Graphical overview for share of total payoffs as a function of the average connectivity in an Erdős-Rényi network.

the exchanging of goods (products or services) began. According to Marx, the exchange value of a commodity presents itself as a quantitative relation, just like "the proportions in which commodities of one sort are exchanged for those of another sort" [135] taking into consideration its price.

In the last couple of decades more and more theories came about on how to determine the value (and consequently, the price) of a given product; however I considered two, totally opposite theories to be relevant and worth implementing them in the TrEcSim simulator, namely *the labor theory of values* and *the marginal utility theory*.

The Labor Theory of Values

A heterodox economic theory, which is usually associated with Marxian economics [135], but also used as guidelines for other, liberal economists such as Adam Smith and David Ricardo. Marx argued that the effective (economic) value of any commodity being bought, sold or exchanged is determined by the total amount of labor (measured in time and effort) required to produce it, rather than by the use or pleasure its owner receives from it [84, 136], and as such it must have the following properties [138]:

- Value: represents a well-defined quantity of human labor resulting in a commodity (product or service) under a given structure of production [136].
- Use-value: represents a given goods' usefulness, as it can satisfy human need or want, physical or ideal, being useful not just to the producer but for the consumer(s) also [138].

- Exchange-value: the relative proportion with which a commodity exchanges for another commodity, giving its owner the benefit of others' labor needed to produce the later.
- Price: the monetary expression of the exchange-value, though this could also be expressed as a direct trading ratio between two commodities without using money [137].

Using a simplified mathematical expression, the general formula to define a commodity's final value can be expressed as follows:

$$W = L + c \tag{5.1}$$

where W is the value expressed — derived from the German word *wert* —, L is the total quantity of labor time (considering an average skill and productivity) required for the production of the given product and c is the constant capital of materials used in production, including the deprecated values of the various tools and machines used [136].

The Marginal Utility Theory

Searching for a concept that would help properly determine the price of a given product, philosophers, economists, physicists and mathematicians like Aristotle, Carl Menger, Gabriel Cramer, Daniel Bernoulli or John von Neumann came to the conclusion that there is a tight relationship between rarity and utility, which would ultimately determine the final price of a product, explaining the discrepancy in the value of goods and services by reference to their secondary, or marginal utility [93, 114, 145]. As such, one of the most notable theories of value in Economy is the theory of marginality, which examines the increase in satisfaction consumers gain from purchasing and consuming an extra unit of a given good.

In the widely-accepted terms of marginalism, we can identify a specific, edge-case value of any commodity, for which it holds true given specific constraints: its marginal value. Adding extra amounts to an already existing good or services undoubtedly causes a decrease of their respective marginal value to the point where it reaches 0 (maximum utility). This economical phenomenon is known as the "law" of diminishing marginal utility [94]. More specifically, this refers to the increase in utility and individual gains from increase in the consumption of a given good or service [134, 146] and states the following:

- The marginal utility of each and every unit decreases as their supply increases, but this also holds true to the opposite; this property denotes the law of diminishing marginal utility [168].
- The marginal utility of a larger unit is greater than those of a smaller unit's, but this also holds true to the opposite; this property denotes the law increasing marginal utility [168].

Even though this theory has given birth to the famous "paradox of water and diamonds" [213], the marginal utility theory, along with all of the related theories, is successfully applied in all related fields.

5.1.3 Economic Simulators

Currently, to the best of my knowledge, there are no simulators to fully implement mainstream economic theories such as the ones previously described, however there are a few niche simulation tools that are employed for visual representation of dynamic economic networks; nonetheless, besides the mentioned flaws, the available simulators have other limiting factors, such as: limited simulation options, the lack of complex algorithms for

BUPT

modeling economic behaviour, or the necessity of supercomputers. Nonetheless, a handful of such simulation software tools worth mentioning are:

- Minsky: the open-source visual computer software for building and simulating dynamic and economic models, which is mainly used in accounting [115].
- Ecolego: a computer simulator used for creating dynamic models and performing probabilistic simulation [19, 41]. By interacting with its GUI (*i.e.* graphical user interface), users can define a handful, but limited parameters and simulation settings. Ecolego also helps to create reports and to plot simulation results.
- EMINERS: a computer simulation software for quantitative mineral resource assessment written in C++ [67]. Though initially it was capable of analyzing data for costs of labor and raw materials, costs for improving mining techniques and their (economical) advantage, as well as several beneficiation methods, it only took a handful of years until EMINERS' algorithms and modules became outdated..
- EURACE: an agent-based model capable of simulating not only single industries, markets or communities, but also the economic activities at the European Union level. The presented model is designed to factor in artificial markets for real commodities (*e.g.* consumption goods, investment goods, labor, *etc.*) when simulating a new economic scenario, as well as financial assets (*e.g.* debt securities, bonds and stocks). EURACE yields results that are identifiable in our day-to-day economical activities, by running large-model simulations, which require massively parallel computing on large supercomputers that are not available to the general public [62, 63].

5.1.4 Limitations

As TrEcSim was created by taking into account existing mathematical models, theories or economic simulators, I took the opportunity of addressing certain limitations of the presented state-of-the-art.

The Rockstar Model

In the rockstar model the number of economic agents and their roles as producer/rockstar or intermediary are predefined and fixed. Also, the agents in the rockstar model only produce one type of product, with a fixed (pricing) value, during the course of a single iteration cycle. As a result, the rockstar model is static one. Another limitation is that the rockstar model assumes only the random topology as the underlying network; however, real-life economic trade networks are not random, as they also exhibit small-world and scale free-network properties [47, 125]. Taken together, the extensions brought to the rockstar model are presented in Table 5.1.

Economic Theories

In what pertains to the mainstream economic theories, most mathematical models or economic simulators, unlike TrEcSim, don't factor them in at all. To this end, TrEcSim can account for certain products as requirements for producing other end-products. To a somewhat extent, TrEcSim also factors in the idioms of the theory of marginality by taking into consideration the products' importance and quality factors.

Feature	Rockstar model	TrEcSim
Agent-role	Predefined	Variable
Network layout	Fixed as random	Evolvable complex topologies
Product type(s)	Unique	Multiple
Product quantity and quality	Not applied	Variable
Product (pricing) value	Fixed	Variable
Profit investment	Not possible	Possible

Table 5.1 Feature-based comparison between the original rockstar model [39] and the extended simulation model in TrEcSim.

Economic Simulators

With reference to the mentioned economic simulators, it is also worth mentioning that most common drawbacks are related to simulation restrictions, or the absence of a complex algorithm capable of realistically driving the behaviour of economic agents. Moreover, some simulators require certain hardware requirements (*i.e.* large supercomputers) which are not available to the general public, rendering them unusable in most situations.

BUPT

5.2 Software Implementation

As previously emphasized, the main goal with TrEcSim is to provide a simple-to-use, web-application with powerful simulation capabilities. These attributes are not only a fundamental necessity for any successful socio-economic simulation framework, but they also ensure that TrEcSim is flexible enough to meet the requirements of tested economic scenarios. Indeed, by providing a wide variety of capabilities, configurable simulation parameters, extensible methods, complex logging, network topology importing/exporting features, and fast visual rendering, ensures that TrEcSim can be used in an effective manner, regardless of user-skill or simulated scenario.

This section focuses on a few key aspects of TrEcSim's implementation, namely: ubiquitous object-oriented design; ASP.NET integration; data manipulation and representation; as well as configurability and logging.

5.2.1 Ubiquitous Object-Oriented Design

TrEcSim is designed with flexibility in mind as its main feature; nonetheless, generally speaking, this is one of the key aspects for any simulation software. Flexibility, however, is fostered and achieved by careful implementation. To this end, I use object-oriented programming which provides for the ability to construct various configuration-instances from individual yet composable objects, thus paving the way for an advanced and fast software application. In order to offer a better insight regarding the relations between classes in TrEcSim, an overview of the class interconnectivity is presented in Figure 5.4.

Another important objective is to make TrEcSim's modules completely extensible via its API; such extensions can easily, properly, and conveniently be achieved by resorting to object-oriented implementation.

5.2.2 ASP.NET Integration

TrEcSim obtains significant flexibility and performance, by using the ASP.NET framework as foundation. From the early development phases I found that ASP.NET is appropriate for my objectives, especially when comparing against other web-specific frameworks. The advantages attributed to ASP.NET are motivated by various reasons: vast class libraries containing a large number of common functions and ready-to-use web-controls, cross-platform support, performance tweaking, smart caching technologies, drastically reduced code needed for large-scale implementations, *etc.* Because it is written in C#, TrEcSim is more powerful and more flexible; it would also be more scalable and extendable in the near future, as mentioned in section 6.3.

Another advantage gained due to the usage of ASP.NET includes the possibility of organizing data by using the popular MVC (Model-View-Controller) design pattern included in the available packages. Also, due to the inherent fact that ASP.NET has cross-language support, we can also make use of SQL (Structured Query Language) to store and retrieve both data stored procedures in and from the Microsoft SQL Server.

5.2.3 Data Manipulation and Representation

TrEcSim takes the input information regarding the underlying economic network topology (*e.g.* topological model/network type, network size, node-to-node link probability, *etc.*) via two methods: either using the built-in basic user interface, or by importing network-related data from a .csv (*i.e.* comma-separated values) file that is created by one of the numerous third-party applications (*e.g.* Gephi [27, 110]). Using the method based on importing

BUPT

topologies from .csv files, users can import any type of network, with any complexity and configuration. A general overview of TrEcSim's main interface can be seen in Figure 5.5. Additionally, in order to better convey the information presented in this section, a sequence diagram is presented in Figure 5.6, showcasing individual actions for creating a new simulation or resuming one, interacting with existing simulations (via drag and drop), as well as exporting the current state of a network.

As with any web-application, once the user issued a new simulation request on the client, this information is passed on and processed on the server. To this end, the better the hardware configuration of the server, the faster TrEcSim will be able to process the required data. Once the simulation has ended, the latest state of the network is stored in a database. The client also receives a copy of this data, where using a lightweight visualization library that is capable of handling large amounts of data (*i.e.* vis.js [11, 12]), the information is rendered for the user to see (as depicted in Figure 5.7) and visually manipulate using the drag and drop idiom.

5.2.4 Configurability and Logging

As previously stated, one of the important features present in TrEcSim is the possibility of customizing the network, as well as the various parameters before each new simulation, arranged into three main groups, as highlighted in Figure 5.8.

The simulation settings-group serves to define the global settings and behaviours of the simulation. To this end, users can configure the basic settings for the current simulation (*e.g.* product-requirement fulfillment order, percentage increase of value for each middleman, furthest distance to search for best producer, *etc.*), network configuration (*e.g.* network topology layout, initial products count, initial producer count, initial product-requirement, *etc.*) and agent behaviour (*e.g.* initial probability of investing in one of the four possible actions — expanding, creating a new product, increasing production quality, quantity or cost, or creating a new link —, enabling/disabling actions, normalization of input-values, *etc.*). TrEcSim permits either individual definition of the above settings, or automatic normalization, based on the ratio of the entered probabilistic value and their total value, essentially applying the following formula:

$$\varphi_{normalized} = \frac{\varphi_{entered}}{\varphi_{total}} \tag{5.2}$$

By using the aforementioned settings, not only does it yield a simulation tailored to one's own needs, but also approximates Borondo's rockstar model using dynamic model-simulation. Furthermore, users can also take advantage of the built-in API in order to overwrite these, and other settings stored in TrEcSim prior to starting the simulation.

All of the generated attributes related to network, economic agents and simulation, as well as individual actions are be logged and stored in a SQL database. This is, essentially, a Big Data set, offering insights regarding the changes to the network — up to the point of the current iteration-cycle —, the decisions taken and their outcome, agent-stance and much, much more. How these results are processed in the end is up to each individual user, but just like other similar applications, TrEcSim offers the possibility of resuming individual simulations or exporting and analyzing the data in its entirety for each iteration of the respective simulation. An example for the listing of current simulations can be seen in Figure 5.9.

BUPT

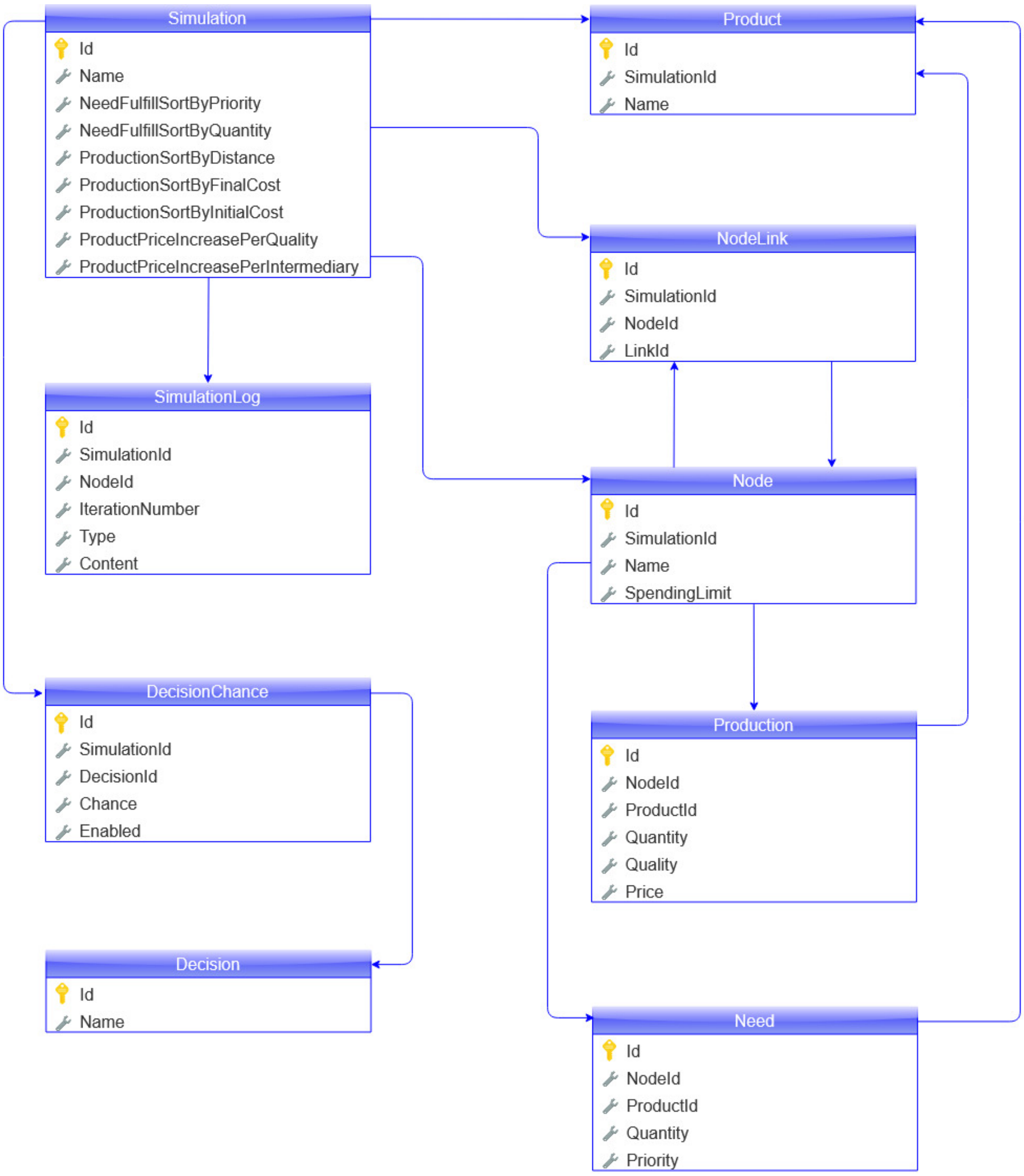


Figure 5.4 Overview of the class interconnectivity in TrEcSim using an UML (Unified Modeling Language) representation.

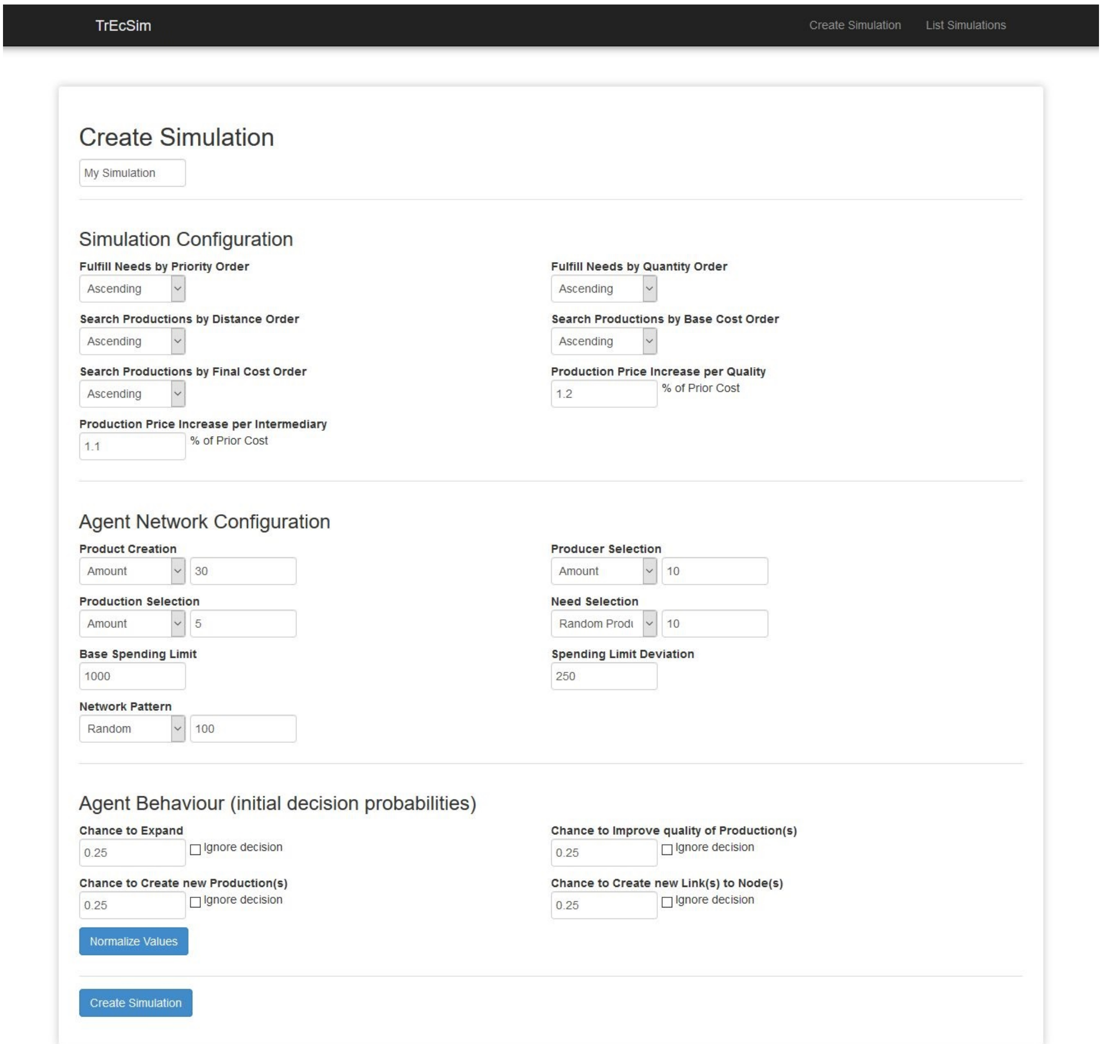


Figure 5.5 Screenshot of TrEcSim's main interface with customizable parameters for creating a new simulation. Visibly present are the three groups of settings responsible for creating a new custom simulation, namely: the overall simulation configurations, the configurations for the economic network, as well as the ones for the agent behaviour.

BUPT

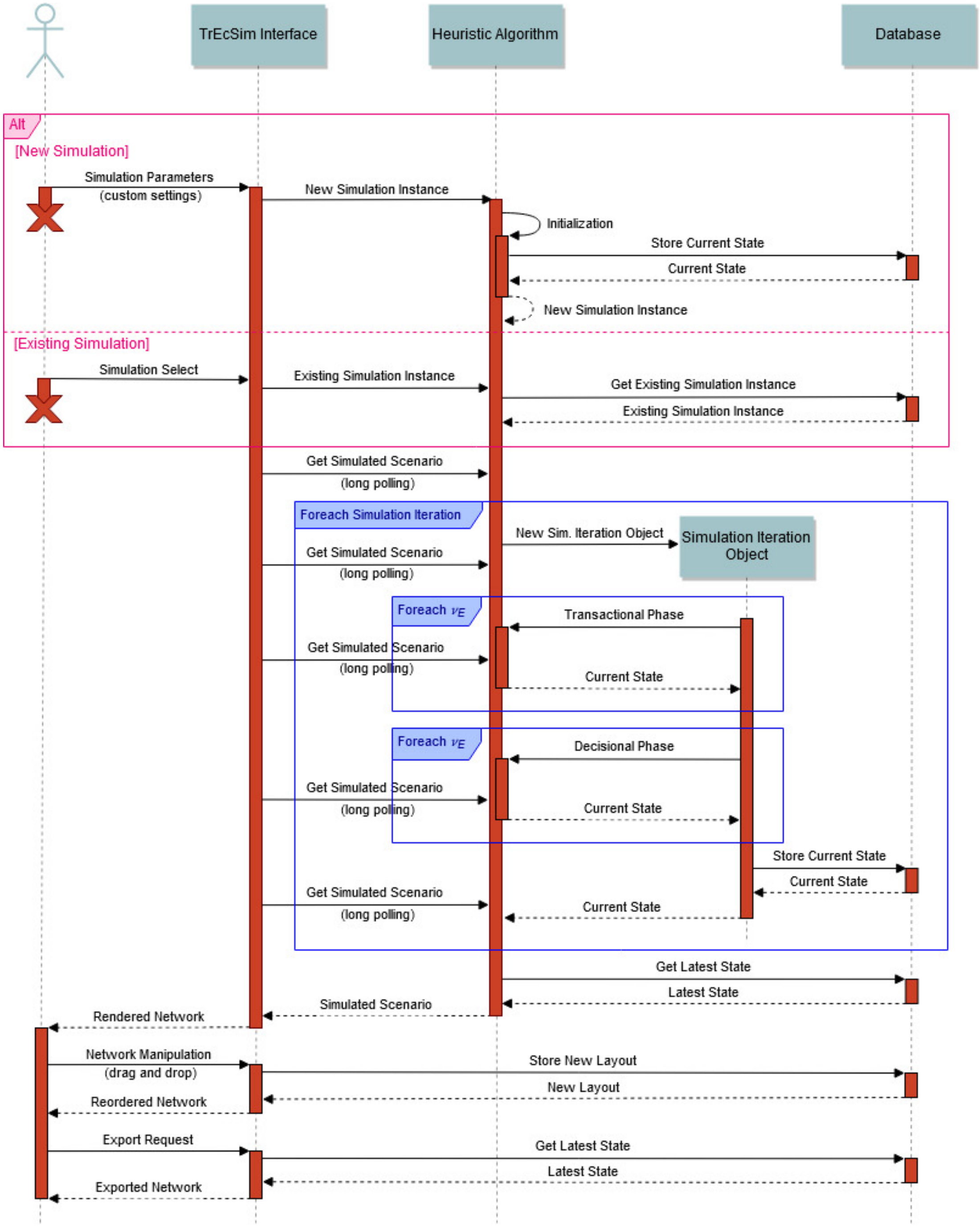


Figure 5.6 Sequence diagram showcasing the interaction between the user and the individual objects in TrEcSim: interface, algorithm and database. Noteworthy actions highlighted (from top to bottom) are: creating a new simulation or resuming one, interacting with existing simulations (via drag and drop), as well as exporting the current state of a network.

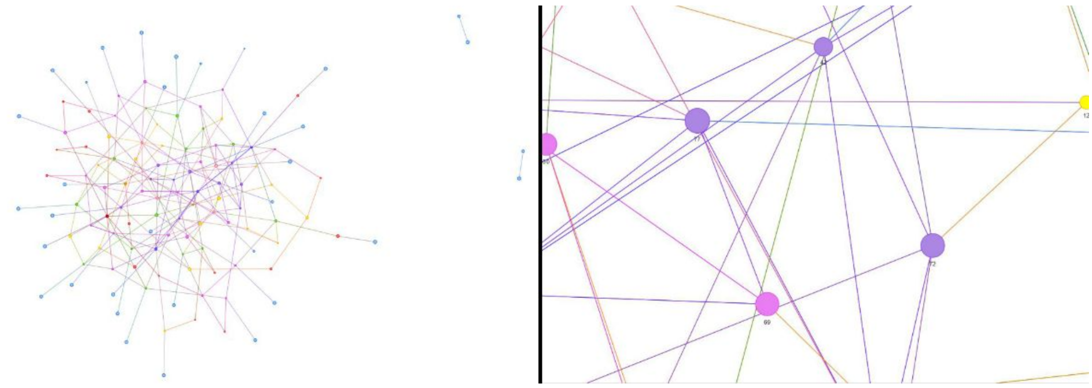


Figure 5.7 Screenshot of TrEcSim's graph-visualization GUI: overview versus close-up.

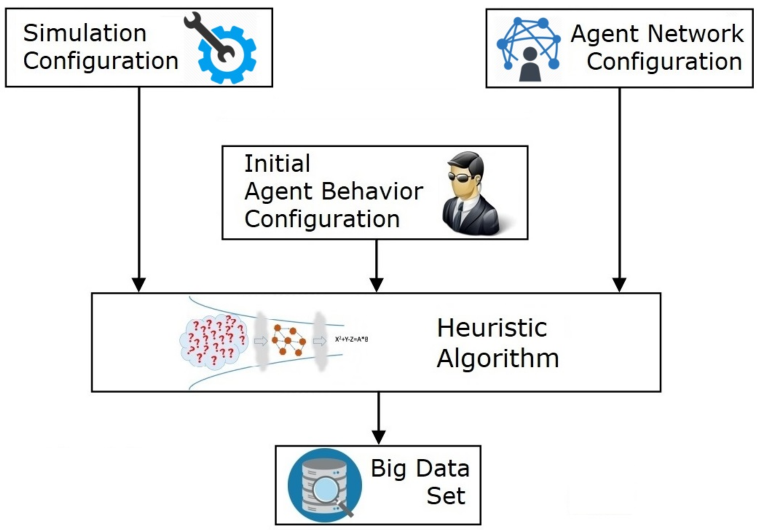
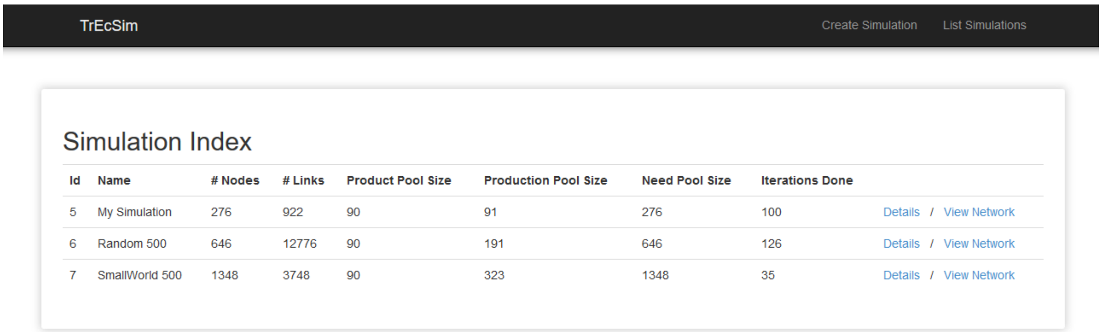


Figure 5.8 Basic diagram of TrEcSim's input and output parameters using basic black-box representation.



Id	Name	# Nodes	# Links	Product Pool Size	Production Pool Size	Need Pool Size	Iterations Done	
5	My Simulation	276	922	90	91	276	100	Details / View Network
6	Random 500	646	12776	90	191	646	126	Details / View Network
7	SmallWorld 500	1348	3748	90	323	1348	35	Details / View Network

Figure 5.9 Screenshot of current simulations, their respective attributes and available GUI options.

BUPT

5.3 Algorithmic Model Description

The main objective of TrEcSim is to provide a simulation framework that is capable of supporting an improved, dynamical model for Borondo's static mathematical model [39]. 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 Figure 5.10, 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 v_e based on network-data or import-data, *createProducts* which defines the products Pr_G and their attributes, *createProductions* which defines producers v_P and their production attributes, and *createNeeds* which defines the sub-set of demands $_G$ for each v_E. Each new iteration of the transactional phase starts with the *getBestProduction* process, to compute the best option for each v_E from where to purchase for satisfying the current needs of each v_E, while the *getAffordableQuantity* process computes the maximum quantity of each demand D. The transactional phase ends with the *finalizeTransaction* process, which finalizes the transaction for each v_E, based on the affordable quantity. To compute the best investment action for the current v_E 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 v_E economic agents. Once an affordable decision has been computed, the *makeDecision* process — the last from the decisional phase — implements the decision. Additionally, Table 5.2 represents a list of the symbols used to describe the framework.

5.3.1 Initialization Phase

TrEcSim's dynamic model starts off with a given, well-defined network topology $G = (v, \varepsilon)$ consisting of $N = |v|$ nodes/agents ($|v|$ is the cardinality of node set v) and a set of unweighted edges ε, representing economic relationships [23]. Since the roles assigned to individual economic agents (*i.e.* producers or middleman) do not exclude the possibility of an agent being both a producer and a consumer, the relationship between any two agents can go both ways, as presented in Figure 5.11. Conclusively, we consider our economic graphs as being *undirected*, regardless of their underlying topology.

The topology, used for individual simulations, is either generated directly within TrEcSim or imported as an adjancy matrix from a third-party application (*e.g.* Gephi [27]). In Figure 5.10 this is represented by the *createNetwork* process. As a result, with a fixed network topology in place, the *createProducts* process defines a set of attributes, required for transactioning various goods between economic agents. To this end, we define the following rules [23]:

- A given product Pr fulfills the demand of economic agent v_E for a given demand D_{Pr}.
- Product Pr is be needed in a certain quantity Qt_{Pr} and with a given importance I (I_{Pr}), both attributes being unique for each agent and each product.
- A product Pr is to be chosen over another based on its quality Q_{Pr}.
- Each product Pr has a specific (pricing) value V_{Pr}, which is defined using equation 5.3 for each product individually, based on the initial cost of the product ($V_{Pr_{initial}}$) and the number of products Pr in the network, Qt_{Pr_n}.

$$V_{Pr} = \frac{1}{V_{Pr_{initial}} Qt_{Pr_n}} \tag{5.3}$$

BUPT

Once the total number of products has been determined and created, the simulator proceeds to instantiate the producers. Each economic agent $v_E \in v$ is a consumer, but has a given chance of also being a producer of a sub-set of products Pr; refer to Figure 5.11 for an illustrative example. Computed during the *createProductions* process from Figure 5.10, the number of producers is determined either by explicitly defining their total number, or by their overall percentage. If the economic agent is a producer, then it is also capable of transacting its produced goods, either directly or indirectly (*i.e.* via one or more intermediate middleman).

In order to have a constant supply and demand within the network of economic agents, the sub-set of demands from the global set of demands D_G needs to be defined for each v_E agent, along with the required product $v_{E_{D_{Pr}}}$, where $v_{E_{D_{Pr}}} \in Pr_G$. This takes place in the *createNeeds* process. The demand D_{Pr_i} ($i \in \mathbf{N}$ and $i = \overline{1,n}$) is defined at the start of the simulation and is not necessarily assigned for production by a producer v_P; however, during the simulation, TrEcSim can assign its production to any agent v_E which would benefit of its production. As such, the demand sub-set is defined using one of the following options:

- Single product: a given node may or may not require a particular product, regardless if it is currently produced in the network or not.
- Single from production: similar with previous option, but with reassurance that this demanded product indeed exists (it is produced) in the network: $\forall\ v_{E_{D_{Pr}}}\ \exists\ P_G$.
- Multiple products: a given node may have several demands, which may or may not be fulfilled.
- Multiple from production: similar to the previous option, except that all products are present in the product-pool: $\forall\ v_{E_{D_{Pr}}} \in Pr_G$.

The overall simulation duration is quantified in terms of iteration cycles — a substitute for time in a real-life scenario. Each such iteration cycle consists of two phases for each agent v_E: a *transactional* phase and a *decisional* phase.

5.3.2 Transactional Phase

In the transactional phase, all existing economic agents v_E identify their current list of demands (*i.e.* products that they want to purchase), based on specific attributes of products: importance factor, quantity needed, quantity available, quality, and pricing. When faced with the possibility of choice, it is innate human behaviour to choose the product Pr with lower pricing value V, if the product has at least a certain quality Q, therefore we use the shortest path from the producer v_P to the consumer v_E agent, as each middlemen v_M influences the final cost of the transaction. As a result, in the *getBestProduction* process (Figure 5.10) each v_E agent is able to choose between settling for a product with both a lower pricing value V and with a lower quality-factor Q over a more expensive product (with a better quality), even if limited by purchasable quantity Qt. Additionally, while the starting-value (pricing) of each product Pr is defined at the start of the simulation, it's subject to change during the decisional phase. Considering the fact that each middleman also retains a certain percentage of the final revenue, it is therefore directly influenced by the number of intermediate economic agents in the current stage of the simulation. As such, by computing the shortest path between the selected producer v_P and consumer v_E — as each middlemen v_M influences the final cost of the transaction —, as well as factoring in the product attributes (importance factor, quantity needed, quantity available, quality and pricing value), possible transactions are identified for each v_E agent in the *getAffordableQuantity* process.

BUPT

Before committing the respective goods, the algorithm TrEcSim is built on validates the following properties of each possible transaction identified: required quantity needed compared to the quantity available, as well as the available currency and — in case of insufficient funds — the maximum quantity obtainable for the given amount. Conclusively, in the *finalizeTransaction* process — the last step of the transactional phase —, all of the validated transactions are finalized for each v_E economic agent.

At the end of the transaction phase, each v_E agent takes turns in initiating the identified transactions, fulfilling their demands for the required products in exchange for a total amount of currency. The amount value is then either split into multiple payments based on payoff percentage for multiple middlemen, or kept in full by the respective producer. In this last step of the transaction phase, the *finalizeTransaction* process also computes the payoff for each economic agent by applying equation 5.4, similar to the one used for computing the payoff of the content producing agent (*i.e.* rockstar) in the rockstar model[1], $\varphi_{v_{P_{total}}}$:

$$\varphi_{v_{P_{total}}} = Q_{Pr_i} \sum_{j=1}^{l} \langle k \rangle^j \, (1 - \varphi_{v_{M_{total_{pct}}}})^{j-1} \tag{5.4}$$

where l and $\varphi_{v_{M_{total_{pct}}}}$ can be computed by using equations 5.5 and 5.6, respectively:

$$l \approx \frac{lnN}{ln\langle k \rangle} \tag{5.5}$$

$$\varphi_{v_{M_{total_{pct}}}} = \frac{\sum_{i=1}^{m} \varphi_{v_{M_i}}}{100} \tag{5.6}$$

where $i \in \mathbf{N}, i = \overline{1, v_{M_{total}}}$ and $m = |v_M|$ is the number of middlemen involved in the chain of the current transaction. Conclusively, in order to compute $\varphi_{v_{M_i}}$ — *i.e.* the payoff collected by middleman v_{M_i}, where $i \in \mathbf{N}$ and $i = \overline{1, m}$ —, we can apply the following equation:

$$\varphi_{v_{M_i}} = \frac{Cf(T)}{1 + Ip_I} \tag{5.7}$$

and follow through with equations 5.8 to 5.10.

$$Cf(T) = Qt_{Pr} Cf(Pr) \tag{5.8}$$

$$Cf(Pr) = Ci_Q(Pr)(1 + Ip_I)^{v_{M_m}} \tag{5.9}$$

$$Ci_Q(Pr) = Cb(Pr)(1 + Ip_Q)^Q \tag{5.10}$$

where v_{M_m} is the last middleman in the current list of transaction. As evidenced from equation 5.9, the final cost of the product is directly influenced by v_{M_n}, where $v_{M_n} \in [0, v_n]$. This means that the closer the buyer and the producer are to each other, the less the product Pr will cost, and at the same time the more the producer can retain its payoff. Hence, in the particular case in which $n = 0$ — meaning that there are no middlemen —, the initial cost of product Pr will also be the price what the buyer will pay for it:

$$Cf(Pr) = Ci_Q(Pr) \tag{5.11}$$

[1]The equation used in the rockstar model described by Borondo *et al.* contained variables under different notations and/or descriptions and, as such, were adapted to fit the attributes implemented in TrEcSim with a similar meaning and/or role: the talent attribute T is substituted with the quality attribute Q, while α — the percentage collected by the middlemen in the current chain of transaction — is substituted by $\varphi_{v_{M_{total_{pct}}}}$

BUPT

Conclusively the payoff of the last middleman in the chain of intermediaries can be determined by applying equation 5.12:

$$\varphi_{v_{M_n}} = Cf(T) - \frac{Cf(T)}{1 + Ip_I} \tag{5.12}$$

5.3.3 Decisional Phase

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 — the essential extension to the rockstar model — where the dynamics of the network and its topology are determined [26]. It is in this phase, where the economic agents probabilistically decide which of the following actions[3] 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 — for each v_E in the network — which economic agent would be best suited to link to, in order to improve the current agent's economical stance. As already mentioned, the payoff that a given producer v_P receives in the transactional phase is directly related to quality, quantity and value of product *Pr*; this, in turn is greatly influenced by the total number of middlemen (v_{M_n}) involved in the transaction. Such a process can be modeled, at the start of the simulation, using either a probabilistic or deterministically calculated value. This value, however, is greatly influenced by its worth during the simulation process compared to all other actions. Settling for this action would also be suitable if, for instance, the demand for new or improved product is much less then the current (global) set of products: $D_{G_n} \ll Pr_{G_n}$.

- **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 ever-growing demands: $D_{Pr_n} \gg Pr_{G_n}$.

- **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.*

5.3.4 Agent Behaviour

In order to implement a realistic process of determining the agent's course of investment by selecting one of the four available actions, I create a new mechanism for modeling the behaviour of economic agents, inspired by the tolerance-based interaction model [196]. The rationale behind this approach is that the tolerance-based interaction model represents a dynamic model of opinion spread/contagion, which produces results that are similar to those

[3] For further information pertaining the implementation of the four actions from a programmatic point of view, please refer to the code snippets presented in appendix A

BUPT

of real-world systems. Hence, I consider that the agent's attitude towards economic action is similar to holding an opinion on a given matter, being influenced by individual's previous experiences but also by other individuals with whom it interacts. Therefore, I implement in TrEcSim the possibility of influencing — and being influenced by — other neighboring v_E economic agents, according to the tolerance-based model (represented by the *getPastDecisionScoresForNeighbors* process in Figure 5.10) [196].

A similar economic behaviour was also observed and documented by others, the most notable of which was Herbert A. Simon, a multi-disciplinary pioneer. He introduced the heuristic decision-making strategy titled "satisficing" [181] — a combination of satisfy and suffice [132] — in order to explain the rationale of economic agents under circumstances in which an optimal solution cannot be established given limited resources. As a result, economic agents will settle either for the first option that meets a given need, or the option that seems to address most needs, rather than the optimal solution [57], which is present only in an ideal world [182]. Similar observations have also been made by Daniel Kahneman and Amos Tversky, notable for their psychology of judgment and decision-making, as well as behavioural economics. Their work explored the biases and failures in rationality, which is systematically exhibited in human decision-making [192], namely that humans are irrational beings, systematically making choices that defy clear logic (by not weighing up the facts), only to improve on them on the long run as a result — sometimes even with the help of other members of society [111].

Starting from the second simulation iteration, at the beginning of the decisional phase for each agent v_E, economic agents interact with a random number of their neighbors and become influenced by the last investment these neighboring agents have made. The number nb of neighboring agents the current economic agent interacts with (from the total of nb_{max} neighbors) is determined randomly. As a result, the economic agent is able to recalculate the probability of investing in a specific action. However, economic agents don't interact (*i.e.* influence economic behaviour of other agents) in every iteration cycle; in fact, after the interaction, each agent is prohibited from further interacting for a random number of iterations. This mechanism is implemented by a random timeout interval between 1 and 50 simulation cycles.

According to the presented scenario, each agent v_E uses the following expression in order to compute the new probabilities for investing in each A_i actions (where $i \in \mathbf{N}$ and $i \in [1, 4]$):

$$A_{i_{new}} = 0.5A_{i_{current}} + \sum_{j=1}^{nb} A_{i_{current_j}} \frac{0.5}{nb} \tag{5.13}$$

Once the *determineCurrentDecision* process has computed the best action for the v_E economic agent to invest in — based on both the *getPastDecisionScores* and the *getPastDecisionScoresForNeighbors* processes —, the *makeDecision* process implements the selected A_i.

5.3.5 Algorithmic Model Complexity

In order to measure the complexity of the heuristic algorithm behind TrEcSim and to determine the average simulation duration, I first have to identify and isolate key network properties, which might impact the simulation run-time. As a result, I consider that the following two scenarios might yield valuable information pertaining the complexity of TrEcSim:

- Simulate networks with distinct densities — the density is the number of links in the networks, normalized by the total number of possible links $N(N-1)$, where N is the number of nodes —, but with a fixed number of economic agents.

BUPT

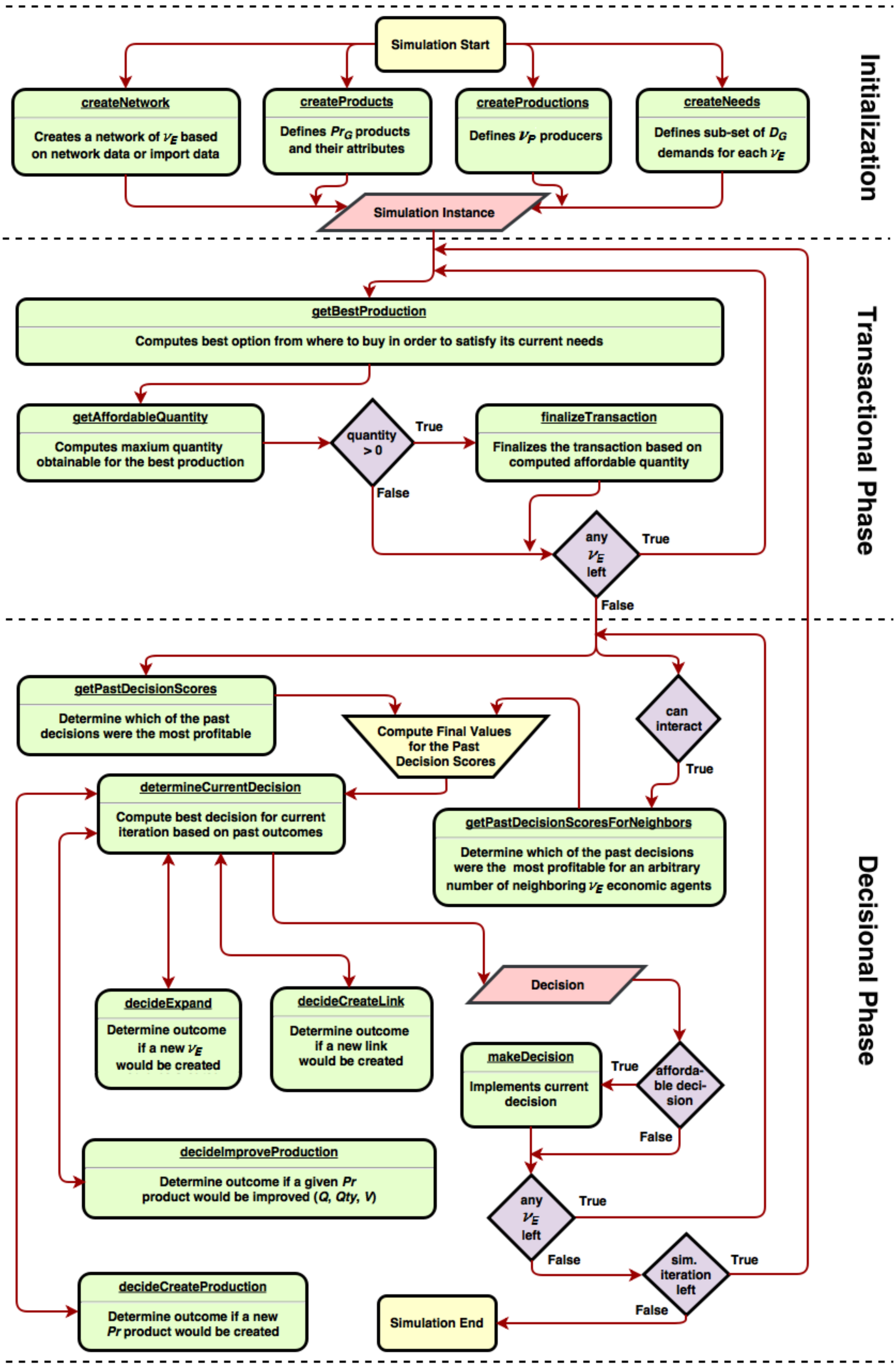
- Simulate networks with linearly increasing number of economic agents (1000 to 1900 nodes), but with the same density.

After running both scenarios for 600 iterations — I determined this threshold through empirical testing needed to obtain a stabilized network, regardless of topology or density —, and averaging the results over 10 independent networks, I obtain the results depicted in Figure 5.12 and Figure 5.13; as such, the evolution of time for the two mentioned scenarios follows a polynomial (*i.e.* cubic) growth and a linear growth, respectively. In both cases, the evolution in time indicates that the current implementation of the algorithm does not require significant computational resources and is indeed scalable.

Symbol	Interpretation
v_E	Economic agent $\in v$
v_P	Producer $\in v$
v_M	Middleman $\in v$
Pr	Product
Pr_G	All products produced globally in the network; $Pr \in Pr_G = \{Pr_1, Pr_2, Pr_3, \cdots, Pr_i, \cdots, Pr_n\}$, where $i \in \mathbf{N}$ and $i = \overline{1, n}$
Qt_{Pr}	Quantity of product Pr
Qt_{Pr_G}	Global product quantities in the network; $Qt_{Pr} \in Qt_{Pr_G} = \{Qt_{Pr_1}, Qt_{Pr_2}, Qt_{Pr_3}, \cdots, Qt_{Pr_i}, \cdots, Qt_{Pr_n}\}$, where $i \in \mathbf{N}$ and $i = \overline{1, n}$
I_{Pr}	Importance of product Pr
I_{Pr_G}	Global product importance (factors) in the network; $I \in I_{Pr} = \{I_{Pr_1}, I_{Pr_2}, I_{Pr_3}, \cdots, I_{Pr_i}, \cdots, I_{Pr_N}\}$, where $i \in \mathbf{N}$ and $i = \overline{1, n}$
Q_{Pr}	Quality of product Pr
Q_{Pr_G}	Global product qualities in the network; $Q_{Pr} \in Q_G = \{Q_{Pr_1}, Q_{Pr_2}, Q_{Pr_3}, \cdots, Q_{Pr_i}, \cdots, Q_{Pr_n}\}$, where $i \in \mathbf{N}$, $i = \overline{1, n}$ and $Q_{Pr_i} \in [0, 100]$
D_{Pr}	Demand for product Pr
D_{Pr_G}	Global product demands in the network; $D_{Pr} \in D_G = \{D_{Pr_1}, D_{Pr_2}, D_{Pr_3}, \cdots, D_{Pr_i}, \cdots, D_{Pr_n}\}$, where $i \in \mathbf{N}$, $i = \overline{1, n}$ and $D_{Pr_i} \in [0, 100]$
$v_{E_{D_{Pr}}}$	Economic agent's (v_E) demand for product Pr
V_{Pr}	Pricing (value) of product Pr; $V_{Pr} \in V_{Pr_G} = \{V_{Pr_1}, V_{Pr_2}, V_{Pr_3}, \cdots, V_{Pr_i}, \cdots, V_{Pr_n}\}$, where $i \in \mathbf{N}$, $i = \overline{1, n}$ and $V_{Pr_i} \in \mathbf{R}_+$
l	Cutoff
$\langle k \rangle$	Average degree of network
T	Transaction
$\varphi_{v_{M_{total_{pct}}}}$	Percentage of payoff collected by the middlemen v_M participating in a given transaction T
$Cf(T)$	Final cost of a given transaction T
$Cf(Pr)$	Final cost of product Pr
Ip_I	Increased price per intermediary (v_M)
$Ci_Q(Pr)$	Initial cost for producing product Pr, taking into consideration its quality Q
$Cb(Pr)$	Base cost of product Pr
Ip_Q	Increased (new) price of the product for each quality value Q
A	Investment action $\in \{A_1, A_2, A_3, A_4\}$
$A_{i_{new}}$	New probability of investing in action A_i; $i \in \mathbf{N}$ and $i \in [1, 4]$
$A_{i_{current}}$	Current probability of investing in A_i; $i \in \mathbf{N}$ and $i \in [1, 4]$
$A_{i_{current_j}}$	Neighboring agent j's current probability of investing in A_i; $i \in \mathbf{N}$ and $i \in [1, 4]$, j is the index of the nodes that are selected neighbors of the agent v_E
nb	Number of neighboring economic agents the current v_E interacts with; $nb \in \mathbf{N_0}$ and $nb = \overline{0, nb_{max}}$ (nb_{max} is the total number of neighbors)

Table 5.2 List of used symbols, along with their associated interpretations used in the TrEcSim model.

BUPT


Simulation Start
createNetwork
Creates a network of $\mathcal{V}_E$ based on network data or import data
createProducts
Defines Pr_G products and their attributes
createProductions
Defines $\mathcal{V}_P$ producers
createNeeds
Defines sub-set of D_G demands for each $\mathcal{V}_E$
Simulation Instance
Initialization
getBestProduction
Computes best option from where to buy in order to satisfy its current needs
getAffordableQuantity
Computes maxium quantity obtainable for the best production
quantity > 0
True
False
finalizeTransaction
Finalizes the transaction based on computed affordable quantity
any $\mathcal{V}_E$ left
True
False
Transactional Phase
getPastDecisionScores
Determine which of the past decisions were the most profitable
Compute Final Values for the Past Decision Scores
can interact
True
determineCurrentDecision
Compute best decision for current iteration based on past outcomes
getPastDecisionScoresForNeighbors
Determine which of the past decisions were the most profitable for an arbitrary number of neighboring $\mathcal{V}_E$ economic agents
Decision
decideExpand
Determine outcome if a new $\mathcal{V}_E$ would be created
decideCreateLink
Determine outcome if a new link would be created
makeDecision
Implements current decision
True
afforda-ble deci-sion
False
decideImproveProduction
Determine outcome if a given Pr product would be improved (Q, Qty, V)
any $\mathcal{V}_E$ left
True
False
sim. iteration left
True
False
decideCreateProduction
Determine outcome if a new Pr product would be created
Simulation End
Decisional Phase

BUPT

Figure 5.10 (previous page) Flowchart describing the implementation of our extended model within the TrEcSim framework. Visually delimiting the three main components (*i.e.* initialization phase, transactional phase and decisional phase), the flowchart highlights the individual processes — which implement the components of the simulator — and the relations between them [23].

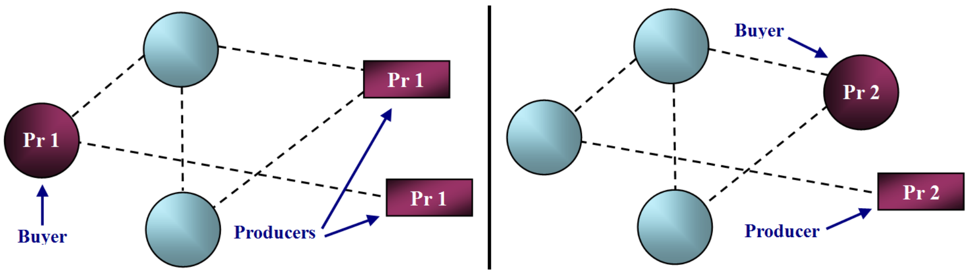


Figure 5.11 Visual representation of different user-roles in the economic network and of how roles can change when taking into consideration different products. An economic agent v_E producing Pr_1 (represented with a violet rectangle) may also be a v_E in need of Pr_2 — produced by the same agent —, while all other agents, who may or may not be middlemen in these two transactions, are represented with a blue circle [25].

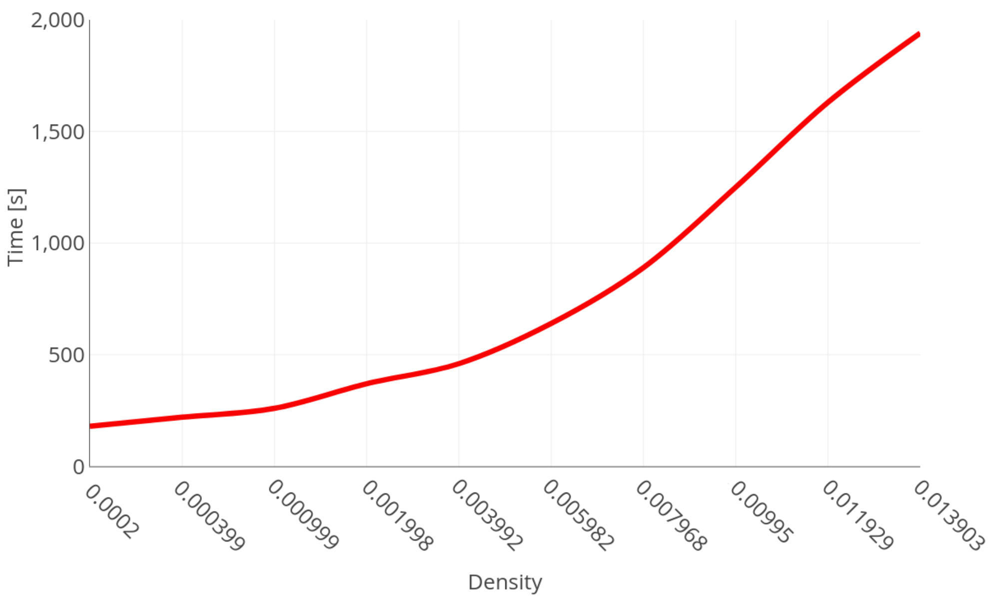


Figure 5.12 Graphical representation of the evolution in time for networks of varying densities.

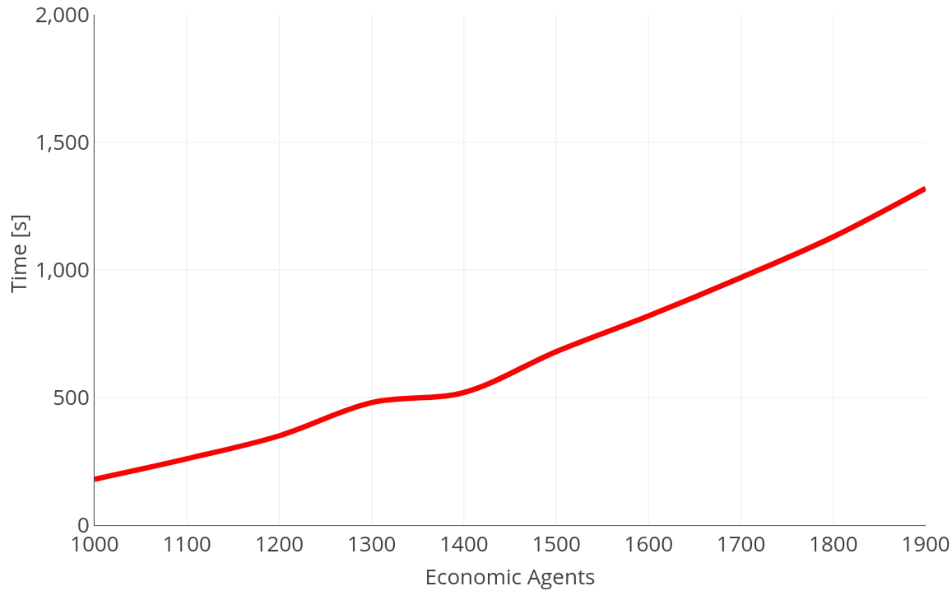


Figure 5.13 Graphical representation of the evolution in time for networks of varying number of economic agents.

BUPT

Network ID	Nodes	Edges	$\langle k \rangle$	Dns
1	100K	1M	20	0.0002
2	100K	2M	40	0.0004
3	100K	5M	100	0.001
4	100K	10M	200	0.002
5	100K	20M	400	0.004
6	100K	30M	600	0.006
7	100K	40M	800	0.008
8	100K	50M	1000	0.01
9	100K	60M	1200	0.01
10	100K	70M	1400	0.014

Table 5.3 Edge-count, average degree $\langle k \rangle$, and density Dns for each of the 10 distinct synthetic networks of size N=100,000 nodes and distinct fundamental topologies. All 10 density configurations are used for each synthetic topology considered in this paper (*i.e.* mesh, small-world, random and scale-free).

5.4 Simulating Trade in Economic Networks

First, in subsection 5.4.1 I present the simulation results and the analysis for economic transactions using the extended rockstar model. Then, in subsection 5.4.2 I observe the preferred choice of investment for each economic agent when face with limited actions at their disposal. In subsection 5.4.3 I analyze the distribution of payoff based on agent role, while subsection 5.4.4 introduces the novel approach to determining and analyzing the fairness of economic exchange networks.

In all four subsections I make use of the same network properties and initial conditions for each new simulation in order to obtain accurate results. As such, I generate 10 distinct networks (*i.e.* with different densities) for each of the following fundamental topologies: 2D mesh [32], small-world (Watts-Strogats) [210], random (Erdős-Rényi) [71] and scale-free (Barabási-Albert) [22] networks; each network has a size of 100,000 nodes.

The generated networks are summarized in Table 5.3 and are created with the intent of simulating two distinct cases for each topology, one where the producers are assigned randomly to the existing v_E 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. The network density Dns represents the ratio between the actual number of edges in the network $E = |\varepsilon|$ and the highest possible number of edges in a network with N nodes (*i.e.* number of edges in a fully connected network):

$$Dns = \frac{E}{\frac{N(N-1)}{2}} \tag{5.14}$$

Each simulation is run with the exact same custom-defined settings. As such, I custom-define the following network properties: the initial number of economic agents (1,000), initial number of links between them (5,328), initial product-types (30) and initial production count (45 units/product type, 1,350 in total). The other parameters are left at their default value, as listed in Table 5.4.

All simulations are run for 600 iterations; I determined this threshold through empirical testing needed to obtain a stabilized network, regardless of topology. Finally, the results

BUPT

Settings-group	Parameter	Value
Simulation configuration	Fulfill needs by priority order	Ascending
	Search productions by distance order	Ascending
	Search productions by final cost order	Ascending
	Productions price increase per intermediary	1.1
	Fulfill needs by quantity order	Ascending
	Search productions by base cost order	Ascending
	Production price increase per quality	1.2
Agent network configuration	Product creation	30 (Amount)
	Production selection	45 (Amount)
	Base Spending limit	1000
	Network pattern	Import .gdf file
	Producer selection	10 (Amount)
	Need selection	10 (Random product)
	Spending limit deviation	250
Agent behaviour	Chance to expand	0.25*
	Chance to create new production(s)	0.25*
	Chance to improve quality of production(s)	0.25*
	Chance to create new link(s) to node(s)	0.25*

Table 5.4 Settings used for each new simulation with TrEcSim. The values marked with (*) change during the simulations described in 5.4.2.

presented in the upcoming charts are obtained by averaging 10 independent simulations for each network topology.

5.4.1 Results Obtained 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) [39], 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, by analyzing the evolution of payoff for both cases, I can clearly identify the

BUPT

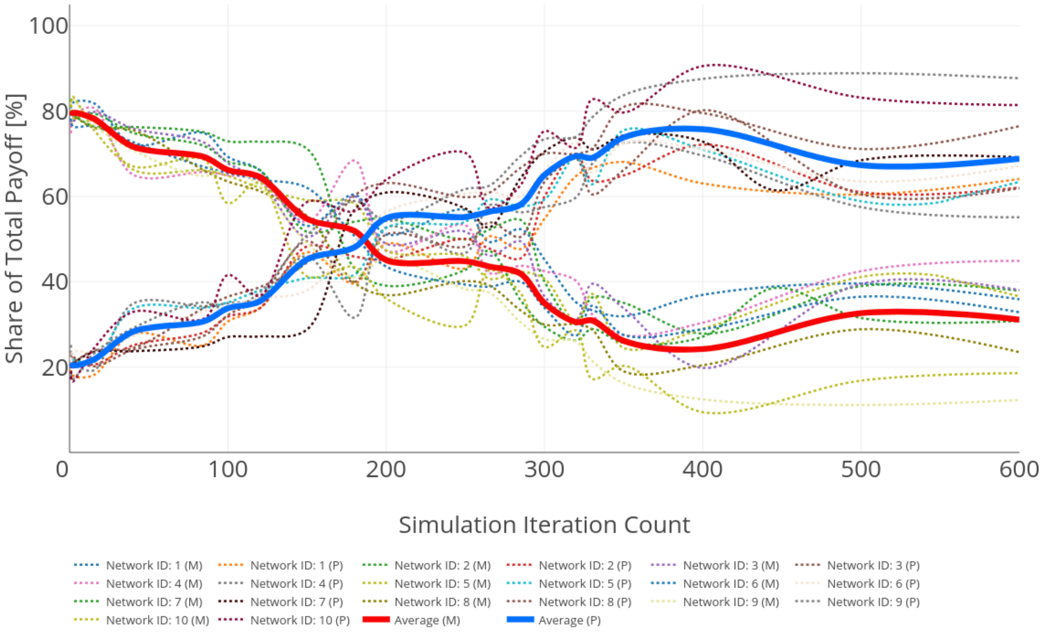


Figure 5.14 Share of the total payoff evolution with iteration count in a mesh topology for producers (P) and middlemen (M). The producers were randomly assigned from the pool of economic agents throughout the network [25].

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 v_E 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 v_E agent.

At first sight, the simulation results obtained for the extended model — especially for the preferential assignment of the agent-roles — contradict the findings presented by Borondo *et al.* in [39], namely that an increased network density alone automatically leads towards a meritocratic environment in economic exchange networks; indeed, while increasing network density might eventually shift the respective network from a topocratic environment to a meritocratic one, it only truly impacts the outcome of the share of total payoff — which from an economic point of view is not that relevant on itself —, and not that of the payoff of each v_E economic agents itself. If, however, we would like to obtain relevant information regarding the topic of meritocracy versus topocracy in economic exchange networks, we have to analyze the given network from other perspectives as well, most important of which is the payoff distribution at agent-level.

5.4.2 Limiting Actions of Investment

Whilst the previous charts accurately depict the share of total payoff for both producers and middlemen for networks of a certain topology, they do so by enabling the v_E economic agents to invest in all available actions, most importantly in creating new links. As such, in order to better differentiate the behaviour of the v_E in the network when limiting certain (emergent) behavioural aspects and analyzing the results, I simulate two distinct scenarios for all network topologies, namely:

BUPT

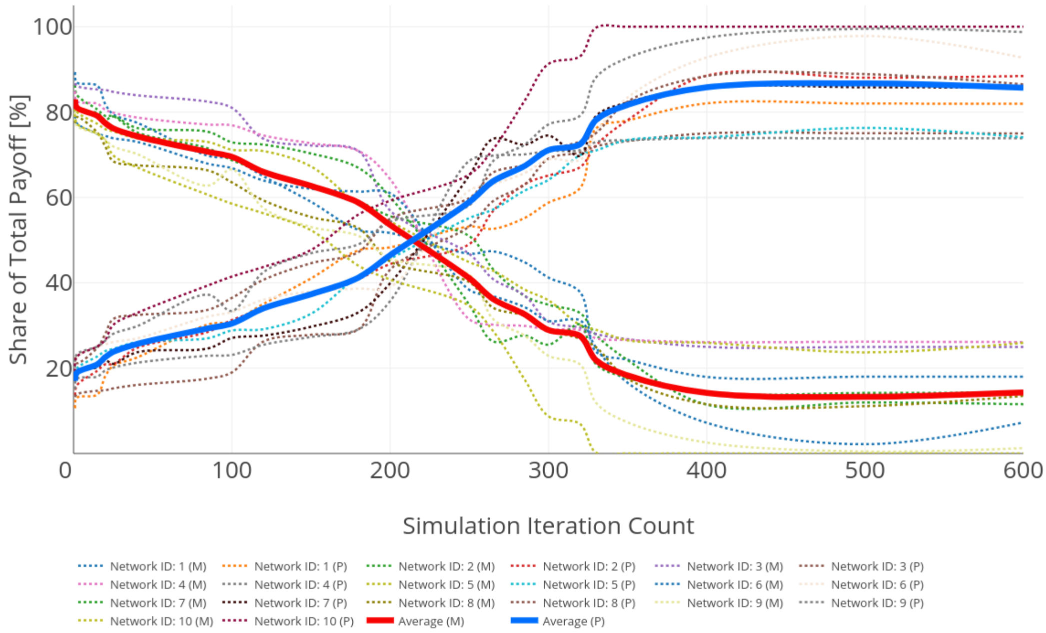


Figure 5.15 Share of the total payoff evolution with iteration count in a small-world topology for producers (P) and middlemen (M). The producers were randomly assigned from the pool of economic agents throughout the network [25].

- **Scenario 1**: all settings are left at default values — as listed in Table 5.4 —, but economic agents are limited to Actions 2, 3 and 4 to invest in. All of these three investment-actions have the same probability ($\approx 33\%$), while Action 1, the possibility of creating new links between economic agents is set to 0%.
- **Scenario 2**: much like the previous scenario, but instead of allowing the possibility of creating new v_E economic agents (*i.e.* Action 4), agents are allowed to create new links, thus enabling Actions 1, 2 and 3.

Mesh Topology

Assigning the roles of producers randomly throughout the mesh network topology, the analysis of the simulation results for Scenario 1 yields interesting results indeed. Conclusively, 27 new products are created, totalling to 57 products (90% increase), while the number of production units increases to 2392 (77% increase). Upon close analysis, the number of increased products and production units is the result of the v_E economic agents not being able to invest in the most efficient option, namely creating new links throughout the network (*i.e.* Action 1). The number of economic agents does increase however to 1335, meaning an approximated increase of 34%. For Scenario 2, 8,723 new links were created (*i.e.* Action 1), effectively increasing by 156% and as a result, being the most preferred option of investment. The number products throughout the network increased also by 40% (to a total of 42), while individual production count increases to 89 units (98%).

For the second case, when producers are assigned preferentially to v_E economic agents, simulating Scenario 1 yields the following results: the total number of products increases to 59 (97% increase), each with 55 individual production units, meaning an increase of

BUPT

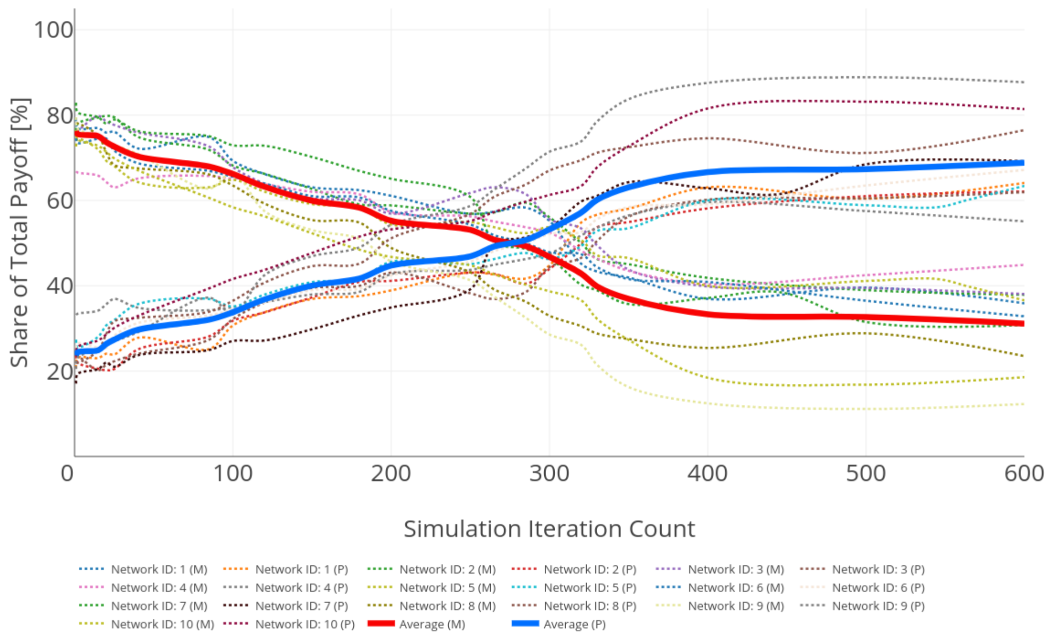


Figure 5.16 Share of the total payoff evolution with iteration count in a random topology for producers (P) and middlemen (M). The producers were randomly assigned from the pool of economic agents throughout the network [25].

22%. In Scenario 2, TrEcSim increases the number of links in the network to 17,109 (221% increase), the number of Pr products to 52 (73% increase), and the total number of units to 1,805 (roughly by 34%). The similar choices of investment and their associated outcomes can be interpreted as follows: having almost the same average degree as any random node in a small-world network, economic agents in a random network take advantage of their increased production quality or lower prices per each unit of production; the middleman, on the other hand, have no other choice but to increase the price for each product intermediated or risk losing their payoff. To this end, production of individual units has also increased in number, namely to 89 units (an increase of 98%).

Small-World Network Topology

After 600 iterations, assigning the producers randomly to the existing v_E economic agents and using the hypothesis for Scenario 1, the simulation creates 294 (30%) new economic agents, implicitly linking them to the agent choosing the expansion. The economic agents, without the possibility of shortening their path to the buyer (*i.e.* investing in Action 1), choose by a very large margin to invest in increasing production count instead, generating a total of 2529 units, representing an increase of 87%. This is of no surprise, as choosing to increase production for the same cost (*i.e.* Action 3) is the only remaining viable action, leaving only a couple of agents to invest in new products (*i.e.* Action 2), increasing total product count to 52 (a 73% increase). Using Scenario 2, I identify 9,335 new links, meaning an increase of 175%. This strongly suggests that the custom algorithm behind TrEcSim — just like in any real-life scenario — deems it more advantageous (from an economic standpoint) to extend each producer's reach in the network. As such, by investing in Action 1, that particular v_E economic agent gains several advantages. One of these

BUPT

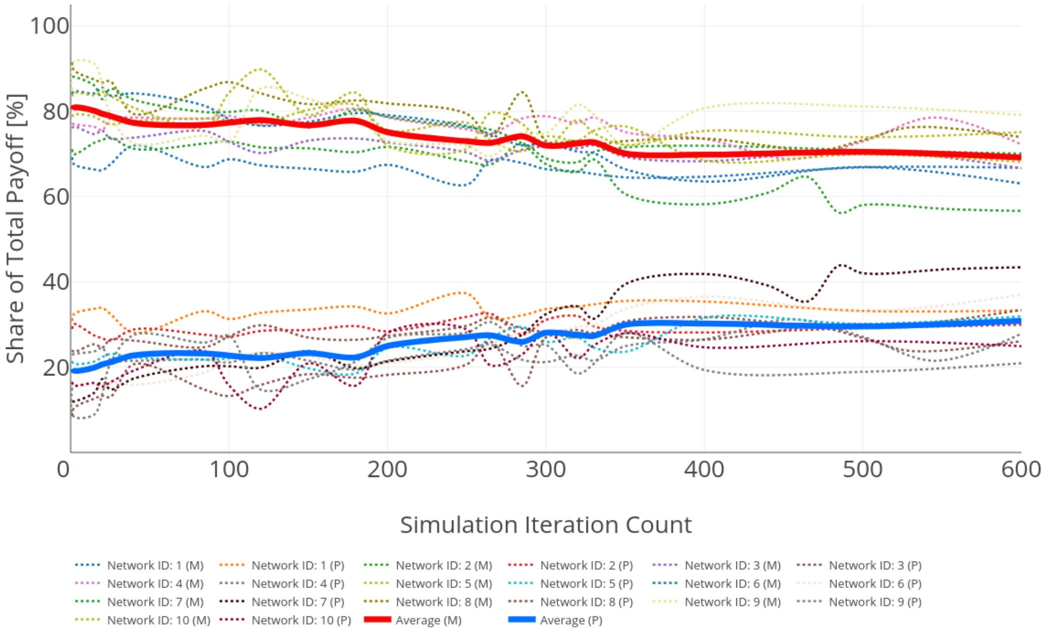


Figure 5.17 Share of the total payoff evolution with iteration count in a scale-free topology for producers (P) and middlemen (M). The producers were randomly assigned from the pool of economic agents throughout the network [25].

advantages is the retention of an ever-growing payoff for the producers, compared to that of the middlemen. Another advantage when agents are allowed to invest in creating new links is the ability to sell the same amount of products cheaper, or — even better from an economic point of view — to sell more products for the same amount of production costs. To this end, production increases to 92 units, representing an increase of 105% for each product, supplemented by an increase of 9 new products being produced (30%).

For the second case, when producers are assigned preferentially and using the hypothesis for Scenario 1, the simulation creates 415 (42% increase) new v_E agents economic agents (*i.e.* Action 4), linking them to the agent choosing the expansion. Similar the previous results, increasing production count (*i.e.* Action 3) was the preferred investment, obtaining a total of 3,729 units (representing an increase of 176%), whereas the total number of Pr products increased to 49 (63% increase, Action 2). For Scenario 2 TrEcSim creates 6,615 new links, increasing the total number of links in the network by 105%(*i.e.* investing in Action 1), 39 Pr products (30% increase) and a total of 1,825 of units (35% increase).

Random Network Topology

After simulating Scenario 1, the number of Pr products increases to 52 (an effective increase of 73%), while the total number of production units increases to 1409 (an estimate increase of 4%). Similar to the simulation corresponding to the small-world topology, the only viable option the agents have is to invest in either Action 2 or Action 3, compared to the benefit a new link would bring by choosing Action 1. Nonetheless, producers eventually start obtaining more after each transaction compared to middlemen, adopting an effective strategy against a topocratic, oligarchic society. Similar to the previous topology, when

BUPT

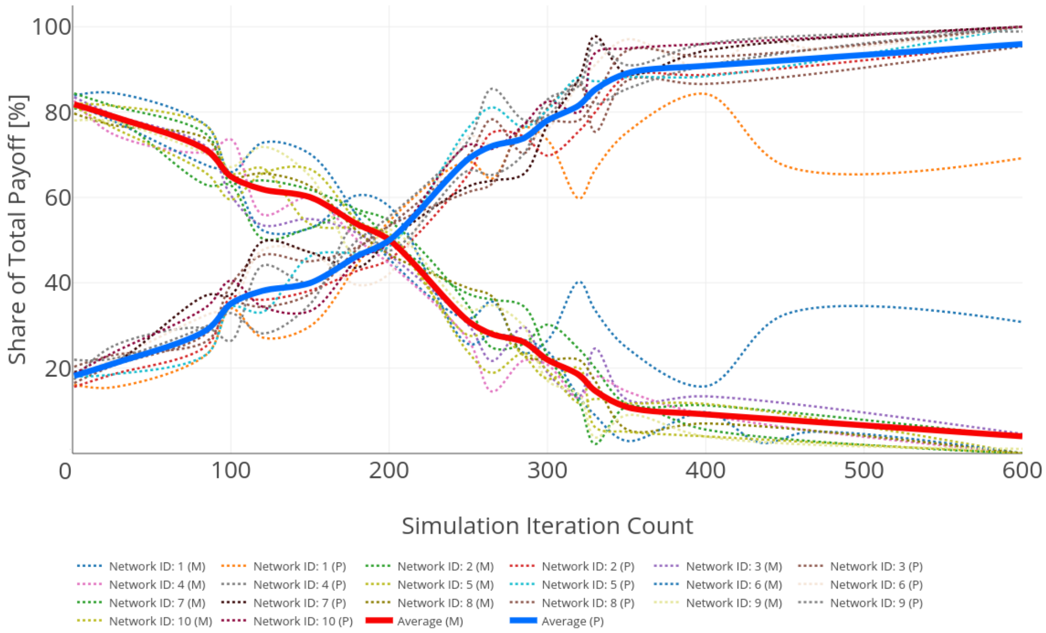


Figure 5.18 Share of the total payoff evolution with iteration count in a mesh topology for producers (P) and middlemen (M). The producers were assigned preferentially to the agents with highest degrees [25].

considering the conditions for Scenario 2, TrEcSim slowly but consistently increases the number of links in the network to 104,072, obtaining an effective increase of 29%.

Assigning the producers preferentially throughout the network and simulating Scenario 1 with the same starting conditions, the number of Pr products increases to 67 (an effective increase of 123%), while the total number of production units — fairly similar to the initial simulations — increases to 1,604 (19% increase). The number of economic agents present increases also, namely to 1271 (27% increase). In Scenario 2, TrEcSim increases the number of links in the network to 17,109 (221% increase), the number of Pr products to 52 (73% increase), and the total number of units to 1,805 (roughly by 34%). The similar choices of investment and their associated outcomes can be interpreted as follows: having almost the same average degree as any random node in a small-world network, economic agents in a random network take advantage of their increased production quality or lower prices per each unit of production; the middleman, on the other hand, have no other choice but to increase the price for each product intermediated or risk losing their payoff. To this end, the total number of units has also increased in number, namely to 1541 units (an increase of 14%).

Scale-Free Network Topology

The most interesting result from all simulations pertains to the scale-free topology. In this case, the evolution of the payoff for the producers and middlemen is of more importance, due to the fact that the individual economic agents of a scale-free network are linked to so-called hubs; as a result, the well-positioned middlemen take advantage of their topocratic position within the network. This fact clearly shows in all generated figures pertaining the scale-free topology, as well as its analysis.

BUPT

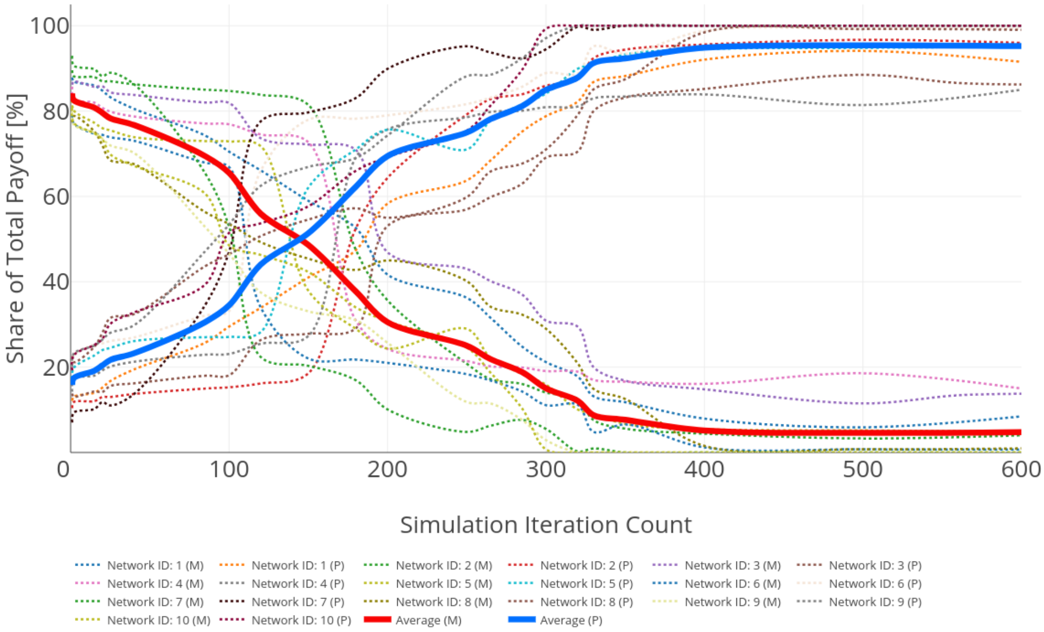


Figure 5.19 Share of the total payoff evolution with iteration count in a small-world topology for producers (P) and middlemen (M). The producers were assigned preferentially to the agents with highest degrees [25].

When assigning producers randomly throughout the network and applying the conditions for Scenario 1, the number of Pr products increases to 34 units, meaning an increase of 13%, while individual product count increases to 2085 units (an increase of 54%). The development of this simulation yields an interesting observation, namely that even after the 600 cycles of simulation the producers do not manage to rival the payoff gained by the middlemen; this, in turn, strongly supports the hypothesis that the clustering is still present in the network, keeping the remaining middlemen in favorable positions. In Scenario 2, where economic agents are allowed to create new links, the number of links increases to 19,224, meaning an effective increase of 522%, representing the preferred choice of investment. Only when they manage to create enough new links within the network does the payoff gained by the producers rival that of the middlemen. To this end, in this scenario, I observe only a relatively small increase of products (14 new Pr products, roughly equivalent to 47%), and with a final production count of 73 units for each Pr product, representing an increase of 62%. Much like in the case of the previous two network topologies, agents do not have the option to expand, thus providing incentives for increasing the number of economic agents in the network.

When the producers are assigned preferentially, the number of Pr products increases to 59 (97% increase, almost double to that of the initial simulation), while the total product count increased to 1,611 units (an increase of 19%). The number of v_E agents present in the network increases only by 7%, to a total of 1,072. This is true when considering the conditions for Scenario 1. For Scenario 2, where economic agents are allowed to create new links, the number of links increases to 21,701, meaning an effective increase of 307% — by far the preferred choice of investment, much like in the initial results, leaving the other two investment options seem less profitable for the v_E: the total number of Pr products

BUPT

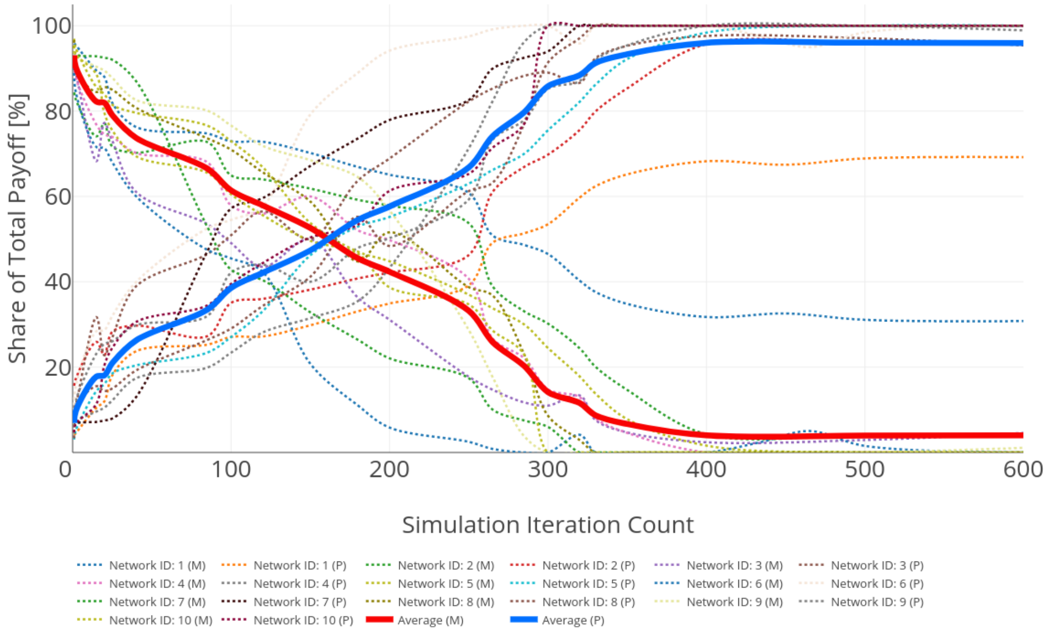


Figure 5.20 Share of the total payoff evolution with iteration count in a random topology for producers (P) and middlemen (M). The producers were assigned preferentially to the agents with highest degrees [25].

increased by 23 new Pr products (meaning an increase of 77%, roughly double the increase of the previous results) and with a final (total) production count of 4,207 units for each product, representing an increase of 212%.

5.4.3 Payoff Distribution for Each Agent Role

Obtaining the payoff distribution for each agent role is done by looking at the results of the simulations performed in subsection 5.4.1 and analyzing the distribution of payoff among both types of agents: producers and middleman. When the producers are assigned randomly I observe that the average payoff (obtained from simulating the same type of topology with 10 different densities, as seen in Table 5.3) of the producers for the mesh, small-world and random network topologies — Figure 5.22, Figure 5.23 and Figure 5.24, respectively — represent a positively skewed distribution of payoff among the rest of the producers. In other words, only a handful of them benefit from an increased payoff, while the rest of the producers earn a lot less. 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 (and averaging the results for all 10 densities) 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. As seen in Figure 5.25, when producers are assigned randomly, the majority of the middlemen tend to earn a lot more than the average.

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

BUPT

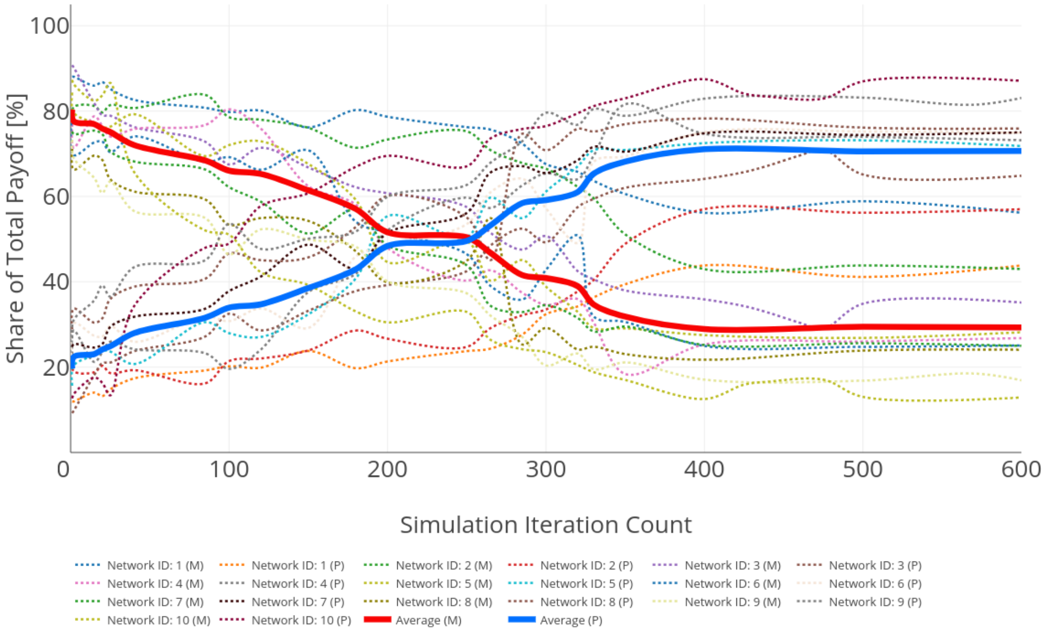


Figure 5.21 Share of the total payoff evolution with iteration count in a scale-free topology for producers (P) and middlemen (M). The producers were assigned preferentially to the agents with highest degrees [25].

presented in Figure 5.26, Figure 5.27 and Figure 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 v_E economic agents, adapting to the way they were assigned by means of selective investment in one of the four actions.

5.4.4 Ergodicity of Payoff Distribution 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 undertakes the right decisions, but it can also be punished for the wrong ones. If the two distributions are substantially different, then the system is non-erogodic (or path-dependent) and considered as unfair [31]. This underlines the assumption that the median behaviour of an economic agent over time on a given trajectory is not influenced

BUPT

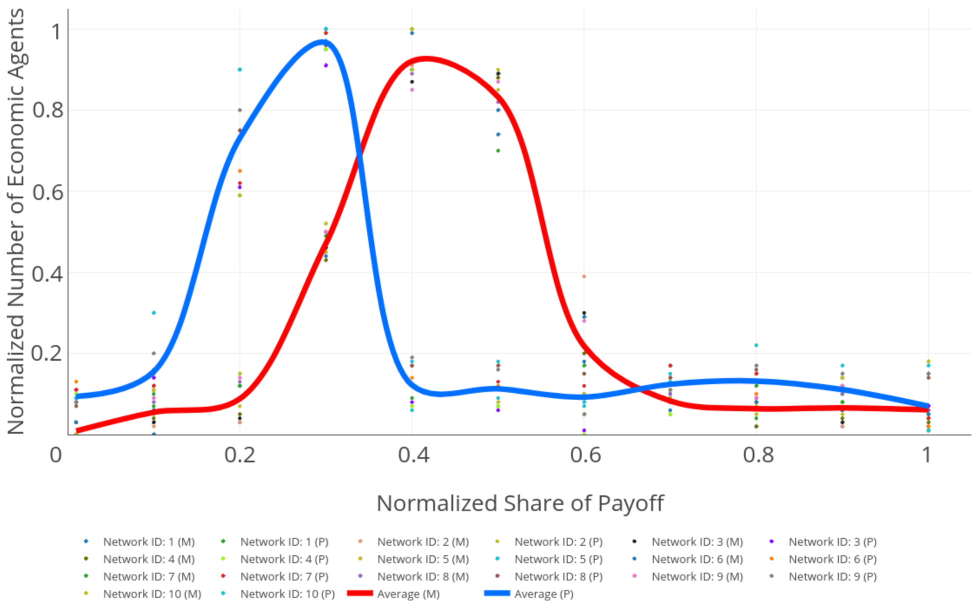


Figure 5.22 Distribution of payoff in a mesh topology, where the payoff and number of economic agents (separately accounted for producers (P) and middlemen (M)) are normalized to the total network payoff and the total number of economic agents, respectively. The producer roles were assigned randomly to the agents throughout the network [25].

by the particular trajectory selected and represents the key analytical device in Statistical Physics [105, 167]. As such, I investigate the ergodicity of complex (economic) systems 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.

In accordance with the simulation scenarios from section 5.4, the role of producer is assigned to economic agents in two ways: randomly and preferentially. By gathering information on the number of iterations each economic agent spends in a given payoff category — *e.g.* 0-5k, 5k-10k, 10k-15k, *etc.* —, as well as the number of total economic agents in each payoff category, I obtain the charts presented in Figures 5.30 — 5.37. Also, after fitting the distributions for each scenario (*i.e.* random and preferential producer-assignment), I obtain the results presented in Tables 5.5 — 5.8. The fitting process was carried out by employing EasyFit, a software system for data fitting in dynamical systems [173].

The numerical analysis highlights the similarities 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 is the only one with significant differences in distributions, while the other three topologies (*i.e.* mesh, random and small-world) have small to very small variation, proving that our economic (extended) model is ergodic. Specifically, for the random assignment of middlemen I measure an absolute difference of 4.25% between space and time distributions on the mesh; 8.37% on the small-world; 6.18% on the random network; 40.6% on the scale-free network. For preferential attachment I obtain differences of 6.06% on the mesh; 1.74% on the small-world; 2.96% on the random network; 31.05% on the scale-free network. When switching over to the

BUPT

Distribution type	Agent-role	Random	Preferential
Time interval	Middlemen	Normal ($\sigma = 6131.4,\ \mu = 13664$)	Chi-squared ($\upsilon = 13604$)
	Producers	Chi-squared ($\upsilon = 14482$)	Lognormal ($\sigma = 0.62544,\ \mu = 8.9898$)
Number of economic agents	Middlemen	Normal ($\sigma = 6404,\ \mu = 14398$)	Chi-squared ($\upsilon = 14482$)
	Producers	Normal ($\sigma = 6191.7,\ \mu = 13518$)	Lognormal ($\sigma = 0.65137,\ \mu = 9.0931$)

Table 5.5 Payoff distributions (for random versus preferential assignment of producers) fitting and associated parameters for the charts presented in Figure 5.30 and Figure 5.31 pertaining to the mesh topology.

Distribution type	Agent-role	Random	Preferential
Time interval	Middlemen	Normal ($\sigma = 6544.4,\ \mu = 13933$)	Chi-squared ($\upsilon = 10529$)
	Producers	Normal ($\sigma = 6848.2,\ \mu = 14337$)	Lognormal ($\sigma = 0.49823,\ \mu = 9.1823$)
Number of economic agents	Middlemen	Normal ($\sigma = 5904.7,\ \mu = 13374$)	Chi-squared ($\upsilon = 10716$)
	Producers	Normal ($\sigma = 6726.2,\ \mu = 13619$)	Lognormal ($\sigma = 0.68406,\ \mu = 9.1128$)

Table 5.6 Payoff distribution (for random versus preferential assignment of producers) fitting and associated values for the charts presented in Figure 5.32 and Figure 5.33 pertaining to the small-world topology.

Distribution type	Agent-role	Random	Preferential
Time interval	Middlemen	Normal ($\sigma = 7313.7,\ \mu = 13935$)	Chi-squared ($\upsilon = 14482$)
	Producers	Normal ($\sigma = 7128.4,\ \mu = 10722$)	Lognormal ($sigma = 0.7826,\ \mu = 8.9259$)
Number of economic agents	Middlemen	Normal ($\sigma = 6861.3,\ \mu = 14147$)	Chi-squared ($\upsilon = 14295$)
	Producers	Normal ($\sigma = 6221.8,\ \mu = 10803$)	Lognormal ($\sigma = 0.7049,\ \mu = 8.7353$)

Table 5.7 Payoff distribution fitting (for random versus preferential assignment of producers) and associated values for the charts presented in Figure 5.34 and Figure 5.35 pertaining to the random topology.

BUPT

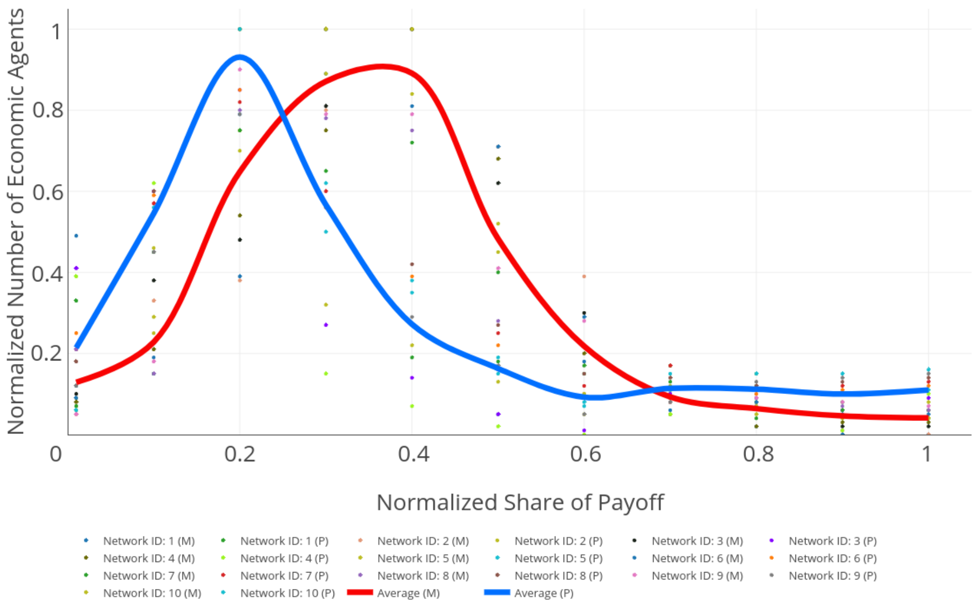


Figure 5.23 Distribution of payoff in a small-world topology, where the payoff and number of economic agents (separately accounted for producers (P) and middlemen (M)) are normalized to the total network payoff and the total number of economic agents, respectively. The producer roles were assigned randomly to the agents throughout the network [25].

preferential assignment of producers, the numerical differences between the distribution fits remain within the same margins, with the exception of the scale-free topology where they rise up to 79.9%. 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, after simulating economic activities according to the prerequisites presented in section 5.4, I obtain the values presented in Table 5.9. The percentages of bankrupt economic agents are indeed within the boundaries 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.

BUPT

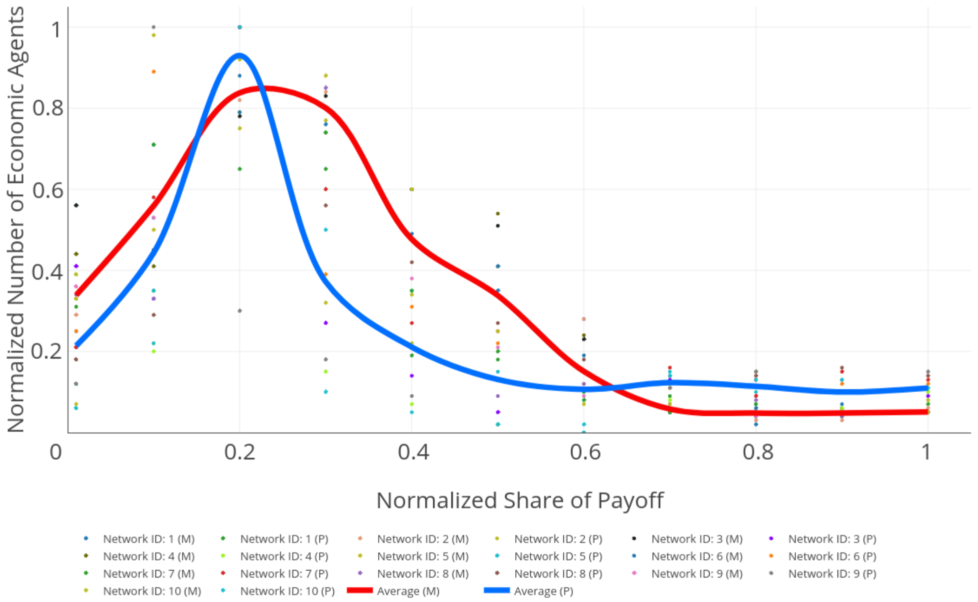


Figure 5.24 Distribution of payoff in a random topology, where the payoff and number of economic agents (separately accounted for producers (P) and middlemen (M)) are normalized to the total network payoff and the total number of economic agents, respectively. The producer roles were assigned randomly to the agents throughout the network [25].

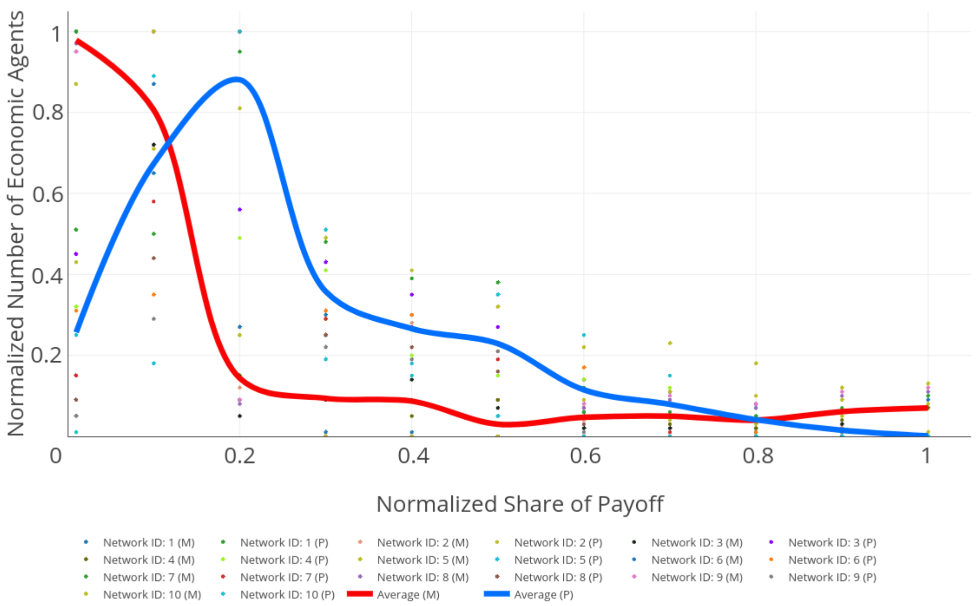


Figure 5.25 Distribution of payoff in a scale-free topology, where the payoff and number of economic agents (separately accounted for producers (P) and middlemen (M)) are normalized to the total network payoff and the total number of economic agents, respectively. The producer roles were assigned randomly to the agents throughout the network [25].

BUPT

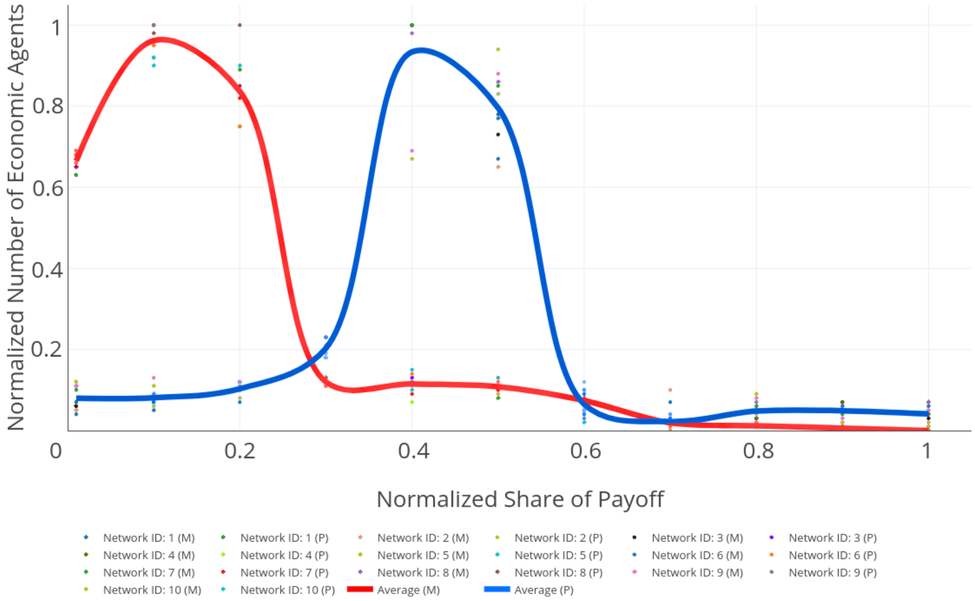


Figure 5.26 Distribution of payoff in a mesh topology, where the payoff and number of economic agents (separately accounted for producers (P) and middlemen (M)) are normalized to the total network payoff and the total number of economic agents, respectively. The producers roles where roles were assigned preferentially to the agents with the highest degrees in the network [25].

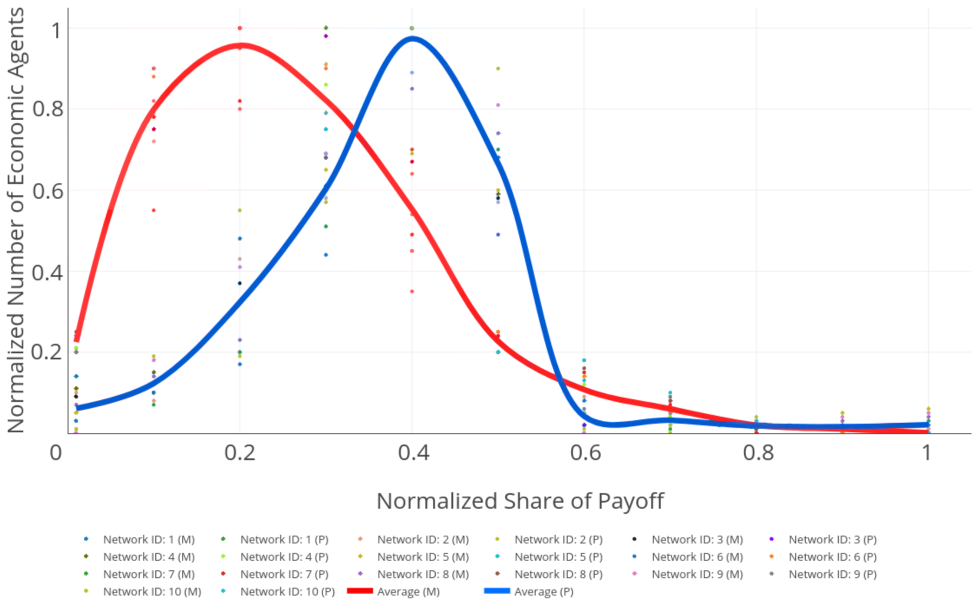


Figure 5.27 Distribution of payoff in a small-world topology, where the payoff and number of economic agents (separately accounted for producers (P) and middlemen (M)) are normalized to the total network payoff and the total number of economic agents, respectively. The producers roles where roles were assigned preferentially to the agents with the highest degrees in the network [25].

BUPT

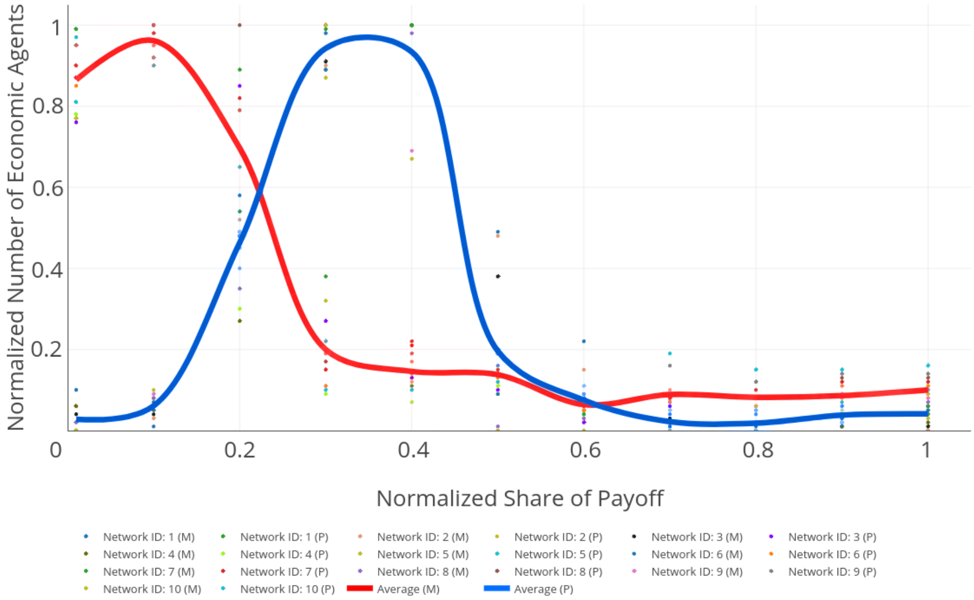


Figure 5.28 Distribution of payoff in a random topology, where the payoff and number of economic agents (separately accounted for producers (P) and middlemen (M)) are normalized to the total network payoff and the total number of economic agents, respectively. The producers roles where roles were assigned preferentially to the agents with the highest degrees in the network [25].

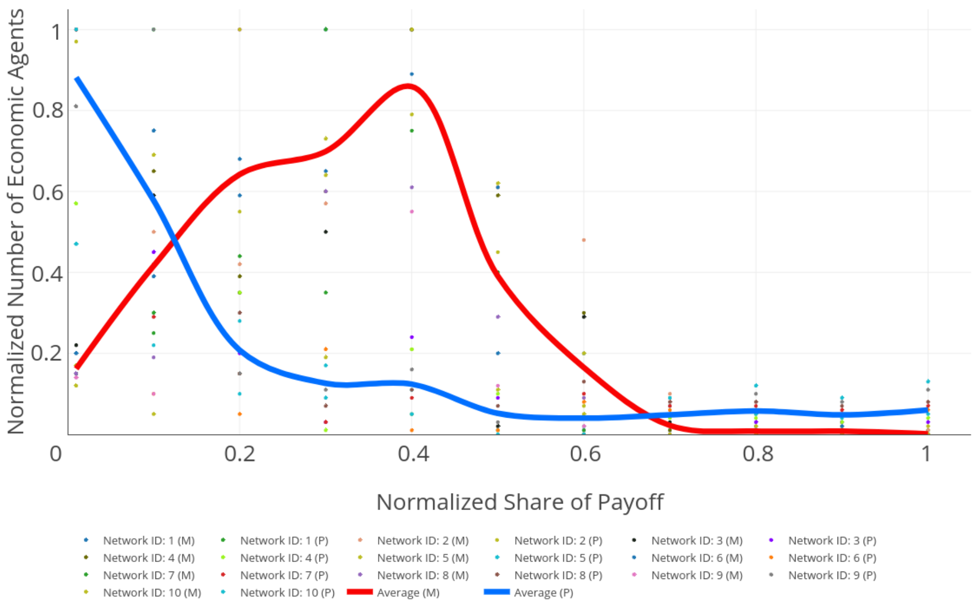


Figure 5.29 Distribution of payoff in a scale-free topology, where the payoff and number of economic agents (separately accounted for producers (P) and middlemen (M)) are normalized to the total network payoff and the total number of economic agents, respectively. The producers roles where roles were assigned preferentially to the agents with the highest degrees in the network [25].

BUPT

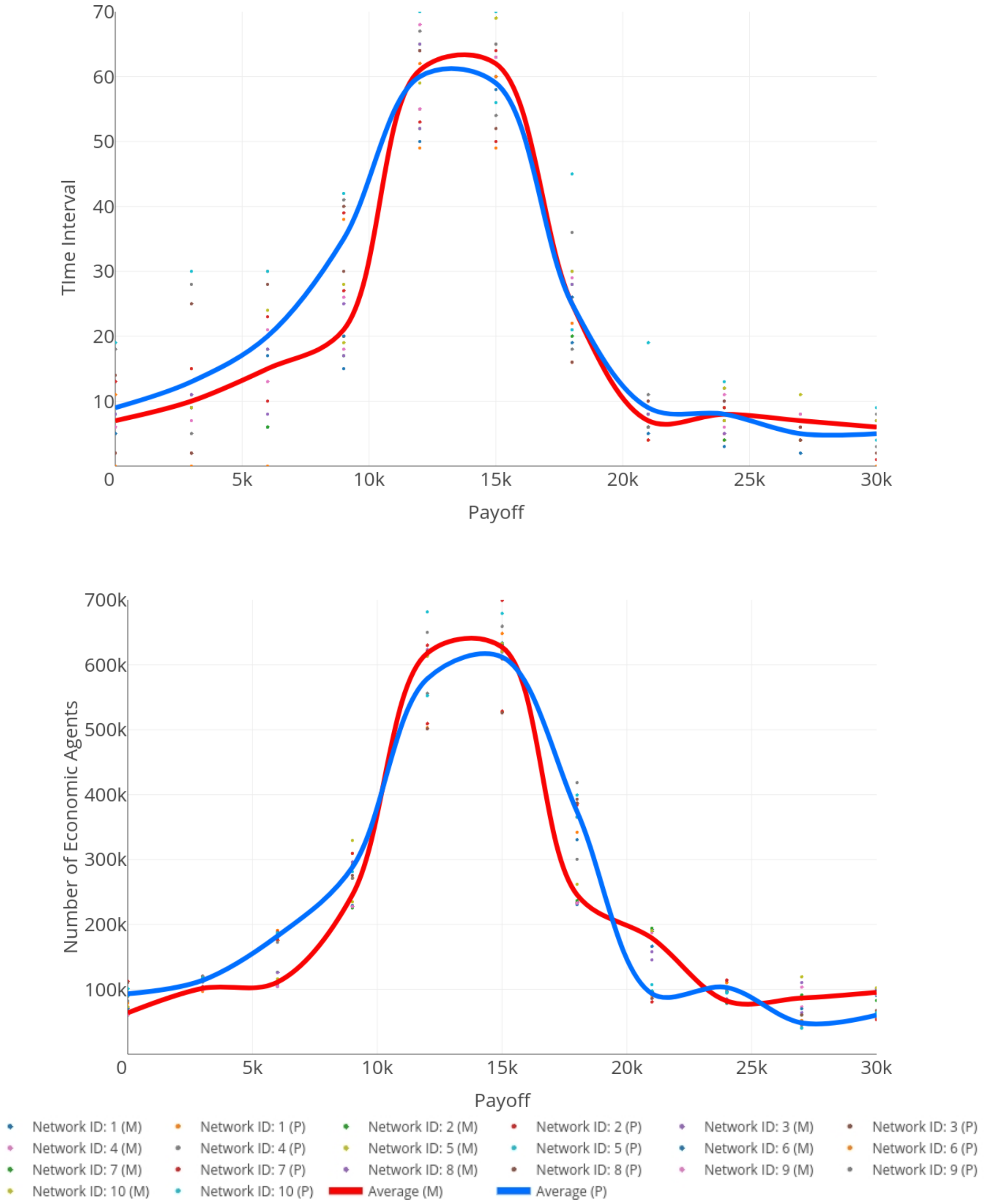


Figure 5.30 Payoff evolution in a mesh network, in relation to both the time spent in each payoff category interval (upper panel) and the number of economic agents — producers (P) and middlemen (M) — in each payoff category (lower panel). The producers were assigned randomly in the network [25].

BUPT

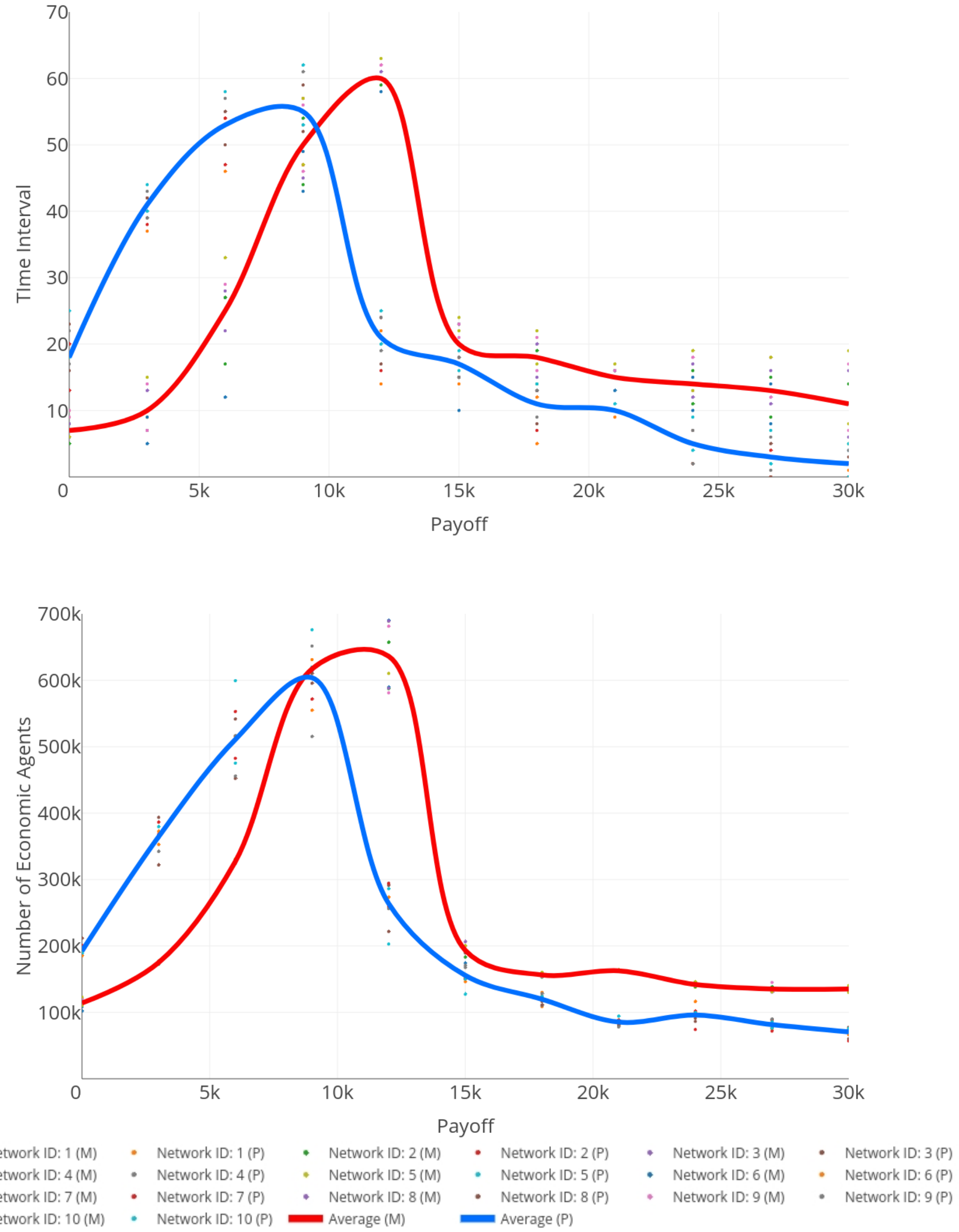


Figure 5.31 Payoff evolution in a mesh network, in relation to both the time spent in each payoff category interval (upper panel) and the number of economic agents — producers (P) and middlemen (M) — in each payoff category (lower panel). The producers were assigned preferentially to the agents with the highest degrees in the network [25].

BUPT

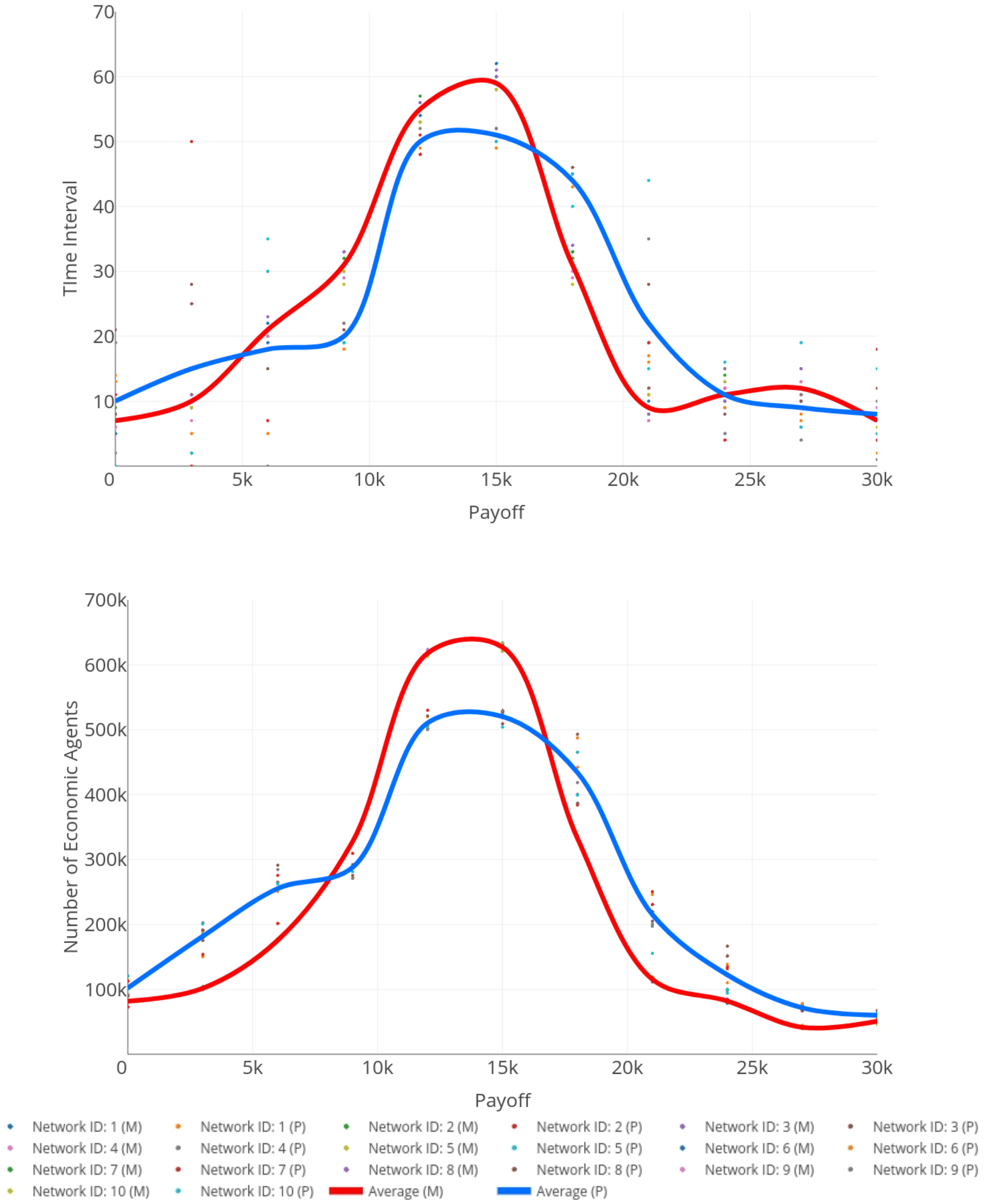


Figure 5.32 Payoff evolution in a small-world network, in relation to both the time spent in each payoff category interval (upper panel) and the number of economic agents — producers (P) and middlemen (M) — in each payoff category (lower panel). The producers were assigned randomly in the network [25].

BUPT

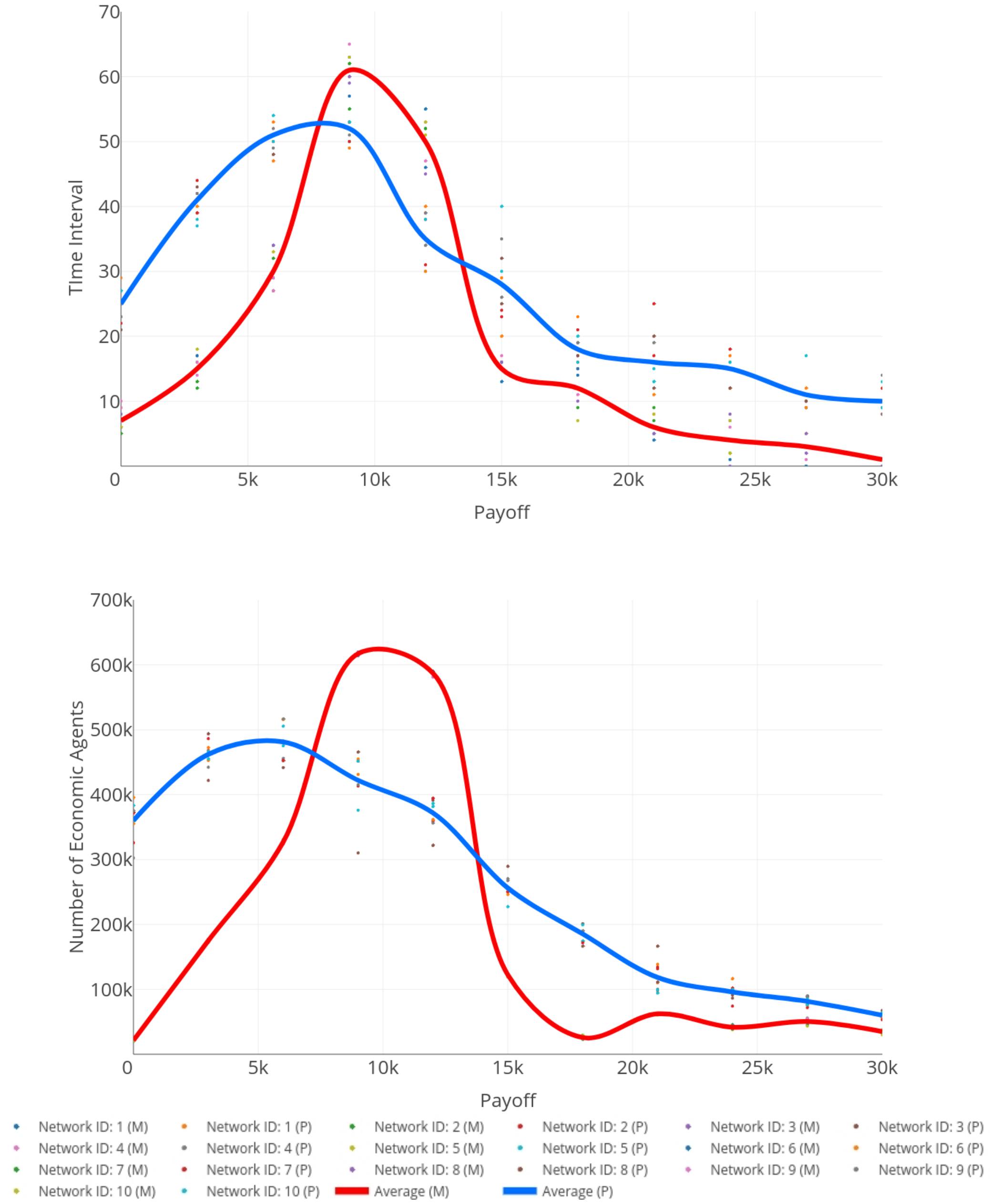


Figure 5.33 Payoff evolution in a small-world network, in relation to both the time spent in each payoff category interval (upper panel) and the number of economic agents — producers (P) and middlemen (M) — in each payoff category (lower panel). The producers were assigned preferentially to the agents with the highest degrees in the network [25].

BUPT

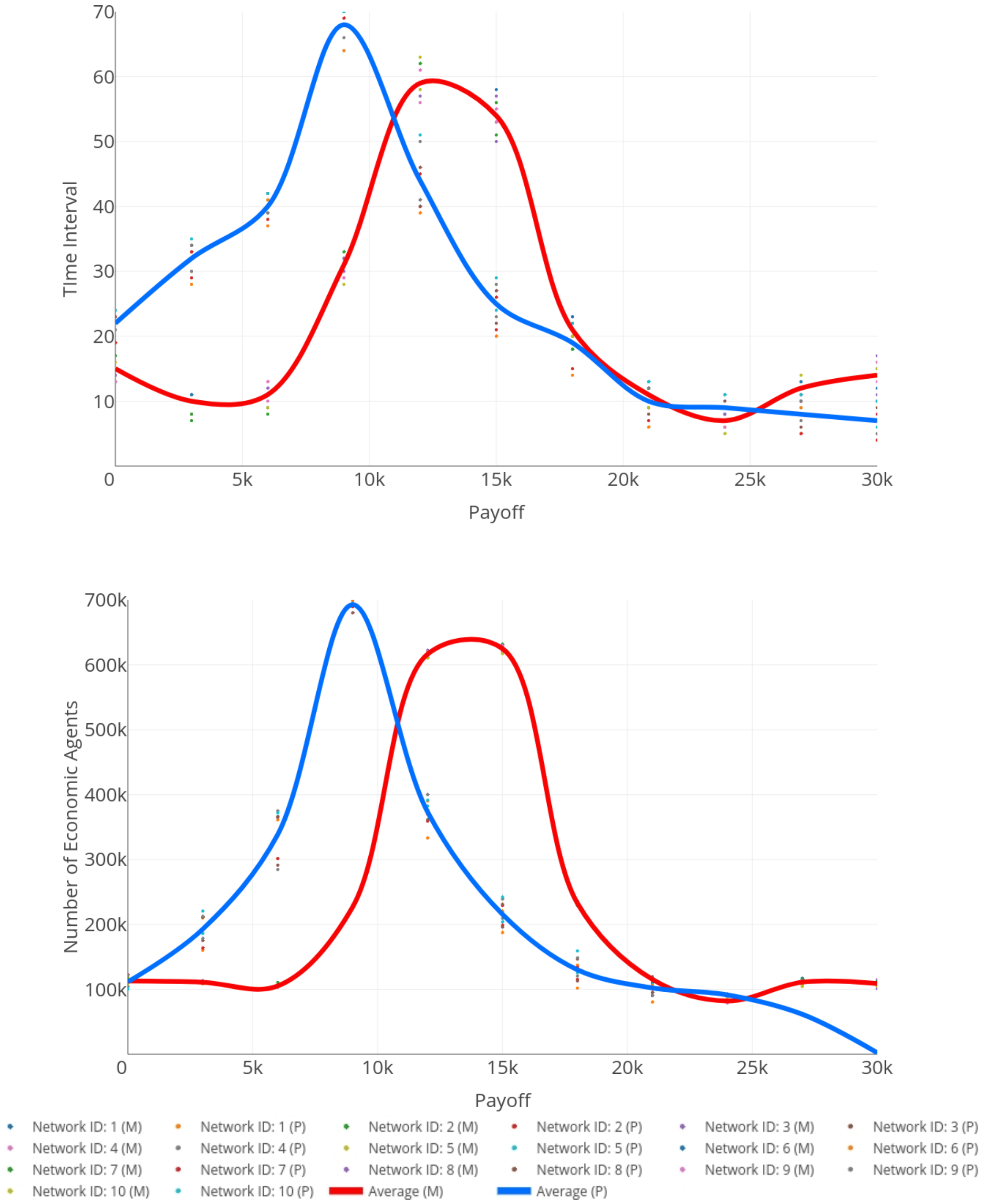


Figure 5.34 Payoff evolution in a random network, in relation to both the time spent in each payoff category interval (upper panel) and the number of economic agents — producers (P) and middlemen (M) — in each payoff category (lower panel). The producers were assigned randomly in the network [25].

BUPT

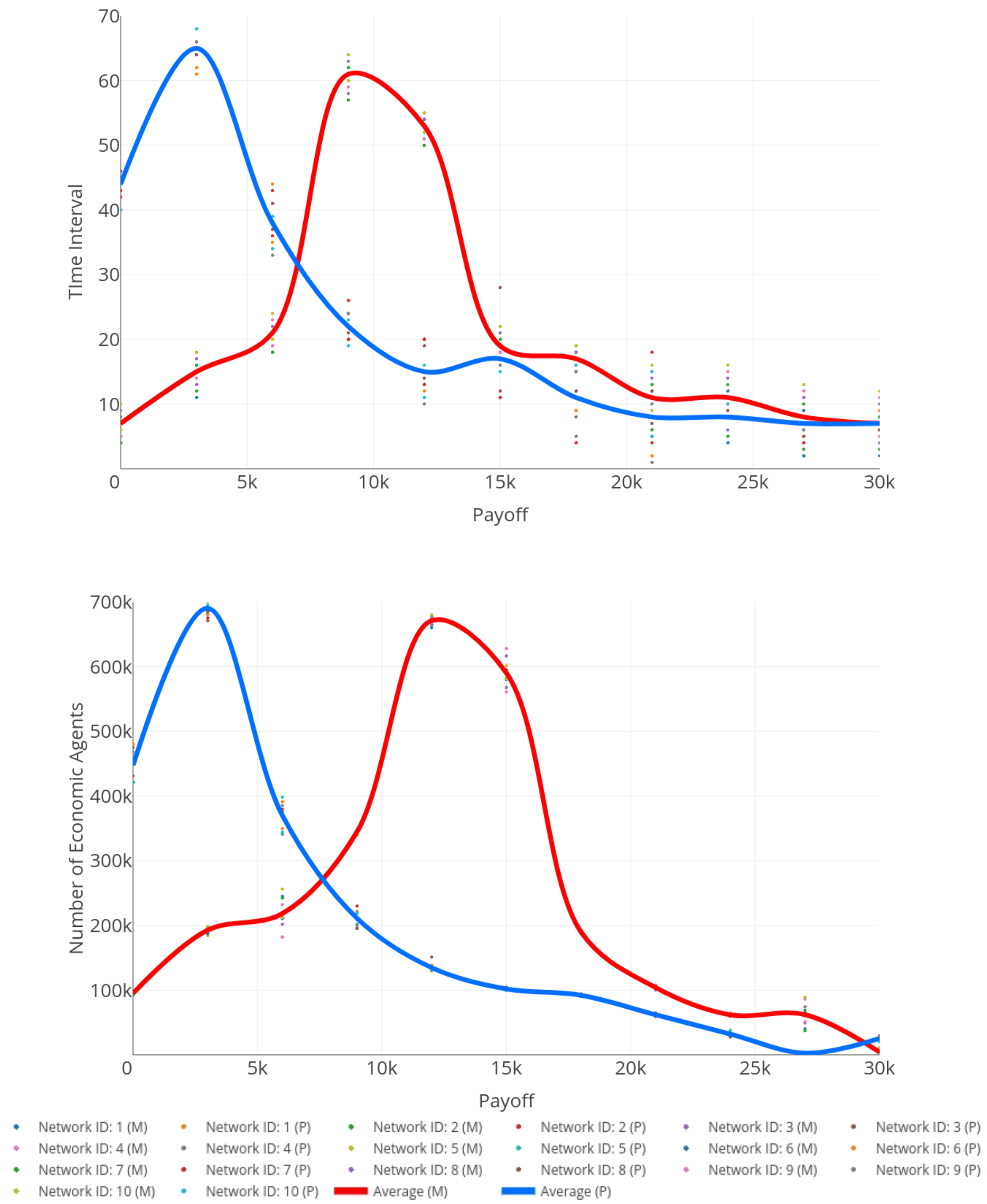


Figure 5.35 Payoff evolution in a random network, in relation to both the time spent in each payoff category interval (upper panel) and the number of economic agents — producers (P) and middlemen (M) — in each payoff category (lower panel). The producers were assigned preferentially to the agents with the highest degrees in the network [25].

BUPT

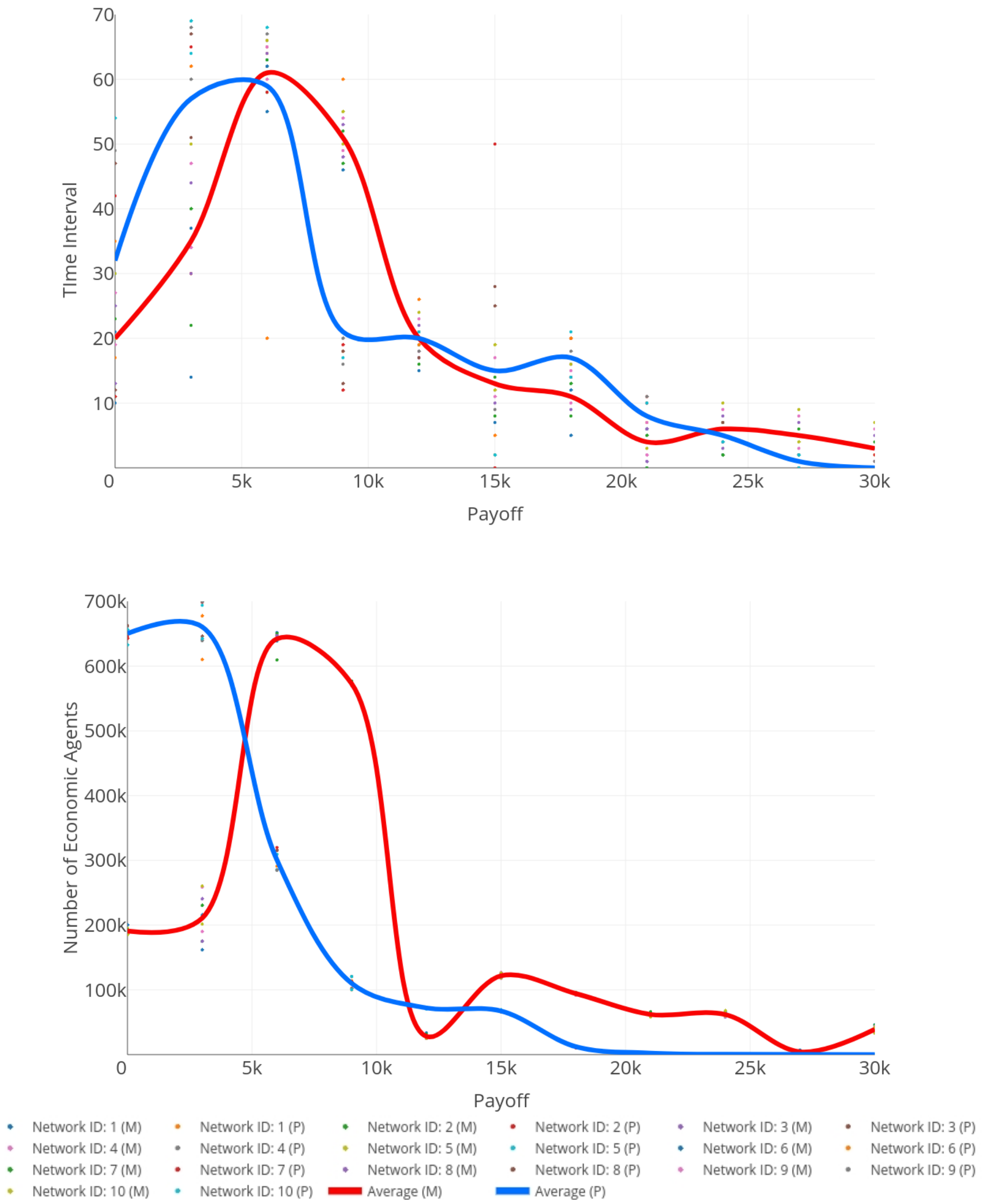


Figure 5.36 Payoff evolution in a scale-free network, in relation to both the time spent in each payoff category interval (upper panel) and the number of economic agents — producers (P) and middlemen (M) — in each payoff category (lower panel). The producers were assigned randomly in the network [25].

BUPT

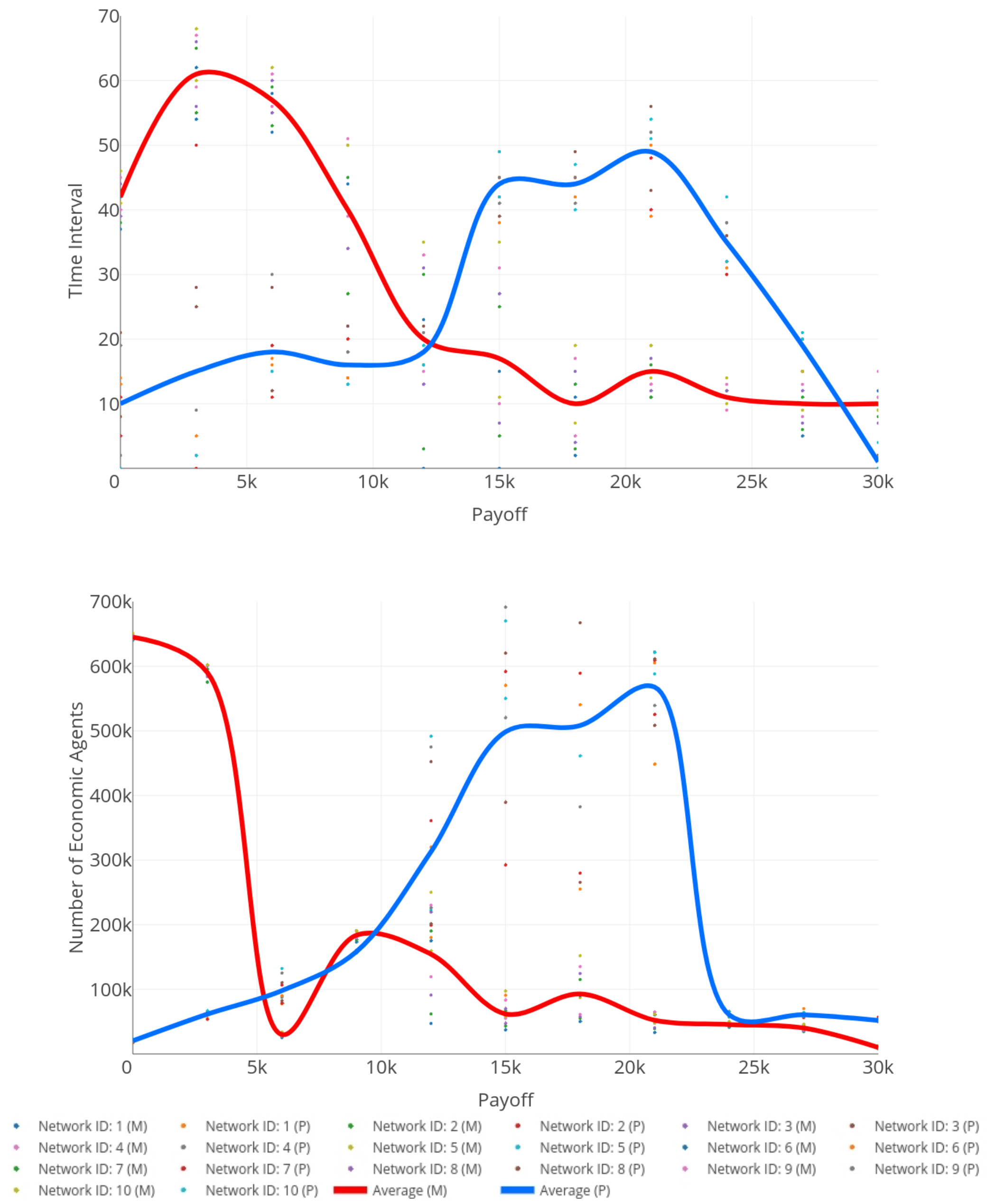


Figure 5.37 Payoff evolution in a scale-free network, in relation to both the time spent in each payoff category interval (upper panel) and the number of economic agents — producers (P) and middlemen (M) — in each payoff category (lower panel). The producers were assigned preferentially to the agents with the highest degrees in the network [25].

BUPT

Distribution type	Agent-role	Random	Preferential
Time interval	Middlemen	Lognormal ($\sigma = 0.6136,\ \mu = 8.9864$)	Lognormal ($\sigma = 0.80805,\ \mu = 8.8074$)
	Producers	Lognormal ($\sigma = 1.2853,\ \mu = 7.6585$)	Normal ($\sigma = 7246.9,\ \mu = 16271$)
Number of economic agents	Middlemen	Lognormal ($\sigma = 0.36445,\ \mu = 10.0009$)	Lognormal ($\sigma = 0.55708,\ \mu = 8.4677$)
	Producers	Lognormal ($\sigma = 0.25708,\ \mu = 8.4677$)	Normal ($\sigma = 5630,\ \mu = 19323$)

Table 5.8 Payoff distribution fitting (for random versus preferential assignment of producers) and associated values for the charts presented in Figure 5.36 and Figure 5.37 pertaining to the scale-free topology.

Topology	Random	Preferential
Mesh	8.2%	10.11%
Small-World	6.58%	8.42%
Random	7.84%	9.03%
Scale-Free	18.72%	21.9%

Table 5.9 Financially bankrupt economic agents in the four distinct fundamental topologies, both for random and preferential agent-role assignment.

BUPT

6. Conclusions

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 [210] or economic activity [3, 122].

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 [206]. 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 [81] — 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 [128].

In light of the study nominating Kevin Bacon as the most influential node in the IMDB actor network [75], 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 [81] and the Jazz musicians network [90], 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 Johnson to 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 [51, 69, 113, 206] —, 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 cannot 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 [39].

At first glance, the findings presented by Borondo *et al.* in [39] 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 network. 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 unfair 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 v_E 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.

6.1 Publications

To this date I have the following publications submitted, accepted and presented at international conferences or journals, relevant to the domain of Computer Science:

6.1.1 International Conferences

1. **Gabriel Barina**, Alexandru Topirceanu, and Mihai Udrescu. "MuSeNet: Natural patterns in the music artists industry." In: 9th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI), pp. 317-322. IEEE, 2014. **Indexed WoS** (Accession Number: WOS:000343400600055, ISBN:978-1-4799-4694-5).

2. Alexandru Topirceanu, **Gabriel Barina**, and Mihai Udrescu. "Musenet: Collaboration in the music artists industry." In: European Network Intelligence Conference (ENIC), pp. 89-94. IEEE, 2014. **Indexed WoS** (Accession Number: WOS:000361480100014, ISBN:978-1-4799-6914-2E).

3. **Gabriel Barina**, Mihai Udrescu, Alexandru Topirceanu, and Mircea Vladutiu. "Simulating Payoff Distribution in Networks of Economic Agents." In: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 467-470. IEEE, 2018. **Indexed WoS** (Accession Number: WOS:000455640600076, ISBN:978-1-5386-6051-5).

4. Maria-Alexandra Barina, **Gabriel Barina**. "From Elusive to Ubiquitous: Understanding Smart Cities." In: 19th International Conference on Informatics and Economy (IE). 2020. **In press; pending WoS indexation.**

6.1.2 International Journals

1. **Gabriel Barina**, Mihai Udrescu, Alexandra Barina, Alexandru Topirceanu, and Mircea Vladutiu. "Agent-based simulations of payoff distribution in economic networks." In Social Network Analysis and Mining (SNAM) 9, no. 1, p. 63. 2019. **Indexed WoS** (Accession Number: WOS:000492592200001, ISSN: 1869-5450).

6.1.3 Book Chapters at International Publishers

1. **Gabriel Barina**, Calin Sicoe, Mihai Udrescu, and Mircea Vladutiu. "Simulating trade in economic networks with TrEcSim". In: Alhajj R., Hoppe H., Hecking T., Bródka P., Kazienko P. (eds) Network Intelligence Meets User Centered Social Media Networks (European Network Intelligence Conference, ENIC), pp. 169-185. Lecture Notes in Social Networks. Springer International Publishing, Cham, 2017. **Indexed WoS** (Accession Number: WOS:000507984600012, ISBN:978-3-319-90312-5; 978-3-319-90311-8).

6.2 Research Milestones

Throughout the entirety of my doctoral studies, I have aimed at following and achieving the milestones presented in Table 6.1; these milestones were planned in advance and approved by my advisers in my first year of studies, however they were adjusted based on both the results obtained and the feedback obtained from the scientific community. Concurrently, in order to provide a meaningful contribution to the field of Social Network Analysis, not only did I have to meet these requirements, but I also had to monitor the scientific activities of other researchers on an international level, ensuring me that the proposals and results presented in this thesis are indeed original, relevant and useful.

#	Milestone	Result	Year
1	Keep track of the state-of-the-art in topics of interest in the field of Social Network Analysis	-	2014 - 2019
2	Using data mining to obtain relevant data, create a state-of-the-art emergent collaboration network based on real-world data in order to analyze and compare its fundamental properties to other, similar networks	MuSeNet, the musical artists' society network	2014
3	Propose a new metric capable of quantifying the sociability of a node in regard to the social features	The sociability $\mathcal{S}$-metric for complex networks	2014
4	Create a simulator based on either a heuristic or a genetic algorithm, that is capable of simulating emergent relationships between (economic) agents and releasing it as a tool	TrEcSim, the Trade and Economical Simulator, available online. It uses a heuristic algorithm	2017
5	Enhance TrEcSim by implementing several improved mechanisms of modeling the interaction of the (economic) agents	Implementation of the tolerance-based interaction model, as it represents a dynamic model of opinion spread, offering results similar to other, real-world systems. Further improvements are pending implementation	2018
6	Publish results at international conferences relevant to the field of Social Network Analysis	6 scientific papers written	2014 - 2020
7	Elaborate and defend thesis	-	2019 - 2020

Table 6.1 Proposed milestones for my doctoral studies and the results obtained.

6.3 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”.

- Added 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.

- Improved interface: an even more customizable GUI 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.

6.4 Closing Thoughts

Complex networks are comprehensively studied due to their multitude of applications throughout many fields of science. As a result, such fields, starting from Medicine and Sociology all the way to Architecture, Engineering and Economy, benefit from the theories and methodologies of Social Network Analysis.

Economy and Marketing, for instance, constantly strive to increase their income by analyzing markets and consumers. As such, they study the ever-changing needs of consumers, as well as their strengths and weaknesses. Medicine applies network theory in order to help determine outbreaks, model the dynamic evolution of diseases and to limit or stop its spread completely, while Sociology applies graph theory in order to analyze and influence the opinion of everyday users. These are just but a few examples of applying Social Network Analysis in order to better understand, shape and improve our everyday life. Even so, network theory is still in its relative infancy. Coupled with the fact that markets, diseases, human behaviour, *etc.* are mostly erratic and somewhat unpredictable, we can see how social network analysis is still deficitary. To counter this detriment, we can combine Social Science with applied sciences, and by using computers as a tool, scientists can simulate and analyze interleaved mathematical and psychological models much faster and with greater ease than ever before. As a results, recent advancement in a multitude of fields have come up with new ways of modeling and analyzing said markets, diseases, human behaviour, *etc.* and to convey results of great theoretical and scientific value.

Evolved gradually from basic network topologies (*e.g.* bus, star, mesh, *etc.*), complex networks have infused our daily lives profoundly; ranging from natural networks (*e.g.* actors network, musicians network, recipe network, *etc.*) to synthetic network (*e.g.* air traffic network, the World Wide Web, *etc.*), have generated interest in all fields of science. Nowadays, due to an increased interest in network theory, fostered by social, economic and computational advancements, complex networks encompass newer and more advanced topologies, which better resemble real-world networks.

Such real-world networks are also studied in this thesis. In chapter 4, by gathering relevant data, I created MuSeNet and investigated how artists work and co-exist together. Similar to an existing study pertaining actors, I analyzed the source of their relationship, how they form communities based on music genre or race and how they all converge to those few agents, who own a recording studio. I also compared the resulting network not only with other similar networks (*i.e.* Jazz musicians, IMDB actors), but also with social networks from platforms like Facebook and Twitter. Due to their popularity, the available data regarding these platforms is near infinite, and as such, they represent the perfect means to map and compare MuSeNet to.

Similarly, modern science, along with Economy and Marketing, are trying to create improved topologies that resemble real-world networks, in order to better analyze behaviour, need and outcome. As such, they make use of graph theory as well as Computer Science, in order to create a dynamic environment that resembles our own. Such is the case of TrEcsim, presented in chapter 5. As a state-of-the-art economic simulator, TrEcSim was used to simulate dynamic behaviour of economic agents, and how network topology affects their payoff.

As a conclusion, Social Network Analysis, coupled with numerous other fields of science, has proved countless times to be of paramount importance in understanding how we, as a society function, how we chose to collaborate with others, how we influence — willingly or unwillingly — the choices of others and how we profit from them.

References

[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] Acemoglu, D. and Ozdaglar, A. (2011). Opinion dynamics and learning in social networks. *Dynamic Games and Applications*, 1(1):3–49.

[3] Agranoff, R. and McGuire, M. (1998). Multinetwork management: Collaboration and the hollow state in local economic policy. *Journal of Public Administration Research and Theory*, 8(1):67–91.

[4] Ahmed, A., Batagelj, V., Fu, X., Hong, S.-H., Merrick, D., and Mrvar, A. (2007). Visualisation and analysis of the internet movie database. In *Visualization, 2007. APVIS'07. 2007 6th International Asia-Pacific Symposium on*, pages 17–24. IEEE.

[5] Akerlof, G. A. (1970). The market for" lemons": Quality uncertainty and the market mechanism. *The quarterly journal of economics*, pages 488–500.

[6] Alberich, R., Miro-Julia, J., and Rosselló, F. (2002). Marvel universe looks almost like a real social network. *arXiv preprint cond-mat/0202174*.

[7] Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47.

[8] Albert, R., Jeong, H., and Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *nature*, 406(6794):378.

[9] Alkemade, F. and Castaldi, C. (2005). Strategies for the diffusion of innovations on social networks. *Computational Economics*, 25(1-2):3–23.

[10] Alma Yadira Quiñonez Carrillo, Cristian Moral Martos, D. F. R. L. E. A. M. B. G. P. G. H. G. M. P. G. Q. C. G. I. T. T. H. R. O. A. J. d. J. F. C. J. A. R. L. L. A. C. S. L. J. V. G. L. M. A. V. M. C. G. A. M. A. L. M. D. M. E. M. A. H. R. M. E. M. R. M. A. M. M. R. A. A. V. R. L. A. (2014). Recibe, revista electrónica de computación, informática, biomédica y electrónica. Accessed 04.2014, available at *recibe.cucei.udg.mx/Recibe/index.php/Recibe*.

[11] Almende, B. (2016). vis. js–a dynamic, browser based visualization library. *URL http://visjs. org/. Acesso em*, 1.

[12] Almende, B. and Thieurmel, B. (2016). visnetwork: Network visualization using'vis. js' library. *R package version 0.2*, 1.

[13] Alon, U. (2007). Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8(6):450–461.

[14] Anderson, R. M., May, R. M., and Anderson, B. (1992). *Infectious diseases of humans: dynamics and control*, volume 28. Wiley Online Library.

[15] Arrow, K. J. and Debreu, G. (1954). Existence of an equilibrium for a competitive economy. *Econometrica: Journal of the Econometric Society*, pages 265–290.

[16] Arthur, W. B. (1999). Complexity and the economy. *science*, 284(5411):107–109.

[17] Aughenbaugh, J. M. and Paredis, C. J. (2004). The role and limitations of modeling and simulation in systems design. In *ASME 2004 International Mechanical Engineering Congress and Exposition*, pages 13–22. American Society of Mechanical Engineers.

[18] Austin, D. (2006). How google finds your needle in the web's haystack. *American Mathematical Society Feature Column*, 10(12).

[19] Avila, R., Broed, R., and Pereira, A. (2003). Ecolego-a toolbox for radioecological risk assessment. Technical report.

[20] Bar-Yam, Y. (2016). The limits of phenomenology: from behaviorism to drug testing and engineering design. *Complexity*, 21(S1):181–189.

[21] Barabasi, A.-L. (2002). Linked: How everything is connected to everything else and what it means. *Plume Editors*.

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

[23] Barina, G., Sicoe, C., Udrescu, M., and Vladutiu, M. (2017). Simulating trade in economic networks with trecsim. In *European Network Intelligence Conference*, pages 169–185. Springer.

[24] 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.

[25] 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.

[26] Barina, G., Udrescu, M., Topirceanu, A., and Vladutiu, M. (2018). Simulating payoff distribution in networks of economic agents. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 467–470. IEEE.

[27] Bastian, M., Heymann, S., Jacomy, M., et al. (2009). Gephi: an open source software for exploring and manipulating networks. *ICWSM*, 8:361–362.

[28] Batagelj, V. and Mrvar, A. (1998). Pajek-program for large network analysis. *Connections*, 21(2):47–57.

[29] Bauer, F. and Lizier, J. T. (2012). Identifying influential spreaders and efficiently estimating infection numbers in epidemic models: A walk counting approach. *EPL (Europhysics Letters)*, 99(6):68007.

[30] Bejan, A. (2016). *The physics of life: the evolution of everything*. St. Martin's Press.

[31] Berman, Y., Peters, O., and Adamou, A. (2017). An empirical test of the ergodic hypothesis: Wealth distributions in the united states. *Available at SSRN 2794830*.

[32] Bern, M. W. and Plassmann, P. E. (1997). *Mesh generation*. Pennsylvania State University, Department of Computer Science and ....

[33] Biglaiser, G. (1993). Middlemen as experts. *The RAND journal of Economics*, pages 212–223.

[34] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

[35] Blume, L. E. and Durlauf, S. N. (2005). *The economy as an evolving complex system, III: Current perspectives and future directions*. Oxford University Press.

[36] Bollobás, B. (1998). Random graphs. In *Modern Graph Theory*, pages 215–252. Springer.

[37] Borgatti, S. P. and Everett, M. G. (2006). A graph-theoretic perspective on centrality. *Social networks*, 28(4):466–484.

[38] Borgatti, S. P., Mehra, A., Brass, D. J., and Labianca, G. (2009). Network analysis in the social sciences. *science*, 323(5916):892–895.

[39] 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.

BUPT

[40] Brin, S. and Page, L. (2012). Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks*, 56(18):3825–3833.

[41] Broed, R. and Shulan, X. (2008). User's manual for ecolego toolbox and the discretization block. Technical report, Swedish Radiation Protection Authority.

[42] Burt, R. S. (2004). Structural holes and good ideas. *American journal of sociology*, 110(2):349–399.

[43] Burt, R. S. (2009). *Structural holes: The social structure of competition*. Harvard university press.

[44] Camarinha-Matos, L. M. and Afsarmanesh, H. (2008). *Collaborative networks: Reference modeling*. Springer Science & Business Media.

[45] Cano, P., Celma, O., Koppenberger, M., and Buldu, J. M. (2006). Topology of music recommendation networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 16(1):013107.

[46] Chang, K. (2020). Marvel universe social graph. Accessed 2020, available at *http://syntagmatic.github.io/marvel/*.

[47] Chatterjee, A. and Chakrabarti, B. K. (2007). *Econophysics of markets and business networks*, volume 27. Springer.

[48] 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.

[49] Chen, M., Kuzmin, K., and Szymanski, B. K. (2014). Community detection via maximization of modularity and its variants. *IEEE Transactions on Computational Social Systems*, 1(1):46–65.

[50] Chen, W., Lakshmanan, L. V., and Castillo, C. (2013). Information and influence propagation in social networks. *Synthesis Lectures on Data Management*, 5(4):1–177.

[51] Chen, W.-K. (1997). *Graph theory and its engineering applications*, volume 5. World Scientific.

[52] Chen, X. (2019). Awesome public datasets. Accessed 2019, available at *github.com/caesar0301*.

[53] Choromański, K., Matuszak, M., and Miekisz, J. (2013). Scale-free graph with preferential attachment and evolving internal vertex structure. *Journal of Statistical Physics*, 151(6):1175–1183.

[54] Christakis, N. A. and Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England journal of medicine*, 357(4):370–379.

[55] Chung, F. (1997). Open problems of paul erdos in graph theory. *Journal of graph theory*, 25(1):3–36.

[56] Cohen, R. and Havlin, S. (2010). *Complex networks: structure, robustness and function*. Cambridge University Press.

[57] Colman, A. M. (2015). *A dictionary of psychology*. Oxford Quick Reference.

[58] Council, N. R. et al. (2005). Network science committee on network science for future army applications.

[59] Coviello, L., Sohn, Y., Kramer, A. D., Marlow, C., Franceschetti, M., Christakis, N. A., and Fowler, J. H. (2014). Detecting emotional contagion in massive social networks. *PloS one*, 9(3):e90315.

[60] Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695(5):1–9.

[61] Da Silva, R. A. P., Viana, M. P., and da Fontoura Costa, L. (2012). Predicting epidemic outbreak from individual features of the spreaders. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(07):P07005.

BUPT

[62] 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.

[63] Deissenberg, C., Van Der Hoog, S., and Dawid, H. (2008). Eurace: A massively parallel agent-based model of the european economy. *Applied Mathematics and Computation*, 204(2):541–552.

[64] Dekkers, R. and Bennett, D. (2009). Industrial networks of the future: review of research and practice. In *Dispersed Manufacturing Networks*, pages 13–34. Springer.

[65] 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.

[66] Durhuus, B. (2017). Accessed 08.2017, available at *math.ku.dk/ durhuus/rgn10/calendar.html*.

[67] Duval, J. S. (2012). Version 3.0 of eminers-economic mineral resource simulator. Technical report, US Gelogical Survey.

[68] 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. *Health Research Policy and Systems*, 14(1):34.

[69] Easley, D. and Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press.

[70] Ellson, J., Gansner, E., Koutsofios, L., North, S. C., and Woodhull, G. (2001). Graphviz—open source graph drawing tools. In *International Symposium on Graph Drawing*, pages 483–484. Springer.

[71] Erdos, P. and Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60.

[72] Eubank, S., Guclu, H., Kumar, V. A., Marathe, M. V., et al. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180.

[73] Fang, J.-q., Wang, X.-f., Zheng, Z.-g., Bi, Q., Di, Z.-r., and Xiang, L. (2007). New interdisciplinary science: Network science (1). *PROGRESS IN PHYSICS-NANJING-*, 27(3):239.

[74] Fararo, T. and Sunshine, M. (1964). A study of a biased friendship net. syracuse, ny.

[75] Fass, C., Ginelli, M., and Turtle, B. (1996). *Six degrees of kevin bacon*. Plume Books.

[76] Ferrara, E., De Meo, P., Catanese, S., and Fiumara, G. (2014). Detecting criminal organizations in mobile phone networks. *Expert Systems with Applications*, 41(13):5733–5750.

[77] 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.

[78] Freeman, L. (2004). The development of social network analysis. *A Study in the Sociology of Science*, 1.

[79] Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41.

[80] Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239.

[81] Gallos, L. K., Potiguar, F. Q., Andrade Jr, J. S., and Makse, H. A. (2013). Imdb network revisited: unveiling fractal and modular properties from a typical small-world network. *PloS one*, 8(6):e66443.

[82] Gao, J., Buldyrev, S. V., Stanley, H. E., and Havlin, S. (2012). Networks formed from interdependent networks. *Nature physics*, 8(1):40–48.

[83] Gao, Z.-K., Small, M., and Kurths, J. (2017). Complex network analysis of time series. *EPL (Europhysics Letters)*, 116(5):50001.

BUPT

[84] Gartman, D. (1978). Marx and the labor process: an interpretation. *Insurgent Sociologist*, 8(2-3):97–108.

[85] Geven, S., Weesie, J., and van Tubergen, F. (2013). The influence of friends on adolescents' behavior problems at school: The role of ego, alter and dyadic characteristics. *Social Networks*, 35(4):583–592.

[86] 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*.

[87] Girvan, M. and Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826.

[88] Gladwell, M. (2006). *The tipping point: How little things can make a big difference*. Little, Brown.

[89] Glaeser, E. L., Sacerdote, B., and Scheinkman, J. A. (1996). Crime and social interactions. *The Quarterly Journal of Economics*, 111(2):507–548.

[90] Gleiser, P. M. and Danon, L. (2003). Community structure in jazz. *Advances in complex systems*, 6(04):565–573.

[91] Golbeck, J. (2013). *Analyzing the social web*. Newnes.

[92] Golbeck, J. and Hansen, D. (2014). A method for computing political preference among twitter followers. *Social Networks*, 36:177–184.

[93] Gordon, B. J. (1964). Aristotle and the development of value theory. *The Quarterly Journal of Economics*, 78(1):115–128.

[94] Gossen, H. H. (1854). *Entwickelung der Gesetze des menschlichen Verkehrs: Und der daraus fließenden Regeln für menschliches Handeln*. F. Vieweg.

[95] Guille, A., Hacid, H., Favre, C., and Zighed, D. A. (2013). Information diffusion in online social networks: A survey. *ACM Sigmod Record*, 42(2):17–28.

[96] Håkansson, H. (1990). Technological collaboration in industrial networks. *European Management Journal*, 8(3):371–379.

[97] Hansen, D., S. B. and Smith, M. (2011). Introduction to social media and social networks. *Analyzing social media networks with NodeXL: insights from a connected world. Burlington-USA: Elsevier. Cap*, 1:3–9.

[98] Hashmi, A., Zaidi, F., Sallaberry, A., and Mehmood, T. (2012). Are all social networks structurally similar? In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 310–314. IEEE.

[99] Heer, J. and Boyd, D. (2005). Vizster: Visualizing online social networks. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 32–39. IEEE.

[100] Herrera, F., Herrera-Viedma, E., et al. (1996). A model of consensus in group decision making under linguistic assessments. *Fuzzy sets and Systems*, 78(1):73–87.

[101] Hodgson, G. M. (2017). Karl polanyi on economy and society: a critical analysis of core concepts. *Review of social economy*, 75(1):1–25.

[102] Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558.

[103] Hosmer, S. T. (1996). Psychological effects of us air operations in four wars, 1941-1991. lessons for us commanders. Technical report, RAND CORP SANTA MONICA CA.

[104] Hu, J. and Zheng, W. X. (2014). Emergent collective behaviors on coopetition networks. *Physics Letters A*, 378(26-27):1787–1796.

[105] Huang, K. (2009). *Introduction to statistical physics*. CRC press.

[106] Hussain, O. A., Anwar, Z., Saleem, S., and Zaidi, F. (2013). Empirical analysis of seed selection criterion in influence mining for different classes of networks. In *Cloud and green computing (CGC), 2013 third international conference on*, pages 348–353. IEEE.

BUPT

[107] Jackson, M. O. (2010). *Social and economic networks*. Princeton university press.

[108] Jackson, M. O. (2011). An overview of social networks and economic applications. In *Handbook of social economics*, volume 1, pages 511–585. Elsevier.

[109] Jackson, M. O. and Yariv, L. (2006). Diffusion on social networks. *Economie publique/Public economics*, 1(16).

[110] Jacomy, M., Venturini, T., Heymann, S., and Bastian, M. (2014). Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PloS one*, 9(6):e98679.

[111] Kahneman, D. and Tversky, A. (1996). On the reality of cognitive illusions. pages 582–591.

[112] Kalamaras, D. (2014). Social networks visualizer (socnetv): Social network analysis and visualization software. *Social Networks Visualizer*.

[113] Kantarci, B. and Labatut, V. (2013). Classification of complex networks based on topological properties. In *Cloud and Green Computing (CGC), 2013 Third International Conference on*, pages 297–304. IEEE.

[114] Kauder, E. (1953). Genesis of the marginal utility theory: from aristotle to the end of the eighteenth century. *The Economic Journal*, 63(251):638–650.

[115] Keen, S. (1995). Finance and economic breakdown: modeling minsky's "financial instability hypothesis". *Journal of Post Keynesian Economics*, 17(4):607–635.

[116] Kellaris, J. J. and Kent, R. J. (1992). The influence of music on consumers' temporal perceptions: does time fly when you're having fun? *Journal of consumer psychology*, 1(4):365–376.

[117] Kempe, D., Kleinberg, J., and Tardos, É. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM.

[118] Kempe, D., Kleinberg, J. M., and Tardos, É. (2005). Influential nodes in a diffusion model for social networks. In *ICALP*, volume 5, pages 1127–1138. Springer.

[119] Kim, J. (2006). Emergence: Core ideas and issues. *Synthese*, 151(3):547–559.

[120] Laumann, E. O., Galaskiewicz, J., and Marsden, P. V. (1978). Community structure as interorganizational linkages. *Annual review of sociology*, 4(1):455–484.

[121] Lawyer, G. (2015). Understanding the influence of all nodes in a network. *Scientific reports*, 5:8665.

[122] Lee, I.-W., Feiock, R. C., and Lee, Y. (2012). Competitors and cooperators: A micro-level analysis of regional economic development collaboration networks. *Public administration review*, 72(2):253–262.

[123] Leon, F. (2012). Emergent behaviors in social networks of adaptive agents. *Mathematical Problems in Engineering*, 2012.

[124] Leskovec, J. and Krevl, A. (2014). SNAP Datasets: Stanford large network dataset collection. Accessed 07.2014, available at *snap.stanford.edu/data*.

[125] Li, X., Jin, Y. Y., and Chen, G. (2004). On the topology of the world exchange arrangements web. *Physica A: Statistical Mechanics and its Applications*, 343:573–582.

[126] Li, Y. (1999). Money and middlemen in an economy with private information. *Economic Inquiry*, 37(1):1–12.

[127] LLC., A. M. N. (2019). Allmusic. Previously known as All Music Guide or AMG, accessed 2019, available at *allmusic.com*.

[128] Loscalzo, J. and Barabasi, A.-L. (2011). Systems biology and the future of medicine. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 3(6):619–627.

[129] Lü, L., Chen, D., Ren, X.-L., Zhang, Q.-M., Zhang, Y.-C., and Zhou, T. (2016). Vital nodes identification in complex networks. *Physics Reports*, 650:1–63.

BUPT

[130] Lü, L., Zhang, Y.-C., Yeung, C. H., and Zhou, T. (2011). Leaders in social networks, the delicious case. *PloS one*, 6(6):e21202.

[131] Makse, H. (2019). Complex networks and soft matter lab. Accessed 2019, available at *www-levich.engr.ccny.cuny.edu/webpage/hmakse*.

[132] Manktelow, K. and Galbraith, N. (2012). *Thinking and reasoning: An introduction to the psychology of reason, judgment and decision making*. Psychology Press.

[133] Manresa, E. (2013). Estimating the structure of social interactions using panel data. *Unpublished Manuscript. CEMFI, Madrid*.

[134] Marshall, A. (2009). *Principles of economics: unabridged eighth edition*. Cosimo, Inc.

[135] Marx, K. (1867). Capital, volume i.

[136] Marx, K. (1910). *Value, price, and profit*. CH Kerr & Company.

[137] Marx, K. (1973). Outlines of the critique of political economy. *Trans, Martin Nicolaus, Harmondsworth: Penguin*.

[138] Marx, K. (2010). A contribution to the critique of political economy. In *Marx Today*, pages 91–94. Springer.

[139] Masoudi-Nejad, A., Schreiber, F., and Kashani, Z. R. M. (2012). Building blocks of biological networks: a review on major network motif discovery algorithms. *IET systems biology*, 6(5):164–174.

[140] Mataric, M. J. (1993). Designing emergent behaviors: From local interactions to collective intelligence. In *Proceedings of the Second International Conference on Simulation of Adaptive Behavior*, pages 432–441.

[141] Matthews, L. and Richard, P. (2017). Who is central to a social network? it depends on your centrality measure.

[142] Maxwell, J. C. (1993). *Developing the leader within you*. Thomas Nelson Inc.

[143] McDonald, M. and Wilson, H. (2016). *Marketing Plans: How to prepare them, how to profit from them*. John Wiley & Sons.

[144] McLaurin, R. D. (1982). *Military propaganda: psychological warfare and operations*. Praeger New York, NY.

[145] Meikle, S. et al. (1997). Aristotle's economic thought. *OUP Catalogue*.

[146] Menger, K. (1892). On the origin of money. *The Economic Journal*, 2(6):239–255.

[147] Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

[148] Miller, H. (2019). Scales in spatial statistical analysis: other definitions, other fields. Accessed 08.2019, available at *rose.geog.mcgill.ca/wordpress/?p=2070*.

[149] Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M., and Alon, U. (2004). Superfamilies of evolved and designed networks. *Science*, 303(5663):1538–1542.

[150] 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.

[151] Newman, M. (2010). Networks: an introduction. *United Slates: Oxford University Press Inc., New York*, pages 1–2.

[152] Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404–409.

[153] Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2):167–256.

[154] Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the national academy of sciences*, 101(suppl 1):5200–5205.

BUPT

[155] Newman, M. E. (2005). A measure of betweenness centrality based on random walks. *Social networks*, 27(1):39–54.

[156] Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582.

[157] Newman, M. E. (2008). The mathematics of networks. *The new palgrave encyclopedia of economics*, 2(2008):1–12.

[158] Nicolis, G. and Nicolis, C. (2012). *Foundations of complex systems: emergence, information and predicition*. World Scientific.

[159] Noack, A. (2009). Modularity clustering is force-directed layout. *Physical Review E*, 79(2):026102.

[160] O'Connor, T. (1994). Emergent properties. *American Philosophical Quarterly*, 31(2):91–104.

[161] Oliveira, J. G. and Barabási, A.-L. (2005). Human dynamics: Darwin and einstein correspondence patterns. *Nature*, 437(7063):1251–1251.

[162] on Network Science for Future Army Applications, C. (2005). *Network Science*. National Academies Press.

[163] Onnela, J.-P., Arbesman, S., González, M. C., Barabási, A.-L., and Christakis, N. A. (2011). Geographic constraints on social network groups. *PLoS one*, 6(4):e16939.

[164] Opsahl, T., Agneessens, F., and Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social networks*, 32(3):245–251.

[165] Parigi, P. and Sartori, L. (2014). The political party as a network of cleavages: Disclosing the inner structure of italian political parties in the seventies. *Social Networks*, 36:54–65.

[166] Pastor-Satorras, R. and Vespignani, A. (2001). Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200.

[167] Peters, O. (2019). The ergodicity problem in economics. *Nature Physics*, 15(12):1216–1221.

[168] Polleit, T. (2011). What can the law of diminishing marginal utility teach us. *Mises Daily, Feb*, 11:2011.

[169] Ravasz, E. and Barabási, A.-L. (2003). Hierarchical organization in complex networks. *Physical Review E*, 67(2):026112.

[170] Rickards, M. and De Beaumont, S. (2000). *The encyclopedia of ephemera: a guide to the fragmentary documents of everyday life for the collector, curator, and historian*. Psychology Press.

[171] Rubinstein, A. and Wolinsky, A. (1987). Middlemen. *The Quarterly Journal of Economics*, 102(3):581–593.

[172] Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, 31(4):581–603.

[173] Schittkowski, K. (2002). Easy-fit: a software system for data fitting in dynamical systems. *Structural and Multidisciplinary Optimization*, 23(2):153–169.

[174] Scholz, M. (2010). Node similarity as a basic principle behind connectivity in complex networks. *arXiv preprint arXiv:1010.0803*.

[175] Scholz, M. (2015). Node degree distribution. Accessed 08.2017, available at *network-science.org/powerlaw_scalefree_node_degree_distribution.html*.

[176] School, P. I. (2017). Superhero connections. Accessed 08.2017, available at *research.prattsils.org/blog/coursework/information-visualization/superhero-connections*.

[177] Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., and Tufano, P. (2012). Analytics: the real-world use of big data. how innovative enterprises extract value from uncertain data. ibm global business services. *IBM Institute for Business Value, o. O*.

BUPT

[178] Sekara, V., Stopczynski, A., and Lehmann, S. (2016). Fundamental structures of dynamic social networks. *Proceedings of the national academy of sciences*, 113(36):9977–9982.

[179] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11):2498–2504.

[180] Šikić, M., Lančić, A., Antulov-Fantulin, N., and Štefančić, H. (2013). Epidemic centrality—is there an underestimated epidemic impact of network peripheral nodes? *The European Physical Journal B*, 86(10):440.

[181] Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, 63(2):129.

[182] Simon, H. A. (1992). Rational decision-making in business organizations. *USA: World Scientific*, 33.

[183] Smith, J. M., Halgin, D. S., Kidwell-Lopez, V., Labianca, G., Brass, D. J., and Borgatti, S. P. (2014). Power in politically charged networks. *Social Networks*, 36:162–176.

[184] Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. (2010). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, 27(3):431–432.

[185] Sonnenwald, D. H. (2007). Scientific collaboration. *Annual review of information science and technology*, 41(1):643–681.

[186] Staudt, C., Schumm, A., Meyerhenke, H., Görke, R., and Wagner, D. (2012). Static and dynamic aspects of scientific collaboration networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on*, pages 522–526. IEEE.

[187] Stepney, S., Polack, F. A., and Turner, H. R. (2006). Engineering emergence. In *11th IEEE International Conference on Engineering of Complex Computer Systems (ICECCS'06)*, pages 9–pp. IEEE.

[188] Strogatz, S. H. (2001). Exploring complex networks. *Nature*, 410(6825):268–276.

[189] Sun, C. (2019). Allimdb 5000 movie datasetmusic. Accessed 2019, available at kaggle.com/deepmatrix/imdb-5000-movie-dataset.

[190] Systems, C. (2019). Centrifuge. Accessed 2019, available at *centrifugesystems.com*.

[191] Teng, C.-Y., Lin, Y.-R., and Adamic, L. A. (2012). Recipe recommendation using ingredient networks. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 298–307. ACM.

[192] Testoni, S. (2017). The undoing project. *LSE Behavioural Science*.

[193] Topirceanu, A., Barina, G., and Udrescu, M. (2014a). Musenet: Collaboration in the music artists industry. In *2014 European Network Intelligence Conference*, pages 89–94. IEEE.

[194] Topirceanu, A., Udrescu, M., and Vladutiu, M. (2013). Network fidelity: A metric to quantify the similarity and realism of complex networks. In *Cloud and Green Computing (CGC), 2013 Third International Conference on*, pages 289–296. IEEE.

[195] Topirceanu, A., Udrescu, M., and Vladutiu, M. (2014b). Genetically optimized realistic social network topology inspired by facebook. In *Online Social Media Analysis and Visualization*, pages 163–179. Springer.

[196] Topirceanu, A., Udrescu, M., Vladutiu, M., and Marculescu, R. (2016). Tolerance-based interaction: a new model targeting opinion formation and diffusion in social networks. *PeerJ Computer Science*, 2:e42.

[197] Truong, K. (2017). Imdb network analysis. Accessed 08.2017, available at *katie-truong.github.io/network-analysis/IMDB*.

[198] Udrescu-Miroslav, M. (2017). *Network Science in Computer Engineering and Information Technology*. PhD thesis, "Politehnica" University of Timisoara, Romania.

BUPT

[199] Uzzi, B. (1996). The sources and consequences of embeddedness for the economic performance of organizations: The network effect. *American sociological review*, pages 674–698.

[200] Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative science quarterly*, pages 35–67.

[201] Valente, T. W., Fujimoto, K., Unger, J. B., Soto, D. W., and Meeker, D. (2013). Variations in network boundary and type: A study of adolescent peer influences. *Social Networks*, 35(3):309–316.

[202] Van Der Hofstad, R. (2016). *Random graphs and complex networks*, volume 1. Cambridge university press.

[203] Van Steen, M. (2010). Graph theory and complex networks. *An introduction*, 144.

[204] Walters, P. (2000). *An introduction to ergodic theory*, volume 79. Springer Science & Business Media.

[205] Wang, P., Lü, J., and Yu, X. (2014). Identification of important nodes in directed biological networks: A network motif approach. *PloS one*, 9(8):e106132.

[206] Wang, X. F. and Chen, G. (2003). Complex networks: small-world, scale-free and beyond. *IEEE circuits and systems magazine*, 3(1):6–20.

[207] Wasserman, S. and Faust, K. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.

[208] Watts, D. J. (1999). *Small worlds: the dynamics of networks between order and randomness*. Princeton university press.

[209] Watts, D. J. (2004). *Six degrees: The science of a connected age*. WW Norton & Company.

[210] Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442.

[211] Wernicke, S. (2006). Efficient detection of network motifs. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 3(4).

[212] Wernicke, S. and Rasche, F. (2006). Fanmod: a tool for fast network motif detection. *Bioinformatics*, 22(9):1152–1153.

[213] White, M. V. (2002). Doctoring adam smith: the fable of the diamonds and water paradox. *History of Political Economy*, 34(4):659–683.

[214] Wright, R. and Wong, Y.-Y. (2014). Buyers, sellers, and middlemen: Variations on search-theoretic themes. *International Economic Review*, 55(2):375–397.

[215] Wuchty, S., Oltvai, Z. N., and Barabási, A.-L. (2003). Evolutionary conservation of motif constituents in the yeast protein interaction network. *Nature genetics*, 35(2):176–179.

[216] Xia, Y., Ren, X., Peng, Z., Zhang, J., and She, L. (2016). Effectively identifying the influential spreaders in large-scale social networks. *Multimedia Tools and Applications*, 75(15):8829–8841.

[217] Yu, W., Wen, G., Yu, X., Wu, Z., and Lü, J. (2014). Bridging the gap between complex networks and smart grids. *Journal of Control and Decision*, 1(1):102–114.

[218] Zajonc, R. B. (1985). Emotion and facial efference: A theory reclaimed. *Science*, 228(4695):15–21.

[219] Zhang, H., Zhang, J., Zhou, C., Small, M., and Wang, B. (2010). Hub nodes inhibit the outbreak of epidemic under voluntary vaccination. *New Journal of Physics*, 12(2):023015.

A. Implementation of Investment Actions

In order to offer a better insight regarding the implementation of the investment actions presented in section 5.3.3 of this thesis, in what follows I disclose code snippets I deem important from a programmatic point of view [1]. That being said, the implementation of each investment is done in the *DecisionManager* static class, extended from the *BL.Node* class; this allows for direct calls for each existing method of the BL.Node class. Additionally, each action of investment — initially enabled via the existing graphical user interface presented in Figure 5.5 — has its own *investmentCost* property, which determines the cost of investment for the given action.

```
public static class DecisionManager{
    private static readonly Random Rng = new Random((int)DateTime.Now.ToBinary());

    public static async Task<bool> DetermineDecisionScore(this Node node, FullSimulation
        currentSim)...

    public static async Task<bool> Expand(this Node parentNode, FullSimulation sim,
        ExpansionPattern pattern)...

    public static async Task<bool> ImproveProductionQuality(this Node node, FullSimulation
        sim)...

    public static async Task<bool> CreateProduction(this Node node, FullSimulation
        currentSim)...

    public static async Task<bool> CreateLink(this Node node, FullSimulation currentSim)...
}
```

Listing A.1 Overview of the DecisionManager class.

A.1 Action 1: Creating a New Link

As a first step, all eligible target nodes are identified, namely the ones which are not directly linked to the node implementing the current action:

```
var validNodes = currentSim.Network.Where(n => n.Id != node.Id && node.Neighbours.All(nb =>
    nb.Id != n.Id);
```

Listing A.2 Snippet for creating a new link.

A random node is chosen from the *validNodes* array as an endpoint for the new link, by determining the shortest path between one such node and the one implementing the decision. This is achieved by using a variant of the breadth-first algorithm. The implementation of the search algorithm (Listing A.3) is done by means of a function which returns a new

[1] The term *node* is used to programmatically represent an economic agent

dictionary *Dictionary <int, int>* object. In this key-value dictionary object, each key is the index of the node through which the current link is created, while each value represents the index of a given node in the network (*i.e. List <Node>*).

```
public static class NetworkManager{
    public static void GetShortestPathsHeap(this Node start, List<Node> network){
        var networkSize = network.Count;
        var bfsResult = new Dictionary<int, int> { { start.Id, -1 } };

        bfsResult = GetShortestPathsByNeighbours(bfsResult, networkSize, new List<Node> { start }, network);

        start.ShortestPathsHeap = bfsResult;
    }

    public static List<Node> GetShortestPathToNode(this Node origin, Node dest, List<Node> network){
        if (origin == null || dest == null){
            return null;
        }

        if (origin.ShortestPathsHeap == null || origin.ShortestPathsHeap.Count == 0){
            return null;
        }

        if (!origin.ShortestPathsHeap.ContainsKey(dest.Id)){
            return null;
        }

        var res = new List<Node>();

        while (true){
            if (!origin.ShortestPathsHeap.ContainsKey(dest.Id)){
                return null;
            }
            if (origin.ShortestPathsHeap[dest.Id] == -1){
                res.Add(origin);
                return res;
            }
            res.Add(dest);
            dest = network.First(node => node.Id == origin.ShortestPathsHeap[dest.Id]);
        }
    }

    private static Dictionary<int, int> GetShortestPathsByNeighbours(Dictionary<int, int> heap, int networkSize, List<Node> startNodes, List<Node> network){
        while (true){
            if (heap.Count == networkSize || startNodes == null || startNodes.Count == 0){
                return heap;
            }
            var nextIterationNodes = new List<Node>();

            foreach (var startNode in startNodes){
                var neighbours = startNode.Neighbours;
                if (neighbours == null){
                    continue;
                }

                foreach (var neighbour in neighbours){
                    if (heap.ContainsKey(neighbour.Id)){
                        continue;
```

```
                    }

                    heap.Add(neighbour.Id, startNode.Id);
                    nextIterationNodes.Add(network.First(node => node.Id == neighbour.Id));
                }
            }

            if (nextIterationNodes.Count == 0){
                return heap;
            }

            startNodes = nextIterationNodes;
        }
    }
}
```

Listing A.3 Implementation of the breadth-first algorithm in TrEcSim.

If the current network is not connected one, creating a link might not be possible, in which case the breadth-first algorithm returns a *null* value; otherwise, the result is stored in the *pathToTarget* variable in order to compute the cost of the investment for the current decision. The snippet for computing this cost can be found in Listing A.4.

```
var investmentCost = 100.0;

if (pathToTarget == null){
    investmentCost *= Math.Pow(1.3, node.Neighbours.Count);
} else {
    investmentCost *= Math.Pow(1.3, pathToTarget.Count);
}
```

Listing A.4 Snippet for the breadth-first algorithm used to create identify the shortest path between any two nodes.

If the investmentCost is less than *node.SpendingLimit* (the available currency for the current node), a link is created between the node initiating the current action and the target-node. The full implementation of Action 1 can be found in Listing A.5.

```
public static async Task<bool> CreateLink(this Node node, FullSimulation currentSim){
    var validNodes = currentSim.Network.Where(n => n.Id != node.Id && node.Neighbours.All(nb
         => nb.Id != n.Id)).ToList();

    if (validNodes.Count == 0){
        return false;
    }

    var targetIndex = Rng.Next(0, validNodes.Count - 1);
    var targetNode = validNodes[targetIndex];

    node.GetShortestPathsHeap(currentSim.Network);

    var pathToTarget = node.GetShortestPathToNode(targetNode, currentSim.Network);

    var investmentCost = 100.0;
    if (pathToTarget == null){
        investmentCost *= Math.Pow(1.3, node.Neighbours.Count);
    } else {
        investmentCost *= Math.Pow(1.3, pathToTarget.Count);
    }

    if (node.SpendingLimit < investmentCost){
```

BUPT

```
            return false;
        }

        node.SpendingLimit -= investmentCost;

        var newLinks = new List<NodeLink>();
        newLinks.Add(new NodeLink { NodeId = node.Id, LinkId = targetNode.Id, SimulationId =
             currentSim.Simulation.Id });
        newLinks.Add(new NodeLink { NodeId = targetNode.Id, LinkId = node.Id, SimulationId =
             currentSim.Simulation.Id });

        var createdLinks = await NodeLinkCore.CreateAsync(newLinks, true).ConfigureAwait(false);
        if (createdLinks == null){
            return false;
        }

        var savedNode = await NodeCore.UpdateAsync(node, true).ConfigureAwait(false);
        if (savedNode == null){
            return false;
        }

        var logEntry = await SimulationLogCore.CreateAsync(new SimulationLog{
            Type = (int)SimulationLogType.Decision,
            NodeId = node.Id,
            Content = $"{(int)Enum.Decision.CreateLinks} {investmentCost}"
        }).ConfigureAwait(false);

        if (logEntry == null){
            return false;
        }

        savedNode.Neighbours = node.Neighbours;
        savedNode.ShortestPathsHeap = node.ShortestPathsHeap;
        node = savedNode;

        return true;
    }
```

Listing A.5 Full implementation of Action 1: the creation of a new link between two nodes.

A.2 Action 2: Creating a New Product

The algorithm searches for the set of products for which are neither produced, nor in demand by the current node (see Listing A.6). If this set is an empty one (*i.e.* no such products have been identified), the decision for investing in Action 2 will be aborted.

```
var validProducts = currentSim.Products.Where(product => !node.Productions.Any(p =>
     p.ProductId == product.Id) && !node.Needs.Any(n => n.ProductId == product.Id))
```

Listing A.6 Snippet for identifying the set of products which the current node can decide to produce.

On the other hand, if at least one such product has been identified, the underlying algorithm chooses a product in a random manner, along with all existing productions and demands from the current network.

```
var productionsForChosenProduct = currentSim.Productions.Where(p => p.ProductId ==
     chosenProduct.Id);
```

```

var needsForChosenProduct = currentSim.Needs.Where(p => p.ProductId == chosenProduct.Id);
```

Listing A.7 Snippet for identifying the set of products which the current node can decide to produce.

The average base cost of the respective product is computed, as well as the average quality, quantity needed and quality available. If there are any demands for the current product, the algorithm will compute the supply-demand ratio using the following snippet:

```
var generalRatio = Math.Sqrt(Math.Pow(1 - (double)neededQuantity / producedQuantity, 2));
```

Listing A.8 Snippet for computing the supply-demand ration for a given product.

The quality, quantity and base cost of the current product are determined by the following snippet:

```
var chosenPrice = averagePrice * generalRatio;

var chosenQuality = (int)(averageQuality * generalRatio);

var chosenQuantity = Rng.Next(0, (int)(neededQuantity * generalRatio));
```

Listing A.9 Snippet for computing the quality, quantity and the base cost of the current product.

The new production will be generated based on the obtained results. The cost of investing in Action 2 is computed using the following snippet:

```
var investmentCost = production.Quantity * production.PriceByQuality(currentSim.Simulation);
```

Listing A.10 Snippet for computing the cost of investment in Action 2.

If the cost of investing in Action 2 (*i.e.* investmentCost) is less than the curent node's available currency, the decision to invest in this actions is aborted. The full implementation of this decision is presented in Listing A.11.

```
public static async Task<bool> CreateProduction(this Node node, FullSimulation currentSim){
    var validProducts = currentSim.Products.Where(
        product =>
        !currentSim.Productions.Any(p => p.ProductId == product.Id && p.NodeId == node.Id) &&
        !currentSim.Needs.Any(n => n.ProductId == product.Id && n.NodeId ==
            node.Id)).ToList();

    if (validProducts.Count == 0){
        return false;
    }

    var chosenProductIndex = Rng.Next(0, validProducts.Count - 1);

    var chosenProduct = currentSim.Products[chosenProductIndex];

    var productionsForChosenProducts = currentSim.Productions.Where(p => p.ProductId ==
         chosenProduct.Id).ToList();
    var needsForChosenProducts = currentSim.Needs.Where(p => p.ProductId ==
         chosenProduct.Id).ToList();

    var averagePrice = 10.0;
    if (productionsForChosenProducts.Count > 0){
        averagePrice = productionsForChosenProducts.Average(p => p.Price);
    }

```

BUPT

```
    var averageQuality = 20;
    if (productionsForChosenProducts.Count > 0){
        averageQuality = (int)productionsForChosenProducts.Average(p => p.Quality);
    }

    var neededQuantity = 0;
    if (needsForChosenProducts.Count > 0){
        try {
            neededQuantity = needsForChosenProducts.Sum(p2 => p2.Quantity);
        }
        catch (Exception){
            //overflow
            neededQuantity = Rng.Next(10, 100);
        }
    }

    var producedQuantity = 0;
    if (productionsForChosenProducts.Count > 0){
        try{
            producedQuantity = productionsForChosenProducts.Sum(p2 => p2.Quantity);
        }
        catch (Exception){
            //overflow
            producedQuantity = Rng.Next(0, 100);
        }
    }

    var chosenPrice = averagePrice;
    var chosenQuality = averageQuality;
    var chosenQuantity = 30;

    if (producedQuantity != 0){
        var generalRatio = Math.Sqrt(Math.Pow(1 - (double)neededQuantity / producedQuantity,
             2));
        chosenPrice = averagePrice * generalRatio;
        chosenQuality = (int)(averageQuality * generalRatio);

        if (neededQuantity != 0){
            chosenQuantity = Rng.Next(0, (int)(neededQuantity * generalRatio) + 1);
        }
    }

    var production = new Production{
        NodeId = node.Id,
        ProductId = chosenProduct.Id,
        Price = chosenPrice,
        Quality = chosenQuality,
        Quantity = chosenQuantity
    };

    var investmentCost = production.Quantity *
         production.PriceByQuality(currentSim.Simulation);

    if (node.SpendingLimit < investmentCost){
        return false;
    }

    node.SpendingLimit -= investmentCost;

    var createdProduction = await ProductionCore.CreateAsync(production,
         true).ConfigureAwait(false);
```

```
    if (createdProduction == null){
       return false;
    }

    var savedNode = await NodeCore.UpdateAsync(node, true).ConfigureAwait(false);
    if (savedNode == null){
       return false;
    }

    savedNode.Neighbours = node.Neighbours;
    savedNode.ShortestPathsHeap = node.ShortestPathsHeap;
    node = savedNode;

    var logEntry = await SimulationLogCore.CreateAsync(new SimulationLog{
       Type = (int)SimulationLogType.Decision,
       NodeId = node.Id,
       Content = $"{(int)Enum.Decision.CreateProductions} {investmentCost}"
    }).ConfigureAwait(false);

    return logEntry != null;
 }
```

Listing A.11 Full implementation of Action 2: the creation of a new product.

A.3 Action 3: Improving Current Production

For the node investing in Action 3, the algorithm iterates over the current node's productions set, namely *node.Productions*:

```
foreach (var production in node.Productions)
```

Listing A.12 Snippet for iterating over a given node's own productions.

For each of the node's own productions, the algorithm computes the cost of investment (*i.e.* investmentCost) by using the snippet in Listing A.13. As previously, if this cost is less than the current node's available total currency, the quality of production will increase by 1 unit. The full implementation of this action can be found in Listing A.14.

```
var investmentCost = production.Quantity * Math.Pow(1 +
     sim.Simulation.ProductPriceIncreasePerQuality, production.Quality);
```

Listing A.13 Snippet for computing the cost of investment.

```
public static async Task<bool> ImproveProductionQuality(this Node node, FullSimulation sim){
   var ownProductions = sim.Productions.Where(p => p.NodeId == node.Id).ToList();

   foreach (var production in ownProductions){
      var investmentCost = production.Quantity * Math.Pow(1 +
           sim.Simulation.ProductPriceIncreasePerQuality, production.Quality);
      if (node.SpendingLimit < investmentCost){
         continue;
      }

      production.Quality++;
      node.SpendingLimit -= investmentCost;

      var savedProduction = await ProductionCore.UpdateAsync(production,
           true).ConfigureAwait(false);
      if (savedProduction == null){
```

BUPT

```
            return false;
        }

        var savedNode = await NodeCore.UpdateAsync(node, true).ConfigureAwait(false);
        if (savedNode == null){
            return false;
        }
        savedNode.Neighbours = node.Neighbours;
        savedNode.ShortestPathsHeap = node.ShortestPathsHeap;
        node = savedNode;

        var logEntry = await SimulationLogCore.CreateAsync(new SimulationLog{
            Type = (int)SimulationLogType.Decision,
            NodeId = node.Id,
            Content = $"{(int)Enum.Decision.ImproveProductions} {investmentCost}"
        }).ConfigureAwait(false);

        if (logEntry == null){
            return false;
        }
    }

    return true;
}
```

Listing A.14 Full implementation of Action 3: the improvement of production quality.

A.4 Action 4: Creating a New Node

The decision to invest in Action 4 implies — as previously stated in this thesis — the creation of a new node, which will inherit, to some extent, the characteristics (*e.g.* demands, productions, *etc.*) of the node initiating this action. In order to implement this mechanism (of inheriting from the parent node), I created the *ExpansionPattern* class (Listing A.15); the members of this class model the way a new node is created in the network.

```
public class ExpansionPatterns{
    public static ExpansionPattern SimpleChild { get; } = new ExpansionPattern{
        WealthPercentage = 10,
        LinkToParent = true,
        AdditionalLinks = 0,
        InheritNeeds = true,
        AdditionalNeeds = 0,
        InheritProduction = true,
        AdditionalProductions = 0
    };
}
```

Listing A.15 Snippet showing the declarations of the members inside the ExpansionPatterns class.

The algorithm starts off with the creation a new instance of the *BL.Node: childNode* class. The available currency available to the newly created node equals the cost of investment (*i.e.* investmentCost) of the current action; to this end, the *WealthPercentage* member plays an important role in computing this cost:

```
childNode.Money = parentNode.Money / ExpansionPattern.WealthPercentage;
```

Listing A.16 Snippet for computing the newly created node's currency.

BUPT

Based on the values of the other members of the ExpansionPattern class (*i.e. InheritNeeds*, *InheritProductions*), the childNode will also possibly inherit both the needs and productions of the parent node as well; similarly, the newly created node might also produce new products or have its own set of demands based on the *AdditionalNeeds* and the *AdditionalProductions* members.

```
public static async Task<bool> Expand(this Node parentNode, FullSimulation sim,
     ExpansionPattern pattern){
   var investmentCost = parentNode.SpendingLimit / pattern.WealthPercentage;

   var childNode = new Node{
      Id = -1,
      SimulationId = sim.Simulation.Id,
      Name = $"Node {sim.Network.Count + 1}",
      SpendingLimit = investmentCost
   };

   childNode = await NodeCore.CreateAsync(childNode).ConfigureAwait(false);
   if (childNode == null){
      return false;
   }

   parentNode.SpendingLimit -= investmentCost;

   var childLinks = new List<NodeLink>();

   if (pattern.AdditionalLinks > 0){
      add additional links
   }

   parentNode = await NodeCore.UpdateAsync(parentNode, true).ConfigureAwait(false);
   if (parentNode == null){
      return false;
   }

   if (pattern.LinkToParent){
      childLinks.Add(new NodeLink { NodeId = parentNode.Id, LinkId = childNode.Id,
           SimulationId = sim.Simulation.Id });
      childLinks.Add(new NodeLink { NodeId = childNode.Id, LinkId = parentNode.Id,
           SimulationId = sim.Simulation.Id });

      var createdLinks = await NodeLinkCore.CreateAsync(childLinks,
           true).ConfigureAwait(false);
      if (createdLinks == null){
         return false;
      }
   }

   if (pattern.InheritNeeds){
      var newNeeds = sim.Needs.Where(need => need.NodeId == parentNode.Id).ToList();
      newNeeds.ForEach(need => need.NodeId = childNode.Id);

      var createdNeeds = await NeedCore.CreateAsync(newNeeds, true).ConfigureAwait(false);
      if (createdNeeds == null){
         return false;
      }
   }

   if (pattern.InheritProduction){
```

BUPT

```
        var newProductions = sim.Productions.Where(prod => prod.NodeId ==
            parentNode.Id).ToList();
        newProductions.ForEach(prod => prod.NodeId = childNode.Id);

        var createdProductions = await ProductionCore.CreateAsync(newProductions,
            true).ConfigureAwait(false);
        if (createdProductions == null){
            return false;
        }
    }

    var logEntry = await SimulationLogCore.CreateAsync(new SimulationLog{
        Type = (int)SimulationLogType.Decision,
        NodeId = parentNode.Id,
        Content = $"{(int)Enum.Decision.Expand} {investmentCost}"
    }).ConfigureAwait(false);

    return logEntry != null;
}
```

Listing A.17 Full implementation of Action 4: the creation of a new node.

DORIN LELEA

Digitally signed by DORIN LELEA
T=Profesor, C=RO, S=TIMIS, L=TIMISOARA, O=UNIVERSITATEA POLITEHNICA DIN TIMISOARA, E=dorin.lelea@upt.ro,
SERIALNUMBER=201104209DL10, G=DORIN, SN=LELEA, CN=DORIN LELEA
Atest integritatea acestui document
Computerul meu
2020.12.04 16:01

BUPT