Babeş–Bolyai University
Cluj-Napoca, Romania
Faculty of Physics

Ph.D Thesis Summary

Statistical Physics Approach to Complex Social Systems

Advisor: Prof. Dr. NÉDA ZOLTÁN

Derzsy Noemi

Referees: Prof. Dr. SIMON VIORICA

Prof. Dr. KÁSA ZOLTÁN

C.P.I. Dr. BENDE ATTILA

Cluj–Napoca

2012

Keywords: complex systems, random networks, social networks, econophysics, Paretos law, income distribution, income dynamics, wealth distribution

Contents

Introduction

In our daily life we are surrounded by complex systems without taking notice of their presence. We go to work where we become part of telephone, e-mail or internet systems. Also, we interact with firms, work colleagues, contributing to the development of business/economical systems. In private life we are members of friendship networks, sexual relation complex systems of social nature. These systems are ranging from global to atomic and molecular extent such as protein exchange, neural activity, virus spread. The study of these complex systems before the 20^{th} century was quite difficult, even impossible. The complexity of these systems at topological level stands in the huge number of elements and the tremendous number of connections (links) between them. The interaction between elements and specific self-organization characteristics reflect the complexity of these systems at behavioral level as well. Studying this large amount of data set can be achieved only with high technology computers and database administration tools. Thus this research field before the booming development of computational technology seemed a less approachable research topic. Nowadays many complex systems have been already studied due to present advanced technology and large database management abilities. The exponential development of computer science played a key role in the study of complex systems. Many information that so far were unreachable, like telephone communication records between humans, bank transactions, neural activity, virus spread, all these information got stored in computer databases. The latest popular social networking services store huge information on social complex systems. Their records could provide valuable information on social complex system topology, organization, interaction, behavior and time evolution. In the 21^{st} century we live in a highly digitalized era. Traffic on streets, in are monitored with video cameras, our social interactions are all recorded on telephone service databases on internet, e-mail service records. These data sets offer unique possibilities to analyze the structure and dynamics of these complex systems and based on these information these systems could be modeled with appropriate programming skills. Having access to all these information, human activity patterns could be revealed. Complex systems belonging to various fields have been successfully studied and many progresses have been reported in scientific literature. However the social complex system study is much more limited in analyzing human interactions. Unlike microscopic creatures (neurons, bacteria) humans wish to protect their privacy. Thus in spite of the enormous number of available digitalized records on social interaction the study of such complex systems proves to be limited. The confidentiality of these data sets restricts the access of scientists to many valuable information.

Despite the limitations of obtaining databases of social nature, the end of the twentieth century's research condensed around the study of complex systems from various fields. More particularly physicists got involved in research topics from fields like economy, biology, sociology by applying statistical physics methods and models aiming to understand these complex systems' structure and dynamics. Thus the statistical physics tools proved to have highly interdisciplinary applicability. Reported results remarkably validate the efficiency of statistical physics approach in various study fields. In this period the Networks research field evolved rapidly.

Networks are a new domain of Statistical Physics, developed mainly from the graph theory. The importance of using networks in science increased strongly, when researchers from different study fields (physicists, biologists, physicians, etc.) discovered that many social, natural or artificial systems are governed by the underlying network structure. Networks are usually the backbone of Complex Systems. This means that finding out the characteristic network structure and time evolution could help us in understanding many aspects of complex systems. The study of complex systems finds its roots in the graph theory. A graph is an abstract representation of a set of objects connected by links. A network is a large graph with numerous elements and links. Networks are the backbone of complex systems. Studying their topology and dynamics are quite straight forward by using classical methods of statistical physics. Successfully determining these features helps in understanding a complex system's behavior and time evolution. With these information the system can be modeled. The creation of a true model is crucial in order to predict a future structure of a complex system.

Prediction raises, however, new questions: if we can predict, can we prevent or change the natural behavior of a complex system with an external tool? For example, if we get to know the trajectory of virus spread, can we localize it and stop it efficiently? Or can we create a biological weapon with it, orienting and focusing the movement and evolution of the virus system in a certain direction?

In this blooming period of interdisciplinary application of statistical physics methods many physicists chose to study complex systems of economic nature, like trade relations, economic transactions, wealth distributions, company interdependencies, etc. These researches lead to the development of econophysics, also a new interdisciplinary research field. In this period of global financial crisis this topic is of great interest for everyone who is preoccupied with understanding the behavior of these complex systems and wants to speculate their future outcome. Econophysics researchers try to understand their structure and governing mechanisms aiming to create models that can reproduce the topology and offer the possibility to analyze dynamical processes on these systems. This way the prediction and control of economic systems can be attained. The past few years of research offered numerous results enlightening characteristics indicating that these systems are not random and unpredictable. This encouraged further investigation in economics by using statistical physics approach. Results revealed certain universal laws that these systems obey, irrespective of the geographical region the analyzed data come from.

The structure of the thesis

In the present thesis two complex social systems are studied. The study aims to investigate two complex social systems by investigating their underlying network structure. The first one reveals the professional connections between universities from Europe. The second network based on a social security data set reveals the network of employers for a district in Romania. Their structures are investigated aiming to understand the behavior of these systems. The thesis also discusses income distributions, a research area of econophysics, another modern research field that uses statistical physics methods to analyze complex social systems of economic nature. Our research in this direction is based on the analysis of a complete real economic data set. Here we study income and wealth distribution and analyze the time evolution of this economic system.

The thesis is structured as follows. The first chapter defines Networks and reviews the main characteristic measures from graph theory used in studying networks. The basic types of networks and their particular features are also synthetized. The second chapter presents the most important statistical physics methods used in studying random networks from structure to dynamics. The third chapter presents a brief review on the development of social network study and enumerates the most important results reported on previously investigated complex social systems. The study of our two social networks is discussed and the significant conclusions of our work are drawn.

The fourth chapter is condensed around income and wealth distribution studies in complex social systems. It briefly reports its evolution and pioneering works in this field. In the last part of the chapter we report our study on the structure of the income distribution and analyze the dynamics of the system. The main results and conclusions are also discussed. The thesis contains original work and results that have been published in several papers. In the last chapter we state the final conclusions of our work and stress the importance of our results for future researches in this direction.

A review of Part I

Networks

Part I contains a review on the development of Networks from graph theory and presents elements of graph theory that are characteristic measures of networks. It contains descriptions of the basic graph types, their properties and reveals the importance of networks approach in complex system study.

To understand a complex system we study its underlying network structure. The network topology helps us visualize the system and its features' time evolution. Networks are a new domain of Statistical Physics that have developed mainly from the graph theory. Graph theory bases were established in the eighteeth century by Leonhard Euler. Since then graph theory evolved a lot and in the twentieth century the main research became focused around the study of random graphs. In that period it was believed that connections between the nodes in a complex system are randomly assigned. The theory of random graphs was founded by Paul Erdős and Alfréd Rényi. Many useful theorems resulted from the observation that the probabilistic methods are helpful in many graph theory problems, so for many years it was believed that all complex systems are driven by randomness [1, 2]. There are certain characteristic measures for networks that enlighten their topological properties such as the degree distribution function, average path length, clustering coefficient. Depending on the values of these quantities, scientists developed three basic category of network models that are representative for the majority of studied natural or artificial complex systems: the random (Erdős-Rényi) network, small-world (Watts-Strogatz) network and scale-free (Barabási-Albert) network classes. While the first two models are based on random attachment rules in the system, the last one was introduced by Barabási and Albert after researches proved that real systems are not driven by randomness [3]. This is the most recent model and research results indicate that this type of network is quite frequent in social and biological systems. The two generic features of the scale-free model: growth and preferential attachment produce interesting dynamics of the nodes. Study of these systems' topology and dynamics represents a key for predicting the topology of complex systems, to control their evolution, to cure a damaged system or to destroy one in the most effective way. In the last few years many complex systems that surround us were studied by their underlying network. The vast research in this field demonstrates that real networks are not simple Erdős-Rényi type graphs as it was believed in the beginnings. They have well outlined features following a specific law. This makes the real system heterogeneous and complex. Preliminary steps in order to understand a complex system's behavior are thus the analysis of the network's topology study and dynamics. The time evolution study is vital because every system is time dependent and suffers changes. The key here is to follow up how the topology characteristics change in time. After determining the rules that govern the dynamics of the system, a model that reproduces this can help in predicting the future evolution of the network. Another advantage of such a model is the possibility to study the system's attack tolerance to external interference or to errors that affect a certain component of the system. Researches have shown that scale-free systems governed by preferential attachment have robust topology, thus a random attack on a node does not influence the system as much as it does in case of a random network. This is understandable, because as it was presented previously, the scale-free network have many nodes with very few links and a small number of hubs that collect the most links in the system. In contrary, the random network's nodes possess almost the same number of connections. Thus, a random attack would have the same devastating impact on each element equally. Destroying a node, the links would be destroyed, as well, and the system would fall apart into several isolated components. This is not the case for scale-free networks, where a random attack wouldn't have a decisive impact. The phenomenon can be understood by analyzing the degree distribution function of scale-free networks: the majority of the nodes have very few links, while contrary to these, there is a small number of hubs that possess a huge number of nodes. A random attack on the network is likely to hit the nodes with few connections, because these nodes are the most frequent ones in the network. Destroying such a node will not affect the connectedness of the system. These robust networks, though, are very easily destroyed with a targeted attack. Targeting the hubs would result in destroying a massive number of connections, leading to an efficient annihilation of the system. The fact that a random network is a giant cluster, a node's removal has an impact on the entire system, while the scale-free network that organizes itself into several connected clusters, makes the system robust against attacks.

Controlling a complex system means to know exactly how the system reacts to certain interferences and how it evolves in time. Unfortunately, this is far from being reached. The complexity of these systems is so remarkable that a total control over them seems impossible, at least for now. Though, many efforts have been made to model complex systems and predict their evolution in time. Because of the variety of the patterns different networks have, a model applicable for a system, won't be valid for another one.

Present research in this field intend to find as much answers as possible regarding the general features of the complex system behavior in a hope to efficiently cure presently incurable diseases, targeting with drugs the key nodes in the protein, tissue, cellular or neural network, to stop virus spread and even prevent it with targeted vaccination, to efficiently annihilate organized crime networks and to control many other complex systems that affect our lives.

Statistical Physics methods in studying random graphs

This part presents by detail the Statistical Physics methods applied for the study of dynamical processes on networks. Also it discusses the researches condensed around network modules. Vast studies of the past few years reported that elements of these heterogenous systems tend to organize themselves into communities (modules). Studies revealed that these modules have different structure from the network studied as a whole. Even more, it was observed that certain communities have higher influence on the entire network's behavior and topology than others. These observations underlined the importance of analyzing the network through its modules. Numerous algorithms were developed for efficient community detection and structure analysis. This strategy contributed to the understanding of many real complex systems from many disciplines including biology, information science and sociology [4, 5].

Statistical Physics provides efficient tools for dynamical evolution study on complex networks. It offers models that help study phase transitions, synchronization properties and critical points in a network. It also helps identify crucial nodes in a system, analyze clustering properties and organizing principles [6].

The scientists gained interest in dynamical processes on Networks that can be discussed only within statistical physics framework. For example, it is a problem of great interest in complex social systems the spread of diseases, rumor and information flow, or consensus formation on social networks. The robustness of networks, effects of external and internal attacks have also been widely investigated. All these studies are based on statistical physics models that were adapted on networks in order to get a better insight on complex system behavior and dynamics of processes that take place inside.

Network scientists are particularly concerned with the understanding of the synchronization phenomena, that corresponds to a state when elements of the system follow the same dynamical pattern. An efficient statistical physics tool for analyzing this phenomena is the Kuramoto model. Many researches rely on this model when studying certain community's influence on the entire network. Studies made so far demonstrated that modules synchronize at different times and in a hierarchical way. Thus the Kuramoto model not only helps to study the synchronization in a network, but it can also identify progressively the basic modules, communities in a network. It is a method that proves to be efficient, for example in rumor spreading study on social systems.

Besides synchronization there is another very important phenomena that occurs in complex systems and plays a crucial role in understanding the network's behavior. This is the phase transition, phenomenon also widely studied in statistical mechanics. The classical Ising model aims to analyze cooperative behavior and different types of orderings. This model applied on networks is used to study interactions between nodes and their impact on the entire network's behavior. The Ising model stands at the bases of many models of social behavior and opinion dynamics.

The standard Potts model in Statistical Physics is used to model and study interaction between spins on a lattice. This model comes as a generalization of the Ising model, more precisely the Ising model is equivalent with the two-state Potts model. The Potts model is often used in Physics for phase transition study. In one dimension it is exactly solvable which makes this model a powerful tool. In network research it is a powerful tool for phase transition and cooperative phenomena study.

The Voter model offers the simplest method to analyze the collective behavior like consensus problem, opinion formation, etc. in social complex systems.

The percolation theory in Network research is used for studying the effect on the network structure when removing a node from the system. Thus the percolation problem considers an arbitrary network topology in which each node is occupied with probability p and links are present only between occupied nodes. As p increases the clusters in the network emerge. This leads to the development of the site percolation problem that studies the properties of the clusters as a function of the p occupation probability.

A review of Part II

Part II contains the original contributions in two different topics: networks approach to complex social systems and income distribution and dynamics study in a social system.

Social networks

A social network defines a network type where the nodes are individuals and the edges reflect a type of social connection like friendship, professional connections [7], sexual relations, etc. The nodes and edges can be defined depending on the form of social interaction one is interested in. Consequently nodes can refer to individuals of groups of people while the edges reflect a type of social connection between the nodes. Thus the social network study offers a large spectrum of analysis on human interaction.

The social network study aims to enlighten the topology and behavior of complex system of human interaction and to analyze dynamical processes on them.

The communication between individuals increased in the 21^{st} century due to the emergence of computer technology. The exponential growth of internet communication allowed social network studies that in the past seemed inconceivable. Besides the vast data sets that are available nowadays, their content is reliable and leads to truthful results, contrarily to previous information collecting methods such as surveys and questionnaires. Even more the digitalized information records allow the construction of networks of impressive scale, performance that could not be accomplished by other information gathering methods. Due to the large number of online databases the variety of the network studies increased rapidly. While in the past the information gathering methods were shallow and constrained the research possibilities to only a certain type of social interaction analysis, the present emergence of electronic records allows the investigation of many social networks of various types.

Many significant research results are reported in literature, such as mobile telephone communication network, e-mail communication networks, social, movie actor collaboration network. The latest researches condense around the most popular social networking services such as Facebook and Twitter. Various type of dynamical processes are captured in these databases like social information flow, opinion formation, etc. that offer exceptional opportunities to enlighten the structure and behavior of complex systems of social nature.

The application of social networks also pervades areas like criminality fight and prevention. The organized crime groups, mafia, drug, weapon, human traffic, are all organized in networks. Understanding the behavior of these systems can lead to their more efficient and durable annihilation. Researches conducted in this direction provide a good sample of the efficiency of the social network approach in criminal activity prevention or demolition.

Another subject of great interest is represented by virus spreading pattern analysis. In this area social network study once again proves to offer rapid advances. The contagious virus spreads can be effectively mapped on a social network that has known structure and behavior. Scientists created computer-based models to realistically simulate the spread of contagious diseases. Useful researches were conducted that aimed to model efficient vaccination campaigns [8].

As prior works in the field demonstrate, the significance of the social network studies rely in the interdisciplinary applicability of the techniques, models and results. Social network approach can enlighten professional connections, can determine the spread of infectious diseases, it can efficiently be used as criminality prevention tool, information flow controlling or opinion formation. The latest orientation in this field is headed to controlling human groups and mobilities [9].

The study of the Erasmus student mobility network

The database that we have studied contains information on the Erasmus student mobility for year 2003, having 134330 total number of involved students and 2333 European universities that took part in the program on the studied year.

Our main purpose was to investigate the network of professional connections. Due to the large number of nodes and links in the network a graphical representation of the entire network was not informative. Thus we have plotted for the non-weighted network in hierarchical representation the hubs with at least 55 connections in Fig. 1.

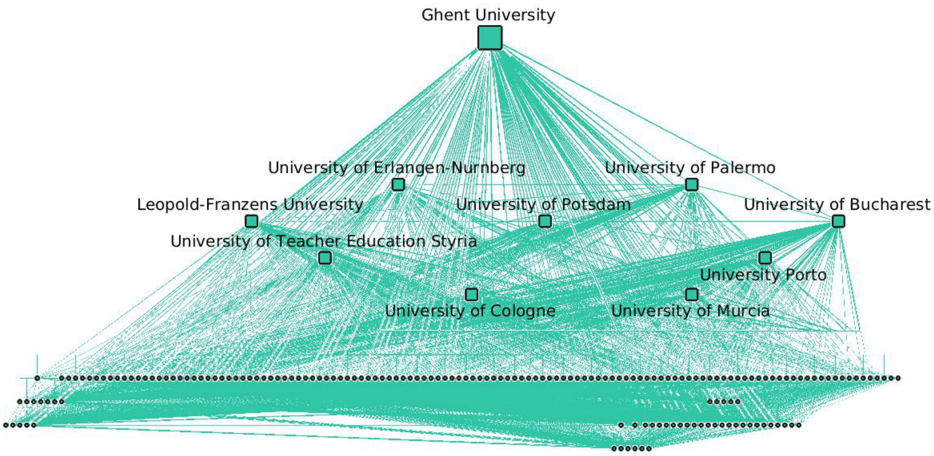

Figure 1: The k-degree component of the NNESM network for k=55. The 180 nodes of this component are present in hierarchical arrangement. The most highly connected 10 hubs are indicated by their names [10].

Fig. 2 visualizes the central part of the weighted version of the network that contains universities that have exchanged at least 15 students in year 2003.

The presence of hubs in the system led us to the assumption that we have a social network with scale-free characteristics (power-law degree distribution, small average path length, high clustering coefficient). Aiming to validate our expectation, we studied the topology of the Erasmus network, first analyzing it as non-weighted and non-directed (NNESM) and secondary as weighted and directed network. To our surprise, we found that the degree distribution function of the number of connections has exponential tail. We obtained for the global clustering coefficient $C_g = 0.183$ and for the local clustering coefficient $C_l = 0.292$. High clustering coefficients are originate from the existence of strong transitivity in human relationships, thus it is a feature present in various types of social networks. Further analysis on the topology concluded that the ESM network exhibits several connected

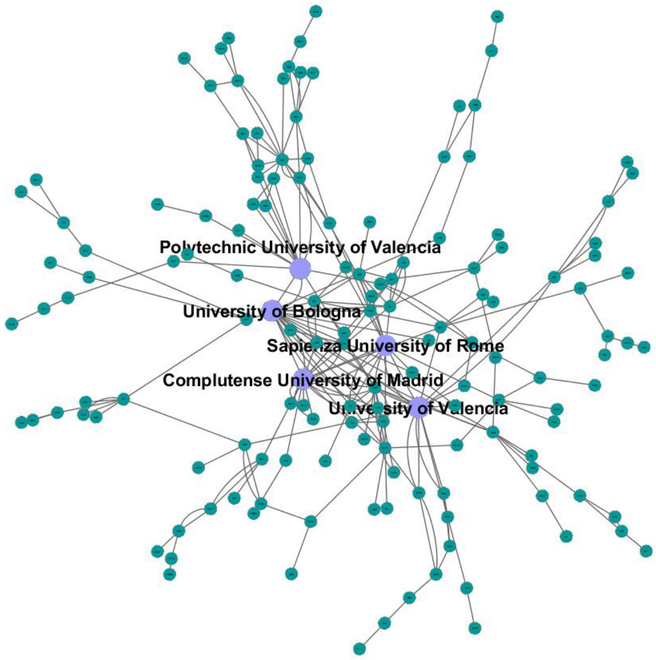

Figure 2: Part of the weighted ESM network: the figure shows the largest connected components obtained by using links on which there are at least 15 students exchanges in year 2003. This component contains 149 elements, the highlighted nodes are the hubs characterized with the most intense student mobilities [10].

components with one giant component that contains 99% of the nodes. The average distance between the nodes in this large component is 2.91 indicating a strongly connected network.

A main characteristic of social networks is the preferential linking or assortative mixing. A relevant selective linking in our network would be a closed interconnectivity of the highly connected universities, that would mean that universities with high number of links prefer to connect with universities that also have large number of connections. Fig. 3(a) is the mean degree of the neighbors represented as the function of the degree of the node (green symbols).

Fig. 3 shows no clear increasing trend, concluding rather a random attachment rule in establishing the connections. Based on this observation we created a simple configuration model that considers random link allocation. The resulted network presented similar topological properties with the

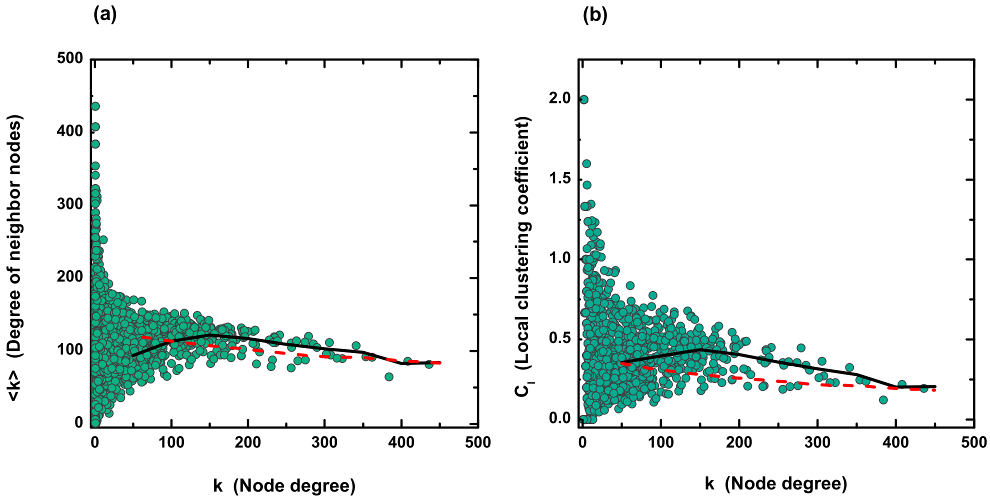

Figure 3: Degree of a node's neighbors (a) and the local clustering coefficients of the nodes (b) plotted as a function of the degree of a node (green circles). The black line on both figures indicate averages of the points taken with a moving average of length 50 on the horizontal axes. Averages calculated for the network obtained from the configuration model appear with dashed red lines.

NNESM network. We assume that the number of Erasmus links a university has is proportional with its size. This idea results from the fact that the number of ESM connections a university has is proportional with the number of professional connections of the professors, which is proportional with the number of professors. But it is known fact that the number of professors of a university reflects the number of the students the university has. We demonstrated that the exponential nature of the size-distribution is a general feature for all geographical regions and it can be explained with the maximum entropy principle.

We also aimed to model the weighted and directed network in order to get a better insight into the development of the observed distribution of weights by adapting the same configuration model to the weighted and directed version of the network. The ESM network had 2/3 unidirectional connections and 1/2 bidirectional ones. Based on this observation in the initial model the links were randomly switched with probability 2/3 into unidirectional and with 1/2 probability into bidirectional links.

It can be observed in Fig. 4 that the IN and OUT degree distribution of the configuration model network is similar with the real ESM IN and OUT

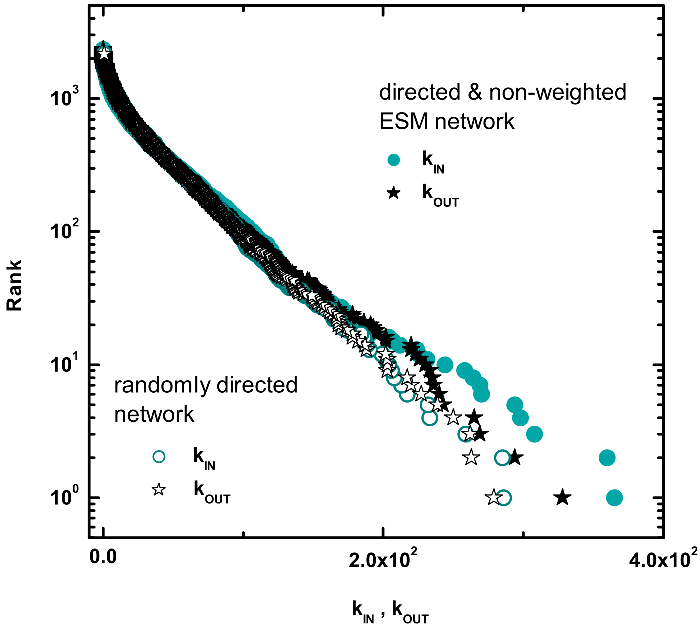

Figure 4: Rank of universities as a function of their IN and OUT degree for the directed ESM network (filled symbols) and for the randomly directed configuration network (open symbols).

degree distributions.

The weighted version of the network was also investigated. The weight gives the strength of a connection, in our case the number of student traveling from the home university to the visited one. Once again we applied the initial configuration model and partitioned the number of students on the previously assigned directed links. In order to do that we analyzed the student flow from universities as a function of the number of OUT connections a university has. Next this scaling was used to partition the total number of students on their home universities. The partitioning raised again the question of preferential linking. Thus we assigned the number of students in two different manners. First we chose randomly the k_{OUT} connections resulting in a partitioning with no preferential linking. The second method was to assume that links that are already more populated with students will be chosen with higher probability resulting a preferential attachment.

Fig. 5 shows the weight distribution on the configuration model using both methods and the weight distribution of the original ESM network. The plot concludes that the preferential linking is indeed important for this allocation process.

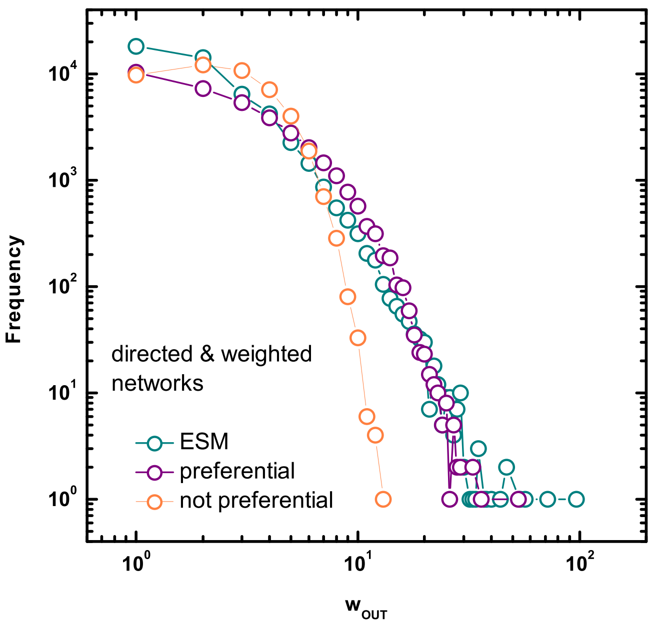

Figure 5: Comparison of weight distributions for the directed and weighted networks. Each network's data points are depicted with different colors for clearer visualization.

Employer network generated by employee mobility

The goal of the present work was to study a real social security database from Romania that contains information about a complex economic trade system from the Cluj district. The data set offers monthly information regarding employers with head-offices registered in Cluj district and their employees for years 2001-2007. This is a codified social security data set that offers unique possibility to study a real and complete complex economic system by analyzing its underlying network structure. The annual employee mobility networks were generated for each year (2001-2007) in particular and our aim was to study their topological properties. These networks were constructed by nodes being the employers and considering a connection between two nodes whether an employee changes its workplace from one to the other. The networks were analyzed, both as directed and undirected networks.

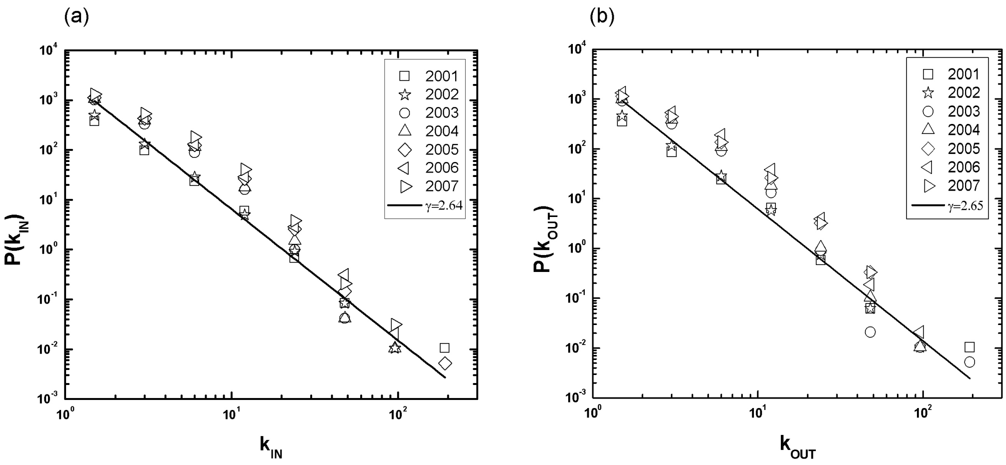

Figure 6: Degree distribution function constructed by a logarithmic bin for the incoming links on log-log axis. The thick continuous line is a power-law fit of the data points with exponent $\gamma = 2.64$ (b) Degree distribution function constructed by a logarithmic bin for the outgoing links on log-log axis. The thick continuous line is a power-law fit of the data points with exponent $\gamma = 2.65$ [11].

The logarithmically binned degree distribution functions plotted in Fig. 6 for the directed networks and Fig. 7 for the non-directed networks exhibit clear straight lines. The power-law fits for the global values of the plots with exponent $\gamma \sim 2.5$ indicate that the nodes are not randomly connected.

To further study the topology of these networks, the clustering coefficient and average path length were also analyzed. The small values obtained for the average path length ($\ell \sim 2.5 - 3.0$) suggest that the nodes in the system are closely connected, specific feature of social networks. This characteristic allows high communication traffic, consequently a better information exchange in the system. For each year, the obtained values of the clustering coefficient are relatively high ($C \sim 3$), suggesting that employers are organized into densely connected clusters. Thus these results confirm that these economical networks, both for the directed and non-directed versions, are of Barabási-Albert network type.

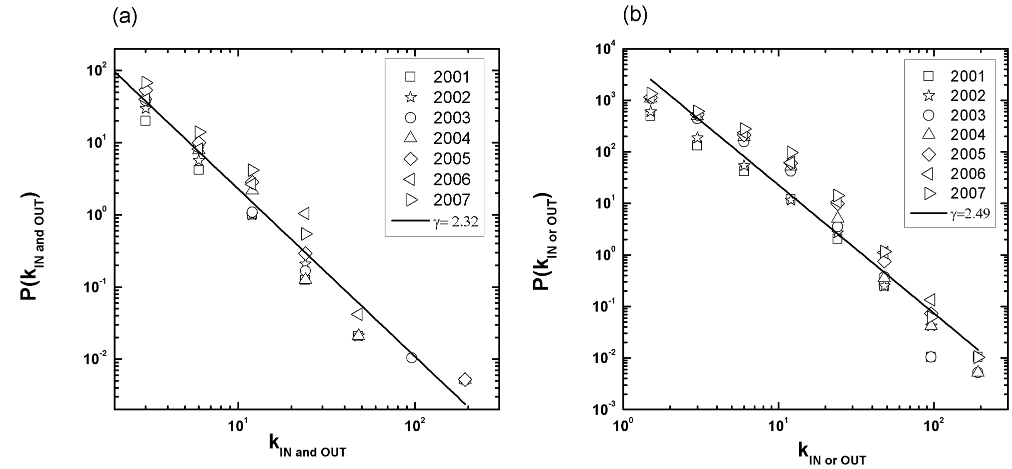

Figure 7: Degree distribution function constructed by a logarithmic bin for the in- and outgoing links on log-log axis. The thick continuous line is a power-law fit of the data points with exponent $\gamma = 2.32$ (b) Degree distribution function constructed by a logarithmic bin for the in- or outgoing links on log-log axis. The thick continuous line is a power-law fit of the data points with exponent $\gamma = 2.49$ [11].

In conclusion we have enlightened the topology of two complex social systems. Thus the network approach in studying complex systems and dynamical processes on them proved to be efficient. Using statistical physics methods, we created a model that truthfully reproduced the features of a social network. Appropriate models can improve further studies and can predict the evolution in time of these complex systems.

Income distribution patterns

Many physicists chose to study systems of economic nature, like trade relations, economic transactions, income and wealth distributions, company interdependencies,etc. These researches lead to the development of econophysics, a new interdisciplinary research field. In this period of global financial crisis this topic is of great interest for everyone who is preoccupied with understanding the behavior of these systems and wants to speculate their future outcome. The rapidly increasing publicly available electronic data records on world-wide economy allowed several studies to be conducted in this area. However, many detailed data sets concerning incomes, wealth and money transactions are kept confidential and are rarely obtained for research purposes. A popular econophysics research topic is the study of money distribution, wealth distribution and income distribution in societies [12]. Several modeling attempts and universal laws were obtained in this area that are applicable to economic societies from different demographic regions [13]. The term *income* refers to an amount of money earned from salaries and investments.

Due to the complex structure of the underlying social networks on which agents interchange money, researchers proposed models that implement network approach that turned out to be helpful in understanding a complex system's structure and behavior from various fields. It also had an important impact on research on economic systems, as well. Economic interactions between agents are considered on small-world or scale-free networks [14, 15] since prior network researches on complex social systems concluded that these types of networks reproduce realistically the structure of social systems. A model based on first-degree family relations networks is a great example of the effectiveness of such approach. The model produces a realistic social network topology and in the mean time generates a truthful wealth distribution [16].

Our research in this direction is based on the income distribution study of an economically significant region in Romania. The investigated data set is the same as the one exploited in our prior work on studying the employer networks generated by the employee mobility. Our data set containing personal income of employees in a district from Romania offers the opportunity to analyze the real income distribution for 9 consecutive years (2001-2009). This huge data set storing information on 535 401 employees and 39 398 employers offers the chance to analyze individual's income distribution. The income in this context is each individual's total earned salary in one year.

Italian economist Vilfredo Pareto observed that wealth distribution in societies obeys a general rule that in forthcoming papers was referred to as Pareto's law [17]. Power-law distributions occur in several real complex systems, natural and artificial ones. Studies from various fields, such as copies of books sold, flowering plants, city sizes, paper citations, link to pages on world wide web, word frequency, diameter of moon craters, etc., all confirm the power-law nature of these distributions [18]. Particularly, its presence is revealed in many studies on the economy of different demographic regions. This economic principle states that 80% of the income goes to 20% of the population, thus in literature is also mentioned as the 80-20 rule. Such a distribution is well described by a power-law with exponent α, named Pareto index.

In Fig. 8 we plotted the income distribution for the upper region ($rank <$

10000) of the distribution function that contains the highest salaries. The Pareto's law is clearly observable for in this part. The power-law fits of the functions resulted a highly stable Pareto exponent over the years with values $\alpha \sim 2.5$.

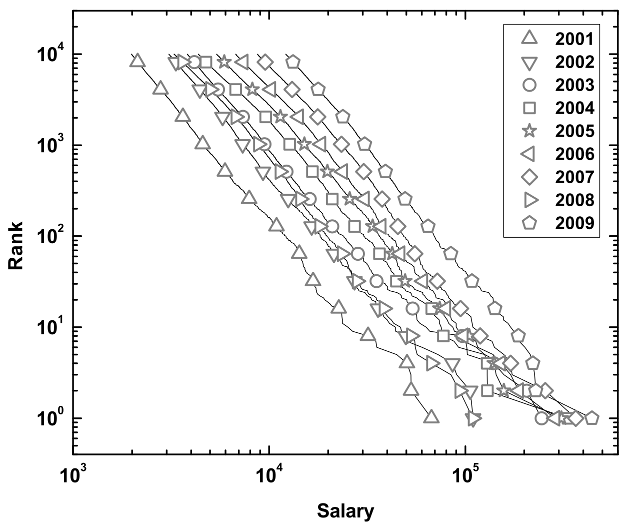

Figure 8: Rank-salary plot on double-logarithmic axis for the top 10000 salaries. Data for Cluj districts (Romania) for all registered workers between years 2001-2009 nominal in Euro. The income distributions are plotted for each year in particular.

In Fig. 9 we analyzed the low and medium part of the income distribution. In this region the income distribution can be fitted by the $P(W) = K \cdot \exp(-W/T_r)$ function, where T_r is the income temperature.

Fig. 9 suggests the validity of the exponential decay for the lower region of income. The income temperature indicate an increasing tendency over the years, except year 2008 when the global financial crisis hit Romania. This year the income temperature registers a strong decrease.

Our primary interest was the governing mechanism behind the relatively constant value of the Pareto exponent during this 9 year time period based on the major fluctuations of the C proportionality constant in the power-law region and the constant change of the income temperature observed in the low or medium income region the stability of the Pareto exponent's values are even more intriguing. As a first assumption we considered that this relative

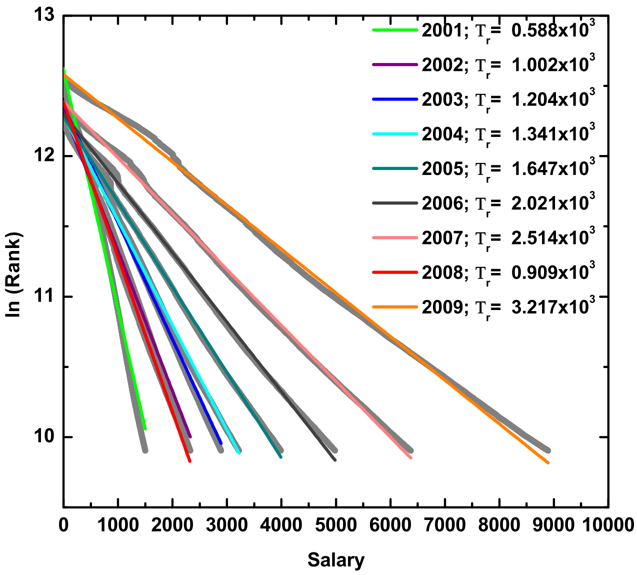

Figure 9: Rank-salary plot on log-normal scale for the $r > 10^4$ incomes (nominal in Euro) in the 2001-2009 time-interval. The lines indicate the exponential fits for the correspondent years. In the legend the values of the T_r income temperature are presented for each year.

stability in the high income region comes as a result of the employees' rank stability. Thus in Fig. 10 we visualized the time evolution of the top 100 ranks.

This plot immediately convinced us that the top part of the income distribution is a highly dynamical region, where constantly new employees appear and disappear. Next we assumed that the stability of the Paretos law and exponent value in such a rapidly changing top list is a consequence of fixed jobs with fixed salaries. This means that the highly payed positions are given and employees hired on these positions may come and go and their mobility would not affect the stability of the high income distribution, nor the value of the Pareto exponent. This conjecture was verified in Fig. 11. However our results suggest that employers are not stable and they constantly change in time.

In conclusion we can state that Pareto law and its corresponding Pareto exponents stability are not results of quasi-stable rank in time for the employees nor the jobs offered by the employers.

Theoretical models aiming to reproduce realistically the income distribu-

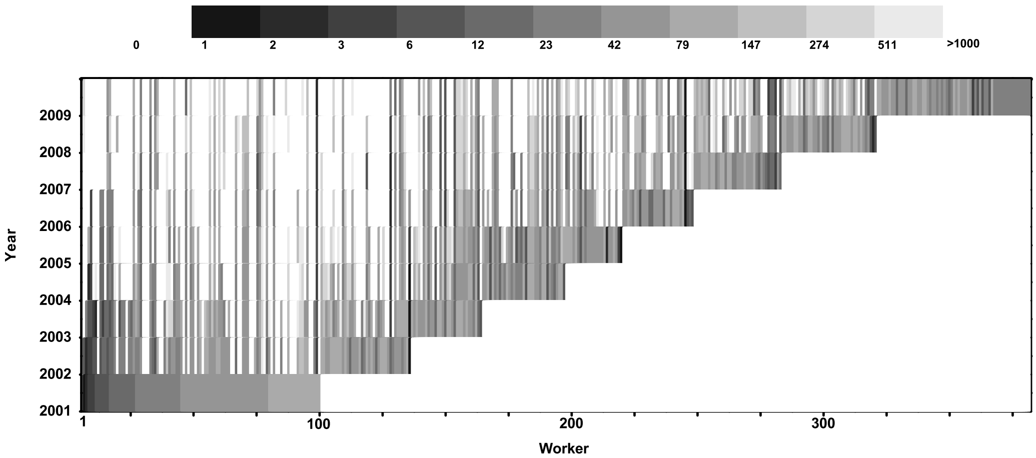

Figure 10: Rank evolution in time for the top 100 employees. The horizontal axis is the workers identification number, while the vertical axis represents the studied year. The gray scales highlight the ranks employees have over the years (see the upper legend). The white color indicates a year with no income for that certain worker or rank above 1000 [21].

tions in societies, assume a multiplicative diffusion in the high income region and additive diffusion in the medium or low income part [19]. Thus it is assumed that in the high income region the time variation of the income ΔW is proportional to the income itself ($\Delta W \propto W$) and this results the power-law function. Meanwhile, in the lower region ΔW is independent of the income, resulting in the low and medium income an exponential distribution [20]. In Fig. 12 we tested the $\Delta W \propto W$ dependence for the top 10^4 taxpayers.

In order to verify the growth rate's independency from the income, conjecture assumed in theoretical models, we computed the probability density for the $\log_{10}[W(t+dt)/W(t)]$ logarithmic growth rate for several years and for several income intervals (bins), based on the work of Fujiwara et al. [22]. Two separate and economically stable years (2002-2003) and (2004-2005) were analyzed in such manner. The results obtained for the different years and different income intervals are presented on Fig. 13(b).

The curves obtained for several years collapse in an approximative manner and have similar shape with the ones illustrated in [22, 23] for the positive logarithmic growth-rate interval. The probability densities obtained in our studies are however more extended in the negative region and the peak is shifted in the positive direction.

Our work offered a complete sampling of income distribution and its time

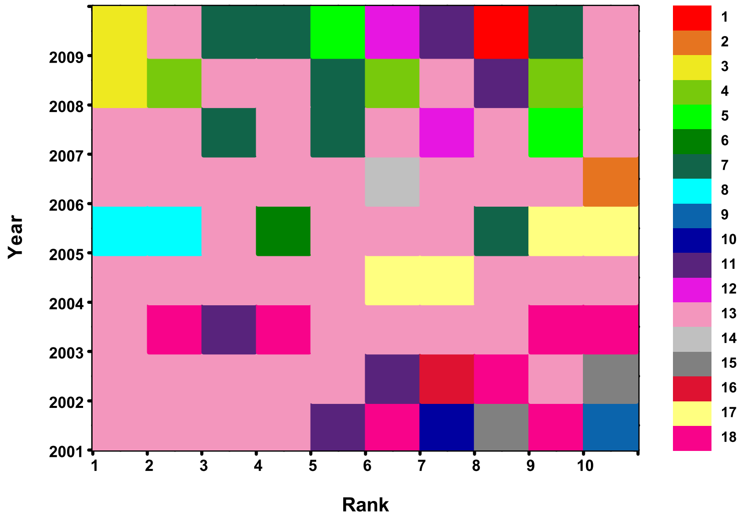

Figure 11: Employers of the top 10 paid jobs for each year. The figure presents for each year the employers of the paid highest 10 salaries. Totally there are 18 such employers, each of them being highlighted with different color and identification number for clarity.

evolution. Our results confirmed the highly dynamical nature of complex social systems and the fact that they are governed by universal laws that can be studied efficiently with statistical physics methods.

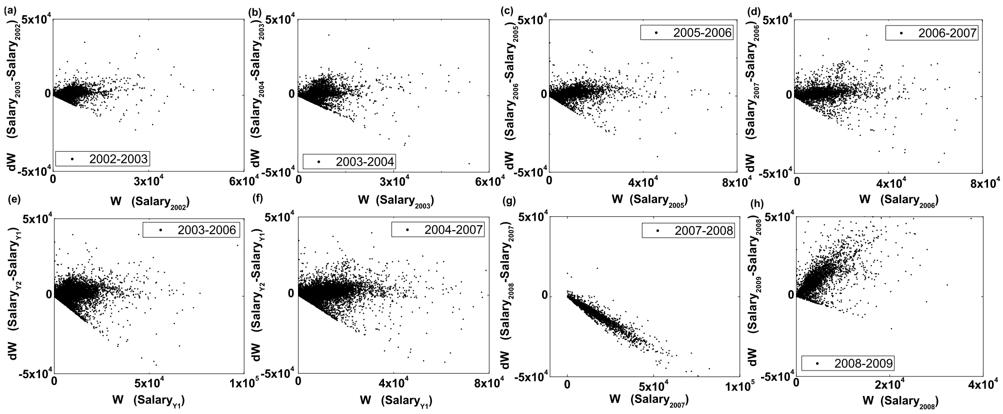

Figure 12: Time evolution of the income distribution. (a)-(d) The figures plot the salary differences between two consequent years as a function of the initial year. (e) and (f) represent data points for 3 years time periods constructed in the same manner. (g) shows a clear decreasing linearity, while the following year (h) an increasing trend can be observed [21].

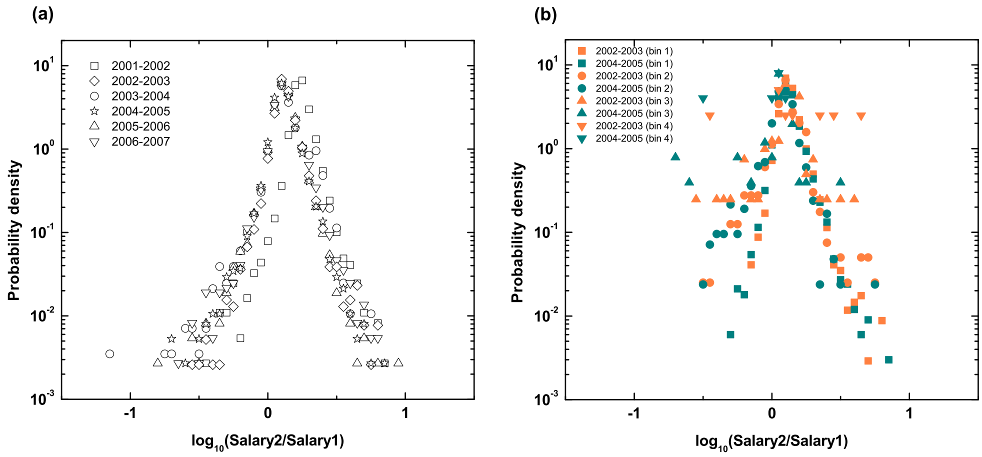

Figure 13: Probability density for the logarithmic growth-rate. (a) Growth rates for the top income region, in consecutive years, excluding the years affected by the economic crisis (2007-2008 and 2008-2009). (b) Binned data points for years 2002-2003 and 2004-2005. The Pareto region is split into four intervals of equal length (bins) in the logarithmic income. The probability density functions are calculated for each interval. Results obtained in such manner are depicted with black color for years 2002-2003 and with gray for 2004-2005. For clearer visualization each bin is represented with a different symbol.

Conclusions

In the present thesis two complex social systems were studied. The first a great sample of social network revealing professional relationships between European universities. The network exhibits exponential degree distribution characteristics, unlike social networks studied in the past. We have demonstrated that this feature comes a result of the fact that universities have exponential size-distributions and their sizes and degree are linearly proportional. In this study we also highlighted the hub universities of the Erasmus student mobility framework. For the directed and weighted network we have proved that the random model introduced by us reproduces realistically the characteristics of the system. Statistical physics methods helped enlighten the topology of this network and its features.

The second analyzed complex social system is of economic nature based on a social security database regarding employers and workers from Cluj district, Romania for 9 consecutive years (2001-2009). The study discusses two distinguish problems. The first problem discusses the topology, characteristics and time evolution of the complex system's underlying network. The results prove that the networks constructed for each year in particular exhibit scale-free characteristics results validating the assumption that social networks are governed by preferential linking.

Using the same database we exploited information available on workers' salaries. This study allowed a complete sampling of income distribution and its time evolution. We have confirmed once again the universal nature of Pareto's law for the top income distribution with a quite stable Pareto exponent over the years. We have revealed the highly dynamical nature of the top region and proved that the Pareto exponent's stability is not a consequence of a stable rank kept by the employees in the income top list, nor the result of the stability of the rank of top jobs. Our results also confirmed the exponential distribution for the low and medium incomes, in agreement with assumptions made by theoretical models. Also we have confirmed a basic conjecture introduced in modeling attempts, that multiplicative processes indeed describe properly the income distribution's structure in the high income region. Thus we have provided experimental evidence for the assumptions of proposed theoretical models. Wealth distribution analysis was also performed. The results did not evidence the presence of Pareto's law. Once again our results were obtained by applying statistical physics tools. We consider that our experimental results will help future modeling attempts for the quite general Pareto-type distribution of the top taxpayers. Also our results offer a better foundation for all models that aim to explain the shape and stability of the income distribution function.

Bibliography

[1] B. Bollobás, Random Graphs, Academic, London (1985)

[2] P. Erdős, A. Rényi, On random graphs, I., Publ. Math. Debrecen 6 (1959)

[3] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Reviews of Modern Physics 74, 47-97 (2002)

[4] R. Milo, S. Shen-Orr, S. Itzkovitz et al, Network motifs: simple building blocks of complex networks, Science 298, 824827 (2002)

[5] G. Palla, I. Derènyi, I. Farkas and T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature, London 435, 814 (2005)

[6] A. Barrat, M. Barthélemy, A. Vespignani, Dynamical processes on complex networks, Cambridge University Press (2008)

[7] A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, T. Vicsek, Evolution of the social network of scientific collaborations, Physica A 311, 90-614 (2002)

[8] P. Bajardi, C. Poletto, D. Balcan, H. Hu, B. Gonçalves, J. J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van den Broeck, V. Colizza, A. Vespignani, Modeling the critical care demand and antibiotics resources needed during the Fall 2009 wave of influenza A(H1N1) pandemic, Emerging Health Threats Journal 2 e11 (2009)

[9] A.-L. Barabási, Bursts. The hidden pattern behind everything we do. Dutton, New York (2010)

[10] A. Derzsi, **N. Derzsy**, E. Káptalan, Z. Néda, Topology of the Erasmus student mobility network, Physica A 390, 2601-2610 (2011)

[11] **N. Derzsy**, Employer mobility network revealed from social security data, Studia UBB Physica 1 (2012)

[12] V. M. Yakovenko and J. B. Rosser, Jr, Colloquium: Statistical mechanics of money, wealth, and income, Rev. Mod. Phys. 81, 1703-1725 (2009)

[13] G. Hegyi, Z. Neda, M. A. Santos, Wealth distribution and Pareto's law in the Hungarian medieval society, Physica A 380, 271-277 (2007)

[14] J. R. Iglesias, S. Gonçalves, S. Pianegonda, J.L. Vega and G. Abramson, Wealth redistribution in our small world, Physica A 327, 12-17 (2003)

[15] D. Garlaschelli, M. I. Loffredo, Wealth dynamics on complex networks, Physica A 338, 113-118 (2004)

[16] R. Coelho, Z. Nèda, J. J. Ramasco, M. A. Santos, A family-network model for wealth distribution in societies, Physica A 353, 515-528 (2005)

[17] V. Pareto, Cours d'economie politique Vol. 2, Macmillan, Paris (1897)

[18] M. E. J. Newman, Power laws, Pareto distributions and Zipf's law, Contemporary Physics, 46:5, 323-351 (2005)

[19] J.-P. Bouchaud and M. Mézard, Wealth condensation in a simple model e of economy, Physica A 282, 536-545 (2000)

[20] V. M. Yakovenko, Econophysics, statistical mechanics approach to, in Encyclopedia of Complexity and System Science, edited by R. A. Meyers,Springer, Berlin (2009)

[21] **N. Derzsy**, Z. Nèda, M. A. Santos, Income distribution patterns from a complete social security database, arXiv:1203-1880v1 [physics.soc-ph] (2012)

[22] Y. Fujiwara, W. Souma, H. Aoyama, T. Kaizoji, M. Aoki, Growth and fluctuations of personal income, Physica A 321, 598-604 (2003)

[23] H. Aoyama, W. Souma, Y. Fujiwara, Growth and fluctuations of personal and company's income, Physica A 324, 352-358 (2003)

List of publications

1. A. Derzsi, **N. Derzsy**, E. Káptalan, Z. Nèda, Topology of the Erasmus student mobility network, *Physica A* 390, 2601-2610 (2011)

2. **N. Derzsy**, Employer mobility network revealed from social security data, *Studia UBB Physica* 1 (2012)

3. **N. Derzsy**, Z. Nèda, M. A. Santos, Income distribution patterns from a complete social security database, arXiv:1203-1880v1 [physics.soc-ph], preprint submitted to *Physica A* (2012)

Acknowledgements

I owe my deepest gratitude to my PhD advisor, Prof. Dr. Zoltán Nèda for the excellent supervision during this work. His patience and continuous support made this thesis possible.

I am grateful to Prof. Dr. Maria Augusta Santos for the warm welcome and excellent working atmosphere at Centro de Física do Porto. Also I would like to thank for the supervision during my stay in Porto and the useful advices she had given.

I would like to thank the referees for their valuable time spent on this thesis.

Thanks should also go to my high school Physics teacher, Mrs. Márta Popa for her excellent classes and special guidance in this field. It is her merit that I became fond of Physics. Appreciation goes to my English language school teachers Emese Nagy and Kornélia Hegedűs, to whom I owe the advanced English knowledge.

I would like to express my gratitude to all my colleagues from university and work who supported, especially to Dr. Luciana Udrescu and Aranka Derzsi for their helpful advices during my PhD study.

I would also like to thank my boyfriend and friends for their continuous support and encouragement.

Besides the above mentioned, I would like to thank my family for their unconditioned support, encouragement and for being the best parents and grandparents a person can have.