UNIVERSITATEA POLITEHNICA TIMIŞOARA

Network Science in Computer Engineering and Information Technology

Habilitation Thesis

Mihai UDRESCU-MILOSAV

2017

Acknowledgements

This thesis stems from my personal research efforts in the fields of complex networks, computer engineering, and algorithm design. However, along the way, this exciting research journey was influenced and supported by many. First, I want to thank my family: my wife Lucreţia, my son Mihai Jr., and my parents Sorina and Mircea. Without their understanding and unconditioned support, none of the contributions presented in this thesis would have been possible. Then, I want to thank my mentors: my former PhD advisor prof. Mircea Vlăduţiu from Politehnica University of Timişoara, and prof. Radu Mărculescu of Carnegie Mellon University. Indeed, the best ideas presented in this thesis can be traced back to my research visit at CMU, hosted by professor Mărculescu's research group in 2011; it was a period of hard and enthusiastic work, in which I was also guided by my good friend Paul Bogdan, who is now a professor at the University of Southern California. When I returned home, many of the seminal ideas from CMU found fertile ground in Timişoara. I was lucky to have on my side all my friends and colleagues from the Advanced Computing Systems and Architectures (ACSA) lab at UPT (Lucian Prodan, Alexandru Topîrceanu, Alexandru Iovanovici, Cristian Ruican, Flavius Opriţoiu, Răzvan Avram, Gabriel Barina, Constantina Gavriliu, Alexandra Duma), as well as with my good friend and UPT postdoc Andrei Lihu. I developed some of the CMU ideas as applications in the field of precision medicine, by opening a fruitful collaboration with dr. Ştefan Dan Mihăicuţă and his colleagues from University of Medicine and Pharmacy Timişoara (UMFT), with the pharmacists from the Faculty of Pharmacy at UMFT (Lucreţia Udrescu, Laura Sbârcea, and Liana Suciu), as well as with prof. Ludovic Kurunczi from the Institute of Chemistry of the Romanian Academy. The field of network science has blessed me with the opportunity of finding new horizons for computing applications, and with the opportunity of discussing novel ideas with outstanding researchers in the fields of medicine and biochemistry, such as my friend Ioan Ovidiu Sîrbu; his dedication for helping people through research is an unfailing source of inspiration. Finally, I want to thank those who financed our research: Linde Healthcare and UEFISCDI.

Contents

Part I

Contributions

Chapter 1

Introduction

Since the beginning of the new Millennia, we have witnessed the emergence of the New Network Science, or the science of Complex Networks [4]; this new field is considered as a stand-alone science although it encompass elements from physics, mathematics, and computer science (see Figure 1.1). Specifically, the science of Complex Networks deals with the structure and behavior of complex systems that can be modelled as graphs, namely mathematical structures consisting of objects, nodes, or vertices, which are connected with lines, links, or edges. The fundamental difference between graphs from conventional graph theory and complex networks consists of the number of nodes (small for conventional graphs, up to several millions for complex networks) and the interconnection topology (regular, Euclidian lattice for conventional graphs, as opposed to complex and irregular for complex networks).

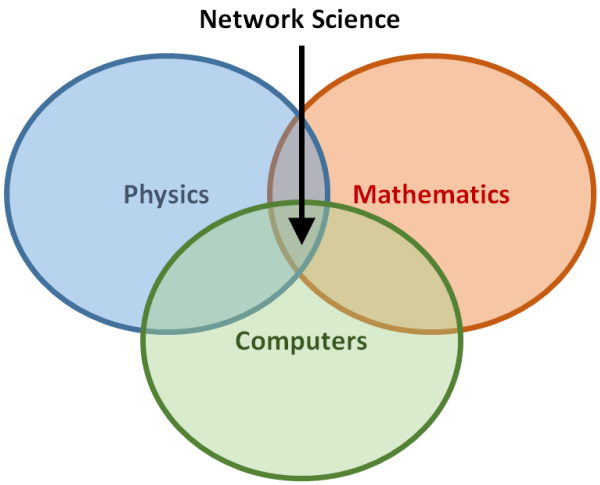

Figure 1.1: A general overview on Network science, highlighting the convergence of Physics (Complex Systems), Mathematics (Graph Theory), and Computer Science and Engineering (Algorithms and Databases).

Although the field of Complex Networks is implicitly multidisciplinary, it is mostly related to physics, specifically to statistical physics and complex systems. The practical applications of such tremendous theoretical developments are extensive: biology,

medicine, economy, social sciences, engineering (including electrical engineering and computers), computer science, and physics. As such, considering the field of application, the vast domain of Complex Networks can be divided in:

- **Biological networks:** disease networks, food networks in given environments, gene networks, pathway networks, metabolic networks, protein interaction networks, drug interaction networks, etc.
- **Technological networks:** computer networks, the world wide web, road and transportation networks, power distribution networks, electronic components network, computer software class networks, etc.
- **Social networks:** online social networks, political networks, economic networks, friendship networks, collaboration networks, etc.
- **Semantic networks:** LISP semantic network, natural language word networks, etc.

As a consequence, technological complex network techniques and methodologies can be used in Computer Engineering applications which entail a big amount of complexity [136]. At the same time, the overarching field of Information Technology includes various approaches where computer algorithms and applications are used for the advancement of biology, medicine, pharmacology, or social physics. Indeed, the last decade has witnessed significant progress in personalized or precision medicine, based on big data techniques and computer technologies such as Complex Network Analysis, Machine Learning (including Deep Learning) [44]. Moreover, the advance in social system physics has gain a lot of momentum since the global dissemination of Online Social Networks [5].

1.1 Motivation

Our main motivation in research is to find solutions for problems pertaining to computer engineering and information technology, by finding inspiration in physics. As such, our previous research, during the PhD program and the period 2004-2012 was focused on quantum computing, namely quantum algorithms, assessing and improving the reliability of quantum circuits.

Since 2012, after being exposed to the seminal ideas of the new science of complex networks, our research started to target the application of network science in computer engineering and information technology. As Figure 1.2 shows, the first application is in the field of Computer Engineering. Specifically, our goal is to find efficient techniques for design space exploration of multi-core, networks on chip (NoC) computer architectures. One of the most critical aspects of such NoC systems is represented by inter-core communication traffic which is mediated by the on-chip network communication infrastructure. The fact that the on-chip networks have regular topologies, as required by physical circuit implementation technology, are making the NoC systems prone to data traffic congestions. We propose fractal topologies for on-chip interconnection in NoC systems as they can be both regular (therefore, amenable to physical implementation)

and efficient in alleviating data traffic congestions. Indeed, we prove the efficiency of fractal NoC topologies by using the methods of complex network analysis.

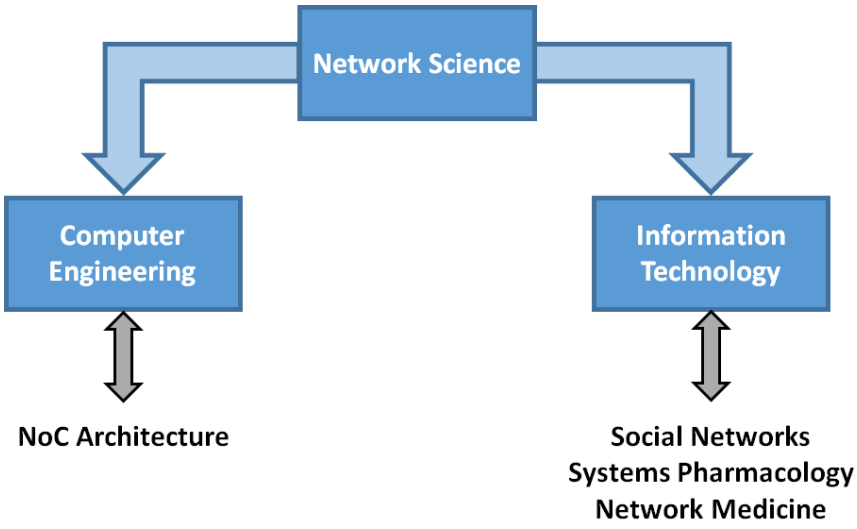

Figure 1.2: Applications of Complex Network Science in Computer Engineering (NoC Fractal Architectures) and Information Technology (Social Networks, Systems Pharmacology, and Network Medicine).

The second application of complex network science targets the field of social networks. As such, we analyze on of the most researched aspects in this field, namely opinion spread models [1][3]. We opted for using computer simulation as research methodology to evaluate and test these opinion spread models in social networks, because a mathematical analysis is not able to provide tools that cope with the huge complexity of social network topologies. For instance, in order to allow for mathematical analysis some social network researchers use simple, regular topologies such as lattice or fully connected networks [3], while others use random networks as this type of complex topologies can be analyzed by using statistics [30]. Therefore, by using the increased analytical power of computer simulation we show that our new model of opinion interaction and spreading in social network generates more realistic patterns. Our model is inspired from the research of Acemoglu *et al.* [1][3] which assumes the presence of so-called stubborn agents (i.e. social network agents which do not change their opinion) and that the degree of being influenced by other opinions is weighted by the *trust factor*, which is a fixed value assigned at the beginning of simulation for each social agent. We add to Acemoglu's model a new type of agents, namely null agents who do not have any opinion, as well as time variability in the trust factor which we call *tolerance*. The rules that govern the evolution of one agent's tolerance toward other agents' opinions are inspired from social psychology [185].

We also applied network science in information technology by aiming at providing computer-based solutions for pharmacological and medical problems. One such approach uses the international public database *drugbank.ca* [205] in order to build a drug-drug interaction network, where nodes represent drugs whereas links represent drug-drug

interaction relationships between the drugs. By employing 2D force-directed layout algorithms [147] and modularity class clustering [144], we automatically segregate drug communities which can be associated to specific drug properties [190]. In turn, drug property associations discovered with our methodology can be employed for predicting new interactions or for performing drug repositioning (repurposing) [171][123]. The methodological framework which uses a dual clustering technique, namely energy-model layouts and modularity classes, can also be used in other medical applications such as finding patient phenotypes in order to foster personalized treatment and precise diagnostics. To this end, we used the aforementioned dual technique in order to define precision clusters in a large database of Obstructive Sleep Apnea Syndrome (OSAS) patients [120]. By analyzing the accurately described patient phenotypes, we are able to introduce a new OSAS prediction score that achieves a drastically improved prediction specificity with respect to the state of the art [192].

1.2 Research Path

Our PhD research activity was related to domains such as Quantum Computing, Computer Reliability, and Evolvable Hardware; this activity was supported by the following research grants:

- Mihai Udrescu (director), Mircea Vladutiu, Oana Boncalo, Alexandru Amaricai, Virgil Petcu, Cristian Ruican, Nicolae Velciov. "Fault Tolerant Design of Quantum and Reversible Circuits (Proiectarea Circuitelor Cuantice si Reversibile Tolerante la Defectare", R&D Grant CNCSIS, Type A, 380/2007 (Total value: 159000 ROL)

- Mircea Vladutiu (director), Mihai Udrescu (member), Lucian Prodan, Oana Boncalo, Alexandru Amaricai, "Bioinspired Computer Architectures for Reversible and Quantum Logic Circuits (Arhitecturi Bioinspirate de Calcul pentru Circuite Logice Reversibile si Cuantice)", R&D Grant, PNII IDEI 17/2007 (Total value: 440000 ROL)

- Mircea Vladutiu (Director), Mihai Udrescu, Lucian Prodan, Oana Boncalo, Alexandru Amaricai, Versavia Ancusa, Nicolae Velciov, Alin Anton, "Bioinspired Design of Applications on Reconfigurable Platforms (Proiectarea Bioinspirata a Aplicatiilor pe Platforme Reconfigurabile)", R&D Grant, CNCSIS Type A 643/2005 (Total value: 67800 ROL)

The main PhD research activity results, as well as results from our post-PhD activity prior to the Complex Network period were reported in the following (selected) publications:

- Books

  B1 Mihai Udrescu, Lucian Prodan, *Emerging Computing Systems: Quantum Computing From a Computer Engineering Perspective*, in Colectia Calculatoare, Editura Politehnica, Timisoara, Romania, 2013, [128 pages], ISBN 978-606-554-684-4.

B2 Mihai Udrescu, *Quantum Circuits Engineering*, in Colectia Calculatoare, Editura Politehnica, Timisoara, Romania, 2009, [227 pages], ISBN 978-973-625-815-2

- Journal papers

J1 Mihai Udrescu, Lucian Prodan, Mircea Vlăduţiu, *Simulated Fault Injection Methodology for Gate-Level Quantum Circuit Reliability Assessment*, Simulation Modelling Practice and Theory, 23, 1, 60–70, 2012 [IF=1.159, Q2 – Computer Science, Software Engineering]

J2 Oana Boncalo, Alexandru Amăricăi, Mihai Udrescu, Mircea Vlăduţiu, *Quantum Circuit's Reliability Assessment with VHDL-Based Simulated Fault Injection*, Microelectronics Reliability, 50, 2, 304–311, 2010 [IF=1.137]

J3 Lucian Prodan, Mihai Udrescu, Oana Boncalo, Mircea Vlăduţiu, *Design for Dependability in Emerging Technologies*, ACM Journal of Emerging Technologies in Computing, 3, 2, Article 6, 2007 [IF=0.759]

- Conference papers

C1 Cristian Ruican, Mihai Udrescu, Lucian Prodan, Mircea Vlăduţiu, *Genetic Algorithm Based Quantum Circuit Synthesis with Adaptive Parameters Control*, IEEE Congress on Evolutionary Computation (CEC), pp. 896–903, Trondheim, Norway, May 18–21 2009 [WoS]

C2 Cristian Ruican, Mihai Udrescu, Lucian Prodan, Mircea Vlăduţiu, *Quantum Circuit Synthesis with Adaptive Parametres Control*, EuroGP2009. European Conference on Genetic Programming (Springer-Verlag Berlin Heidelberg, LNCS), LNCS 5481, pp. 339–350, Tubingen, Germany, Apr 15-17, 2009 [WoS]

C3 Mihai Udrescu, Lucian Prodan, Mircea Vlăduţiu, *Implementing Quantum Genetic Algorithms: A Solution Based on Grover's Algorithm*, 3rd ACM International Conference on Computing Frontiers (CF'06), pp. 71–82, Ischia, Italy, May 2–5, 2006

C4 Mihai Udrescu, Lucian Prodan, Mircea Vlăduţiu, *Simulated Fault Injection in Quantum Circuits with the Bubble Bit Technique*, in Proceedings ICANNGA, Springer Adaptive and Natural Computing Algorithms, pp. 276–279, Coimbra, Portugal, Mar 21–23, 2005 [WoS]

C5 Lucian Prodan, Mihai Udrescu, Mircea Vlăduţiu, *Multiple-Level Concatenated Coding in Embryonics: A Dependability Analysis*, Proceedings GECCO (ACM/SIGEVO), pp. 941–948, Washigton DC, USA, Jun 25–29, 2005

C6 Mihai Udrescu, Lucian Prodan, Vlăduţiu, *Improving Quantum Circuit Dependability with Reconfigurable Quantum Gate Arrays*, Proceedings 2nd ACM International Conference on Computing Frontiers (CF'05), pp. 133–144, Ischia, Italy, May 4-6, 2005

C7 Mihai Udrescu, Lucian Prodan, Vlăduţiu, *The Bubble Bit Technique as Improvement of HDL-Based Quantum Circuits Simulation*, Proceedings IEEE 38th Annual Simulation Symposium, pp. 217–224, San Diego CA, USA, Apr 2–8, 2005 [WoS]

C8 Mihai Udrescu, Lucian Prodan, Vlăduţiu, *Using HDLs for Describing Quantum Circuits: A Framework for Efficient Quantum Algorithm Simulation*, Proceedings 1st ACM Conference on Computing Frontiers, pp. 96–110, Ischia, Italy, Apr 14-16, 2004

Also, as a recognition of our activity in the fields of digital design, electronic design automation, quantum and reversible computation, computer reliability we were invited to join ICT COST Action IC1405 "Reversible computation - extending horizons of computing" as substitute Management Committee member. Moreover, we served as reviewer for several prestigious computer and software engineering journals:

Rev1 IEEE Transactions on Computers (IEEE TC), ISSN 0018-9340, during 2010-2014.

Rev2 IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (IEEE TCAD), ISSN 0278-0070, during 2010-2016.

Rev3 IEEE Transactions on Evolutionary Computation (IEEE TEVC), ISSN ISSN 1089-778X, during 2006.

Rev4 ACM Transactions on Embedded Computing Systems (TECS), ISSN 1539-9087, during 2012.

Rev5 Simulation Modelling Practice and Theory, Elsevier, ISSN 1569-190X, during 2012-2015.

Rev6 International Journal of Computer Mathematics, Taylor & Francis, ISSN 0020-7160, during 2012-2015.

Rev7 Microelectronics Journal, Elsevier, ISSN 0026-2692, during 2009-2014.

Rev8 ACM Journal on Emerging Technologies in Computing Systems (JETC), ISSN 1550-4832, during 2007.

Rev9 Mathematical Problems in Engineering, ISSN 1563-5147, during 2016.

We also served as Program Committee member for some prestigious international conferences, specialized in the fields that were targeted by our post-PhD research period, prior to our research visit at Carnegie Mellon University:

PC1 IEEE/ACM International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS), during 2013-2014.

PC2 IEEE Congress on Evolutionary Computation, during 2009.

PC3 IEEE International Conference on Computer and Information Technology, during 2007-2009.

PC4 IEEE International Symposium on Design & Diagnostics of Electronic Circuits & Systems, during 2008-2016.

However, as quantum computers are still a long way from practical implementation and the number of quantum algorithms is still limited, we aimed at finding new horizons at the frontier between physics and computation. As presented in Figure 1.3, in order to find new momentum for our research, in 2011 we made a research visit at the System Level Design (SLD) Group from the Department of Electrical and Computer Engineering, Carnegie Mellon University. During the visit and afterwards, we kept the same underlying research approach of analyzing and advancing scientific fields, and them apply the results in engineering applications. Our scientific and engineering contributions contributions in the field of Complex Networks pertain to: technological networks (NoC communication), social networks (opinion spread models), and biological networks (drug-drug interactions, patient disease networks).

In order to pursue of research goals, we opened durable collaborations with professor Radu Mărculescu (Carnegie Mellon University) and Paul Bogdan (University of Southern California); also, we assembled a multidisciplinary local team with people from University Politehnica of Timişoara, Department of Computer and Information Technology and "Victor Babeş" University of Medicine and Pharmacy from Timişoara (Department of Pulmonology and Faculty of Pharmacy).

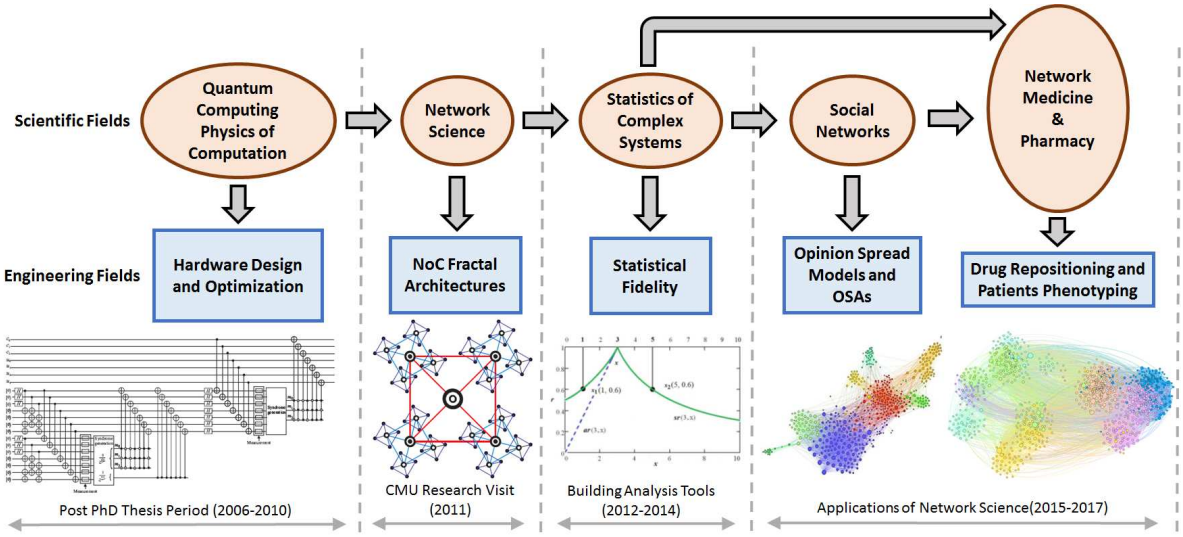

Figure 1.3: The overview of post-PhD research underpinning this thesis, where the upper panels represent scientific fields and the lower panels represent engineering applications. The approached topics are represented chronologically, from left to right.

1.3 Contributions

The results of our latest research period (2012 – present), which is related to Complex Networks applications in Computer Engineering and Information Technology, are published since 2013. Also, since 2014 our activity is supported by the research grants that

resulted as collaborations with "Victor Babeş" University of Medicine and Pharmacy Timişoara:

G1 Mihai Udrescu (director), Alexandru Topîrceanu, Alexandru Iovanovici, Andrei Lihu, Constantina Gavriliu, Ştefan Mihăicuţă (representative of partner organization), Rodica Dan. Daniela Resz. Carmen Ardelean, "Internet of Things meets Complex Networks for early prediction and management of Chronic Obstructive Pulmonary Disease", R&D Grant PNCDI III, P2, PED xxx/2017 (Total value: 514479 ROL)

G2 Ştefan Mihăicuţă (director), Mihai Udrescu (assistant manager), Alexandru Topirceanu, Alexandru Iovanovici, et al., Linde Healthcare RealFund R&D Grant: "Morpheus: A Screening and Monitoring System for Sleep Apnea Syndrome" 2014 (Total value: 75000 EUR)

The publication list linked to our research in Complex Networks is organized in sections: book chapters (BC), journal (J), and conference (C) papers.

- Book chapters

BC1 Alexandru Topîrceanu, Mihai Udrescu, and Mircea Vlăduţiu, *Genetically Optimized Realistic Social Network Topology Inspired by Facebook*, Online Social Media Analysis and Visualization, Lecture Notes in Social Networks, Springer, pp. 163–179, 2014.

- Journal papers

J1 Alexandru Topîrceanu and Mihai Udrescu, *Statistical fidelity: a tool to quantify the similarity between multi-variable entities with application in complex networks*, International Journal of Computer Mathematics 0(0):1–19, 2016. [IF=0.577]

J2 Lucreţia Udrescu, Laura Sbârcea, Alexandru Topîrceanu, Alexandru Iovanovici, Ludovic Kurunczi, Paul Bogdan, and Mihai Udrescu, *Clustering drug-drug interaction networks with energy model layouts: community analysis and drug repurposing*, Scientific Reports 6:32745, 2016. [corresponding and coordinating author, IF=5.228, Q1 – Multidisciplinary Sciences]

J3 Alexandru Topîrceanu, Alexandra Duma, and Mihai Udrescu, *Uncovering the fingerprint of online social networks using a network motif based approach*, Computer Communications 73(B):167–175, 2016. [IF=2.099, Q2 – Computer Science, Information Systems]

J4 Alexandru Topîrceanu, Mihai Udrescu, Mircea Vlăduţiu, and Radu Mărculescu, *Tolerance-based interaction: a new model targeting opinion formation and diffusion in social networks*, PeerJ Computer Science 72:e42, 2016. [corresponding author, IF=2.183, Q1 – Multidisciplinary Sciences]

J5 Liana Suciu, Carmen Cristescu, Alexandru Topîrceanu, Lucreţia Udrescu, Mihai Udrescu, Valentina Buda, and Mirela Cleopatra Tomescu, *Evaluation of patients diagnosed with essential arterial hypertension through network analysis*, Irish Journal of Medical Science (1971 -) 185(2):443–451, 2016. [IF=1.158]

- Conference papers (WoS)

C1 Alexandru Topîrceanu, and Mihai Udrescu, *FMNet: Physical Trait Patterns in the Fashion World*, Proc. 2nd European Network Intelligence Conference (ENIC), Karlskrona, Sweden, pp. 25–32, September 21–22, 2015. [Best paper award]

C2 Cristian Coşariu, Alexandru Iovanovici, Lucian Prodan, Mihai Udrescu, Mircea Vlăduţiu, *Bio-inspired redistribution of urban traffic flow using a social network approach*, Proc. IEEE Congress on Evolutionary Computation, Sendai, Japan, pp. 77–84, May 25–28, 2015.

C3 Alexandru Topîrceanu, Mihai Udrescu, *Measuring Realism of Social Network Models Using Network Motifs*, Proc. 10th Jubilee IEEE International Symposium on Applied Computational Intelligence and Informatics, Timişoara, Romania, pp.443–447, May 21–23, 2015.

C4 Mihai Udrescu, Alexandru Topîrceanu, *What drives the emergence of social networks?*, Proc. 20th International Conference on Control Systems and Computer Science (CSCS), Bucharest, Romania, pp. 999–999, May 27–29, 2015.

C5 Alexandru Topîrceanu, Dragos Tiselice, Mihai Udrescu, *The Figerprint of Educational Platforms in Social Media: A Topological Study Using Online Ego-Networks*, Proc. International Conference on Social Media in Academia: Research and Teaching (SMART), Timişoara, Romania, pp. 355–360, Sep 18–21, 2014.

C6 Alexandru Topîrceanu, Cezar Fleşeriu, Mihai Udrescu, *Gamified: An Effective Approach to Student Motivation Using Gamification*, Proc. International Conference on Social Media in Academia: Research and Teaching (SMART), Timişoara, Romania, pp. 41–44, Sep 18–21, 2014.

C7 Alexandru Topîrceanu, Gabriel Barina, Mihai Udrescu, *MuSeNet: Collaboration in the music artists industry*, Proc. 1st European Network Intelligence Conference (ENIC), Wroclaw, Poland, pp. 89–94, Sep 29–30, 2014.

C8 Gabriel Barina, Alexandru Topîrceanu, Mihai Udrescu, *MuSeNet: Natural Patterns in the Music Artists Industry*, Proc. 9th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI), Timişoara, Romania, pp. 317–322, May 15–17, 2014.

C9 Alexandru Topîrceanu, Mihai Udrescu, Mircea Vlăduţiu, *Network Fidelity: A Metric to Quantify the Similarity and Realism of Complex Networks*, Proc. 3rd IEEE International Conference on Cloud and Green Computing (CGC), Karlsruhe, Germany, pp. 289–296, Sep 30-Oct 02, 2013.

- Conference papers (other databases)

C10 Alexandru Topîrceanu, Mihai Udrescu, Răzvan Avram, Ştefan Mihăicuţă, *Data analysis for patients with sleep apnea syndrome: a complex network approach*, In Soft Computing Applications, Advances in Intelligent Systems and Computing 356, Springer, pp. 231–239, 2016. [SpringerLink]

C11 Alexandru Topîrceanu, Alexandru Iovanovici, Cristian Coşariu, Mihai Udrescu, Lucian Prodan, Mircea Vlăduţiu, *Social cities: redistribution of traffic flow in cities using a social network approach*, In Soft Computing Applications, Advances in Intelligent Systems and Computing 356, Springer, pp. 39–49, 2016. [SpringerLink]

C12 Alexandru Iovanovici, Alexandru Topîrceanu, Mihai Udrescu, Mircea Vlăduţiu, *Heuristic optimization of wireless sensor networks using social network analysis*, In Soft Computing Applications, Advances in Intelligent Systems and Computing 356, Springer, pp. 663–671, 2016. [SpringerLink]

C13 Alexandru Iovanovici, Alexandru Topîrceanu, Mihai Udrescu, Lucian Prodan, Ştefan Mihăicuţă, *A high-availability architecture for continuous monitoring of sleep disorders*, Studies in health technology and informatics Volume: 210, pp. 729–733, 2015. [Scopus]

C14 Alexandru Topîrceanu, Alexandru Iovanovici, Mihai Udrescu, Mircea Vlăduţiu, *Social cities: quality assessment of road infrastructures using a network motif approach*, Proc. 18th International Conference System Theory, Control and Computing (ICSTCC), Sinaia, Romania, pp. 803–808, Oct 17–19, 2014. [IEEE Xplore]

C15 Alexandru Iovanovici, Alexandru Topîrceanu, Mihai Udrescu, Mircea Vlăduţiu, *Design space exploration for optimizing wireless sensor networks using social network analysis*, 18th International Conference System Theory, Control and Computing (ICSTCC), Sinaia, Romania, pp. 815–820, Oct 17–19, 2014. [IEEE Xplore]

As recognition for our contributions to the field of complex networks, we received the Best Paper Award for our paper at the 2nd European Network Intelligence Conference, ENIC, Karlskrona, Sweden, 21-22 Sep, 2015: Alexandru Topîrceanu and Mihai Udrescu, "FMNet: Physical Trait Patterns in the Fashion World" [C1].

Moreover, as a result of multiple contributions, we were awarded research grant PN-III-P2-2.1-PED-2016-1145: Mihai Udrescu (Director), Alexandru Topirceanu, Alexandru Iovanovici, Andrei Lihu, Constantina Gavriliu, Stefan Mihaicuta, Rodica Dan. Daniela Resz. Carmen Ardelean, "Internet of Things meets Complex Networks for early prediction and management of Chronic Obstructive Pulmonary Disease" (value 514479 ROL).

1.4 Complex Network Science

This section introduces the basic theoretical elements and taxonomy that is used for complex network analysis. As the thorough review on this vast field is far beyond the scope of the present thesis, we indicate further references for a thorough analysis [16][75][143][200].

1.4.1 Basic Parameters

Definition 1 A *complex network* is a graph $G = (V, E)$ with V being the set of vertices (or nodes) $V = (v_i, w_i^v)$ and E the set of edges (or links) $E = (e_j, w_j^e)$ which connect some of the vertices in V, where v_i and e_j are individual vertices and edges respectively, along with their corresponding vertex and edge weights w_i^v and w_j^e, for $i = \overline{1, N}$ and $j = \overline{1, |E|}$ (N is the total number of vertices and $|E|$ is the total number of edges in the network).

Definition 2 An *unweighted complex network* is a complex network G where all vertex and edge weights are equal, therefore $w_i^v = w_j^e$ for all i and j. As such, we do not take into account weights, so that $G = (V, E)$, with $v_i \in V$, and $e_j \in E$.

Usually, N is also referred as the *network size*, while E (i.e. all edges in G) are referred as the *network topology*. In order to describe how the complex networks and their topologies are graphically represented, we introduce the following definition:

Definition 3 Given a complex network $G = (V, E)$ and an Euclidean d-dimensional space $\mathbb{R}^d$, a *layout* maps each vertex $v \in V$ to a position $x_v \in \mathbb{R}^d$ and assigns an Euclidean distance $\|x_v - x_w\|$ to each edge $[v, w] \in E$ (v and w are vertices in V).

The distance between two vertices v and w from complex network G is given by the edge number of the shortest path between v and $w \in G$ (See Figure 1.4 for an illustrating example). In order to analyze complex network topologies and behavior, we use the parameters that are described by the following definitions.

Definition 4 The *average path length* L is the typical distance, obtained by averaging the distances between all possible pairs of nodes $(v, w) \in G$. Also, the *diameter* of a network Dmt is the biggest distance between two vertices in graph G.

Definition 5 The *degree* of node v is the number of edges that are incident to this node, k_v. For network G, the average degree is $\langle k \rangle = \frac{1}{N} \sum_{v \in V} k_v$, namely the average value for the degrees k_v pertaining to all N vertices in graph G.

Definition 6 The *clustering coefficient* of a node is given by the fraction of existing edges between the $n(v) = k_v$ neighbors of node v in graph G, from the $\frac{1}{2}k_v(k_v - 1)$ possible edges: $C_v = \frac{2E[n(v)]}{n(v)[n(v)-1]}$ (see Figure 1.5). The clustering coefficient of the entire

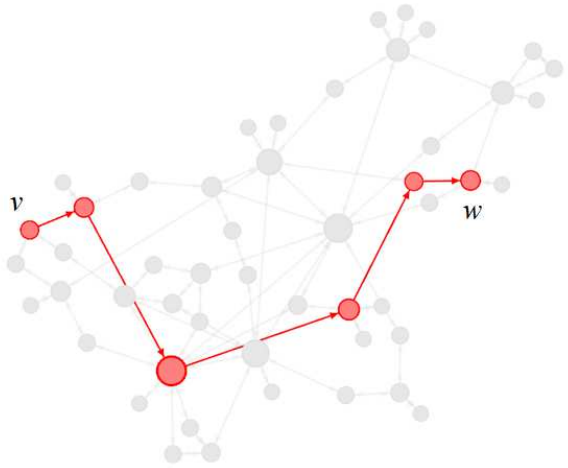

Figure 1.4: As highlighted in this figure's graph, the distance between nodes v and w is 5.

graph G is $C = \frac{1}{N}\sum_{v \in V} C_v$.

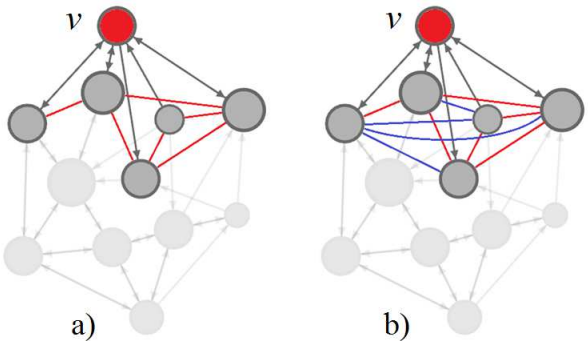

Figure 1.5: The clustering coefficient of a given node (highlighted in red) is the ratio between the $E\left[n(v)\right] = 6$ existing edges – highlighted with red in panel a) – which connect the $n(v) = 5$ neighbors of node v, and the number $\frac{1}{2}n(v)\left[n(v)-1\right] = \frac{1}{2}\cdot 5 \cdot 4 = 10$ of all possible edges – red and blue edges in panel b). As such, in this example, $C_v = \frac{6}{10} = \frac{3}{5}$.

Network clustering consists of classifying all the vertices $v \in V$ in one of the disjoint vertex subsets (i.e. clusters) C_i, pertaining to the set of disjoint subsets $C = \{C_1, C_2, \ldots, C_m\}$, so that $\cup_{i=1}^{m} C_i = V$ [88][147]. Several network parameters are used for network clustering in complex networks [66][78][79], but one of the most useful is the modularity which was advocated by Newman and Girvan [93].

Definition 7 In an unweighted network, the modularity of a clustering C is defined as $Mod_C = \sum_{C_i \in C}\left(\frac{|E_{C_i}|}{|E|} - \frac{\frac{1}{2}{k_{C_i}}^2}{\frac{1}{2}k^2}\right)$, where $|E_{C_i}|$ is the number of edges in cluster C_i, $|E|$ is the total number of edges in the network, k_{C_i} is the total degree of nodes in cluster

C_i, while k is the total degree of nodes in the entire network. As such, $\frac{|E_{C_i}|}{|E|}$ represents the fraction of intra-cluster edge density relative to the density of the entire network (which is assumed to be uniform), while $\frac{\frac{1}{2}{k_{C_i}}^2}{\frac{1}{2}k^2}$ is the expected such fraction. Therefore, modularity grows as clustering produces clusters with edge densities that are larger than expected.

Definition 8 The *network density Dst* is defined as the ratio of all edges in the network to the total number of possible edges $\frac{2|E|}{N(N-1)}$.

1.4.2 Network Centralities

Network centralities represent certain network parameters that can characterize the importance of vertices (nodes) in a complex network. Although such parameters represent individual nodes, and computing average centrality values seems to be the most appropriate way of characterizing the entire graph, it turns out that the *distribution* of these centrality values is paramount when classifying and characterizing complex networks.

One of the most important centrality distributions that are used by network scientists to analyze complex graphs is the *degree distribution*, namely the probability $P(k)$ that, in a given graph, the degree k has a certain value. In that respect, the degree distribution a a histogram which depicts how many nodes in the graph have degree k, with k ranging from 1 to the highest degree in the graph.

Another very important centrality metric which renders useful distributions is the *node betweenness*. The concept of betweenness, which characterizes any vertex's ability of being in-between, can be extended to edges (i.e. links) as well, thus obtaining the *link* or *edge betweenness*.

Definition 9 The *betweenness* $b(v)$ of vertex $v \in V$ from graph $G = (V, E)$ is defined as $\sum_{i,j \in V} \frac{\sigma_{ij}(v)}{\sigma_{ij}}$, where $i \neq j \neq v$, σ_{ij} is the total number of shortest paths from node i to node j, and $\sigma_{ij}\,(v)$ is the number of shortest paths from node i to node j which include node v in their sequence.

Other centralities are also used for the analysis of complex networks by representing their distributions: *closeness*, *eigenvector*, and *page rank*. Computing the inverse of the sum of shortest path lengths between the reference node and all other vertices gives the *node closeness*. The *eigenvector* centrality computes relative scores for all nodes by considering that the connections to high influence vertices are mode important than the connections to low-influence vertices; *page rank* is merely a variant of eigenvector centrality which is used by Google Search in order to rank websites [75][145].

In order to measure the similarity between two complex networks in terms of network parameters and centralities, the network fidelity metric φ was introduced [183]:

$$\varphi^j = \begin{cases} \frac{1}{n} \sum_i \frac{\overline{m_i}}{2\overline{m_i} - \overline{m_i^j}} & if\ \overline{m_i^j} < \overline{m_i} \\ \frac{1}{n} \sum_i \frac{\overline{m_i}}{\overline{m_i^j}} & if\ \overline{m_i^j} \geq \overline{m_i} \end{cases} \tag{1.1}$$

In equation 1.1, j represents the index of the network being compared to the reference network. The index of the network metric which describes the two compared models (e.g. average path length, average degree etc.) is denoted by $i = \{1, 2, ...n\}$, where n is the total number of common metrics taken into consideration. Fidelity takes values between 0 and 1 (or as percentiles), with 1 representing perfect similarity. The metric measurements on the reference model are m_i, respectively m_i^j on the model being compared.

1.4.3 Energy Layouts and Network Clustering

Energy model layouts are layout algorithms that can be represented as force systems. Many energy model layouts are developed as attraction-repulsion or a-r force systems [147]. In a-r layouts, adjacent vertices attract whereas all other pairs of vertices repulse, thus forming groups of vertices with dense connections (i.e. communities or clusters). The a-r forces are proportional to the attraction and repulsion powers (a and r respectively) of the Euclidean distances between the nodes: the attraction between adjacent vertices v and w is $\|x_v - x_w\|^a \cdot \overrightarrow{x_v x_w}$ and the repulsion between any two vertices $v, w \in V$ is $\|x_v - x_w\|^r \cdot \overrightarrow{x_v x_w}$ (with $\overrightarrow{x_v x_w}$ as the unit vector from v to w). Normally, $a, r \in \mathbb{R}$, are chosen so that $a \geq 0$ and $r \leq 0$, so that attraction is not decreasing and repulsion is not increasing with the Euclidean distance. The most popular force-based layout systems are the model of Fruchterman and Reingold (a=2, r=-1) [87], and the LinLog model (a=0, r=-1) [146].

For all a-r energy models, the resulted layout corresponds to the situation where local energy minimum is attained [147]. As such, total energy for the a-r layout ($a > r$) is:

$$T = \sum_{[v,w]} \left(\frac{\|x_v - x_w\|^a}{a+1} - \frac{\|x_v - x_w\|^r}{r+1} \right), \text{where } v \neq w. \tag{1.2}$$

Noack has demonstrated that, when $a > -1$ and $r > -1$, force-directed layout algorithms produce topological clusters which are equivalent with those rendered by modularity-based network clustering [147]. However, force-directed algorithms provide additional topological information about clusters, which leads to recommending the usage of both modularity clustering and a-r force directed layouts for more accurate network analysis [108].

1.4.4 Network Models

The main purpose of complex network theory is to provide models that characterize real-world systems in an accurate manner. To this end, the analysis of complex systems which can be modeled as graphs has revealed that real-world networks (technological, biological, social or semantic) are generally sparse (i.e. with low Dst), characterized by a low average path length, high clustering, and complex degree distribution such as power-law or Gaussian [200]. Also, many real-world complex networks also exhibit complex distributions of other centralities: betweenness, closeness or eigenvector [75].

Regular Networks

Generally in traditional graph theory, analysis can only be performed on regular graphs like meshes or Euclidian lattices such as the graph in Figure 1.6.

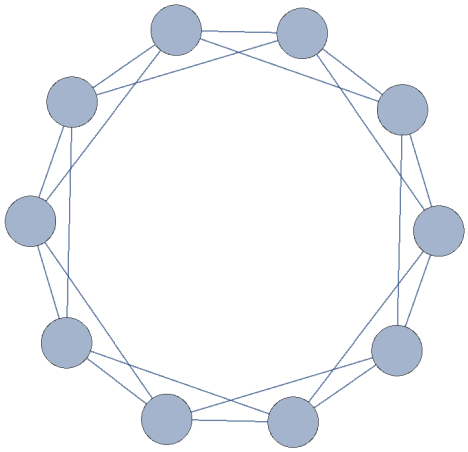

Figure 1.6: Regular graph example with $N = 10$ vertices.

However, it is obvious that real-world systems do not exhibit the regular structure of an Euclidean lattice. Indeed, unlike real-world networks have big average path lengths (because L grows linearly with N) and simple sequences for degree distributions. Generally, regular networks have high clustering coefficients C, similar to real-world networks [200]: in Figure 1.6, $C = \frac{3}{6} = \frac{1}{2}$.

Random Networks

The first step in describing and modeling complex graphs/networks was made at the end of the 1950s, by Hungarian mathematicians Paul Erdös and Alfréd Rényi who came out with the **Random Network** (or Erdös-Rényi) model. According to the Random Network (RN) model, when a network consists of $|V| = N$ vertices there are $\frac{N(N-1)}{2}$ potential edges between these vertices, with each edge having the same probability p of actually being placed (see Figure 1.7 for an illustrative example). Consequently, the average degree in RNs is $\langle k \rangle = pN$, $L \sim \frac{\ln N}{\langle k \rangle}$, and $C = p = \frac{\langle k \rangle}{N} \ll 1$; this means that in RMs we have low average path lengths (similar to real-world networks), but also low clustering coefficients (unlike real-world networks). As opposed to regular networks where degree distributions are simple sequences, RNs exhibit complex structure because they have Poisson degree distributions with the peak corresponding to $\langle k \rangle$.

The main contribution of the RN model is the so-called *network effect*. As such, Erdös and Rényi have proven the phase-transition from a disconnected to a connected random network when probability p becomes bigger than the threshold probability $p_c \sim \frac{\ln N}{N}$. From a fundamental standpoint, RNs are particularly significant, because they bridge the gap between graph theory and probability theory [200].

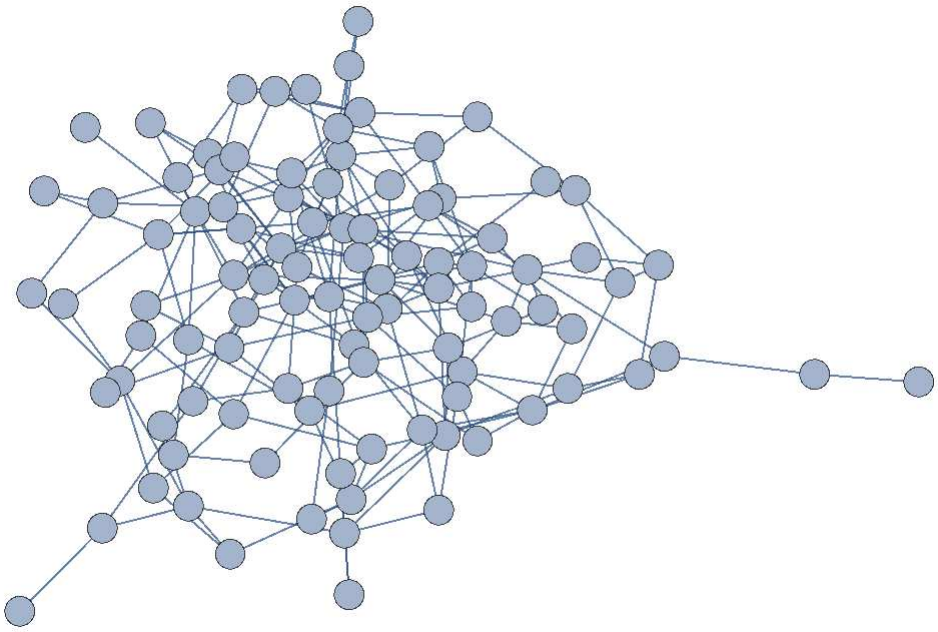

Figure 1.7: Random network example with $N = 100$ vertices.

Small-World Networks

Both regular and random network models fail to accurately describe real-world systems, therefore better models were sought. To this end, mathematicians Duncan Watts and Steven Strogatz noted that a more realistic model will lie somewhere between regular and random models, by taking the best of both worlds: a low L from random networks, and a high C from regular networks [199].

The objective of exploring the space between completely regular and completely random networks is attained by performing a simple algorithm. To this end, we consider a regular network such as the lattice from Figure 1.6 as starting point, then randomly pick with probability p an edge to be "rewired". To this end, one of the two edges connected by the rewired link will be randomly chosen with probability p to remain as connected to the new link; however, the new destination of the newly rewired link will also be picked randomly from the other nodes/vertices. The only constraint consists of the fact that there are no connections from a node to itself, nor there are double connections or links between two nodes. This simple step is run for all N nodes from the initial lattice. As illustrative example, Figure 1.8 presents the network obtained with the algorithm of Watts and Strogatz applied over the Figure 1.6 lattice, for a $p = 0.05$.

When increasing p from 0 to 1, it can be observed that after a relatively small p, the average path length L decreases significantly, whereas clustering coefficient C decreases only when p becomes significantly bigger. Therefore, for a certain range of p values, we can create a network which is somewhere in between regular and random networks, such that the newly created complex graph has low L and big C, similar to the values that real-world complex networks exhibit. In this particular region of low L and high C, the so-called **small-world** effect appears, namely the property of having a relatively

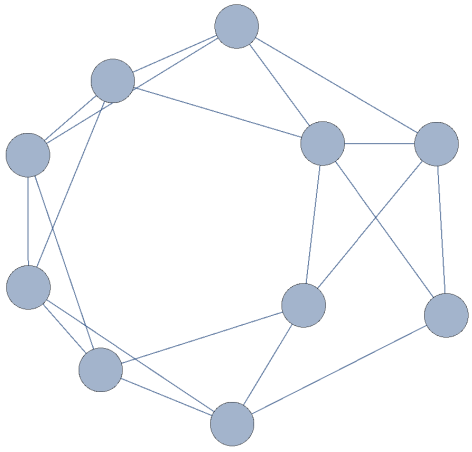

Figure 1.8: Small world network of $N = 10$ nodes, obtained with Watts and Strogatz algorithm, which starts with the lattice from Figure 1.6, with a rewiring probability of $p = 0.05$.

short distance between any two nodes, despite the fact that the network is significantly clustered.

Figure 1.9 presents this exploration methodology of finding the small-world (SW) effect zone, when the starting regular network is a lattice with $N = 1000$ nodes, and the representation is in log-normal coordinates for the evolution with p of the normalized clustering coefficient and average path length values ($\frac{C(p)}{C(0)}$ and $\frac{L(p)}{L(0)}$). $C(0)$ and $L(0)$ represent the clustering coefficient and average path length of the regular network, because these values correspond to probability of rewiring $p = 0$; when $p = 1$, the network becomes fully random, therefore $C(p)$ and $L(p)$ pertain to a random network.

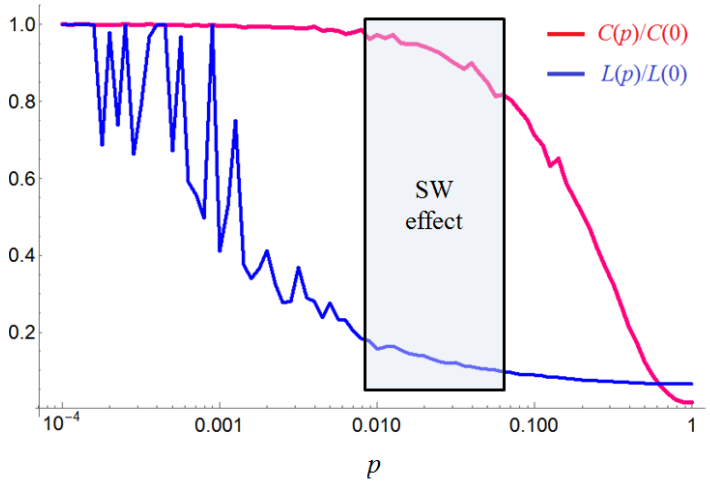

Figure 1.9: Representation in log-normal coordinates for the evolution of normalized clustering coefficient and average path length values $\frac{C(p)}{C(0)}$ and $\frac{L(p)}{L(0)}$, when p grows from 0 to 1. The figure also highlights the window where the SW effect is generated, because L decreases significantly and C remains high.

However, in most real-life networks degree distribution $P(k)$ abides by a power law, and it is not normally distributed (i.e. does not have a Poisson distribution). But in small world networks built with Watts and Strogatz algorithm, the degree has a normal distribution (see Figure 1.10); this situation is explained by the fact that each rewiring used for the small world network algorithm is made with the same probability p.

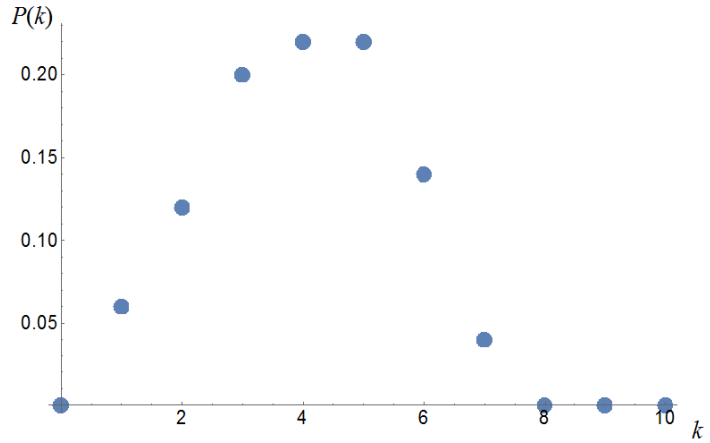

Figure 1.10: Normal degree distribution in a small-world network with 100 nodes.

Scale-Free Networks

The latest fundamental network model in complex network science is the scale-free network, introduced by Albert-László Barabási and Réka Albert in 1999 [18]. The main feature of this model is that it is dynamic, therefore the network grows with each step of the algorithm. A brief description of this network is given as follows:

1. Start with an arbitrary network with M nodes; this M-node network can be small world or random.

2. A number of m new nodes are added. The probability of linking each of the m new nodes to node v from the M existing nodes is $p_v = \sum_{i=1}^{M} \frac{k_v}{k_i}$ (k_v and k_i are the degrees of nodes v and i).

3. The new M is updated as $M + m$. If the new $M \neq M_{max}$ then go to the previous step, else exit.

An example scale-free network, generated with Barabási-Albert algorithm for $M_{max} = 1000$ and $m = 2$ is presented in Figure 1.11. In such a scale-free network, the degree distribution is power-law $P(k) = k^{-\gamma}$, as presented in Figure 1.12.

The main ingredient which produces the power law distribution of degree in scale free networks is the preferential attachment mechanism, which assures that the nodes with already high degree in the network will have an even higher degree, because the probability of attachment $p_v = \sum_{i=1}^{M} \frac{k_v}{k_i}$ and the degree k_v are obviously directly proportional. Oftentimes, it is said that in scale free networks the degree is distributed according to the "rich gets richer principle".

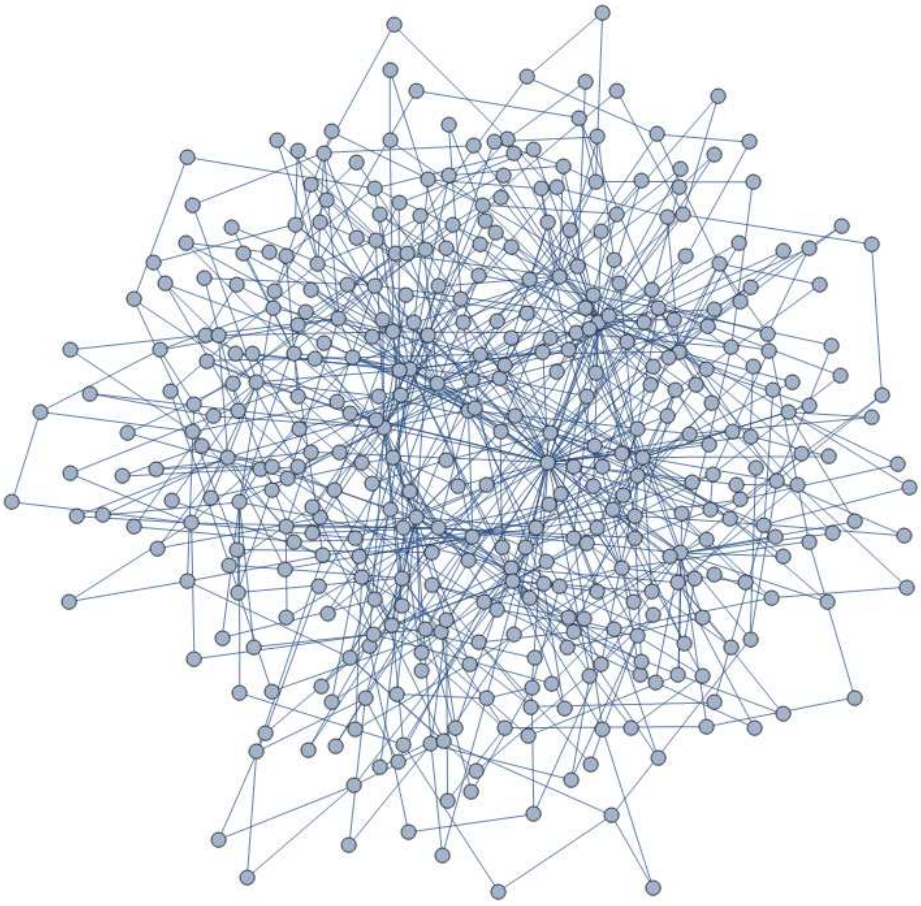

Figure 1.11: Scale-free network created with Barabasi-Albert algorithm, with $N = 1000$ nodes when 2 nodes are added at each step.

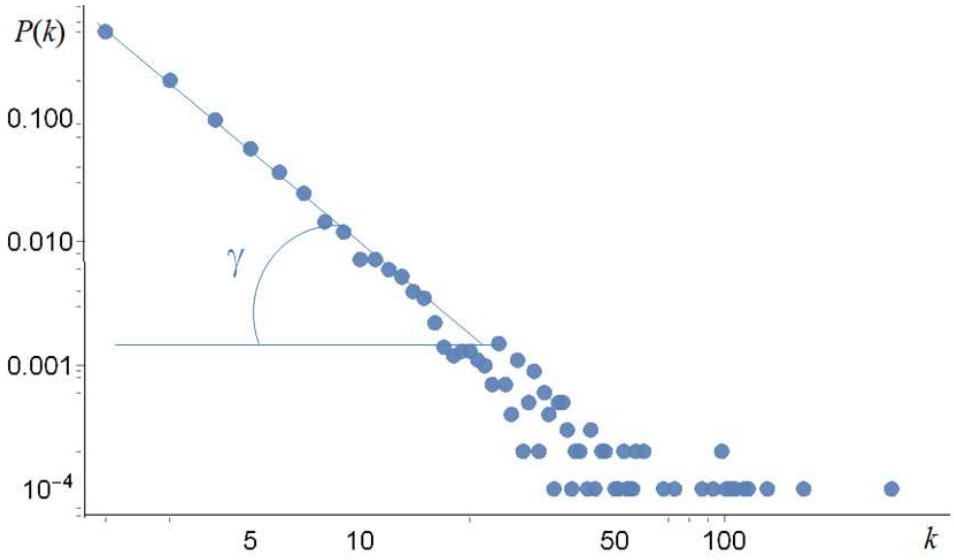

Figure 1.12: Log-log representation of the power-law degree distribution in a $N = 10000$ nodes scale free network, when 2 nodes are added at each step of the Barabasi-Albert algorithm. Up to the cutting point indicated by the vertex of the added angle, the distribution abides by the power-law $P(k) = k^{-\gamma}$, where γ is the slope, namely the angle suggested in the figure.

In conclusion, scale free networks are realistic in terms of degree distribution and low average path length L. However, scale-free networks tend not to be clustered enough to resemble real-life networks. As a consequence, a lot of research has been done recently in order to combine the fundamental network models, by adding randomness to scale free networks or by adding the small world property to scale free models [48],[109]. Also, one of the weaknesses of the scale free network model is that, even if it is very robust to random attacks, it is very vulnerable to targeted attacks. Specifically, if nodes are randomly removed from a scale free network, the network remains connected (i.e. with one component) even if many nodes are removed. Conversely, if the process of removal specifically targets high-degree nodes, the network rapidly becomes disconnected after a small number of removed nodes.

1.5 Thesis Outline

The rest of the thesis is organized as follows. The second chapter presents the application of algorithmic modelling and computer simulation of social network behavior under the form of opinion formation and dissemination. The third chapter introduces two important applications of complex network algorithmic methods in precision medicine: drug repurposing and patient phenotype definition. Next, in the second part of the thesis, we present our future research path. Thus, the fourth chapter presents the prospective application of network science in computer engineering, namely building efficient communication infrastructure for Network on Chip (NoC) multiprocessor systems; indeed, we consider the optimization of NoC communication as a very important emerging re-

search topic that our lab is going to pursue. The last chapter draws the conclusive remarks and sketches future research opportunities for computer engineering in the context of complex systems and big data. As such, we introduce our new research project entitled "Internet of Things Meets Complex Networks for Early Prediction and Management of Chronic Obstructive Pulmonary Disease", or Internet of thiNgs Complex nEtworks PredicTION (INCEPTION). The last part consists of listing all references that define our research, as well as other relevant related works, from state-of-the-art literature.

Chapter 2

Social Network Analysis

2.1 Introduction

In the field of social networks, one of the most important research topics is to model, replicate and analyze how opinion spreads, fluctuates and percolates [94][91][195]. The opinion dynamics in social networks uncovers the positions of socially important agents who have the greatest influence [139]. This aspect is crucial when attempting to model realistic social networks; indeed, without influencing agents who act as drivers, any society would have a rather erratic, unpredictable evolution.

The state of the art in social network opinion formation and evolution models are mostly based on the ideas of social contagion and opinion spread [12][3][207][195][98] [160]; nonetheless, these works are rather limited in terms of accuracy towards real-life situations. The lack of realism in the available models can be explained by the fact that the underpinning opinion interaction models are mostly based on fixed thresholds [61][107]. In fairness, some opinion spread models do not have fixed thresholds, but the evolution of the threshold values is made in accordance with some simple external-state probabilistic processes [80][63].

In order to address the drawbacks entailed by hitherto models, we targeted the missing dynamical traits of opinion spreading, by introducing a new mathematical model. We rely on the fact that most real-world social network observations can be explained using the concept of tolerance, therefore our new social interaction model takes into consideration the individual's internal tolerance state [185], namely the individual's capacity of accepting and adopting other social agents' opinions. We compared our model against big-size, real-life empirical data from Yelp, Twitter and MemeTracker, and then validated it with our social network opinion spread simulator SocialSim [1].

Social networks are particular instances of complex networks, which can be used to describe collective social behavior and emerging complex social phenomena. In a social network $G = (V, E)$, vertices $v \in V$ represent social agents/individuals, and edges $e \in E$ represent social links such as friendship, collaboration, or interaction. For $G = \{V, E\}$, the direct neighborhood of agent $i \in V$ is represented by vertices directly connected to i: $N_i = \{ j \mid (i, j) \in E\}$. According to the models proposed by [3][207], within the set

[1] https://sites.google.com/site/alexandrutopirceanu/projects/socialsim

of all social agents there are two disjoint sets of so-called *stubborn agents* V_0, $V_1 \in V$ who never change their opinion (which can be either positive or negative – 1 or 0); these stubborn agents can be interpreted as social influencers or drivers of opinion. The social agents that are not stubborn (i.e. regular agents) $V \setminus \{V_0 \cup V_1\}$ update their opinion, according to the underlying model, by using the opinion of their direct neighbors (which may include stubborn agents).

We denote the opinion of agent i at time t with $o_i(t)$. The initial state of all social agents, namely their opinions at time $t = 0$ expressed as $o_i(0) \in [0, 1]$, can be allocated by random distribution. The updating of opinion values for regular agents is randomly triggered and can be performed by adopting the opinion of a randomly chosen neighbor, or by averaging the opinions of all direct neighbors.

If we assume an opinion spread model with continuous/analog opinion representation, the state of agent i which holds continuous opinion $o_i(t)$ at moment t is given by $s_i(t)$. If we assume discrete opinion representation, then $o_i(t) = s_i(t)$. Also, if we also assume that we have null/undecided agents in the continuous opinion representation, then $s_i(t)$ has the expression from equation 2.1.

$$s_i(t) = \begin{cases} 0 & if \ \ 0 \leq o_i(t) < 0.5 \\ NONE & if \ \ o_i(t) = 0.5 \\ 1 & if \ \ 0.5 < o_i(t) \leq 1 \end{cases} \tag{2.1}$$

In the next time step $t+1$, a regular social agent a updates its current opinion $o_a(t)$, when interacting with a randomly picked direct neighbor n that has opinion $o_n(t)$. In $G = (V, E)$, agents a and n are direct neighboring nodes if there is a direct link between them. We also assume that there are agents that do not hold a clear opinion (so-called NULL agents); interacting with NULL agents will have no effect on opinion updates. Conversely, if a regular agent a interacts with a regular or stubborn agent n that holds an opinion, the new opinion of agent a at the next time step $t + 1$ is:

$$o_a(t + 1) = \theta_a o_n(t) + (1 - \theta_a)\, o_a(t). \tag{2.2}$$

In equation 2.2, for regular agent a, we consider the degree of accepting other people's opinions or – in other words – the tolerance towards other agents' opinions θ_a as being fixed. Nonetheless, we argue in [185] that θ_a evolves according to individual's traits and experiences. As such, an agent that interacts with a high diversity of opinions becomes more tolerant. Conversely, an agent that faces a low opinion diversity tends to becomes intolerant. Nonetheless, the complex process of evolution towards both tolerance or intolerance is nonlinear. Therefore, the tolerance model we propose in [185] and [191] employs a non-linear tolerance evolution function, as opposed by models from [102] and [203]. To this end, we consider socio-psychological individual traits to be used in our dynamical opinion interaction model, so that tolerance for any regular agent a evolves according to:

$$\theta_a(t) = \begin{cases} max\ (\theta_a(t-1) - \alpha_0 \varepsilon_0,\, 0) & if \ \ s_a(t-1) = s_n(t) \\ min\ (\theta_a(t-1) + \alpha_1 \varepsilon_1,\, 1) & otherwise \end{cases} \tag{2.3}$$

According to equation 2.3, $\theta_a(t)$ evolves according to $\alpha_0\varepsilon_0$ if $s_a(t-1)$ is equal with the state of the neighbor $s_n(t)$ which interacts with a. If $s_a(t) \neq s_n(t)$, then $\theta_a(t)$ is rectified with the non-linear quantity $\alpha_1\varepsilon_1$. Scaling factors α_0 and α_1 values are initially set as 1 and then evolve as follows:

$$\alpha_0 = \begin{cases} \alpha_0 + 1 & if \ \ s_a(t-1) = s_a(t) \\ 1 & otherwise \end{cases} \tag{2.4}$$

$$\alpha_1 = \begin{cases} 1 & if \ \ s_a(t-1) = s_a(t) \\ \alpha_1 + 1 & otherwise \end{cases} \tag{2.5}$$

Scaling parameters α_0 and α_1 are represent bias, so that when an opinion interaction happens (whether the interaction type is one that confirms the previous agent opinion or opposes it), the α parameter that corresponds to the other type of event is reset. As such, the α parameters have the role of increasing the non-linear magnitude of tolerance modification ratios ε_0 (weight of modification towards intolerance) and ε_1 (weight of modification towards tolerance). We choose the fixed values of $\varepsilon_0 = 0.002$ and $\varepsilon_1 = 0.01$ after rigorous and extensive SocialSim simulations [185]. As a result, intolerance grows with ε_0 and decreases with ε_1.

By simulating the social network opinion evolution with SocialSim [185], according to the tolerance-based opinion interaction network, we confirm that opinion disagreement is constant and never ceases, even when we use a complex network topology that drastically degrades tolerance (i.e. the scale-free model, see Figure 2.1).

2.2 Probabilistic modeling of opinion spread models

We provided a probabilistic model for our tolerance-based model in [191]. To this end, we make the assumption that, for any social agent, we have: one completely tolerant state, one completely intolerant state, and at least one intermediary state (i.e. only partially intolerant). Transition occurs between these tolerance states of the social agent, at each moment t of social interaction; however, the interaction with NULL agents will not result in tolerance state transitions. Actually, all tolerance states are interpreted as corresponding to an ordered hierarchy of tolerance levels, from 100% tolerance to 0% tolerance. Tolerance state transitions are made such that any social interaction with an agent having a different opinion determines a transition to the next more tolerant state. Conversely, the interaction with an agent that holds the same opinion generates a transition to the next less tolerant state.

Nonetheless, Markov chain modeling is intractable for a sufficient large number of intermediary states [168]. As such, our study only comprises tolerance modeling with 1 and 2 intermediary states (3 and 4 states). Our Markov chain analysis characterizes all agents in a given social network with two parameters: λ as the rate of encountering the same opinion, and μ as the rate of encountering a different opinion via opinion interaction, with $\lambda + \mu \leq 1$; when there are no NULL agents in the social network, we have $\lambda + \mu = 1$. Therefore, if we consider ρ as the rate of interaction with NULL agents, then $\lambda + \mu + \rho = 1$. We also assume that λ and μ are distributed in an exponential

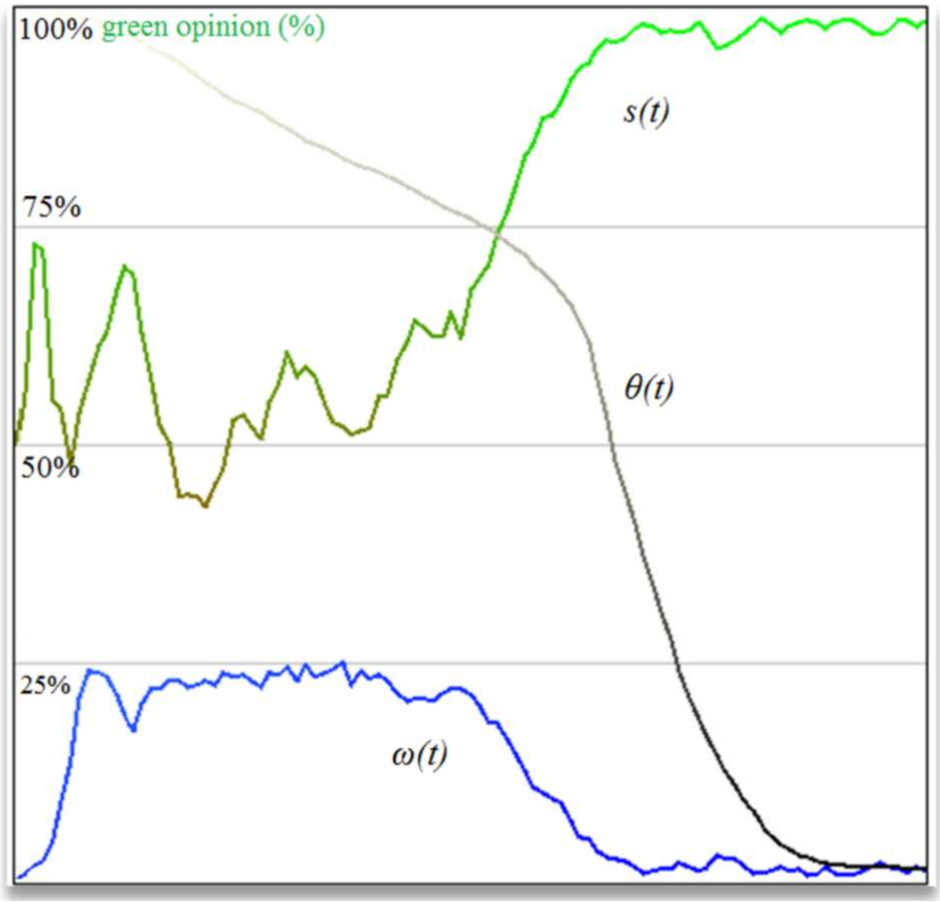

Figure 2.1: Simulation of opinion dynamics with SocialSim [185], in a scale-free social network with 100,000 agents (32 stubborn agents holding opinion '0', 32 stubborn agents holding opinion '1', and no null agents) assuming the tolerance model of opinion interaction. opinion state $s(t)$='1' is represented with green, opinion $s(t)$='0' is represented with red, and intermediary values with intermediary shades between green and red. Because the underlying topology of the social network is scale-free, leading to a high influence from the high degree agents and, consequently, the tolerance θ drastically degrades. Even if the social network is inherently intolerant, there is constant disagreement between social agents, as opinion never ceases to change (as indicated by the evolution of $\omega(t)$).

fashion, so that the probability of having the social agent interacting with the same opinion will be $1 - e^{-\lambda t}$; consequently, the probability of having an interaction with a different opinion will be $1 - e^{-\mu t}$.

2.2.1 3-state tolerance model

If we consider three tolerance states for each social agent, then we have S_0 as the tolerant state, S_1 as the undecide or itermediate state, and S_2 as the intolerant state. In Figure 2.2 we present transitions in the Markov diagram that describe tolerance state evolution in social agents.

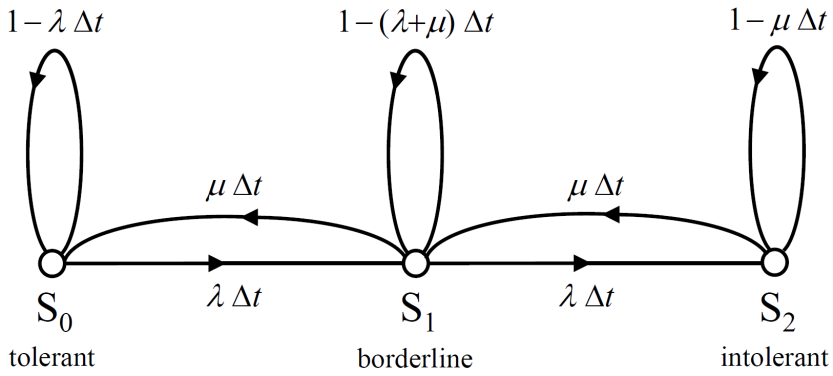

Figure 2.2: Markov diagram corresponding to the 3-state tolerance model, where S_0 is the tolerance state, S_1 is the undecided or intermediate state, and S_2 is the intolerance state.

From Figure 2.2 we derive the state probability expressions at $t + \Delta t$, assuming that we know the current state at t:

$$\begin{cases} P_{S_0}(t+\Delta t) &= (1-\lambda\Delta t)\, P_{S_0}(t) + \mu\Delta t P_{S_1}(t) \\ P_{S_1}(t+\Delta t) &= \lambda\Delta t P_{S_0}(t) + \mu\Delta t P_{S_2}(t) + \\ & \quad + [1-(\mu+\lambda)\,\Delta t]\, P_{S_1}(t) \\ P_{S_2}(t+\Delta t) &= \lambda\Delta t P_{S_1}(t) + (1-\mu\Delta t)\, P_{S_2}(t) \end{cases} \tag{2.6}$$

$P_{S_i}(t)$ and $P_{S_i}(t+\Delta t)$, for $i \in 0, 1, 2$, represent the probabilities of a social agent being in tolerance state S_i at times t and $t + \Delta t$; initially, at moment $t = 0$, $P_{S_0}(0) = 1$ and $P_{S_1}(0) = P_{S_2}(0) = 0$.

By applying Laplace transformation, so that variable t is substituted by s, we obtain state expressions:

$$P_{S_0}(s) = \frac{s^2 + (2\mu+\lambda)\, s + \mu^2}{s^3 + 2\,(\mu+\lambda)\, s^2 + (\mu^2+\mu\lambda+\lambda^2)\, s} \tag{2.7}$$

and

$$P_{S_1}(s) = \frac{\lambda\,(s+\mu)}{s^3 + 2\,(\mu+\lambda)\, s^2 + (\mu^2+\mu\lambda+\lambda^2)\, s} \tag{2.8}$$

Therefore, the probability of not getting to the intolerance state is given by:

$$\begin{aligned} P_{tol}(s) &= P_{S_0}(s) + P_{S_1}(s) = \\ &= \frac{s^2+2(\mu+\lambda)s+\mu^2+\mu\lambda}{s^3+2(\mu+\lambda)s^2+(\mu^2+\mu\lambda+\lambda^2)s} \end{aligned} \tag{2.9}$$

From 2.9, we get the probability of tolerance state at infinity, which can be interpreted as the expected stable tolerance state of the social agent:

$$\lim_{t\to\infty} P_{tol}(t) = \lim_{s\to 0} sP_{tol}(s) = \frac{\mu^2+\mu\lambda}{\mu^2+\mu\lambda+\lambda^2} \tag{2.10}$$

If we do not have NULL agents, or if their number is small enough, then $\mu + \lambda \simeq 1$ and consequently:

$$\lim_{t\to\infty} P_{tol}(t) = \frac{\mu^2+\mu\lambda}{\mu^2+\mu\lambda+\lambda^2} = \frac{\mu\,(\mu+\lambda)}{(\mu+\lambda)^2-\mu\lambda} \simeq \frac{\mu}{1-\mu\lambda} \tag{2.11}$$

The probability of tolerance for $t \to \infty$ (interpreted as corresponding a mature, stable society) can be represented as a function of λ (i.e. the rate of a social agent interacting with another agent with the same opinion). For a convenient graphical representation, ρ is fixed, so that the expression from equation 2.10 becomes function of λ: $P_{tol-3}(\lambda)$ as presented in Figure 2.3).

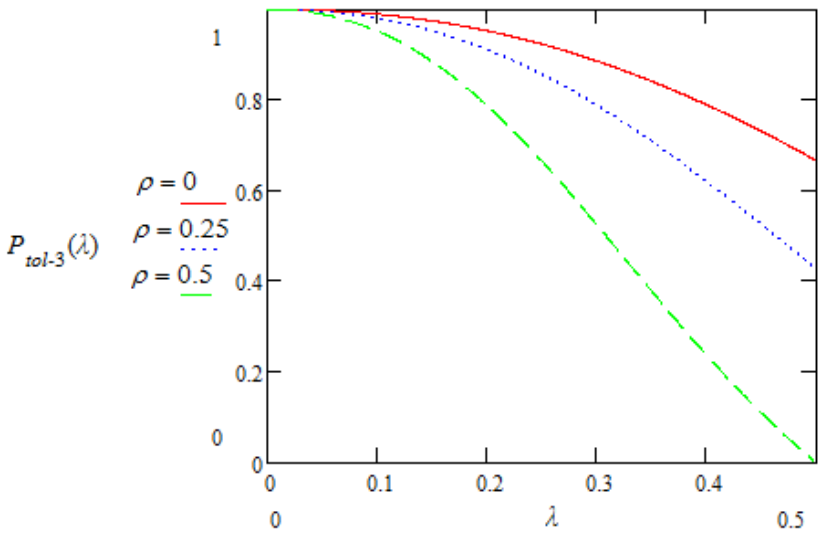

Figure 2.3: Representation for the probability of tolerance $P_{tol-3}(\lambda)$, when the 3-state model is assumed, for 3 illustrating values of ρ (0, 0.25, and 0.5).

2.2.2 Tolerance model with 4 states

If we assume a model with two intermediate tolerance levels (meaning a total of four tolerance levels), then we have the following states: S_0 as the tolerant state, S_1 as the intermediate mostly tolerant state, S_2 as the intermediate mostly intolerant state, and

S_3 as the intolerant state. A transition from state to state only happens when the social agent interacts with another social agent that holds an opinion; conversely, any interaction with a NULL agent results in no transition (see Figure 2.4 for all possible transitions in the 4-states probabilistic tolerance model).

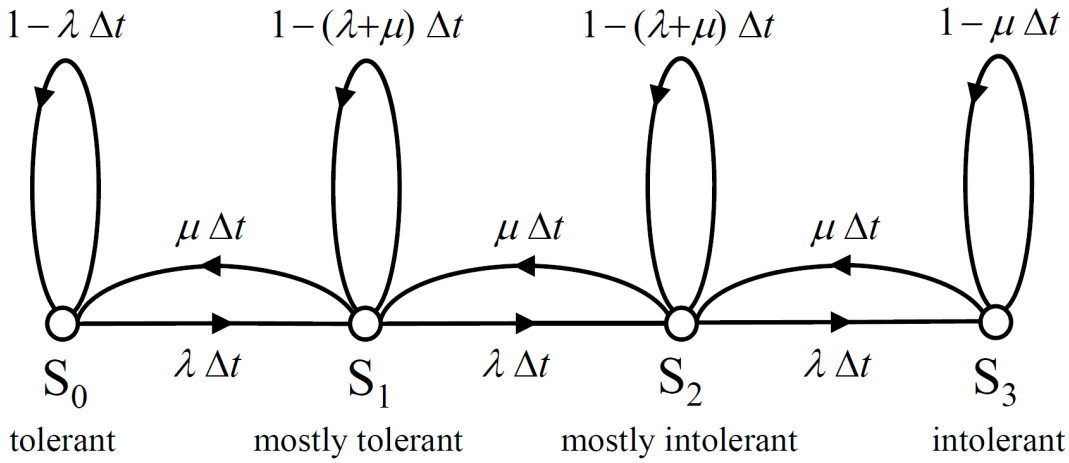

Figure 2.4: Markov diagram corresponding to the 4-state probabilistic tolerance model, where S_0 stands for tolerance, S_3 for intolerance, and states S_1 and S_2 represent intermediary states (meaning mostly tolerant and mostly intolerant).

In accordance with the transitions in Figure 2.4 we have the probabilities of an gent being in one of the 4 states at time $t + \Delta t$:

$$\begin{cases} P_{S_0}(t+\Delta t) &= (1-\lambda\Delta t)\, P_{S_0}(t) + \mu\Delta t P_{S_1}(t) \\ P_{S_1}(t+\Delta t) &= \lambda\Delta t P_{S_0}(t) + \mu\Delta t P_{S_2}(t) + \\ & \quad + [1-(\mu+\lambda)\,\Delta t]\, P_{S_1}(t) \\ P_{S_2}(t+\Delta t) &= \lambda\Delta t P_{S_1}(t) + \mu\Delta t P_{S_3}(t) + \\ & \quad + [1-(\mu+\lambda)\,\Delta t]\, P_{S_2}(t) \\ P_{S_3}(t+\Delta t) &= \lambda\Delta t P_{S_2}(t) + (1-\mu\Delta t)\, P_{S_3}(t) \end{cases} \tag{2.12}$$

At $t = 0$, the probabilities of having the social agent in one of the four states are: $P_{S_0}(0) = 1$ and $P_{S_1}(0) = P_{S_2}(0) = P_{S_3}(0) = 0$. Therefore, by applying Laplace transormation, we solve 2.12:

$$P_{S_0}(s) = \frac{s^3+(3\mu+2\lambda)s^2+\left(3\mu^2+2\mu\lambda+\lambda^2\right)s+\mu^3}{s^4+3(\mu+\lambda)s^3+(3\mu^2+4\mu\lambda+3\lambda^2)s^2+(\mu^3+\mu^2\lambda+\mu\lambda^2+\lambda^3)s} \tag{2.13}$$

and

$$P_{S_1}(s) = \frac{\lambda s^2+\left(2\mu\lambda+\lambda^2\right)s+\mu^2\lambda}{s^4+3(\mu+\lambda)s^3+(3\mu^2+4\mu\lambda+3\lambda^2)s^2+(\mu^3+\mu^2\lambda+\mu\lambda^2+\lambda^3)s} \tag{2.14}$$

The probability of having a relative tolerant state (in other words, the social agent is either tolerant or mostly tolerant) is given by $P_{tol}(s) = P_{S_0}(s) + P_{S_1}(s)$. By taking $P_{tol}(s)$ to infinity, we get social agent's expected stable state:

$$\lim_{t\to\infty} P_{tol}(t) = \lim_{s\to 0} sP_{tol}(s) = \frac{\mu^3+\mu^2\lambda}{\mu^3+\mu^2\lambda+\mu\lambda^2+\lambda^3} \tag{2.15}$$

When we do not have NULL social agents or the number of NULL agents is sufficiently small, $\mu + \lambda \simeq 1$, therefore having:

$$\begin{aligned} \lim_{t\to\infty} P_{tol}(t) &= \frac{\mu^3+\mu^2\lambda}{\mu^3+\mu^2\lambda+\mu\lambda^2+\lambda^3} \\ &= \frac{\mu^2(\mu+\lambda)}{(\mu+\lambda)^3-2(\mu^2\lambda+\mu\lambda^2)} \\ &\simeq \frac{\mu^2}{1-2(\mu^2\lambda+\mu\lambda^2)} \end{aligned} \tag{2.16}$$

In order co conveniently provide a graphical representation of tolerance (see Figure 2.5), we consider the tolerance expression in equation 2.15 as being a function of λ: $P_{tol-4}(\lambda)$.

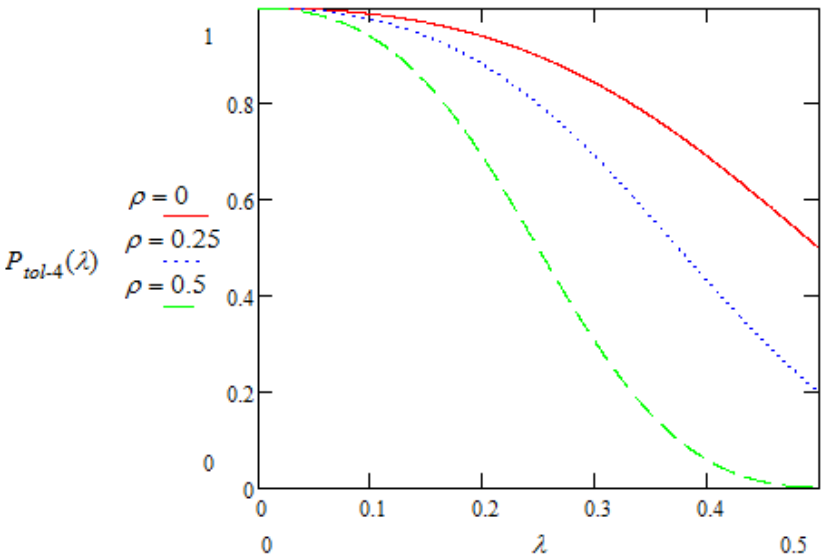

Figure 2.5: $P_{tol-4}(\lambda)$ illustration, when fixed values are used for ρ (0, 0.25, and 0.5). As $\lambda \leq 1 - \rho$ and $\rho \leq 0.5$, it occurs that $\lambda \in [0, 0.5]$.

To provide graphical comparison between the 3-state model of tolerance ($P_{tol-3}(\lambda)$) and the more complex and realistic 4-state model ($P_{tol-4}(\lambda)$), we opt for augmenting the contrast between the two models by considering that there are no NULL agents (so that $\rho = 0$ and $\mu + \lambda = 1$), see Figure 2.6.

2.2.3 Experimental results

Our probabilistic analysis is expressing the social agent's tolerance state in terms of λ, μ, and ρ rates. Therefore, we need to link λ, μ, and ρ with the topological characteristics of the underlying social network where opinion interaction occurs [3]. To this end, we perform computer simulations on distinct topology types, so that empirical values for λ are rendered. Thus, we take four representative social network topologies [200]: mesh, random [74], small-world [199], and scale-free [18].

To perform relevant simulations, we generate 1000-node networks for each of the four representative topologies, by using the available plugins from the Gephi software

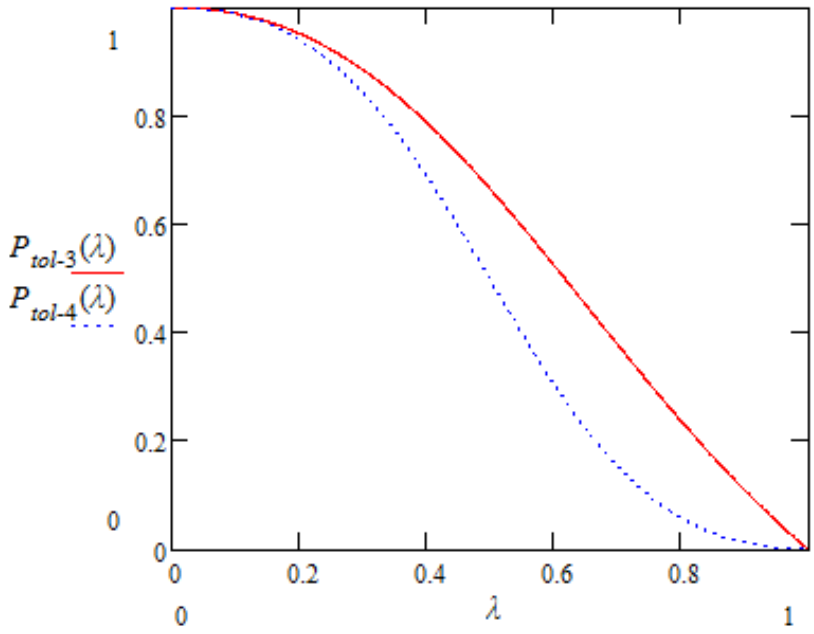

Figure 2.6: Graphical comparison of tolerance probabilities at stable state: $P_{tol-3}(\lambda)$ vs. $P_{tol-4}(\lambda)$ when $\rho = 0$ and $\mu + \lambda = 1$.

package[19]. We adopt the same strategy as in [185] to place stubborn agents [3][207], namely according to a random distribution. Hence, all network that we generate are contain randomly opinionated agents, which correspond to an equal number of stubborn agents having opinion 0 (depicted in red) and 1 (depicted in green), see figure 2.7 for an instance where we have a mesh network with 10 stubborn agents (5 green and 5 red).

In this context, parameter λ is calculated as s given in equation 2.17, where λ_i, is the rate at which node n_i encounters the same opinion can come in contact with the same opinion, for cardinal $|V_i^*| <$ cardinal $|V_i|$ (see equation 2.18).

$$\lambda = \frac{1}{n}\sum_{i=1}^{n}\lambda_i \tag{2.17}$$

$$\lambda_i = \frac{|V_i^*|}{|V_i|} \tag{2.18}$$

For our simulations, we obtain a time evolution for λ as given in Figure 2.8. The simulation results that are presented in Figure 2.8 were stopped only when the difference between the instantaneous λ and the updated median value of λ was ¡3%; however, at least 50,000 iterations were run.

To summarize results, we present maximum, minimum and average λ values, as resulted from the simulation series, in Table 2.1. The results in Table 2.1 further confirm the empirical observations from our previous research work [185]. Indeed, the new results emphasize that:

- Intolerance thrives in regular mesh and scale-free networks. The explanation lies in the specificity of these topologies; as such, local clusters emerge where a certain

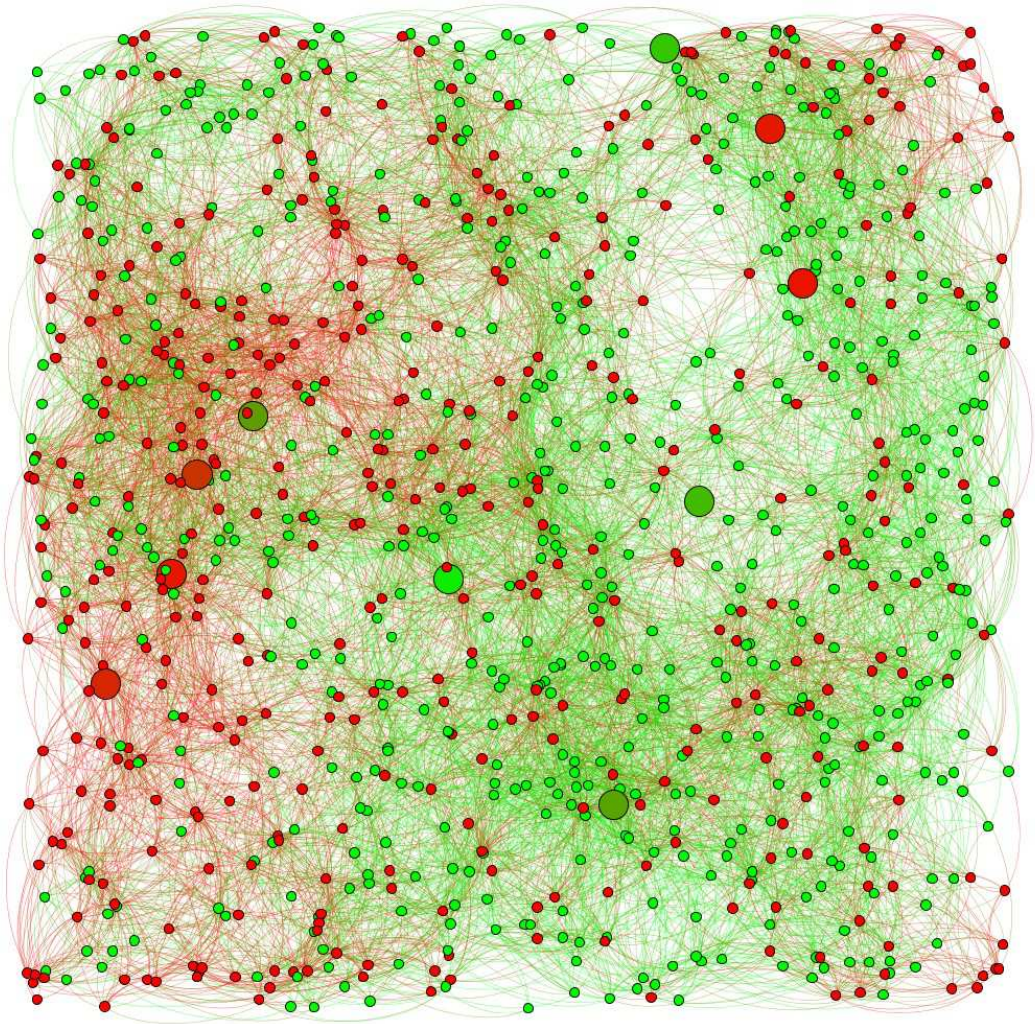

Figure 2.7: Simulation of opinion simulation dynamics in a 1000-node mesh-topology social network. Social agents or nodes have associated colors (red – opinion 0, and green – opinion 1), whereas stubborn agents are emphasized by their larger size.

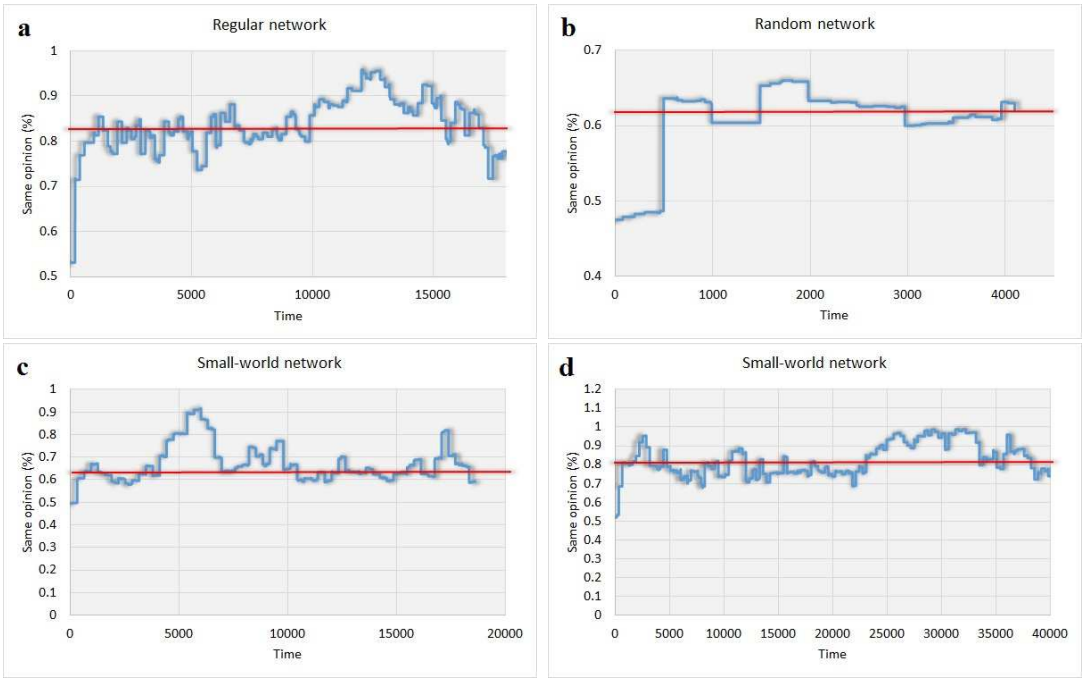

Figure 2.8: Time variation of λ during opinion interaction simulation for: **a.** Regular mesh network topology. **b.** Random network topology. **c.** Watts-Strogatz small-world topology. **d.** Barabasi-Albert scale-free network topology. The red horizontal lines represent the median value for λ.

Table 2.1: Maximum, minimum and average λ values; for a social agent in a given social network, λ represents the rate of opinion interaction with the same opinion. The considered social networks have the following fundamental theoretical topologies for 1000 nodes: mesh, random, small-world and scale-free.

	Regular	Random	SW	SF
λ_{min}	0.72	0.59	0.57	0.67
λ_{med}	0.82	0.62	0.63	0.81
λ_{max}	0.96	0.65	0.91	0.98

opinion begins to dominate, so that social agents only know about the opinion of neighboring nodes. In meshes, these neighbors are the adjacent agents/nodes, whereas in scale free networks the neighbors are oftentimes the nodes with very high degree (a so-called hub). As a result, mesh and scale-free topologies favours the interaction with the same opinion, which in turn create more intolerant social networks, which are less inclined to change and accept different opinions or ideas. Numerical results support this conclusion, because of the high values of λ: $\lambda_{regular} = 0.82$, and $\lambda_{scale-free} = 0.81$.

- Tolerance is promoted in mesh and small world networks, because links are distributed according to a Poisson process. Therefore, opinion clusters rarely form due to the randomly placed long-range links that connect local clusters to clusters holding other types of opinions. As a result, the random interconnections in social networks increase tolerance over time, and fosters a society that is more dynamic, and opened to new opinions/ideas. Indeed, the quantitative results obtained by simulations support these conclusions, because of the lower $\lambda_{random} = 0.62$, and $\lambda_{small-world} = 0.63$.

2.3 Discussion and Conclusions

Our fundamental contribution to the field of social network interaction and opinion spread models consists of endowing social agent with more human-like traits like adaptability and inner predisposition to accept or to reject other people's opinions and ideas. As a result, our model is able to replicate real-world dynamic opinion phenomena, a feature that is not accessible to hitherto social interaction models such as [61][107][124][37][60][80][122]. Quantifying and modeling human action and inner motivation is a very difficult endeavor, therefore simulating and analyzing individual's trust, faith, or tolerance states is still in its infancy. Our approach to this problem is to find inspiration in social psychology, in order to enhance the existing fixed-threshold social interaction models. In social psychology, *individual tolerance* is considered as an important factor in the opinion dynamics of the entire society; au such, the concept of egocentrism, is considered to be linked to individual's internal emotional status [72]. Our approach was to use this model due to the fact that egocentrism is a trait that connects to individual tolerance towards other opinions [204][185].

Moreover, in [191], we have analyzed our tolerance-based opinion interaction model from a probabilistic standpoint, and correlate the findings with simulation results. Our probabilistic analysis explains opinion dynamics obtained by simulation or observed in real-world systems, especially when the society stabilizes, namely at $t \to \infty$. As presented in figures 2.3 and 2.5, one of the most important results in our probabilistic assessment of the tolerance model is that tolerance is higher for the average social agent with small ρ (i.e. the agent connects with a small number of NULL agents – social agents that have no opinion).

Our probabilistic analysis in [191]assumes a 3-state Markov model (when we have just one intermediary state between pure tolerance and pure intolerance) and a much realistic 4-state Markov model (with two intermediary states between pure tolerance and intolerance). Our assessment finds that the probability of tolerance decreases with λ (i.e. the rate of opinion interactions with the same opinion in the social network) in an almost linear fashion; this dependency becomes exponential for the 4-state model. These results suggest that real-life social interaction phenomena are non-linear; moreover, linear interpretations are mere overly simplistic approaches that are not reliable and cannot be used as prediction tools.

In [185] we show that by considering the tolerance-based opinion interaction model, we are able to reproduce dynamical features of opinion formation such as phase transitions and opinion formation phases. We also show that these non-linear dynamic opinion phenomena are influenced by underlying social network topology. In fact, we find that the topology has a stronger influence on opinion spread phenomena than, for instance, network size or stubborn agents placement and distribution. As such, mesh and scale-free networks correspond to conservative, oligarchic societies, whereas random and small world networks correspond to decentralized, democratic societies [185][191].

Chapter 3

Network Pharmacology and Network Medicine

3.1 Network Pharmacology

Over the last decade, the concept of *drug repositioning* (sometimes called *drug repurposing* has been intensely researched and developed within the field of *drug design*. Drug repositioning means uncovering new pharmaceutical properties and indications for already existing drugs [171]. The motivation for the growing interest in drug repositioning strategies is twofold. First, there are the notable advances in scientific fields such as bioinformatics, genomics, physics, complex networks, as well as technological fields such as machine learning and database mining [123][57]. Second, there is a massive marked demand for new drugs, although on the other hand the drug design process is slow and evermore expensive; as a result, entirely new approved pharmaceutical formulations are very hard to get [65]. In this context, it seems like drug repositioning can be a more affordable alternative from both technical and economic standpoints [38][140][153][167]. Indeed, a recent study has revealed that, as of now, we are already taking about 20% of the new drugs brought on the market as being drug repositionings [95]. Moreover, the available repositionings are made as part of personalized (or precision) medicine initiatives [164].

Conventional drug repositioning is performed by using traditional-experimental approaches stemming from biochemistry and genetics; however, many drug repositionings are found by mere serendipity [28]. Nonetheless, our approach to this very important pharmacological problem is powered by computational tools, which are linked to the complex network approach. Over the last decade, many such computational approaches were developed, by taking advantage of the tremendous advances in big data gathering and machine learning; these computational solutions encompass a wide array of issues from pharmacology and drug design, which include drug repositioning. As such, there are computational solutions for uncovering new drug interactions that were unaccounted during clinical trials [178]. Also, there are computational models used to predict the degree of drug safety during therapy [71][121]. For drug repositionings, computational approaches are processing and analyzing multiple comprehensive databases such as drug, genomic, transcriptomic, and phenotypic databases [123]. However, the

results of all these computational methods are mere predictionss, which will have to be validated by traditional *in vivo* and *in vitro*) experimental methods [123].

Many computational methods for drug repurposings are based on recent advances in the new science of complex networks as well as on the fact that we have access to an evermore increasing volume of data about medication (including approved, investigational and experimental drugs). Some of the most available and easy-to-get data on medication consists of drug-drug interactions, namely data about the way drug effects interact when taken together according to some therapeutic schema. Consequently, one of the most important tool in translational, systems and computational pharmacology is based on *drug-drug interactomes* (DDI); DDIs are complex network with nodes/vertices representing pharmaceutical substances (i.e. drugs) and links representing drug interactions (e.g. common mediation by a some enzyme, or synergistic effects). There are many advantages in using drug interactomes (DDI) for analysis, but the most important applications are:

- Predicting new potential interactions [178][105]; based on this principle, many software products were developed in order to issue drug interaction alerts [177].
- Avoiding certain types of drug-drug interactions even from the drug design stage [47][165].
- Exploring the links between pharmaceutical properties and drug interactions; previously, this idea was mostly applied for predicting new drug-drug interactions when we have verified information about confirmed interactions [156][111].

Recent developments aim at using interaction information from DDIs to uncover new drug effects and properties, thus paving the way for drug repositioning [125]. As such, [210] processes a DDI with the Markov Clustering Algorithm, in order to predict new drug functions. Another research takes information about drug side effects from social media (Twitter), in order to build a corresponding DDI and then predict possible drug repurposings [148].

3.1.1 Drug-drug interaction network analysis

We build a Drug-Drug Interaction (DDI) network by processing data on drug-drug interaction from the comprehensive database DrugBank 4.1 [205]; as such, in our DDI each drug is represented as a node, while drug interactions are represented as links between corresponding drugs. At first, such processing made using the Gephi software package [19] will result in a raw DDI (see the top panel of Figure 3.1). When we build our DDI, we do not use any kind of functional information, therefore we do not know *a priori* the usage or properties of these drugs. Also, even if the drug interactions pertain to two basic types, namely antagonistic and synergistic, we do not use this dichotomy, due to the fact that we need all interactions types to contribute at defining drug's functional profile.

Another vital aspect is related to validating our findings, because any predicted or conjectured new drug properties have to be confirmed by medical and pharmacologic

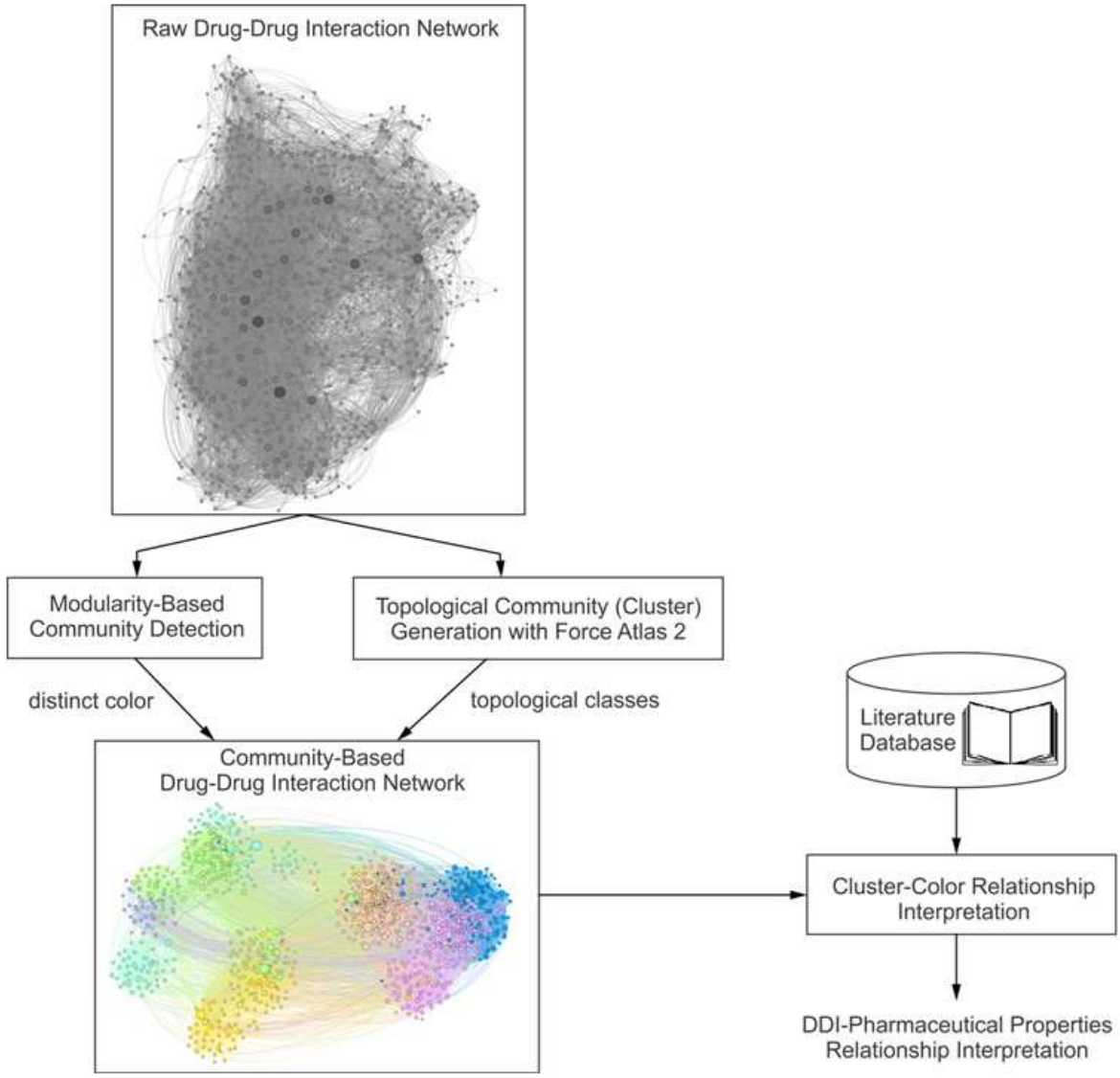

Figure 3.1: Drug-drug interactome (DDI) computer processing methodology for clustering drugs based on modularity classes and force-directed topological clustering. The clustering result is correlated with several pharmacological properties, thus creating incentives for predicting drug repositionings.

tools (i.e. *in vivo* and *in vitro*). Therefore, we find the following validation procedure as convenient. We use the older 4.1 DrugBank version for building and analyzing our DDI, in order to spare the newer DrugBank 4.3 for validating the predicted new pharmacological properties.

In order to cluster the drugs in the DDI we use the methodology from Figure 3.1: we classify network nodes to modularity classes that are depicted with distinct colors [147] and, at the same time, by using the force-directed layout algorithm Force Atlas 2 [108] we generate well-defined topological communities which also act ass clusters. The visual result is what we call Community Based Drug-Drug Interaction Network (CBDDIN) from Figure 3.2.

The novelty of our methodology [190] is that it uses for the first time in the field of drug interactomes the network analysis tools which are conventionally deployed only for social networks. Such an approach is further motivated by the fact that there is a good similarity in terms of network parameters between our DDI and a typical social network: low average path length of $L = 2.978$ given the network diameter $Dmt7$, a relatively high clustering coefficient $C = 0.2$, high average degree $< k >= 20.031$, typical values for modularity $Mod = 0.452$ and density $Dst = 0.017$.

In terms of network centrality metrics, our DDI is characterized by the distributions from Figure 3.3: power-law degree, betweenness and eigenvector distributions, normal closeness distribution. In qualitative terms, these distribution types indicate a scale-free network.

In DrugBank Version 4.1 [205] there is a total of 7739 drug entries (including, besides approved drugs, experimental and investigational drugs). As some of these drug entries have no information about interactions, we remove them because they cannot be connected to the DDI; after the removal process, 1141 drugs will remain. However, additional data is required to verify and validate our predictions; to this end, we cross-check with functional properties that are listed in other databases such as *Drugs.com*, *RxList*, and *DrugBank 4.3*. Also, a good amount of validation is performed by simply searching scientific literature from online article databases.

3.1.2 Interpreting network centralities

Our CBDDIN is shaped by the collection of drug-drug interactions relationships, hence drugs that have the highest centralities (e.g. degree, betweenness) are the most predisposed to other drug-drug interactions. As such, we can scorecard the interaction potential of drugs from CBDDIN; table 3.1 introduces only the top 10 drugs in terms of degree, betweenness, closeness, page rank, and eigenvector.

Description of drug clusters

Topological clusters from Figure 3.2 are generated with the Force Atlas 2 [108] layout algorithm, while the modularity clusters (which are correspondingly labeled with distinct colors) are automatically generated in Gephi [93]. We identify an underlying, generally common property for all topological communities; this labeling process is made by individualizing a pharmacological property which best describe the biggest

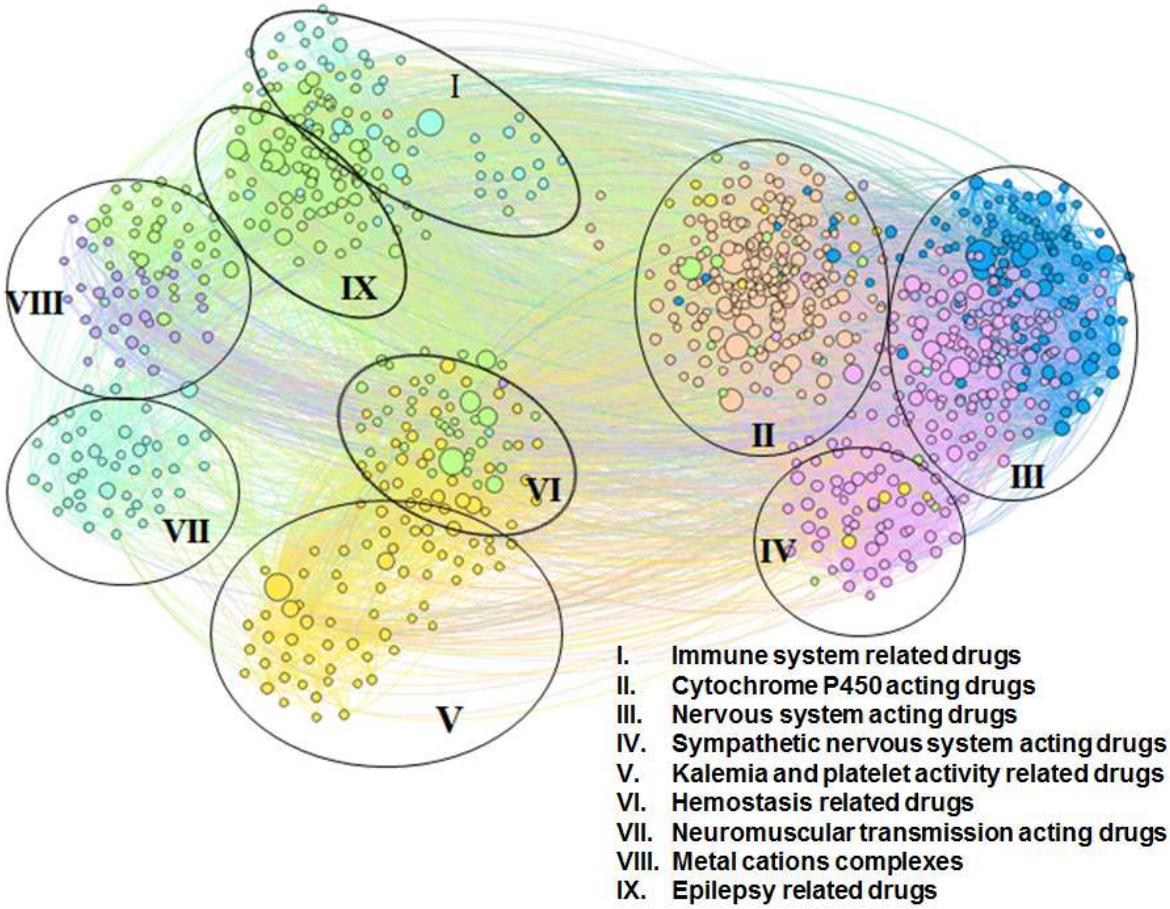

Figure 3.2: Our visual DDI clustering, resulted as the Community-Based Drug-Drug Interaction Network (CBDDIN). Both modularity classes (labeled with colors) and force-directed (Force Atlas 2) clusterings are generated in Gephi, by processing the drug-drug interaction data taken from DrugBank 4.1 (consisting of 1141 drugs and 11688 drug-drug interactions).

Table 3.1: Top 10 hierarchies of drugs from Community-based drug-drug interaction network (CBDDIN) in terms of degree, betweenness, closeness, page rank, and eigenvector values.

	Degree		Betweenness		Closeness		Page rank		Eigenvector	
	Drug	Value	Drug	Value	Drug	Value	Drug	Value	Drug	Value
1.	Voriconazole	250	Triprolidine	47.570	Voriconazole	2.077	Triprolidine	0.009	Voriconazole	1
2.	Triprolidine	198	Trastuzumab	37.522	Ketoconazole	2.121	Voriconazole	0.009	Telithromycin	0.851
3.	Telithromycin	198	Treprostinil	37.168	Phenytoin	2.128	Treprostinil	0.008	Trimipramine	0.845
4.	Warfarin	181	Warfarin	34.129	Cyclosporine	2.141	Warfarin	0.008	Rifampin	0.683
5.	Trimipramine	174	Cyclosporine	33.217	Telithromycin	2.142	Trastuzumab	0.007	Tramadol	0.682
6.	Ketoconazole	161	Tacrolimus	30.491	Tacrolimus	2.155	Telithromycin	0.007	Tacrolimus	0.680
7.	Rifampin	157	Tacrine	27.734	Warfarin	2.168	Cyclosporine	0.006	Trazodone	0.656
8.	Cyclosporine	150	Phenytoin	22.018	Rifampin	2.171	Ketoconazole	0.006	Ketoconazole	0.656
9.	Phenytoin	147	Ketoconazole	21.598	Trimipramine	2.180	Tacrine	0.006	Quinidine	0.623
10.	Tacrolimus	147	Rifampin	21.288	Fosphenytoin	2.184	Rifampin	0.006	Clarithromycin	0.609

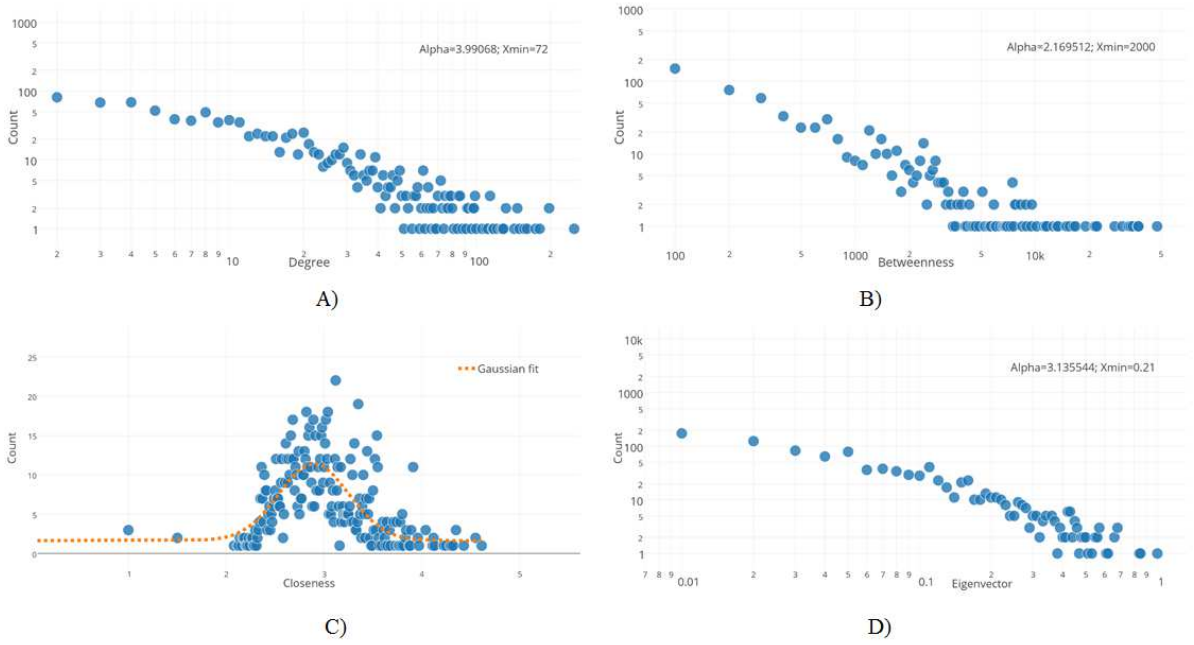

Figure 3.3: Centrality distributions for our CBDDIN: A) degree, B) betweenness, C) closeness, and D) eigenvector. We also provide the slope α and cutoff point $X_{\min}$ values for all power-law distributions. For the normal distribution of closeness, we provide the best Gaussian fit as $f(x) = 1.654 + 9.768 \exp\left(\frac{-(x-2.896)^2}{2 \cdot 0.36^2}\right)$.

majority of nodes/drugs within each cluster. For some clusters, the label targets an organ or system, for others a medical indication, or a well-defined chemical structure. Nonetheless, we do not put functional limits to our labeling process, because we need to obtain the specific property which better identifies the biggest number of drugs from each cluster/community.

The labeling of modularity classes from Table 3.2 is very consistent, as the tags characterize very well almost all drugs in CBDDIN. This result is confirmation of previous results which indicate modularity as an outstanding predictor of functionality in biological systems [144]. Indeed, because CBDDIN is shaped by the distribution of link density, in our approach modularity is very well connected with the concept of drug interactions.

As presented in Figure 3.2, we identify and label accordingly a number of 9 topological clusters; the accuracy of Figure 3.2's labeling is given, in Table 3.3 as confirmation percentages. Each community contains nodes/drugs for which their assigned modularity class actually confirms the topological cluster label. The percentage of topological cluster label confirmation with modularity classes is given in the column "Conf. by mod. [%]". Also, by cross validating with state of the art literature in the fields of pharmacology and biology and by confronting with other databases (Drugs.com [211], RxList [212], DrugBank 4.3 [205], we further explain topological community labels for lots of other drugs. The overall result of cross-validating and label confirmation is presented in the 5th column of Table 3.3.

Table 3.2: Interpretation and labeling og CBDDIN modularity classes, according to modularity clustering algorithm. The allocated labels correspond to pharmacological properties (see column "Modularity class interpretation"). We also indicate the number of drugs in each modularity class, as well as the percentage of drugs that correspond to the allocated label (columns "No. of drugs" and "Consistency [%]").

Modularity class (color)	Code	Modularity class interpretation	No. of drugs	Consistency [%]
Dark blue	DB	Central and peripheral nervous system acting drugs	232	96
Velvet maroon	VM	Substrates, inhibitors and inducers of specific CYP enzymes	210	91
Green	G	Drugs that interfere in different phases of hemostasis, anticonvulsant and epileptogenic drugs	191	85
Magenta	M	Drugs acting on sympathetic nervous system	166	93
Light blue	LB	Drugs targeting cancer, auto-immune disorders (i.e. rheumatoid arthritis), and musculoskeletal system	156	88
Golden brown	GB	Drugs interfering with platelet activity and plasma potassium levels	155	92
Purple	P	Bi-and trivalent cations, chelating agents	31	100

Table 3.3: CBDDIN labels for each of the nine topological clusters. The 2nd column present the modularity classes (identified with colors) which confirm the topological cluster tag. The 3rd column gives the total number of drugs in each topological cluster, which is then broken to the following categories: correctly described by the assigned modularity class (Conf. by mod.), explained by properties listed in drug databases and state of the art literature (Expl. by propr.), and not explained yet (Not expl.)

Comm.	Colors	No.	Conf. by mod. [%]	Expl. by propr. [%]	Not expl. [%]
I	LB, VM	80	77	19	4
II	VM, GB, G, DB, LB, M	271	80	8	12
III	DB, M, LB, GB	307	84	12	4
IV	M, GB, G, DB	81	51	22	27
V	GB	54	33	63	4
VI	GB, G, LB, M, P	125	35	38	27
VII	LB	58	26	31	43
VIII	P, G, LB, GB	69	56	9	35
IX	G, LB	96	30	57	13
		1141	63	22	15

Overall, the most important result is that our methodology can extract pharmacological properties only from drug-drug interaction data. As such, 63% of the assigned drug property labels are in agreement with the drug properties already listed in DrugBank 4.1; additionally, 22% of included drugs are not listed with the assigned property in DrugBank 4.1 but are confirmed as correctly labeled, by an extensive cross-checking process which employs literature survey and other public drug databases. As a result, the remaining 15% of the drugs seem to be out of place according to their respective topological cluster labeling; such situations represent opportunities for predicting drug repositionings. Indeed, repositioning assumptions can be made by simply analyzing the cases which seem as not compliant with topological or modularity cluster labels. We listed all seemingly non-compliant drug cases in column Not expl. [%] (not explained) from Table 3.3. Consequently, we predict that drugs listed as not explained can be repurposed in accordance with a pharmacological property that is given by either their modularity or their topological cluster labels.

Apart from the most obvious approach based on label non-compliance, there is another valid repurposing strategy, namely analyzing drugs which lie at the frontiers between two or mode topological clusters; the assumption is that such borderline situations may indicate multiple pharmacological properties pertaining to the neighboring clusters.

According to our contributions [190], a short functional, pharmacological description of the topological clusters generated with our methodology is given below:
Cluster I Antineoplastic agents, immuno-stimulants and immunosuppressants.
Cluster II Substrates, inhibitors and inducers of specific cytochrome P450 enzymes which mostly have a CYP-related activity.
Cluster III Drugs that affect the metabolism of neurotransmitters, thus incurring central and peripheral nervous effects.
Cluster IV Drugs which act on alpha- and beta- adrenoreceptors, which mostly include the cases of drugs related to sympathetic nervous system (SNS) effects [104][117][196][112]).
Cluster V Most drugs from this cluster consist of renin-angiotensin system acting drugs and diuretics, but we can also find some platelet aggregation inhibitors [114].
Cluster VI Drugs which interfere with different phases of hemostasis.
Cluster VII Drugs that have neuromuscular-blocking activity, from both pharmacodynamic and pharmacotoxicologic perspectives [152][53].
Cluster VIII This cluster mostly consist of bivalent and trivalent metal cations, as well as of their corresponding chelators [141][205][182][135].
Cluster IX Drugs employed for treating various types of epilepsy and seizure, as well as epileptogenic drugs [175][49].

Examples of finding new properties

To illustrate how our network-based is able to predict multiple pharmacological properties and already known repositionings, using only drug-drug interaction data, we present the cases of Zafirlukast and Thalidomide:

- Zafirlukast is a drug used in asthma therapy; its positioning in Cluster II (Figure 3.4A) is motivated by the fact that it is CYP3A4 substrate and inhibitor [205][212]. However, as opposed to most other drugs from Cluster II, zafirlukast is

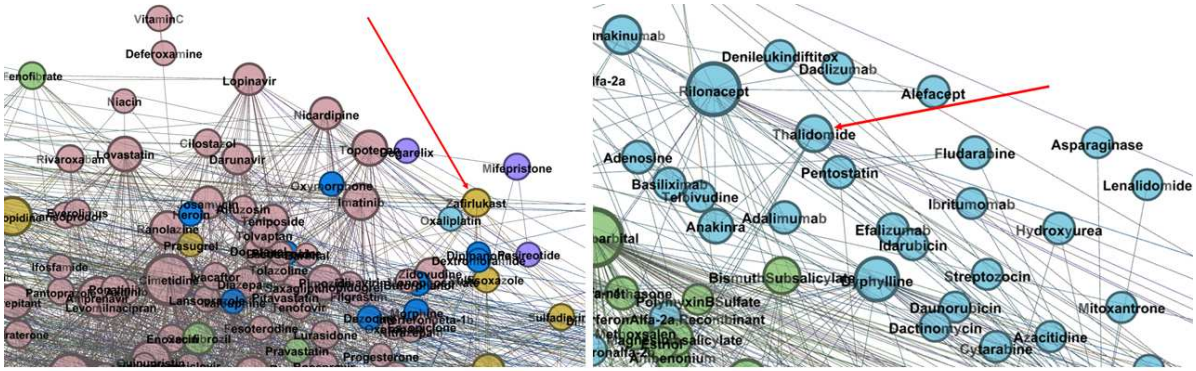

Figure 3.4: CBDDIN details that present: A) the position of Zafirlukast in Cluster II; B) the position of Thalidomide in Cluster I.

pertains to the golden brown (GB) modularity class; this property is confirmed by zafirlukast's platelet activity [116]. Indeed, this example shows how our dual clustering methodology (modularity class plus energy-model layout) is able to recover multiple properties.

- Thalidomide is a recent example of drug repositioning. Initially, thalidomide was designed for thwarting pregnant women's morning sickness, but it was later withdrawn because of its adverse effects on embryos. However, as of today thalidomide is employed for treating immunological and inflammatory disorders [180]. This newly found property is also predicted by our method, as presented in Figure 3.4B: thalidomide pertains to the light blue LB modularity class (Anti-cancer drugs) and is placed within Cluster I (Immune system related drugs).

3.1.3 Results Interpretation

The fact that our drug-drug interaction network clustering methodology generates 9 well-defined pharmacological characteristics (as presented in Figure 3.2) may suggest that the 9 properties have essential roles in drug-drug interaction mechanisms. As such, the fact that 9 pharmacological properties have a stronger influence on drug-drug interactions is in itself a significant contribution.

In terms of accuracy, our prediction of pharmacologic properties is validated for 85% of the drugs with functional information taken from DrugBank 4.1., from other drug databases, and from literature surveys. Hence, motivated by this high prediction accuracy, we argue that it is extremely likely that – in the future – the predicted properties will also be confirmed for the remaining 15%.

In our CBDDIN there are many possible repositionings. For instance, the drug chlorzoxazone is predicted by our method as haing immunological properties, as it is positioned in Cluster I. Also, cefalosporin antibiotics such as cefalotin, cefamandole, or cefixime, are predicted as acting on the immune system or neuromuscular junction, because they pertain to Cluster VII. Moreover, predicting new properties may lead to

predicting new interactions. Indeed, drugs bevacizumab and lorazepam are placed in Cluster II, so that can be predicted as interfering with CYP enzymes. Nonetheless, computational models can only make useful repositioning hints, which may be subsequently confirmed by *in vivo* and *in vitro* biomedical tools. To this end, we indicate that these drugs be further investigated to confirm new repositionings or interactions.

From a fundamental perspective, drug-drug interactions actually express the way that drug behaviors interfere, constructively or destructively. Consequently, it can be said that our network clustering repositioning methodology is acting at behavioral level. Such an approach is contrasting with most hitherto network-based repositioning strategies, which rely on chemical structure similarity relationships or drug-target relationships. We consider our work as a strong argument for the multi-level network approach to drug repurposing, an approach that integrates the behavioral and structural perspectives [209]. In this context, we identify DARPA's Big Mechanism project [43] as an appropriate platform for multi-level integration.

Our methodology from Figure 3.1 can also be used for clustering patients in medical databases, for instance in cardiovascular disorders [176], sleep apnea syndrome [186], or defining endophenotypes [92]. These studies prove that disease risk factors do not associate at random, they rater converge towards well-defining patient phenotypes, which in turn provide valuable information for precision medicine approaches.

3.2 Network Medicine

Obstructive Sleep Apnea Syndrome (OSAS) is a serious clinical disorder caused by abnormal breathing pauses that occur during sleep; this results in sleep fragmentation and excessive daytime somnolence [170][84][120]. There are studies reporting the epidemic incidence of OSAS, with worrying increasing rates over the last 20 years [208][157][155]. If not properly diagnosed and treated, OSAS increases the morbidity and perioperative risks [132][131][159][193][161].

Apnea severity is indicated by the Apnea-Hypopnea Index AHI; this represents the number of breathing pauses of at least 10 seconds, recorded over one hour of sleep. As such, any patient can be classified in one of the following AHI categories: normal or low-risk apnea (L) for $AHI < 5$, mild sleep apnea (Mi) for $5 \leq AHI < 15$, moderate sleep apnea (Mo) for $15 \leq AHI < 30$, and severe sleep apnea (Se) when $AHI \geq 30$.

Usually, polysomnography (PSG) is used as the reference method (i.e. gold standard) of OSAS diagnosis, which is based on measuring AHI. However, PSG is expensive, time-consuming, and generally not adequate for population screening. [170][163]. Since OSAS has a significant prevalence and the PSG-based exhaustive investigation is not feasible when screening a large population, OSAS predictors are preferred for monitoring [192][184].

In current practice, there are three major predictive models based on questionnaires, namely Berlin, STOP, and STOP-BANG [142][169][81][50][51]. Published studies indicate STOP-BANG as the best available predictive score, due to its high sensitivity: 83.6% for $AHI > 5$, 92.9% for $AHI > 15$, and 100% for $AHI > 30$. However, STOP-BANG has a low specificity (56.4% for $AHI > 5$, 43% for $AHI > 15$, and 37% for

$AHI > 30$) [50][51][52] which prevents the usage of this score for population screening. Although there are notable attempts for improving STOP-BANGs specificity [52], they are mainly targeting narrow-type cohorts such as perioperative patients.

Consequently, our paper aims at analysing the general case, with all patient categories being taken into account for screening, and not just some specific cohorts. To this end, our research is underpinned by a complex network perspective on uncovering OSAS phenotypes. Indeed, network science is already successfully used in medicine at disease-level [17], including respiratory applications [66][78][67]. Our network-based approach on OSAS risk factors allows for better, more accurate OSAS phenotype identification, which in turn leads to a new predictive score (SAS_{Score}). In comparison with the state-of-the-art, our OSAS risk prediction score achieves significantly better specificity in predicting actual AHI categories, which makes our SAS_{Score} very appropriate for screening big populations as part of preventive medicine programs.

3.2.1 Processing Medical Databases

In order to use complex network tools for OSAS research, we need real-world OSAS patient datasets. Unfortunately, OSAS patients datasets are scarce and not public; such a situation is justified by multiple aspects: big data techniques were only recently considered as tools for respiratory medicine and OSAS, all patients must undergo hospital polysomnography (which entails a complex, expensive and time-consuming process), while coordinated research efforts for gathering data were only recently introduced.

For instance, the biggest such OSAS database, namely European Sleep Apnea DAtabase - ESADA [101], is not public and it gathers data from 15,956 patients in 24 sleep centers from 16 countries, since 2007. Also, a recent OSAS study [129] where the validation is similar to our approach, uses only one (private) validation database, comprising 1101 patients [162].

As a result, in order to perform network investigation on OSAS, we built our own Apnea Patients Database (APD), consisting of consecutive patients with suspicion of sleep breathing disorders, which were evaluated at Victor Babes Regional Hospital from Timisoara (Western Romania) between March 2005 and March 2012, under the supervision of the hospital's Ethics Committee (internal briefing note no. 10/12.10.2013). At the initial visit, the study protocol was clearly explained, to obtain the patients consent and the acceptance of referral physicians. Subsequently, respiratory polygraphy was performed using both Philips Respironics Stardust polygraph (2005) and MAPs POLYMESAM IV (1998). PSG was carried out with Philips Respironics Alice 5 Diagnostic Sleep System, according to the appropriate guidelines [158]. The polygraphy was performed both at home and at the hospital, whereas PSG measurements were performed at the hospital under medical supervision. To preserve the information accuracy, all collected data were carefully verified; throughout this process, we have ensured complete data confidentiality. Our observational, retrospective study employs only procedures that are standardized and non-invasive, by excluding all useless investigations. Moreover, visits did not entail additional effort for the patients or supplemental budget for the clinic.

All 1371 patients with completed sleep study protocol and signed informed consent

are included in the APD, each with corresponding 108 breathing parameters and anthropometric measurements. The APD distribution of measured AHI is presented in Figure 3.5.

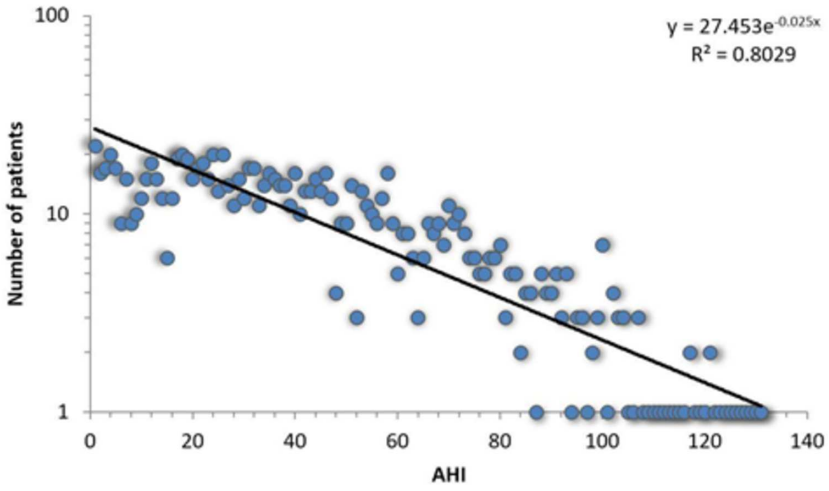

Figure 3.5: Log-linear histogram representing the number of patients with a certain Apnea-Hypopnea Index (AHI) value, along with the corresponding distribution fit. The best fit is the exponential function $f(x) = 27.453 \cdot e^{-0.025x}$.

In order to verify if there is any difference between apnea and non-apnea populations in terms of how risk factors associate and converge, we built a 611 people non-OSAS database NAD (using the same procedure as for the APD). Also, to evaluate the prediction score derived from our study, we gathered a distinct test database TD (fall of 2013) consisting of 231 patients, by following the same procedure. Figure 3.6 presents the distinct roles of our 3 databases, as well as the relationship between them.

Analysis of APD and TD

As patients within TD are used to validate our OSAS prediction with SAS_{Score}, which was obtained by processing patients from APD, we analyse if the distribution of parameters in TD is not too close to the corresponding distributions in APD. Such an investigation is required considering that, although data for the two databases were gathered over distinct periods of time, all measurements were performed in a given geographical region, with subjects that mostly pertain to Caucasian anthropological characteristics.

To this end, we present the distributions of the most relevant parameters in our research (Age A, Body Mass Index BMI, Neck Circumference NC, High Blood Pressure HBP, and Epworth Sleepiness Score ESS), within the validation population (TD) and the apnea patients database (APD) in Table 3.4, under the form of measured averages and their corresponding standard deviations, as well as Gini coefficients. We rely on Gini coefficients for a quantitative measure of data dispersion.

We also provide a visual comparison of AHI and relevant risk factor parameters distributions in APD and TD (see Figure 3.7). All these results show that, given that databases contain subjects from a well-delimited geographical area and were randomly

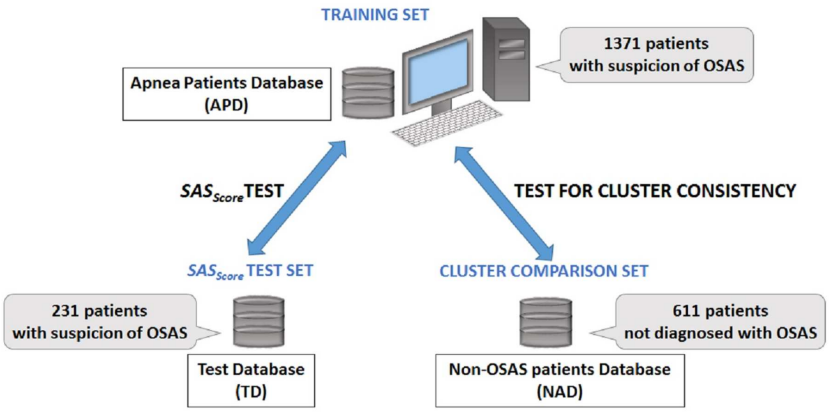

Figure 3.6: Description of databases used in our research. The main Apnea Patients Database (APD), comprising 1371 consecutive patients which arrive at the hospital with suspicion of OSAS between 2005 and 2012, is used to build patient phenotypes and to render the SAS_{Score}. The distinct Test Database (TD), comprising 231 consecutive patients which arrive at the hospital with suspicion of OSAS in 2013, is used to verify the sensitivity and specificity of predicting patient's AHI and OSAS categories. The Non-OSAS patients Database (NAD) uses consecutive assessed people which are not diagnosed with OSAS during the spring 2015 – summer 2016 period, in order to test for cluster consistency (i.e. compare how risk factors converge in clusters for OSAS patients in comparison with people without OSAS).

gathered, the main parameter distributions are somehow similar - mostly normally distributed. However, Gini coefficients (especially for A, BMI and ESS) indicate an important difference between APD and TD distributions. Moreover, Figure 3.7 shows a significantly different AHI histogram for TD in comparison with APD. As such, in APD there are many patients with $AHI > 120$, whereas in TD there is none such patient. Also, in APD, the largest number of patients associated to an AHI value correspond to AHI values < 20; in contrast, in TD, the largest number of patients with a given AHI value correspond to AHI values around 40.

Building the patient network

An unweighted network is a graph (V, E), which consists of a set of vertices (or nodes) V and a set of edges (or links) E that represent connections $[v, w] \in E$ between certain pairs of vertices $v, w \in V$. We build the unweighted Apnea Patients Network (APN), by assigning vertices and edges: each node corresponds to a distinct patient in our OSAS patients database APD, while an edge (link) is created between two vertices if there is a risk factor compatibility between the patients represented by the two vertices (nodes).

The risk factor compatibility is a binary function $f_{RFC} \in \{0, 1\}$ (0 means incompatibility and 1 means compatibility) based on six parameters with high relevance for OSAS: age, gender, BMI, neck circumference, blood pressure (systolic and diastolic), and Epworth Sleepiness Score. We build our APN by considering that $f_{RFC} = 1$ if at least 4 out of 6 parameters are identical; otherwise $f_{RFC} = 0$.

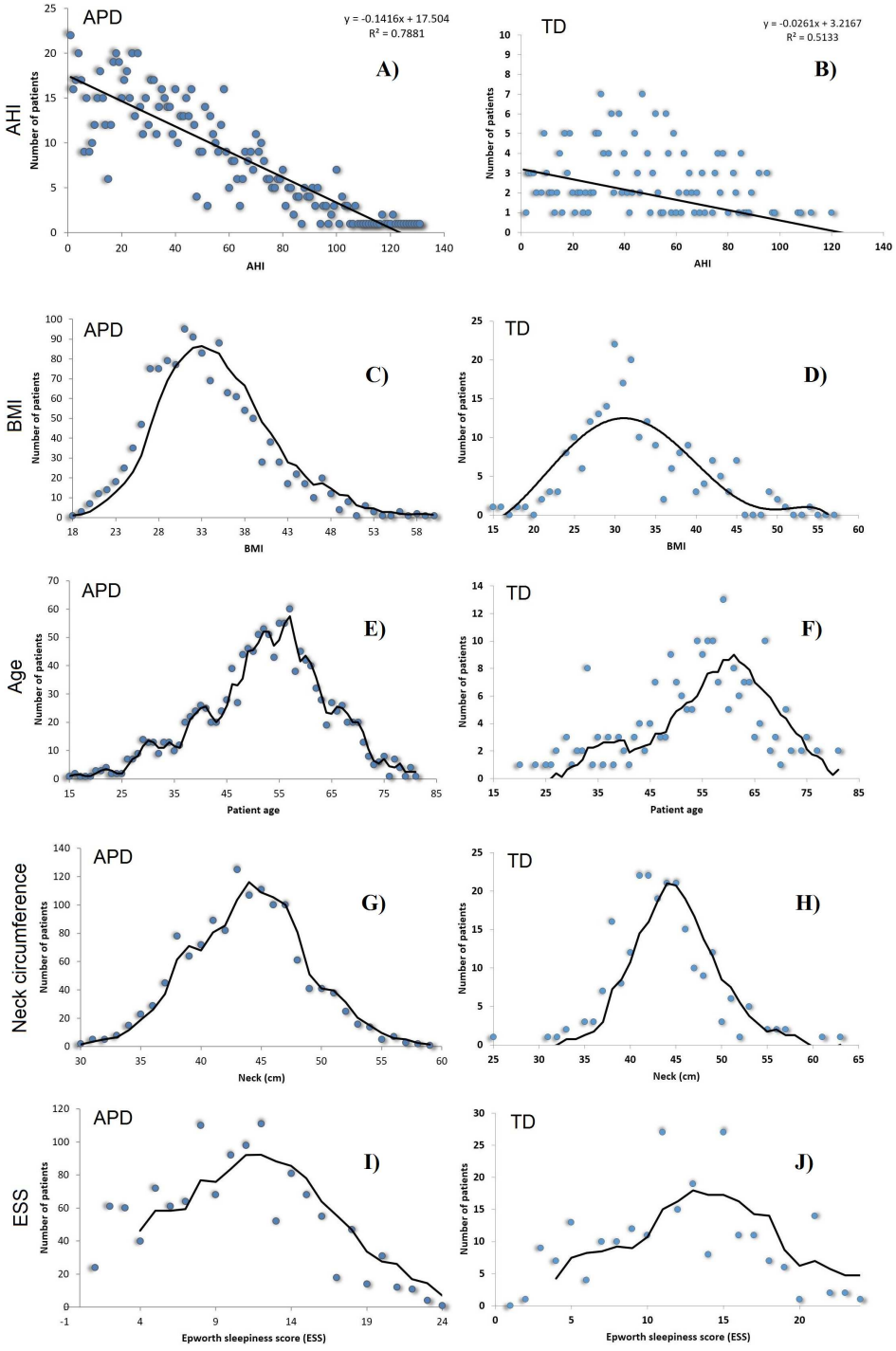

Figure 3.7: AHI distributions in APD (panels A, C, E, G, I) and TD (panels B, D, F, H, J), as well as the normal distributions of BMI, Age A, Neck Circumference NC and ESS, for patients in both APD and TD.

Parameter	APD		TD	
	Average	Gini	Average	Gini
AHI	40.04 ± 27.83	0.388	44.39 ± 26.13	0.334
BMI	33.01 ± 7.94	0.576	32.62 ± 7.41	0.468
A (yrs.)	51.73 ± 12.44	0.539	52.01 ± 13.61	0.397
NC (cm.)	42.01 ± 9.94	0.51	42.64 ± 5.12	0.482
HBP	67.47%	N/A	64.93%	N/A
ESS	10.63 ± 5.38	0.362	11.38 ± 5.19	0.422

Table 3.4: AHI and relevant risk factor parameters distribution in the Apnea Patients Database (APD) and Test Database (TD), given as average values plus standard deviation, as well as Gini coefficients. We only considered boolean values for the High Blood Pressure (i.e. if the patient has high blood pressure or not); in this case, we provided the percentages of people with high blood pressure.

The six parameters are selected from the pool of all relevant risk factors (all measured parameters can be found in the supplementary file *Apnea-Patients-Database-(APD)-and-Test-Database-(TD).xlsx*), because they can be measured easily and objectively; such objective measurements can be performed anywhere, and are widely accepted in the medical literature [120]. In contrast, other scores consider snoring and witnessed apnea episodes as factors, but these are parameters which cannot be observed or measured objectively.

The reason for adopting the 4-out-of-6 criterion is that it assures the right amount of link density in the APN, meaning that there are enough links so that the APN is connected, but not too many links so that communities (i.e. clusters) can be rendered with energy model layouts [147]. As Figure 3.8 shows that the 4 out of 6 link filtering represents the best alternative, we use this criterion to build the APN. To the best of our knowledge, this link filtering procedure is original and hasnt been used before in such network-based approaches.

APN clustering

We clustered the APN, by using a dual clustering methodology: energy-model layouts plus modularity classes, similar to the approach from [190]. Energy-models are force directed network layout algorithms, namely visual tools that assign certain positions in the Euclidian space to both nodes and edges [147]. To this end, we used the Force Atlas 2 algorithm [108] as network layout; this new layout is very effective in clustering various types of complex networks, as it is based on previous theoretical foundation of force directed attraction-repulsion algorithms [87][146]. Indeed, Force Atlas 2 is clustering complex networks by producing well-defined topological clusters. The overview of the entire clustering process, including testing cluster convergence with non-OSAS control patients and validation of SAS_{Score}, is presented in Figure 3.9. In addition to the layout algorithm, we used modularity-based network clustering [93], a method that was proven to be effective in network medicine [66][78].

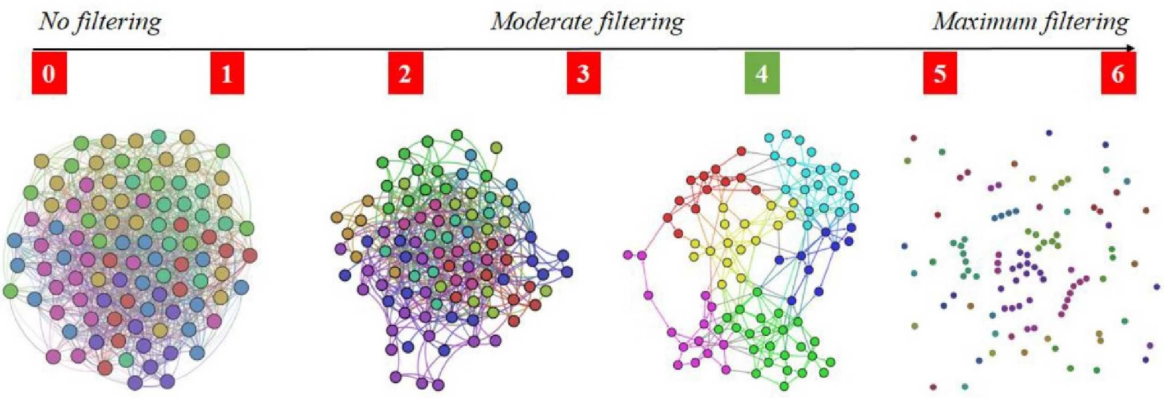

Figure 3.8: Apnea Patients Network (APN) edge filtering, by considering different definitions for $f_{RFC} = 1$, when we adopt the x-out-of-6 criteria ($x = 1, 2, 3, 4, 5$, and 6). The visual result indicate $x = 4$ as the best solution, because the edge density is convenient for rendering topological clusters with energy model layouts. Other values for x will generate too dense or too sparse networks.

3.2.2 Network Analysis for OSAS Patients Phenotype Definition

APN Analysis

The APN representation resulted from our clustering methodology is presented in Figure 3.10, where the distinct colors correspond to distinct modularity classes, and the well-defined topological clusters are explained accordingly. In Figure 3.10, we interpret the 8 topological clusters as distinct phenotypes, and provide the risk factors prevalence as percentages (L, Mi, Mo, Se)% for each such cluster/phenotype.

Non-OSAS Patients Network (NPN) analysis

Using the information from the 611 people non-OSAS database (NAD), we employ the same procedure as for the APN from Figure 3.10. The NAD represents the control population, consisting of people that are not diagnosed with OSAS. The result of applying our methodology on NAD patients is presented in Figure 3.11, where the colors correspond to distinct modularity classes; at the same time, topological communities rendered with the energy-model layout Force Atlas 2 are indicated and explained.

Upon visual inspection, Figure 3.11 suggests that in the non-OSAS control population there are more patterns of risk factors association, which leads to a number of 12 topological clusters and modularity classes that are not correlated with OSAS or AHI risk groups. As such, according to our network-based methodology, it occurs that the 6 considered risk factors consistently converge only for the individuals with OSAS.

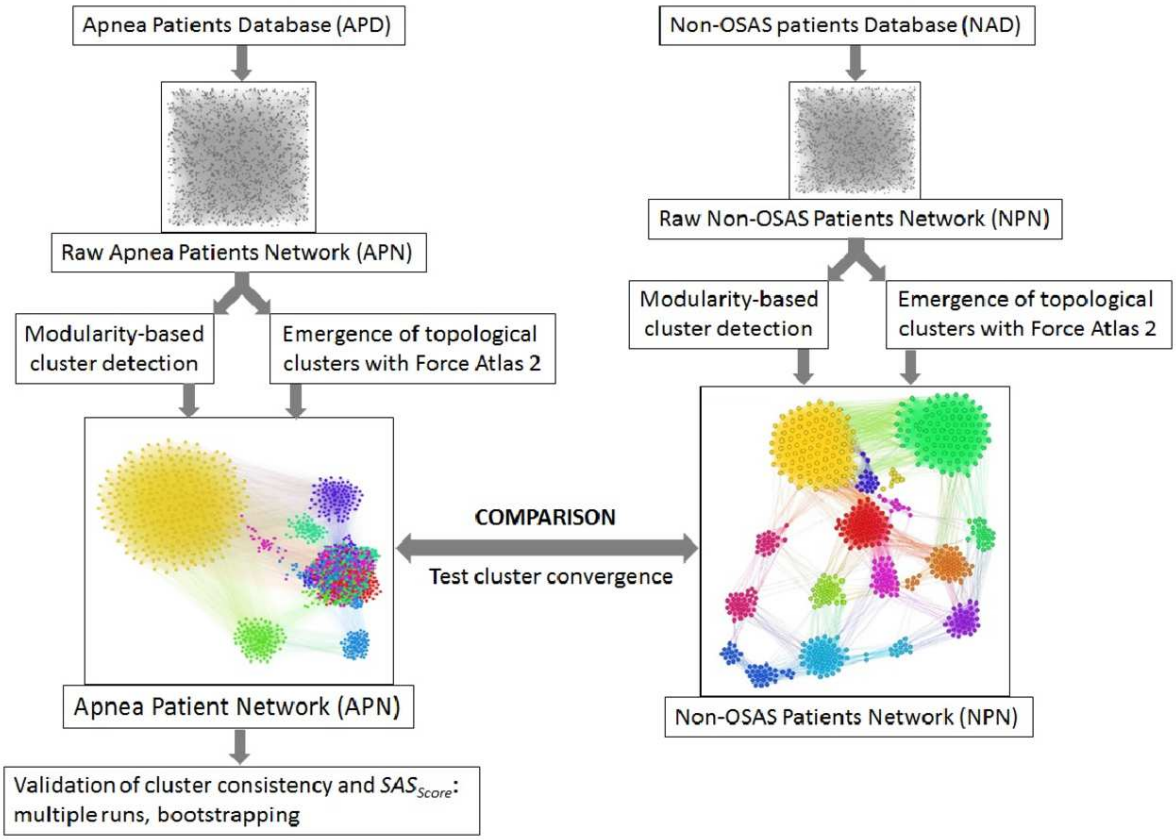

Figure 3.9: Overview of the proposed dual clustering methodology (modularity classes represented with distinct colors and topological clusters emerged from running the force directed layout Force Atlas 2), along with testing cluster formation against non-OSAS control patients (i.e. showing that risk factors converge differently for OSAS and non-OSAS patients), cluster consistency and SAS_{Score} validation.

Description of Phenotypes

In order to have a clear characterization of our rendered phenotypes, we are tracking the OSAS comorbidities (as recorded in the APD) within the APN. To this end, we consider the comorbidity types: *cardiovascular* (e.g. hypertension or stroke), *nutritional* (e.g. obesity or diabetes), and *respiratory*-related (e.g. COPD or asthma). Figure 3.12 presents the highlighted comorbidities within the APN, by using distinct colors for comorbidity types that appear individually, as well as for comorbidity type overlaps (cardiovascular + nutritional, cardiovascular + respiratory, nutritional + respiratory, cardiovascular + nutritional + respiratory), and patients without known comorbidities.

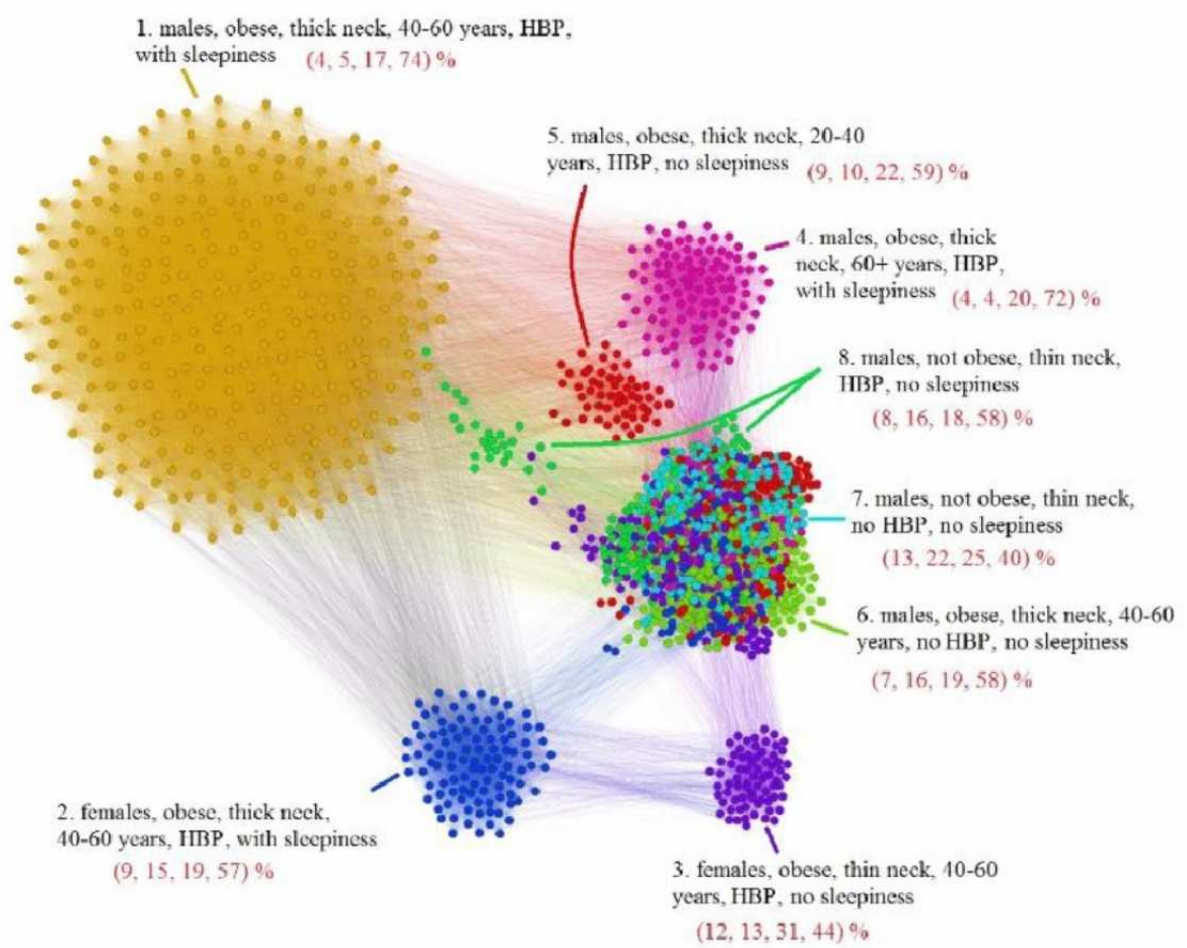

Figure 3.10: Apnea Patients Network (APN) obtained with data from the Apnea Patients Database (APD), according to the risk factor compatibility relationship, using our dual network clustering methodology (i.e. modularity classes and energy-model layouts). The assigned colors correspond to modularity classes, and the 8 topological clusters are indicated. For each topological cluster, statistics are provided in red (as percentages) for all AHI risk groups: low, mild, moderate, and severe, using the format (L, Mi, Mo, Se)% (e.g. in Cluster 2 the patients are distributed on risk groups as follows: 9% L, 15% Mi, 19% Mo, 57% Se). The risk group classification is made with AHI values that are obtained by actually performing polysomnography (PSG) and polygraphy.

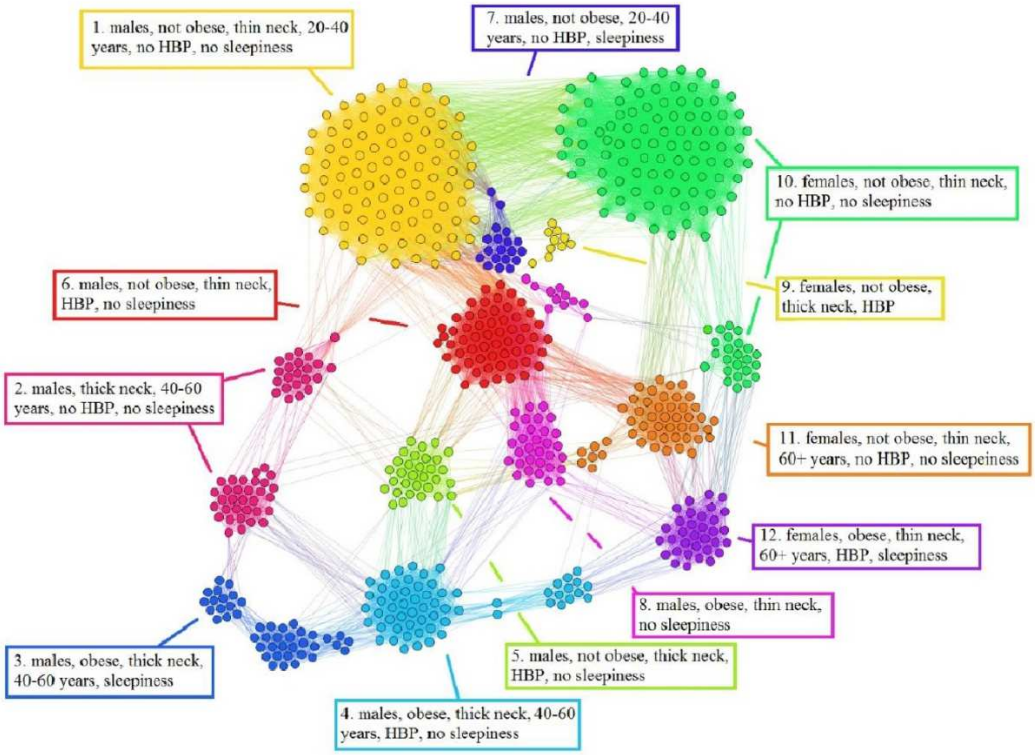

Figure 3.11: Non-OSAS Patients Network (NPN), obtained with data from the non-OSAS database (NAD) control population database. The colors correspond to 12 modularity classes. Also, the energy-model layout Force Atlas 2 generates 12 topological clusters, which are described in terms of risk factors and their correspondence with the modularity classes.

In light of comorbidity and AHI risk groups statistics provided in Table 3.5 for each cluster resulted from our network analysis (as illustrated in Figures 3.10 and 3.12), we characterize the phenotypes as follows:

- **Phenotype 1:** Mostly patients within the *Se* AHI risk group, which are generally obese males with thick neck, high blood pressure, sleepiness, and age between 40 and 60 years. For a large majority of these patients, all comorbidity types overlap.
- **Phenotype 2:** The large majority of these patients have *Mo* and *Se* apnea forms; they are obese females with thick neck, high blood pressure, sleepiness, age between 40 and 60 years. In this phenotype there are no patients with only respiratory comorbidities and only few of them have single nutritional comorbidities.
- **Phenotype 3:** The patients have mostly *Mo* and *Se* apnea, but there are less *Se* forms in comparison with other phenotypes; they are obese females with thin neck, high blood pressure, no sleepiness, and age between 40 and 60 years. This phenotype does not contain patients with only respiratory comorbidities.
- **Phenotype 4:** Mostly *Se* patients; however, there is a significant number of *Mo* individuals, which are generally obese males with thick neck, high blood pressure,

sleepiness, and over 60 years. In this phenotype only a few patients have the single respiratory comorbidities type.

- **Phenotype 5:** Mostly *Se*, *Mo*, and *Mi* patients, which are obese young males with thick neck, high blood pressure, no sleepiness, and age between 20 and 40. In this phenotype, almost all patients have nutritional comorbidities or comorbidity overlaps that include the nutritional type.

- **Phenotype 6:** Consists of mostly *Mo* and *Se* apnea; the patients within this phenotype are generally obese males with thick neck, no high blood pressure, no sleepiness, and middle aged (40-60 years old). Their comorbidities are mostly nutritional-related (either single nutritional comorbidity or an association of comorbidities that contains the nutritional type).

- **Phenotype 7:** Patients with mostly *Mo* and *Se* apnea, but with less *Se* forms in comparison with other phenotypes; this phenotypes patients are generally males of all ages with thin neck, no high blood pressure, and no sleepiness. The majority of these patients have no comorbidities; however, those who have a comorbidity tend to have respiratory-related problems.

- **Phenotype 8:** Patients mostly within *Se*, *Mo*, and *Mi* AHI risk groups; they are males from all age groups with thin neck, no sleepiness, but with high blood pressure. These patients tend to have a single cardiovascular comorbidity type or an association of comorbidities that include the cardiovascular type.

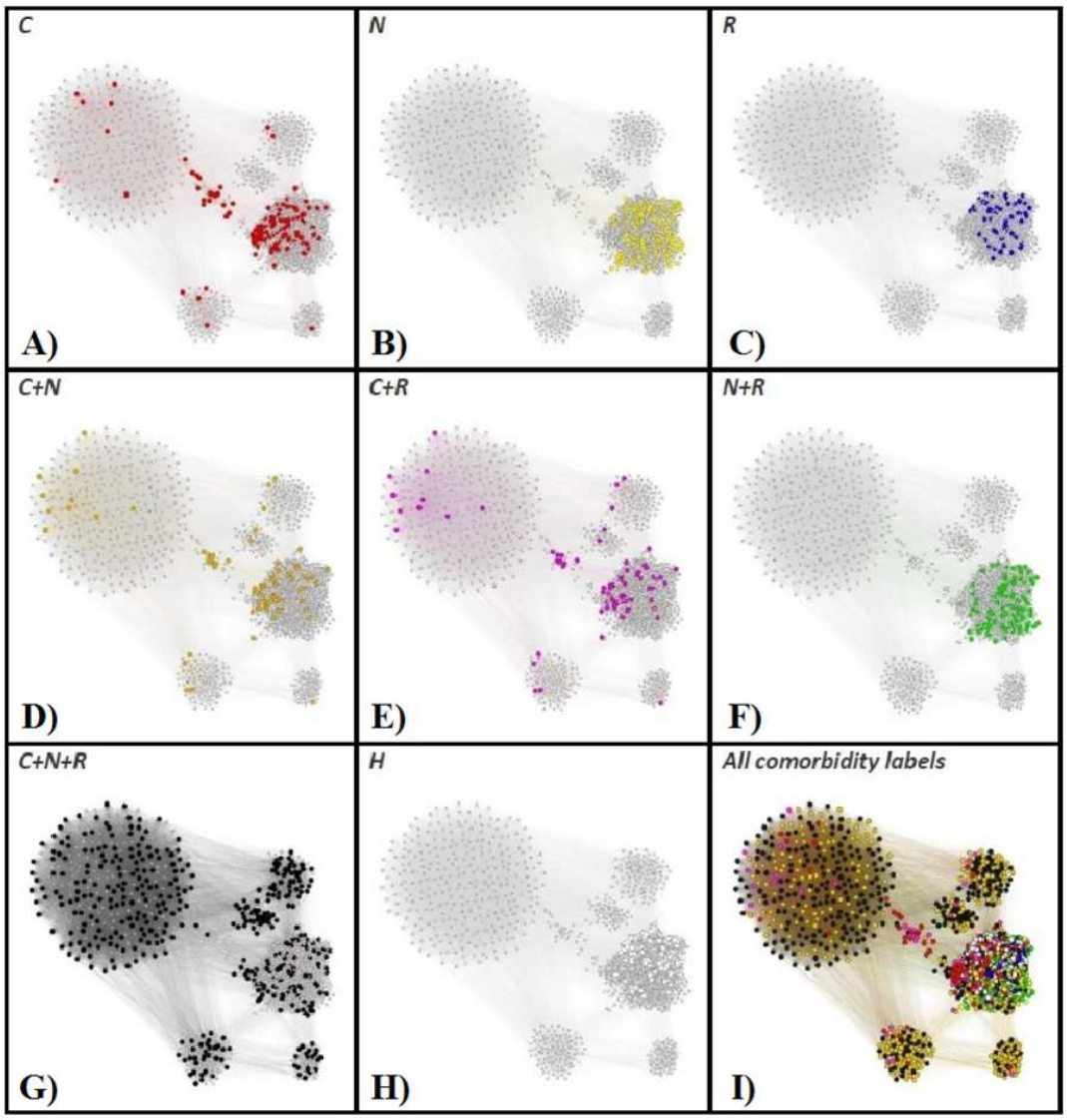

Figure 3.12: The Apnea Patients Network (APN), with highlighted individual comorbidities and associations of comorbidities. The red nodes correspond to patients that have only cardiovascular comorbidities (C in panel A), the yellow nodes to only nutrition-related (N in panel B), and blue nodes represent OSAS patients with only respiratory comorbidities (R in panel C). Patients with overlapping comorbidity types are represented within the APN as follows: orange nodes correspond to cardiovascular plus nutritional comorbidities ($C + N$ in panel D), purple nodes to cardiovascular plus respiratory comorbidities ($C + R$ in panel E), green nodes to nutritional plus respiratory comorbidities ($N + R$ in panel F), and black nodes to the superposition of cardiovascular, nutritional and respiratory comorbidities ($C + N + R$, panel G). The OSAS patients without known comorbidities are highlighted in the APN as white nodes (H in panel H). We also provide the APN where all nodes are labeled according to their comorbidity or comorbidities overlapping in panel I.

OSAS risk prediction with SAS_{Score}

Normally, classifying any new patient in one of the phenotypes can be performed by adding the new patient to the APN and then running the modularity class and force-directed layout algorithms in Gephi one more time. However, in clinical practice, oftentimes physicians are unable to perform these rather complex and time consuming computational steps (i.e. manipulation of databases and managing Gephi plugins), because of deadline commandments and resource constraints.

In order to deal with this problem, we propose a simplified solution for classifying *de novo* patients, using a computer algorithm that is implemented as a web-based application.

As such, for each such newly arrived patient, we measure the relevant data, namely age A (expressed in years), gender G (male/female), body mass index BMI (rational number), neck circumference NC (centimetres), systolic blood pressure SBP and diastolic blood pressure DBP (integer numbers), Epworth sleepiness score ESS (integer number); these data are introduced and recorded in our computer application using a visual interface questionnaire. The measured data is used to further compute the parameters that are employed at classifying patients in one of the 8 network clusters (corresponding to the 8 described phenotypes): high blood pressure HBP, thick neck TN, obesity Ob, age group AG, sleepiness SLP, composite risk factor CRF, reduced composite risk factor $RCRF$. The parameter computation is performed automatically by our computer software, according to Equations 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, and 3.7:

$$HBP = \begin{cases} 1 & \text{if} \quad SBP \geq 140 \text{ and } DBP \geq 90 \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

$$TN = \begin{cases} 1 & \text{if} \quad (G = \text{female and } NC \geq 40) \text{ or } (G = \text{male and } NC \geq 43) \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

$$Ob = \begin{cases} 1 & \text{if} \quad BMI \geq 30 \\ 0 & otherwise \end{cases} \tag{3.3}$$

$$AG = \begin{cases} 1 & \text{if} \quad A < 20 \\ 2 & \text{if} \quad A \in [20, 40) \\ 3 & \text{if} \quad A \in [40, 60) \\ 3 & \text{if} \quad A \geq 60 \end{cases} \tag{3.4}$$

$$SLP = \begin{cases} 1 & \text{if} \quad ESS \geq 11 \\ 0 & \text{otherwise} \end{cases} \tag{3.5}$$

$$CRF = HBP + Ob + TN \tag{3.6}$$

$$RCRF = Ob + TN \tag{3.7}$$

Our computer application employs the simplified algorithm for classifying new patients in one of the 8 phenotypes, as presented in Figure 3.13 under the form of a flowchart. The cluster classification algorithm from Figure 3.13 simply follows the de-

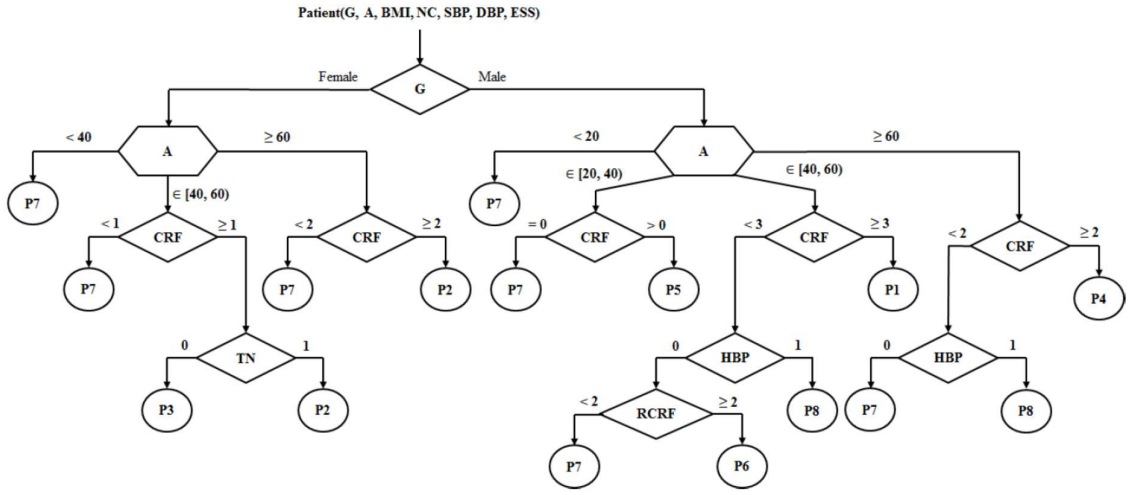

Figure 3.13: The algorithm for classifying new patients in one of the 8 phenotypes, denoted as P_1 to P_8. Decisions are made according to the measured parameters (gender-G, age-A, body mass index-BMI, systolic blood pressure-SBP, diastolic blood pressure-DBP, Epworth sleepiness score-ESS), as well as according to the computed parameters (high blood pressure-HBP, thick neck-TN, composite risk factor-CRF, reduced composite risk factor-$RCRF$).

scription in terms of dominant (i.e. majority) anthropometric characteristics, as presented in Table 3.6.

The diagram in Figure 3.13 is only used for classifying a patient into one of the 8 phenotypes, while our SAS_{Score} formula is used afterwards to compute the quantitative risk of OSAS. For this purpose, our score uses the mean parameter values from each corresponding phenotype. Therefore, our simplified OSAS risk assessment procedure works in the following order:

1. Anthropometric measurements on new patient.

2. Classification of patient in one of 8 phenotypes (using the algorithm in Figure 3.13).

3. Refer to cluster normalization averages from Table 3.7.

4. Computation of SAS_{Score} using Equation 3.8.

To implement the procedure from above, our computer application automatically computes the average parameter values for each cluster in our APN from Figure 3.10 (see Table 3.7). Therefore, having the six recorded parameters for any new patient that has to be evaluated, and the average values $BMI_a^{Cluster}$, $NC_a^{Cluster}$, $SBP_a^{Cluster}$, $DBP_a^{Cluster}$, $ESS_a^{Cluster}$ with $Cluster \in \{1, 2, \ldots, 8\}$ as presented in Table 4, SAS_{Score} is automatically computed by our computer program according to Equation 3.8, then

correspondingly displayed by the web-based interface.

$$SAS_{Score} = \frac{BMI}{BMI_a^{Cluster}} + \frac{NC}{NC_a^{Cluster}} + \frac{1}{2}\left(\frac{SBP}{SBP_a^{Cluster}} + \frac{DBP}{DBP_a^{Cluster}}\right) + \frac{ESS}{ESS_a^{Cluster}}. \tag{3.8}$$

SAS_{Score} values are ≥ 1 and generally < 7; this range of values can further serve at classifying patients according to a risk severity level that corresponds to the actual OSAS severity (i.e. as defined by the AHI risk groups: L, Mi, Mo, Se). To this end, we analyse the correlation between SAS_{Score} and the actual AHI values for all patients within the APN, by approximating the predicted AHI as linear dependence: $AHI_p = 11.227 \cdot SAS_{Score} + 1.9387$. Accordingly, SAS_{Score} can predict the OSAS severity, due to its correspondence with AHI risk groups:

$$OSAS_{Severity} = \begin{cases} L & \text{if } SAS_{Score} < 3 \\ Mi & \text{if } 3 \leq SAS_{Score} < 3.5 \\ Mo & \text{if } 3.5 \leq SAS_{Score} < 4 \\ Se & \text{if } 4 \leq SAS_{Score} \end{cases} \tag{3.9}$$

In order to evaluate SAS_{Score}, we compute the SAS_{Score} on de novo patients from our test database TD. The evaluation results indicate that sensitivity and specificity values for SAS_{Score} applied on TD patients are 0.887 and 0.381 respectively. Prior to our network analysis, we also computed STOP-BANG on each of the TD patients and measured their actual AHI; therefore we are able to compute the STOP-BANG sensitivity and specificity on TD as 0.875 and 0.179 respectively. Taken together, these results obtained on the Test Database, consisting of de novo patients only, suggest that our SAS_{Score} significantly outperforms STOP-BANG in terms of specificity (i.e. it is 2.13 times better), and remains slightly better than STOP-BANG in terms of sensitivity (i.e. an increase of 1.37%). We consider these results as relevant, as long as the distribution of AHI in the TD is notably different from the distribution of AHI in APD (see Figure 3.7 on the top panel).

We also create a test patient network (TPN), similar to building the APN, and then apply our dual clustering methodology. The result is presented in Figure 3.14; upon visual inspection it can be noticed that the clusters emerged in TPN are similar to the clusters from Figure 3.10s APN, even if the number of patients is significantly smaller in TD; this result suggests that the association and convergence of risk factors in OSAS patients is indeed a non-random, consistent process.

3.2.3 Validation of OSAS Risk Prediction with SAS_{Score}

Clustering Consistency Validation

In this subsection, we verify that rendering the clusters in Figure 3.10 is not mere serendipity, and it is not induced by some fortunate heterogeneity of patients. To this end, we perform random shuffling and bootstrapping test investigations.

Because our clustering methodology starts with a random state, namely it starts with the raw network where nodes are randomly placed and the links have corresponding lengths (see Figure 3.9), our first shuffling test consists of running the procedure in

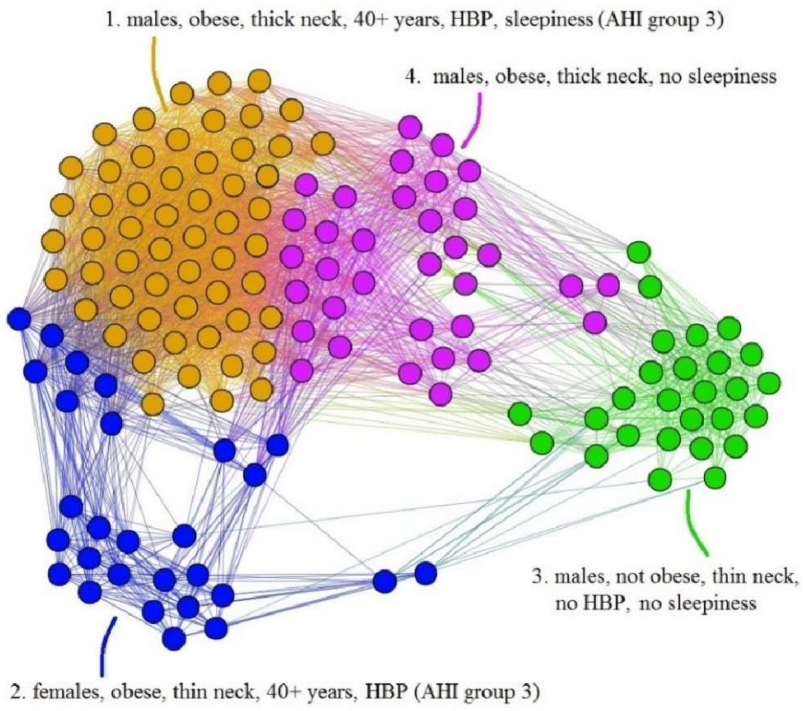

Figure 3.14: Representation of Test Database-TD patients, according to the risk factor compatibility, using the proposed network-based methodology. Although the number of patients is significantly smaller in comparison with the primary Apnea Patients Network-APN, the visual inspection reveals that the emerged clusters are similar to clusters from Figure 3.10 (APN).

Figure 3.9 many times in order to see if we get statistically consistent results. Therefore, we run our dual clustering procedure 100 times on the same APD, and then measure the distribution of anthropometric values for each phenotype. The result of our random shuffling is given in Table 3.8 which presents average anthropometric measurements with standard deviations (expressed as percentages) for the each APN cluster/phenotype after 100 runs. Indeed, the deviations from the average values are very small, emphasizing the consistency of our clustering procedure in Figure 3.9.

Our second test approach entails generating test APDs from the original patient dataset, in order to perform bootstrapping. To do so, we generate 10 new APD datasets with the same number of patients as the original APD, by randomly selecting patients from the original APD database. Therefore, in the test APDs, some of the original patients may be missing, while others may be present two or more times. Next, we apply the same clustering methodology from Figure 3.9 and find that the same phenotypes emerge.

The characteristics of the original APN phenotypes from Figure 3.10 are provided in Table 3.9; Table 3.10 shows the averaged characteristics, for each cluster, over the randomized 10 APNs obtained by bootstrapping. In Table 3.10 distinct character types suggest how close (bold and normal characters) or how far (grey italics) are the pheno-

type characteristics resulted from bootstrapping from the original APN values.

The values from Table 3.10 show that the dominant characteristics of each cluster in the test APNs are very similar to the cluster characteristics from the original APN. As a result, even if we create new APDs with corresponding APNs by shuffling the patients from our dataset, the bootstrapping procedure yields the same phenotypes. The bootstrapping procedure reveals that some phenotypes emerge as more stable when randomized (e.g. clusters 1, 3, 4, 5), with small offset compared to the original APN. The lack of convergence over multiple randomizations could be interpreted as a lack of representativeness as an OSAS phenotype. As such, clusters 2 and 6 are slightly less representative, and clusters 7 and 8 are notably variable, suggesting that the last two phenotypes may not be so well characterized, due to the reduced amount of patient data.

As our APN will grow over time, the less convergent clusters 7 and 8 might become more representative. Indeed, the fact that even in the original APN from Figure 3.10 clusters 7 and 8 present significant overlapping and that their topological segregation from other clusters is somehow fuzzy confirms the conclusion of our bootstrapping investigation.

Validating the Cluster Classifying Algorithm

In order to validate the simplified phenotype/cluster classifying algorithm from Figure 3.13, we propose the following methodology: we classify the patients in the TD, according to the algorithm from Figure 3.13. This procedure assigns each patient from the TD a matching cluster between 1-8.

To confirm the consistency of the assignment, we also add the TD patients in the APN resulting a new APN* (consisting of patients from both APD and TD). We run the force-directed cluster segregation algorithm Force Atlas 2 in conjunction with modularity classes on the APN* using Gephi, and then we compare the number of matches between phenotypes assigned by our algorithm in Figure 3.13 – on one hand –, and Force Atlas 2 in Gephi – on the other hand. The results are that, out of 231 TD patients, 212 (91.77%) are assigned to the correct phenotype with the algorithm in Figure 3.13, thus proving that we can rely on the algorithm for OSAS diagnosis with SAS_{Score} which is described in Section *OSAS risk prediction with* SAS_{Score}.

Comparison with Other Approaches

We compare our SAS_{Score} prediction method against other possible big data approaches. To this end, we apply machine learning on the entire patient population from the APD. As such, we use Wekas [100] classifying and clustering features, in order to obtain an alternative prediction model for our OSAS patients. We rely on the simple K-means [166] and farthest first [64] classifiers; these are popular algorithms integrated in Weka, which were also used in classifying lung cancer [64].

First, we run the simple K-means algorithm with the same attributes (anthropometric measurements) as used in our SAS_{Score}. The algorithm finds a diagnostic model based only on two clusters:

- **Cluster 1** (436 patients): males, age < 50 yrs, low obesity (BMI < 32), thin neck (< 43cm), no HBP;
- **Cluster 2** (935 patients): males, age > 50 yrs, obese (BMI > 32), thick neck ($>$ 43cm), with HBP.

In conclusion, the K-means algorithm assigns any patient into extreme phenotypes: either a severe OSAS phenotype (**Cluster 2**), or a marginally healthy phenotype (**Cluster 1**).

By running the farthest first algorithm, we also obtain a model which assigns patients in one of two possible phenotypes. The first such phenotype (consisting of 983 patients) characterizes all patients with severe OSAS, and the second (388 patients) contains patients with very severe OSAS. Unfortunately, the output of Weka indicates this classifier as having very low performance, because 49.23% of the nodes are assigned incorrectly, meaning that half of the population is not well described by any of the two clusters.

Consequently, we find that machine learning is not well suited for this kind of data analysis, possibly due to the relatively small number of patients in the APD. Moreover, machine learning seems not to able to classify patients on more than 2 OSAS severity levels.

3.2.4 Discussion

The proposed method is not the first to cluster apnea patients [110][198][206], but to the best of our knowledge it is the first network-based approach used for clustering apnea patients. Another important feature is that our network-based methodology employs only easy-to-measure, objective clustering parameters (AHI is used only for phenotype evaluation). This way, our clustering methodology emphasizes the high complexity of OSAS phenotypes, from typical (cluster 1 yellow) to the less obvious ones (clusters 6, 7, 8).

When defining the 8 apnea phenotypes, besides the force-directed layout, we also use modularity class clustering. In Figure 3.10, the phenotypes based on modularity classes are generally consistent with the topological clusters resulted from applying the Force Atlas 2 layout. However, the visual inspection of Figure 3.10 reveals that phenotypes 5, 6, 7, and 8 tend to spatially overlap; this tendency is much stronger for phenotypes 6, 7, and 8. Such a tendency for overlapping phenotypes that characterize patients with generally mild and moderate OSAS is also suggested by Joosten *et al.* [110]. This observation may indicate that these phenotypes are interrelated and generally hard to distinguish in clinical practice. Still, some of the nodes in these clusters (e.g. cluster 8) have a clear tendency towards separation from the overlapping; this indicates that we probably need more patients/nodes, in order to completely segregate Cluster 8. Nonetheless, we preferred to use the distinct modularity classes in conjunction with the topological clusters because they bring more information, i.e. more detail which can be useful for medical analysis.

From a medical standpoint, we note that our dual clustering method renders distinct male and female clusters; this observation is consistent with the state of the art medical

literature which holds gender as a very important predictor of OSAS. For instance, in a 2009-2013 study on 272,705 patients from North America, referred for home sleep apnea testing, clinical OSAS features are found to be more common in males than females [32]. Other studies performed on 23,806 [90], and 1,010 [194] patients respectively, show clear differentiation between the two genders in terms of AHI distribution and severity.

In current practice, the commonly used score for predicting sleep apnea is STOP-BANG. In comparison with STOP-BANG, our SAS_{Score} significantly improves the prediction specificity (2.13 times better than STOP-BANG), and slightly improves its sensitivity. STOP-BANG already has high sensitivity because it was especially designed for perioperative patients, where it is essential to identify all potential risks associated with anaesthesia (including OSAS). To further emphasize the higher specificity of SAS_{Score}, we mention that by using our score, all non-OSAS subjects from NAD were correctly predicted as L (low risk). Also, as opposed to previous scores, SAS_{Score} yields a much more accurate 4-level classification; indeed, for the available training databases, even machine learning approaches cannot render more than a 2-level classification. Moreover, as opposed to STOP-BANG (a fixed questionnaire that cannot be adjusted to specific patients characteristics), SAS_{Score} represents an adaptive methodology. Therefore, as the database grows, better sensitivity and specificity are expected. However, the procedure which leads to rendering SAS_{Score}, as described in section *OSAS risk prediction with* SAS_{Score} represents a simplified application of our patient clustering/phenotyping method; this method can also be applied in offline conditions, which makes it amenable to clinical practice. Otherwise, if the clinician uses the more sophisticated Gephi-based approach, by adding each new patient to the APN and subsequently running the force-directed layout in conjunction with modularity classes, then the classifying algorithm from Figure 3.13 becomes useless.

Although the network-based method was applied on patients from a given geographical area, having mostly Caucasian anthropometric characteristics, it can be directly employed for other targeted populations. As such, the network analysis will render new, specific cluster average values (such as $BMI_a^{Cluster}$, $NC_a^{Cluster}$, $SBP_a^{Cluster}$, $DBP_a^{Cluster}$, $ESS_a^{Cluster}$). Subsequently, SAS_{Score} values that are specific to the targeted population, can be rendered with the SAS_{Score} equation. Eventually, due to its higher specificity, the SAS_{Score} can be integrated into a large area apnea monitoring procedure, which aims at specifically discovering typical severe cases (easy to investigate with portable devices), as well as isolating borderline cases (that might be investigated in a sleep lab with a polysomnography device). This way, efficient personalized patient processing can be achieved by making use of prioritization according to the predicted severity level. For instance, this method can be a useful tool for sleep apnea screening in large population categories, such as professional drivers since, at the European level, the new 2014/85/EU directive regarding professional drivers is recommended from January 2016 [1]. In this context, our website `sasscore.appspot.com` is a good example of a large-area, accessible OSAS risk prediction tool. Indeed, SAS_{Score} can be conveniently computed in both clinical and population-monitoring practices, due to the fact that it is implemented as

[1] Commission Directive COMMISSION DIRECTIVE 2014/85/EU of 1 July 2014 amending Directive 2006/126/EC – European Parliament and the Council on driving licences.

easy-to-use smartphone and web-based applications [2] www.pneumoresearch.ro). To this end, processing data and obtaining the prediction score requires less than 1 minute per individual.

[2] https://play.google.com/store/apps/details?id=aerscore.topindustries.aerscore&hl=en

Ph.	Description	Comorbidity types and associations [%]								AHI risk groups [%]			
		C	N	R	$C+N$	$C+R$	$N+R$	$C+N+R$	H	L	Mi	Mo	Se
1	Mostly Se with $C+N+R$ comorbidities	2.38	-	-	32.65	3.4	-	**61.57**	-	4	5	17	74
2	Mostly Mo and Se, thick neck females, no R, few N	6.49	0.65	-	**61.04**	3.9	0.65	27.27	-	9	15	19	57
3	Mostly Mo, Se, thin neck females, no R	13.66	10.24	-	**48.29**	1.95	2.93	22.93	-	12	13	31	44
4	Mostly Mo and Se, elderly males with thick neck, few R	6.52	3.62	-	**36.96**	5.8	3.62	**44.48**	-	4	4	20	72
5	Mostly Mi, Mo, Se, obese young males, mostly N	2.47	**18.52**	1.23	19.75	3.09	**20.99**	32.72	1.23	9	10	22	59
6	Mostly Mo, Se, 40-60 yrs. obese males, no HBP, mostly N	2.58	**33.55**	5.81	10.97	1.29	**31.61**	7.74	6.45	7	16	19	58
7	Less Se, non-obese, thin neck, no HBP males, mostly H, some R	8.84	0.68	27.21	0.68	4.76	-	-	**57.83**	13	22	25	40
8	Less Se, non-obese, thin neck, HBP males, mostly C	**38.79**	2.59	-	9.48	**28.45**	2.59	18.1	-	8	16	18	58

Table 3.5: Description of the eight relevant Apnea Patients Network (APN) phenotypes. The phenotypes (Ph) are listed with a short description in terms of most or least predominant AHI risk groups (low risk-L, mild-Mi, moderate-Mo, and severe-Se), significant combinations of the 6 objective parameters, and most/least predominant comorbidity types (cardiovascular-C, nutritional-N, respiratory-R, and without comorbidities-H) or comorbidity types overlaps ($C+N$ = cardiovascular + nutritional, $C+R$ = cardiovascular + respiratory, $N+R$ = nutritional + respiratory, $C+N+R$ = cardiovascular + nutritional + respiratory). For each phenotype, we provide the corresponding percentages for comorbidity types and comorbidity type associations, as well as the percentage of patients pertaining to one of the AHI risk groups; the boldface entries correspond to representative values, in terms of simple majority. In phenotype descriptions, HBP stands for high blood pressure.

Cluster	Size [no.]	G	Ob	TN	AG	HBP	SLP
1	294	M	1	1	3	1	–
2	116	F	1	1	3	1	–
3	178	F	1	0	∼3	1	∼0
4	208	M	1	∼1	4	1	–
5	76	M	1	1	2	1	–
6	218	M	1	1	–	0	–
7	219	M	∼0	0	–	0	∼0
8	62	M	0	–	3	1	–

Table 3.6: Dominant anthropometric characteristics of each phenotype. A dash entry (–) means that the metric is not statistically relevant for the corresponding cluster; a tilde (∼") means partially relevant (60-80% have the indicated characteristic), and no special sign means highly relevant (>80% of patients in the cluster have the indicated characteristic).

	Cluster							
	1	2	3	4	5	6	7	8
$BMI_a^{Cluster}$	36.83	37.54	31.88	33.75	34.99	32.93	22.74	28.67
$NC_a^{Cluster}$	47.66	43.9	35.91	46.79	43.17	43.66	33.39	37.36
$SBP_a^{Cluster}$	143.96	145.32	139.48	140.56	133.74	125.46	119.66	134.55
$DBP_a^{Cluster}$	89.46	87.61	85.80	84.28	84.43	80.75	73.16	84.41
$ESS_a^{Cluster}$	11.49	10.77	7.74	10.10	8.58	9.50	6.28	8.90

Table 3.7: Average values for the relevant parameters (body mass index-BMI, neck circumference-NC, systolic blood pressure-SBP, diastolic blood pressure-DBP, Epworth sleepiness score-ESS), which are computed for each of the 8 clusters in Figure 3.10.

Cluster	A	BMI	NC	HBP	ESS	AHI
1	51.89 ± 0	36.83 ± 0	47.66 ± 0	$100\% \pm 0$	11.49 ± 0	50.75 ± 0
2	54.11 ± 3.51	37.73 ± 0.86	44.13 ± 0.47	$92\% \pm 5.22$	10.59 ± 0.99	38.3 ± 0.49
3	56.14 ± 1.87	31.88 ± 1.27	35.87 ± 1.13	$87\% \pm 5.77$	7.8 ± 2.45	30.92 ± 3.09
4	66.47 ± 0.28	33.84 ± 1.55	44.15 ± 2.46	$90\% \pm 6.78$	10.02 ± 1.64	42.99 ± 3.3
5	34.54 ± 1.34	35.09 ± 1.53	44.52 ± 2.39	$51\% \pm 5.27$	8.94 ± 3.26	49.66 ± 4.94
6	51.23 ± 2.45	32.94 ± 1.66	45.23 ± 3.08	$0\% \pm 0$	9.85 ± 2.82	44 ± 4
7	43.59 ± 2.62	24.6 ± 5.77	33.01 ± 1.35	$7\% \pm 10.57$	6.46 ± 2.51	25.93 ± 2.9
8	55.92 ± 4.47	27 ± 3.55	39.96 ± 4.51	$96\% \pm 10.42$	8.75 ± 3.64	34.29 ± 2.66

Table 3.8: Average anthropometric measurements with standard deviations (expressed as percentages [%]) for the anthropometric parameters (Age A, BMI, Neck Circumference NC, High Blood Pressure HBP, Epworth Sleepiness Score ESS, Apnea-Hypopnea Index AHI) in the APN clusters/phenotypes after 100 runs (for instance, the average age in cluster 2 is 54.11 years with a standard deviation of 3.51%).

Cluster	G	Ob	TN	AG	HBP	SLP
1	M (100%)	1 (100%)	1 (100%)	3 (100%)	1 (100%)	1 (61%)
2	F (100%)	1 (96%)	1 (100%)	3 (92%)	1 (100%)	1 (56%)
3	F (80%)	1 (81%)	0 (100%)	3 (69%)	1 (96%)	0 (70%)
4	M (100%)	1 (82%)	1 (77%)	4 (100%)	1 (99%)	1 (52%)
5	M (100%)	1 (96%)	1 (83%)	2 (100%)	1 (100%)	0 (64%)
6	M (88%)	1 (87%)	1 (100%)	3 (57%)	0 (100%)	0 (57%)
7	M (67%)	0 (61%)	0 (99%)	3 (50%)	0 (100%)	0 (77%)
8	M (98%)	0 (100%)	0 (55%)	3 (84%)	1 (100%)	0 (58%)

Table 3.9: Defining phenotype characteristics given as majority values with corresponding percentages, as measured in the original APN communities.

Cluster	G	Ob	TN	AG	HBP	SLP
1	**M (100%)**	**1 (100%)**	**1 (100%)**	**3 (100%)**	**1 (100%)**	**1 (60%)**
2	**F (100%)**	**1 (96%)**	**1 (100%)**	*3 (74%)*	**1 (100%)**	**1 (60%)**
3	F (90%)	**1 (81%)**	**0 (100%)**	**3 (63%)**	1 (87%)	**0 (69%)**
4	**M (95%)**	**1 (86%)**	1 (88%)	**4 (100%)**	**1 (97%)**	**1 (53%)**
5	**M (97%)**	**1 (97%)**	**1 (86%)**	**2 (100%)**	1 (79%)	**0 (62%)**
6	M (95%)	**1 (87%)**	**1 (96%)**	*3 (73%)*	**0 (100%)**	**0 (58%)**
7	*M (81%)*	**0 (64%)**	0 (89%)	*3 (75%)*	**0 (99%)**	0 (70%)
8	*M (82%)*	0 (84%)	*0 (80%)*	3 (68%)	**1 (100%)**	0 (71%)

Table 3.10: Characteristic features of phenotypes averaged over 10 randomized APNs. Majority average percentages for the dominant values are rounded to the nearest integer, while the colors represent how close these averages are to the measurements on the original APN from Table 6: bold entries correspond to very close matches (percentage difference $\leq$6%), normal character entries correspond to a good match (6% $<$ percentage difference $\leq$ 15%), while grey-italics table entries correspond to significant differences ($>$15%).

Part II

Future Research Developments

Chapter 4

On-Chip Communication Networks

4.1 Introduction

The electronic technology has already reached the crucial point where the complexity of resources is no longer easy to deal with. If, in the past, the main challenges in computer systems design were related to the resource scarcity, today we are rather confronted with the problems brought by their abundance [179]. Under these circumstances, besides computation, communication becomes a pivotal design aspect [138].

The Network-On-Chip paradigm creates incentives for coping with the communication needs of multi and many-core systems. However, the design and analysis of such a structure will have difficulties in performing optimization in a space spanned by network topology, routing function, and node communication resources (i.e. buffer space) [127]. Optimization becomes even more difficult, as network's behavior is characterized by a non-equilibrium [69], multi-fractal process [26][27]. Moreover, it seems that the workload, generated by software processes, has the same fractal characteristics [96]. As such, dealing with dynamic critical phenomena such as congestion becomes rather cumbersome if we are to rely on conventional solutions only [127].

In this context, the primary objective of this paper is to offer a framework for developing silicon chips that can accommodate 1000 and over cores. Therefore, besides the paramount importance of efficient communication, the topological solution proposed herein would have to cope with other demands, as – for instance – power consumption, heat dissipation, controllability or reliability.

A legitimate question would be: what is the topology that will provide means for dealing with the communication-related problems and – at the same time – cope with scalability demands of accommodating thousands of cores, while preserving a regular structure that is amenable to VLSI implementations? In order to deal with these challenges, computer engineering has found inspiration in Nature back to its dawn [179]; indeed, these solutions provide correct and robust answers for many complexity issues, including communication [143]. This paper proposes such a solution inspired by Nature's fractal shape design [126], for generating underlying topologies in NoC many-core (i.e. more than thousand cores) communication systems. It is based on the brain connection model from [137], which provides communication support for "a billion-agent society".

The rest of the paper is organized as follows: Section 4.2 presents the status quo in the

field of multi-core NoC and the perspectives of building thousand of cores communication systems, Section 4.3 describes the framework for fractal NoC architecture generation, and Section 4.4 details how resources can be allocated within the proposed architectural framework.

4.2 Background and Motivation

4.2.1 State-of-the-Art

State-of-the-art architectures in NoC-based system design mostly rely on regular mesh-like topologies. There are many advantages in using such a uniform structure: it simplifies the design complexity, the application mapping and the VLSI physical implementation details [127]. But, on the other hand, this view is tributary to a rather static approach, that does not take into account the extremely complex dynamic phenomena like phase transition, that occurs especially when the network size grows [62][69][197].

Of course, this problem was already anticipated and dealt with in a variety of ways. For instance, [181] proposes a nature-inspired, heuristic architectural synthesis framework. However, using this approach, unless some specific communication protocols, mapping and scheduling techniques are used, the design complexity will be high. Other techniques with topological implications are relying on long range links, which are inserted such that critical dynamic phenomena, such as congestion, are avoided [150].

Another approach is to use an appropriate theoretical framework that allows for characterizing the dynamical parameters of the regular network, so that the resource allocation can be performed accordingly [26][27][151]. The theoretical analysis for complex networks is generally based on 3 important parameters or measures [200]: average path length (L), clustering coefficient (C) and degree distribution (P). In short, the average path length is the average number of edges between any 2 nodes in the given topology. The clustering coefficient of a node is the fraction of the node's neighbor pairs that are neighbors between themselves. The average of clustering coefficients pertaining to all the nodes will give the network clustering coefficient, which captures the characteristics of a small-world network. The degree of a node l represents the number of its neighbors (connections). $P(l)$ is the degree distribution of all nodes in the network, and it has a big importance in characterizing the scale-free property of natural networks [143][200].

In a mesh topology, network's performance decreases with its size, due to the significant increase in the average path length. Of course, there is a plethora of other regular topologies (e.g. torus, butterfly, quadtree) that are considered for NoC architectures, but they normally share the same properties. Generally, regular topologies will have bigger L and smaller C in comparison with natural, self-organized networks [200].

A regular topology where the path length grows only logarithmically with the network size is the fat-tree structure (equivalent to the multi-level star topology) [97][119]. However, this type of network has a clustering coefficient of 0, thus generating a high demand for communication resources (bandwidth, virtual channels, buffers) at the higher structure levels, therefore seriously affecting the scalability.

4.2.2 Perspectives

In the context presented in Section 4.2.1, we present a topological framework that tries to satisfy most of the thousand-core NoC systems demands:

- scalability;
- ease in VLSI implementation;
- performance;
- power consumption and heat dissipation support.

In order to make such an assessment, we will use the complex network parameters such as the average path length, clustering coefficient and degree distribution [86]. These parameters will help in identifying the network topological patterns that characterize most of the real-world, natural and artificial networks: small world (SW) and scale free (SF) [143][200].

The topologies used for the regular Networks on Chip (NoC) systems [118][127][82] have several disadvantages. The mesh structure can accommodate a limitless number of cores, but the average path length has to suffer very much when the number of cores is increasing; also for any type of unmodified mesh the Clustering Coefficient is zero. On the other hand, the fat-tree/star structure provides a low average path length, but cannot accommodate many cores without overly complicating the communication structure at the top levels [97].

4.3 On-Chip Fractal Architectures

In order to overcome the disadvantages entailed by using conventional topologies, we propose a solution called *fractal architecture*, that can accommodate a limitless number of cores. Also, such a solution will have a Clustering Coefficient of around 0.5-0.6, depending on the structural version (this will be further discussed in this section), and it provides a very short average path length, i.e. proportional with $\log N$, where N is the total number of cores in the network. The shapes of the proposed topologies are fractal [126] because:

- exhibit self-similarity;
- have a fractal dimension that is bigger than the topological dimension;
- can be generated by using simple recursive rules;

Parallel computers exhibit self-similarity [149], but this is not sufficient to support the high communication complexity. The approach presented herein is further motivated by the evidence that fractal topology model captures the functionality of brain communication patterns ([137] Appendix 4: Brain Connections), which represent Nature's design to support communication complexity.

A noteworthy observation is that in all the fractal graph topologies presented herein, a node consists of a router with a connected processing element or core (see Figure4.1).

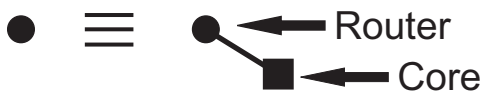

Figure 4.1: Basic notation: a node in the fractal topology.

4.3.1 Topology Generation

2D Topology

The first step in creating the 2D fractal architecture is to decide the number c_0 of cores of the *b*asic level. For the first level, denoted as *level* 1, we connect the $c_0 \geq 3$ (with $c_0 \in \mathbb{N}$) cores in a ring structure (which is a genuine 2D shape). Besides these c_0 cores, we will need an extra core with a special-purpose router that concentrates communication with the upper levels, which is distinctively represented in Figure 4.2 as a dot within a circle. For the sake of simplicity, unless a special discussion is required, we will refer to this central node as *router*. The router is connected to each of the c_0 nodes as presented in the leftmost part of Figure 4.2.

In order to create fractal architecture with two levels, we proceed the same way we did with the nodes in the first step (i.e. for level 1). Therefore, we need a number of c_0 level 1 structures (4 in the example considered in Figure4.2) plus another central router. Because this new router supports communication at level 2, we represent it as a dot within two concentric circles. We then connect the level 1 central routers in a ring and subsequently connect each of these routers to the level 2 router, see the middle part of Figure4.2. As a rule, in order to create level n we will need a number of c_0 level $n-1$ structures and a level n central router; we then connect the c_0 routers in the center of level $n-1$ structures in a ring, and subsequently connect each of these routers to the newly added, level n central router. The shapes presented in Figure4.2, meaning the exact angles and the relative positioning of the clusters, are not necessarily identical with those generated by the on-chip physical implementation.

The total number of nodes of the proposed fractal architecture is $\frac{{c_0}^{n+1}-1}{c_0-1}$ where c_0 is the number of cores of the basic level (level 1) and n is the number of levels. The maximum path length in such a fractal architecture is $2n$. As the number of levels grows logarithmically with the total number of cores $n \propto \log N$, we render that the average path length of the presented fractal structure is logarithmic with the total number of cores.

3D Topology

The generation of a 3D network topology follows the same principle as for the 2D network. It starts with a basis 3D topology, consisting of c_0 nodes. In order to maintain simplicity, we will restrict ourselves to using regular basic topologies like the regular tetrahedron ($c_0 = 4$), cube ($c_0 = 8$), regular dodecahedron ($c_0 = 12$). The common feature of all these considered 3D topologies is that they are symmetric, i.e. any node will have the same number of neighbor (directly linked) nodes. These basic topologies will also have a central router that concentrates the traffic that uses the upper levels.

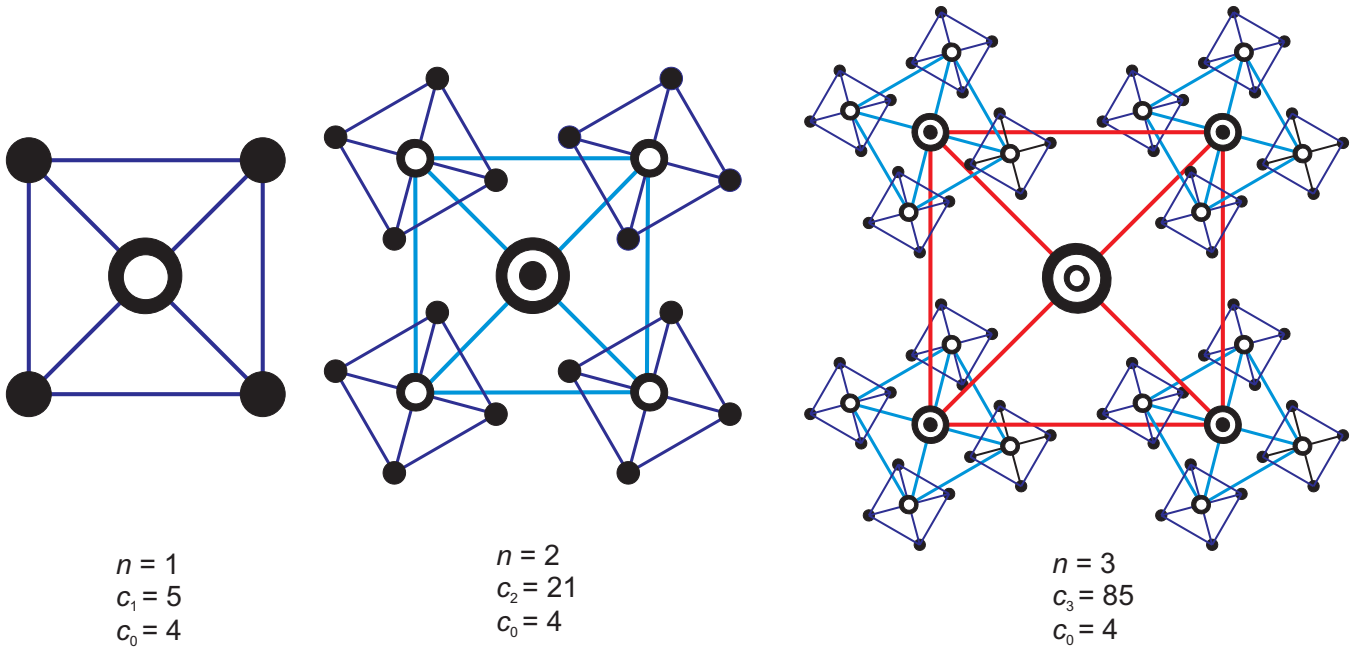

Figure 4.2: Fractal architecture creation procedure when $c_0 = 4$, for $n = 1$, 2 and 3 levels.

However, the most favorable case for the local traffic, in the sense that the shortest paths will not go through the central routers (therefore only non-local traffic will use the central routers), is the case of the regular tetrahedron.

The next level will be built by linking the c_0 basis levels, through their central routers, by employing the same topology as for the basis, and a central upper-level router which will be linked to every lower central router. The next level will be generated using the same recursive procedure, as presented in Figure4.3 for the case of the regular tetrahedron fractal. The number of nodes in the basis level, according to the level number is the same as for the 2D square-based fractal topology. The only difference is that the local connectivity is higher, meaning a shortest path of 1 between any 2 local nodes (i.e. nodes at the same fractal level which are connected at the same central node). In the case of a cube, the shortest path between any 2 local nodes is 3 if we exclude the central node. If the central node is used for local traffic (a situation that we want to avoid), then the shortest path length becomes 2.

Static Analysis

For the fractal structures presented in Sections 4.3.1 and 4.3.1, the static characteristics are presented under the form of average path length L, clustering coefficient C and degree distribution P.

For the p-D fractal structures ($p = 2, 3$) with n levels and with c_0 nodes in the basic

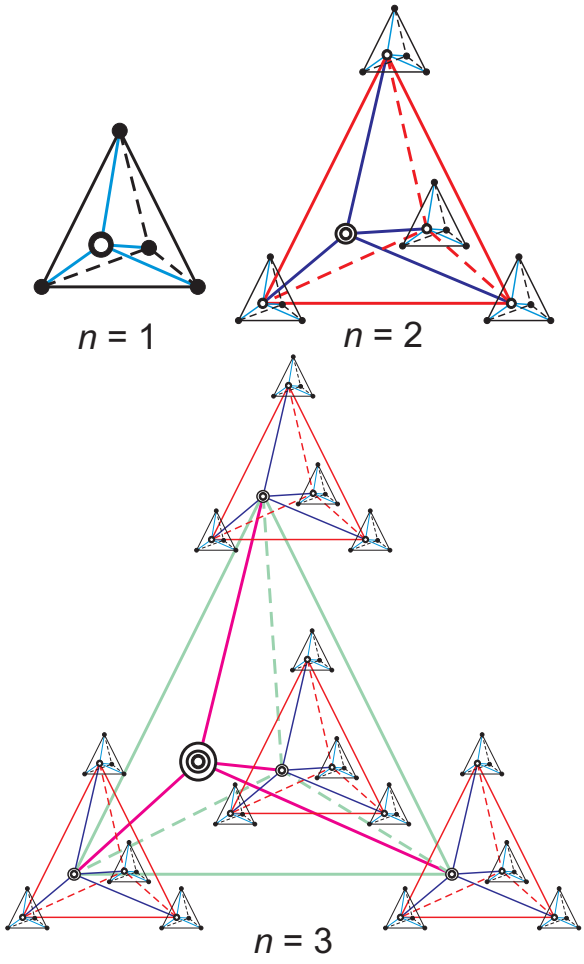

Figure 4.3: 3D fractal architecture creation procedure: the regular tetrahedron topology with 1, 2, and 3 levels.

ring, we can express the number of nodes at distance i from a given node, $M_f^{p\mathrm{D}}(i)$:

$$M_f^{p\mathrm{D}}(i) \simeq \begin{cases} pc_0^{k-1} & \text{for } i = 2k-1 \\ c_0^k - \sum_{j=0}^{2k-1} M_f^{p\mathrm{D}}(j) & \text{for } i = 2k \end{cases} \tag{4.1}$$

Having this approximation formula that excludes central nodes, we can define an upper bound for the average path length in all these fractal 2D topologies:

$$L_f^{p\mathrm{D}}(n) < \frac{1}{{c_0}^n} \sum_{i=1}^{2n} i \cdot M_f^{p\mathrm{D}}(i) \tag{4.2}$$

For the clustering coefficient it becomes more appropriate to express it for the 2D fractal structure as in Equation 4.3.

$$C_f^{2\mathrm{D}}(n) = \frac{\frac{2}{3}c_0^n(c_0-1) + \frac{2c_0}{c_0+3}\left(c_0^{n-1}-1\right) + 2}{c_0^{n+1}-1} \tag{4.3}$$

At the same time, generalizing an expression for 3D clustering coefficient would be impracticable, because performing generalization for all considered 3D shapes would be too difficult. Therefore, we present the clustering coefficient for the 3D regular tetrahedron fractal structure ($c_0 = 4$, Equation 4.4) and for the 3D cube fractal topology ($c_0 = 8$, Equation 4.5).

$$C_{f-4}^{3\mathrm{D}}(n) = \frac{\left(\frac{{c_0}^n}{2}+1\right)(c_0-1) + \frac{2c_0}{7}\left({c_0}^{n-1}-1\right)}{{c_0}^{n+1}-1} \tag{4.4}$$

$$C_{f-8}^{3\mathrm{D}}(n) = \frac{\frac{{c_0}^n}{2}(c_0-1) + \frac{1}{4}\left({c_0}^{n-1}-1\right) + \frac{6}{c_0}}{{c_0}^{n+1}-1} \tag{4.5}$$

The degree distribution is given by equation 4.6:

$$P_f^{p\mathrm{D}}(l) = \begin{cases} \frac{c_0^n(c_0-1)}{c_0^{n+1}-1} & \text{for } l = p+1 \\ \frac{c_0-1}{c_0^{n+1}-1} & \text{for } l = c_0 \\ \frac{c_0\left(c_0^{n-1}-1\right)}{c_0^{n+1}-1} & \text{for } l = c_0+p+1 \\ 0 & \text{for all other } l \end{cases} \tag{4.6}$$

Equations 4.2–4.6 $c_0 \geq 3$ present the number of nodes in the basic-level ring, $n \in \mathbb{N}$ is the number of levels when indexing the levels starting with 1. In equation 4.6, $l \in \mathbb{N}$ is the degree parameter of a node.

This assessment indicates that the proposed structure is regular and clustered, while achieving low average path length (small-world effect, SW) [143]. Also, because this topology is self-similar it has scale-free (SF) characteristics. This aspect is further emphasized by the fact that if we reroute a relatively small number of long-range links, the degree distribution is brought closer to a power-law [68], as suggested in Figure4.4.

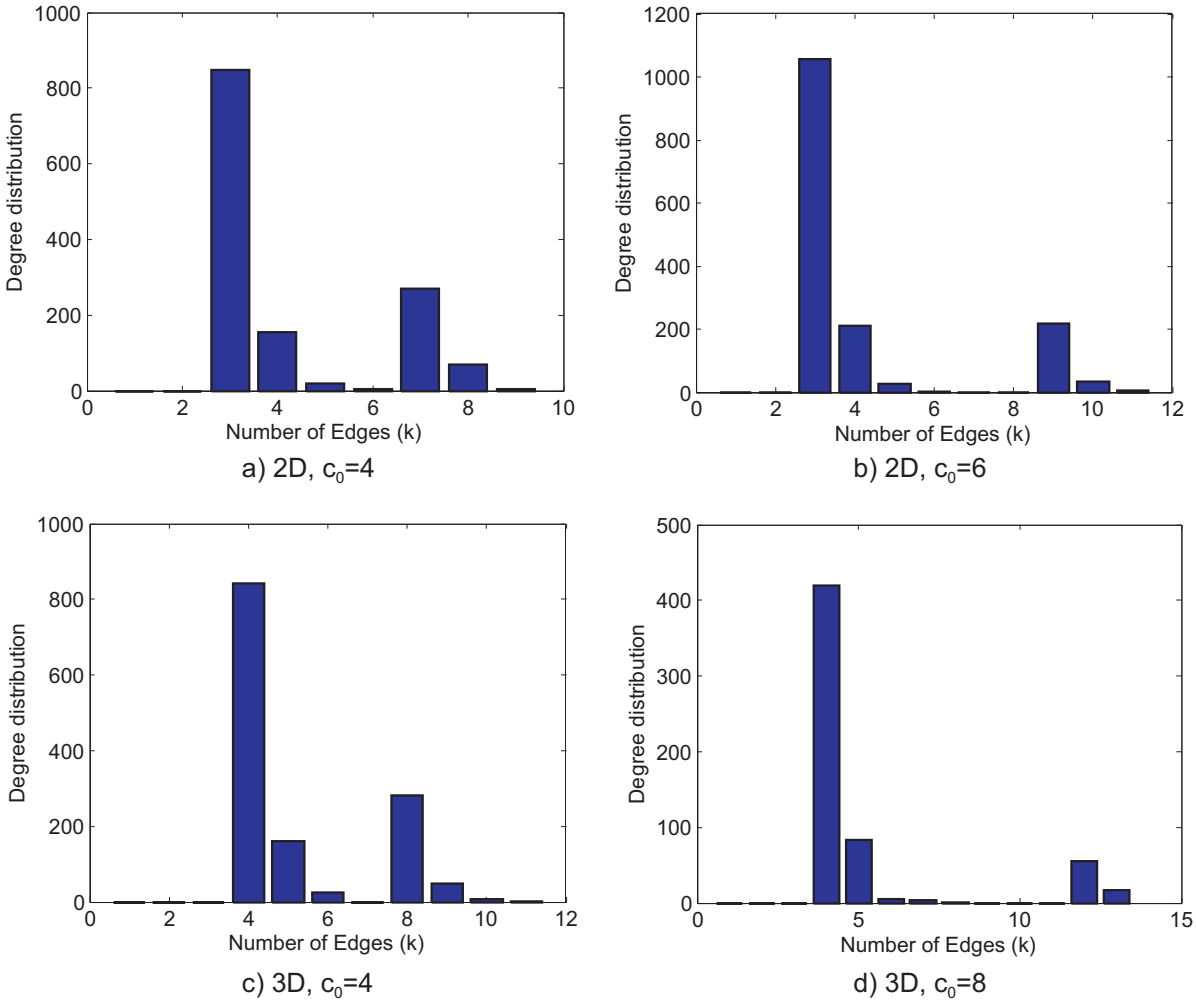

Figure 4.4: Degree distribution when inserting 10% long range links in fractal topologies: a) 2D square; b) 2D hexagon; c) 3D regular tetrahedron; d) 3D cube.

4.4 Resource Management in Fractal Networks

4.4.1 Routing

In a regular cartesian topology, like mesh or torus, the node indexing used for routing purposes is performed according to cartesian node position. For our fractal topology we have allocated indexes for each node by a simple breadth-first search in the graph, having the starting point at the top of the graph, and the address sequence being assigned from 1 to $\frac{{c_0}^{n+1}-1}{c_0-1}$.

Also, when dealing with the fractal network, the conventional routing mechanisms such as the x-y routing cannot be used. Therefore, we have to look for a suitable routing mechanism, inspired by the routing strategies that were used for the somehow similar conventional structures, like quad-tree and fat-tree [187][97].

Our routing strategy first performs *u*pward routing from the source node towards a node that pertains to an upper cluster (i.e. c_0 nodes with a router in the middle, as described in Section 4.3.1), whose central node represents a common ancestor for both source and destination nodes. Then, when the packet arrives in this cluster, it is routed within the cluster (*l*ocal routing) towards a node that represents an ancestor for the destination. Finally, *d*ownward routing is performed from the upper cluster node towards the destination node.

Local Routing

First, we have to determine the shortest path from the source to the destination. In order to illustrate this, we consider the general case from Figure 4.5. We first simply check if $address_{destination} = 1$; if this condition is true then the destination is the central router (1) which can be reached in one hop from any other cluster node. If the condition is not true, then we compute $(address_{source} - address_{destination}) \bmod c_0 - \frac{c_0}{2}$ which can give one of the following outcomes:

a) 0: it means that the destination is on the opposite side, i.e. the distance from the source to the destination is the same in both directions;

b) $-x$: the destination is x hops to the right, x can be any number from 1 to $\left(\frac{c_0}{2}-1\right)$;

c) $+x$: the destination is x hops to the left, x can be any number from 1 to $\left(\frac{c_0}{2}-1\right)$.

After determining the position of the destination relative to the source pertaining to the same cluster (consisting of all c_0 local nodes, plus their central router), we route the packet with the following probability:

$$p_l = \begin{cases} \left(1-\frac{1}{\kappa}\right)\cdot\left(1-\frac{\delta_{min}}{c_0}\right), & \text{shortest path} \\ \left(1-\frac{1}{\kappa}\right)\cdot\frac{\delta_{min}}{c_0}, & \text{longest path} \\ \frac{1}{\kappa}, & \text{through the router} \end{cases} \tag{4.7}$$

Where δ_{min} is the minimum distance from the source to the destination ($\delta_{min} = \frac{c_0}{2} - |\,(address_{source} - address_{destination}) \bmod c_0 - \frac{c_0}{2}|$) and $\kappa > 1 \in \mathbb{R}$. The central node

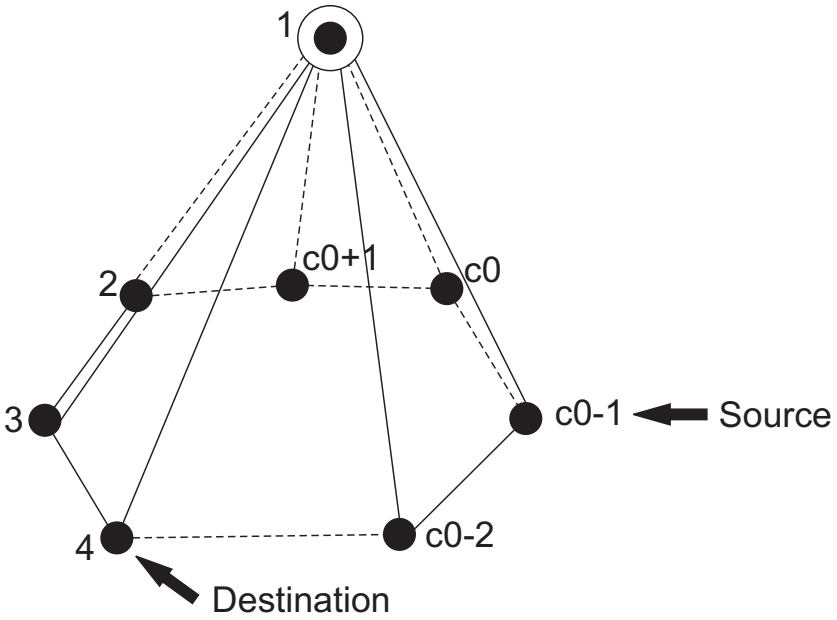

Figure 4.5: Local routing within the basic cluster of the fractal topology. The cluster consists of a ring of c_0 nodes, linked to a central node, that allows for communicating with the upper levels.

is always used when communicating with the upper levels. Therefore, when dealing with local traffic (i.e. within the cluster), it is desired not to use the central node, by maximizing κ. Thus, we avoid congesting this central router with unnecessary local traffic.

However, when $c_0 > 4$, excluding the central router for local traffic leads to performance loss because the average path length within the cluster increases. This means that when the NoC system runs an application with highly clustered traffic, κ can be smaller because the central routers are not solicited with intense inter-level traffic.

We have 2 special cases when $c_0 = 3$ or 4, i.e. the ring becomes a triangle and a square respectively. For $c_0 = 3$, the minimum distance between any two local nodes is 1, without using the central node. When $c_0 = 4$ the minimum distance between any two local nodes is 2, and the route can always avoid the central node. In both these cases, p_l can be computed as:

$$p_l^* = \begin{cases} 1 - \left(1 - \frac{1}{\kappa}\right) \cdot \frac{\delta_{min}}{c_0}, & \text{shortest path} \\ \left(1 - \frac{1}{\kappa}\right) \cdot \frac{\delta_{min}}{c_0}, & \text{longest path} \end{cases} \tag{4.8}$$

Upward Routing

Given a node x, its parent is given by:

$$\text{parent}\,(x) \Leftarrow \lfloor \left(\frac{x-2}{c_0} \right) + 1 \rfloor$$

In order to determine a route between two nodes x and y in the network, we apply the following deterministic algorithm:

if $(x + y) \neq 2$ **then**
  while parent $(x) \neq$ parent (y) **do**
    if parent $(x) <$ parent (y) **then**
      $y \Leftarrow$ parent (y)
      add to route y
    else
      $x \Leftarrow$ parent (x)
      add to route x
    end if
  end while
end if

This algorithm finds a common ancestor for the source and destination, and at the same time determines the number of levels that need to be passed from the source to the common ancestor; we denote this number by l_s. Thus, we can describe the probabilistic routing procedure from the source to the common ancestor, by expressing the routing probabilities as follows in Equation 4.9:

$$p_u = \begin{cases} \frac{1}{2n \cdot l_s} & \text{same cluster (left)} \\ \frac{1}{2n \cdot l_s} & \text{same cluster (right)} \\ \frac{n-1}{n}\left(1 - \frac{1}{l_s}\right) & \text{center router (up)} \end{cases} \tag{4.9}$$

Note that l_s decreases each time we go up a level, and it will be the indicator that we have reached the cluster of the common ancestor when $l_s = 1$. In this situation, we perform a local routing, as described in Section 4.4.1(within the cluster that contains the common ancestor of the source and destination).

Downward Routing

After reaching the common ancestor, we compute l_d (the hop distance between the common ancestor and the destination node), and then start going down on the hierarchy levels, following a probabilistic route that has a similar description with the upward routing, as in Equation 4.10:

$$p_d = \begin{cases} \frac{1}{2n \cdot l_d} & \text{downward to the left} \\ \frac{1}{2n \cdot l_d} & \text{downward to the right} \\ \frac{n-1}{n}\left(1 - \frac{1}{l_d}\right) & \text{straight down} \end{cases} \tag{4.10}$$

In contrast with the upward direction, here l_d increases from 1 to the difference of levels determined by the destination and the common ancestor. In the equation above, downward to the left or right means routing downward to the left or right of one of the destination's ancestors, whereas routing straight down means route directly to one of the destination's ancestors.

Total Routing

If we want to perform total routing from anywhere in the graph, first we have to determine the relative position of the destination from the source, which can be: up, down,

or in the same cluster. If it is up, then we act as described in Section 4.4.1, if it is down then we act as described in Section 4.4.1, i.e. the downward routing, and last, if it is in the same cluster we follow the procedure described in Section 4.4.1.

3D Routing

In the case of the 3D structures, the basic cluster is not organized as a ring, therefore local routing cannot work the way it is described in Section 4.4.1. However, taking into consideration the discussion from Section 4.3.1, the regular tetrahedron 3D fractal topology is favorable for shortest path routing, because local traffic can always use the shortest paths without going through the central node. In this case, given the fact that the upper levels of the structure are connected in the same way, the most favorable solution for regular tetrahedron fractal topology is to use deterministic routing, by determining the shortest path with Dijkstra's algorithm [45] or some other similar efficient algorithms.

4.4.2 Dimensioning of Communication Channels

In order to preserve the spatial fractality within the overall structure, the size of the links between the nodes in the upper levels will have to be increased according to the number of that level. This observation also helps in accommodating the heavier traffic that exists at the higher levels. The central nodes are required by the proposed fractal architectures, but are prone to congestion because they concentrate the traffic generated at the lower-level nodes (and the number of lower-level nodes grows exponentially with the number of levels).

In order to avoid these congestion situations, the natural solution is to allocate the number of virtual channels for links at level i according to the number of nodes at level $i-1$. These level $i-1$ nodes communicate outside their cluster through their corresponding central nodes that pertain to level i. This number of level $i-1$ nodes is equal with the number of nodes within a cluster, which is c_0, while the total number of lower-level nodes communicating with upper levels through one level-i central node is $N_i = \frac{(c_0)^i - 1}{c_0 - 1}$. Therefore, if any link at level $i-1$ has v virtual channels, then the size of a link at level i will have $v \cdot s$ virtual channels, where s (depending on c_0) is the scaling factor which assures that the number of virtual channels grows towards the top of the proposed fractal architecture. Figure4.6 presents the scaling of the virtual channels on a 2-level fractal architecture example with clusters having $c_0 = 4$ nodes.

The problem of selecting the scaling factor can be approached in an ad-hoc manner; if the traffic is not clustered, then the probability of packets leaving the node clusters is low, and we can avoid allocating a large number of virtual channels. However, not all applications entail clustered traffic, and so the topology has to offer support for inter-level communication, by allocating a substantial number of virtual channels at the upper levels.

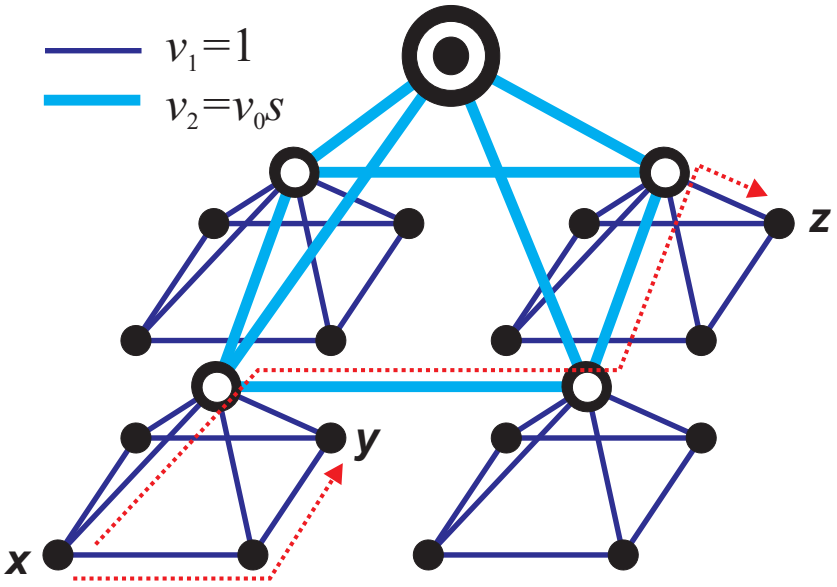

Figure 4.6: Allocation of channel width in terms of virtual channels, on a 2-level fractal architecture for $c_0 = 4$. The number of virtual channels for the links on fractal level i is v_i.

Allocating for scaling factor of c_0

From a fractal geometry standpoint, this scaling factor can be related to the Hausdorff-Besicovitch (HB) fractal dimension. The HB dimension of a fractal is given by $\frac{\log n_c}{\log s_f}$, where n_c is the number of self-similar copies at each level, and s_f is the scaling factor when generating the shapes at level $i+1$, from level i. In the case of our fractal topology, the number of self-similar copies is $n_c = c_0$. When we allocate a number of $v \cdot c_0$ virtual channels for each link at level i, and each of the links at level $i-1$ has v virtual channels, the scaling factor will be $s = s_f = c_0$. Therefore, in these conditions, the Hausdorff dimension will be $\frac{\log c_0}{\log c_0} = 1$. It is a virtual channel scaling factor that corresponds to a situation where there is a distinct set of virtual channels dedicated to communicating between any two nodes from the fractal topology. This case is equivalent to a totally connected network.

This virtual channel allocation – that corresponds to a fractal dimension of 1 – is ideal, because it avoids packet competition for virtual channels. However it is using an exponential number of virtual channels. If the fractal topology is built on m levels and each link at the lowest level has just 1 virtual channel, then at the highest level each link will have ${c_0}^{m-1}$ virtual channels. Nonetheless, the probability of really needing this huge communication infrastructure is small when the number of fractal levels is small.

We call *local traffic* communication that occurs between nodes within the same cluster on the same fractal level, as presented in the example form Figure4.6 when node x communicates with node y. As shown in Figure4.6, when the communication nodes do not pertain to the same cluster (e.g. node x communicates with node z), the *non-local traffic* will use the upper-level links. Assuming uniform random traffic, as the probability

of having non-local traffic (and therefore using the upper-level links) is approaching 1 when the number of levels $n > 3$, it is expected that most of the traffic will use the virtual channels on the upper levels. As such, in these conditions, the scaling factor will have to correspond to a Hausdorff dimension that is closer to 1. The probability of non-local traffic is given in equation 4.11.

$$P_{nl}(n) = \frac{(\sum_{i=1}^{n} {c_0}^i) - c_0 - 1}{(\sum_{i=1}^{n} {c_0}^i) - 1} \tag{4.11}$$

The probability of non-local traffic in fractal architectures with a basic level having 3, 4, 6 and 8 nodes is presented in Figure4.7.

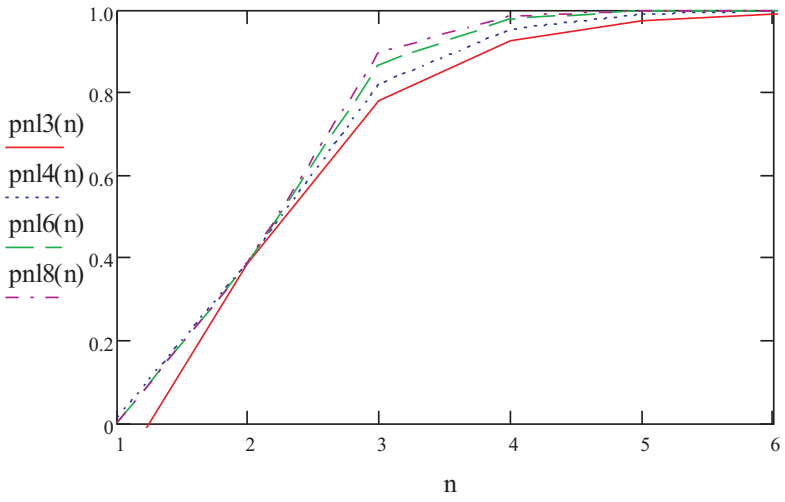

Figure 4.7: Probability of non local traffic in fractal architectures with the number of levels n, for $c_0 = 3$, 4, 6, and 8 nodes: $pnl3(n)$, $pnl4(n)$, $pnl6(n)$, and $pnl8(n)$ respectively.

Allocating for logarithmic scaling factor

When aiming at allocating a linear amount of virtual channels with the number of nodes, instead of looking for highly-efficient communication infrastructure, the scaling factor at level i can be set as $s = \lfloor \ln(N_i) + 1 \rfloor$, where $\lfloor a \rfloor$ denotes the largest integer that is smaller than a. Nevertheless, this allocation procedure will use a linear number of virtual channels with the number of fractal levels n, as N (the total number of nodes) is exponential with n (fractal level number) and the number of virtual channels is fixed as logarithmic with N.

Equivalent mesh allocation

However, when allocating virtual channels, another valid approach consists of using approximately the same number of virtual channels as the equivalent mesh architecture

(i.e. the mesh that will accommodate the same number of nodes as the fractal architecture). This total number of virtual channels for the mesh equivalent, VC_m, is given in equation 4.12, where v_m is the number of virtual channels allocated for each link in the mesh.

$$VC_m(n) = 2 \cdot \sqrt{\frac{{c_0}^{n+1} - 1}{c_0 - 1}} \left(\sqrt{\frac{{c_0}^{n+1} - 1}{c_0 - 1}} - 1 \right) \cdot v_m \tag{4.12}$$

In the proposed fractal architecture, the total number of allocated virtual channels (VC_f) with the number of levels n, when the scaling factor $s = c_0$, is given in equation 4.13.

$$VC_f(n) = 2 \cdot n \cdot {c_0}^n \tag{4.13}$$

When $s \neq c_0$, the total number of virtual channels in the fractal network, VC_{fs}, with the number of levels n is:

$$VC_{fs}(n) = 2 \cdot c_0 \cdot s^{n-1} \cdot \frac{\left(\frac{c_0}{s}\right)^n - 1}{\frac{c_0}{s} - 1} \tag{4.14}$$

At the same time, applying the virtual channel allocation according to the logarithmic scaling factor, entails a total number of virtual channels as given in equation 4.15:

$$\begin{aligned} &VC_{ah}(n) = 2 \cdot {c_0}^n + \\ &+2 \cdot \textstyle\sum_{i=1}^{n-1} \left[\lfloor \ln \left(\frac{{c_0}^{i+1}-1}{c_0-1} - 1 \right) + 1 \rfloor {c_0}^{n-i} \right] \end{aligned} \tag{4.15}$$

Equations 4.13 to 4.15 can be used for somputing the scaling factor so that the total number of virtual channels in the fractal network is approximatively the same as for the equivalent mesh. Figure4.8 presents the case of $c_0 = 4$, where the scaling factor $s = 3$ will approximate the number of virtual channels from the equivalent mesh network with $v_m = 2$.

Therefore, the virtual channel allocation for the fractal architecture (when $c_0 = 4$), so that the total number of virtual channels is approximately the same as for the equivalent mesh topology, is obtained for a scaling factor $s = 3$. This means that the corresponding Hausdorff fractal dimension becomes $\frac{\log 4}{\log 3}$. Because the fractal dimension > 1, it means that the network is no longer fully connected (i.e. there are no dedicated sets of virtual channels for connecting any two nodes), and competition for virtual channels is possible. In fact, as the fractal dimension grows, virtual channel contention also grows, thus affecting the communication efficiency that can be achieved when the Hausdorff dimension is 1.

4.4.3 Surface Allocation

It is expected that the clustered nature of fractal topology shapes entails sparse utilization of silicon die. As such, in order to be able to assess the heat dissipation capabilities of such a fractal architecture, knowing how it is occupying the silicon die is of paramount

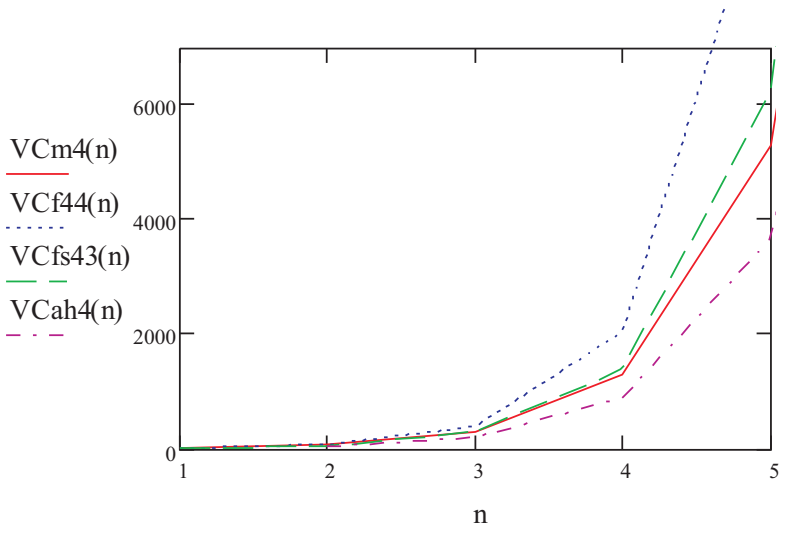

Figure 4.8: Network comparison in terms of total virtual channel number: VC_{f44} virtual channels for the fractal architecture where $c_0 = s = 4$, VC_{f43} for $c_0 = 4$ and $s = 3$, VC_{ah4} for the logarithmic scaling factor allocation procedure when $c_0 = 4$, and VC_{m4} for the mesh equivalent of a fractal network where $c_0 = 4$ and $v_m = 2$.

importance. To this end, we will use four parameters for the n-level fractal architecture: one that indicates the area required by the nodes or vertices (router+core) (V_n), another that indicates the area of the links (K_n), one that indicates the area used for both node logic and links (A_n), and the last one expressing the total silicon area (S_n). An obvious observation is that $A_n < V_n + K_n$, as some of the node and link surfaces are overlapping from a geometric point of view.

This analysis starts from the base level 1 of a 2D fractal topology with $c_0 = 4$ from Figure4.2, as presented in Figure4.9, where the square allocated for the node has a $\frac{\lambda}{3}$ edge, the thickness of the link is γ and the entire block is included within a $\lambda \times \lambda$ square. The values for the 4 parameters in our analysis are given, for the first level, in Equation

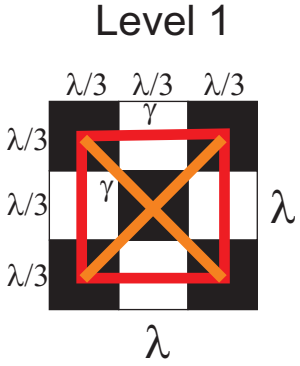

Figure 4.9: Level 1 fractal square architecture: the black squares represent the surface allocated to the nodes, whereas the colored lines represent the links.

4.16.

$$\begin{aligned} V_1 &= 5\left(\tfrac{\lambda}{3}\right)^2 \\ K_1 &= \tfrac{4}{3}\left(2+\sqrt{2}\right)\gamma\lambda \\ A_1 &= 5\left(\tfrac{\lambda}{3}\right)^2 + \tfrac{4}{3}\gamma\lambda \\ S_1 &= \lambda^2 \end{aligned} \tag{4.16}$$

For the second level, see Figure4.10, we considered that the thickness of links connecting the nodes from the upper level will grow with a factor α. Also, the center nodes pertaining to the upper levels will occupy half of the center square surface (the black $\frac{\lambda}{2} \times \frac{\lambda}{2}$ in the middle of Figure4.10. The values of the 4 parameters, in a 2-level fractal

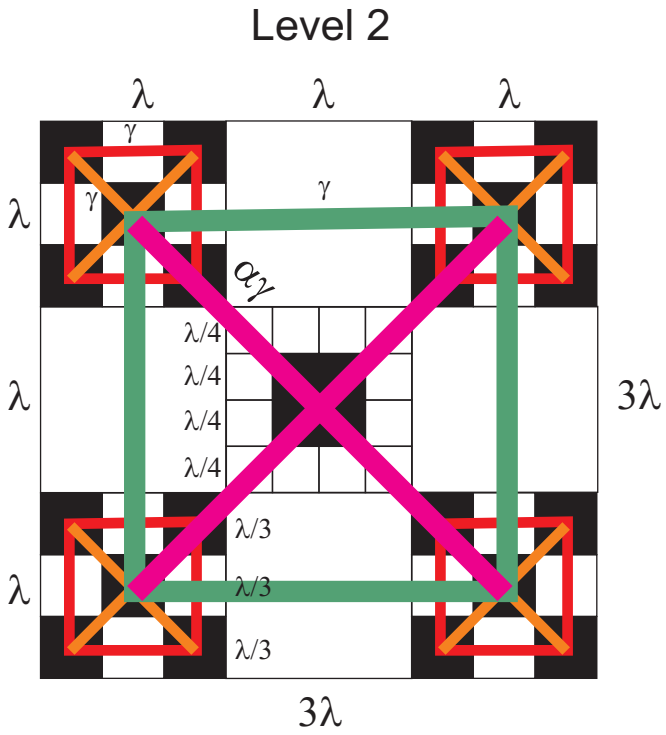

Figure 4.10: Surface allocation for level 2 fractal square architecture.

structure corresponding to Figure4.10 are given in Equation 4.17.

$$\begin{aligned} V_2 &= 4V_1 + \left(\tfrac{\lambda}{2}\right)^2 \\ K_2 &= 4K_1 + 8\gamma\lambda + 4\sqrt{2}\alpha\gamma\lambda \\ A_2 &= 4A_1 + 4\tfrac{5}{3}\gamma\lambda + \sqrt{2}\alpha\gamma\lambda + \left(\tfrac{\lambda}{2}\right)^2 \\ S_2 &= (3\lambda)^2 \end{aligned} \tag{4.17}$$

It can be proven that for any $i > 1$ we have:

$$\begin{aligned} V_{i+1} &= 4V_i + \left(3^i\tfrac{\lambda}{2}\right)^2 \\ K_{i+1} &= 4K_i + 4\cdot 3^i\sqrt{2}\alpha^{i+1}\gamma\lambda + 8\cdot 3^i\alpha^i\gamma\lambda \\ A_i &= 4A_i + \sqrt{2}3^i\alpha^{i+1}\gamma\lambda + 4\cdot 3^i\alpha^i\gamma\lambda + \\ &\quad +8\cdot 3^{i-1}\alpha^{i-1}\gamma\lambda + \left(3^i\tfrac{\lambda}{2}\right)^2 \\ S_{i+1} &= (3^{i+1}\lambda)^2 \end{aligned} \tag{4.18}$$

Having first level parameters V_1, K_1, A_1 and S_1 from Equation 4.16, and the recursive relations from Equation 4.18, we obtain that at level n we have:

$$\begin{aligned} V_n &= 4^n \left\{ V_1 + \tfrac{\lambda^2}{20} \left[\left(\tfrac{9}{4}\right)^n - 1 \right] \right\} \\ K_n &= 4^n \left[K_1 + \left(2+\sqrt{2}\alpha\right) \tfrac{\left(\frac{3}{4}\alpha\right)^n - 1}{\frac{3}{4}\alpha - 1} \gamma\lambda \right] \\ A_n &= 4^n A_1 + 4^{n-1} \left(4 + \alpha\sqrt{2}\right) \tfrac{\left(\frac{3}{4}\alpha\right)^n - 1}{\frac{3}{4}\alpha - 1} \gamma\lambda + \\ &\quad + 2 \cdot 4^{n-1} \tfrac{\left(\frac{3}{4}\alpha\right)^{n-1} - 1}{\frac{3}{4}\alpha - 1} \gamma\lambda + \\ &\quad + \tfrac{8}{3} \cdot 4^{n-1} \gamma\lambda + \tfrac{\left(\frac{9}{4}\right)^n - 1}{5} \cdot 4^{n-1} \lambda^2 \\ S_n &= \left(3^n \lambda\right)^2 \end{aligned} \tag{4.19}$$

The parameters from equations 4.19 can be used to compute the ratios given by the surface required by the nodes (core+router), the links and nodes+links, versus the total surface: V_n/S_n, K_n/S_n, and A_n/S_n respectively. These ratios are represented in Figure4.11, for the example case of $\lambda = 3$, $\gamma = \frac{\lambda}{20}$, and $\alpha = 2$.

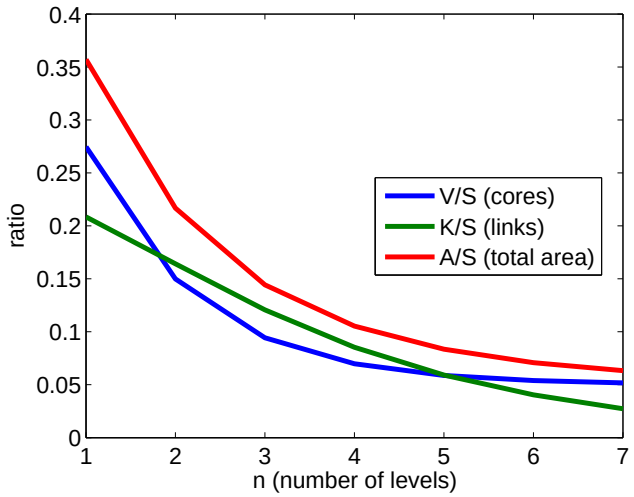

Figure 4.11: Ratio of surface occupancy for nodes V/S, links K/S, and nodes+links A/S versus the number of levels n.

The plot in Figure4.11 shows that the area utilization decreases exponentially with the number of fractal levels, as the fractal structure becomes more sparse, thus facilitating heat dissipation.

4.5 Conclusion

Using complex network science in order to optimize on-chip multiprocessor communication networks is indeed a very promising research topic. As NoC communication entails

dealing with complex emerging phenomena, the solutions for optimizing NoC architectures will have to find inspiration in the new science of complexity, including statistical physics and complex networks.

We note that there are still many difficulties ahead, as current fabrication technology is fundamentally limiting what can actually be implemented in Silicon (including fractal or complex network topologies). Indeed, from a technological standpoint, only regular structures are supported, while long-range interconnections are hard to build. However, new on-chip communication technologies such as optical [201][202] or wireless communication [89][115] offer new hope for overcoming the above-mentioned difficulties. As such, we intend to further pursue research on this topic, by adapting the complex network tools to the new NoC communication technologies.

Chapter 5

Conclusions

The present thesis argues that there is fertile ground for research at the frontiers between computer engineering, information technology, physics, medical and social sciences. Indeed, the advent of *big data* techniques such as *machine* and *deep learning* or *complex networks* have spurred significant advances in scientific fields like *physics*, *biology*, *precision medicine*, *genomics*, *social networks* and *social physics*, *emergent human behavior modeling*, etc. Moreover, there are strong reasons which indicate the new network science as a reliable source of solutions for engineering problems, ranging from managing complexity in NoC communication to designing power grids or controlling road traffic. Our thesis shows how the new science of complex networks can be used for optimizing NoC communication, modeling opinion dynamics in social networks, and precision medicine (drug repositioning and uncovering disease phenotypes).

However, in order to perform big data approaches, we need lots of relevant data on physical processes, human activity, or human physiological signals. To this end, computer engineering can provide solutions and tools such as Internet of Things (IoT) or Cyber Physical Systems (CPS): a lot of sensors gather big amounts of data to be analyzed and then used for computation, communication and control. Following this trend, it can be said that we entered the new era of *inexact computation* [128], which is characterized by integrating the computer in complex physical, biological, social and processes.

To this end, within UEFISCDI's PN-III-CERC-CO-PED-2016 framework, we proposed as coordinator the research project entitled "Internet of Things Meets Complex Networks for Early Prediction and Management of Chronic Obstructive Pulmonary Disease", or Internet of thiNgs Complex nEtworks PredicTION (INCEPTION). The motivation for our proposal is that recent research indicates Chronic Obstructive Pulmonary Disease (COPD) as the third cause of death and one of the main impediments for the quality of living in todays society. COPD is defined as the clinical condition which reduces pulmonary capacity; it is not reversible, however if diagnosed at an early phase, its evolution can be controlled. Unfortunately, early detection of COPD is a difficult task, and often time people are diagnosed when they are already in an advanced stage. Capitalizing on recent research results which indicate sensor systems, mobile, and Internet of Things solutions as very useful for monitoring and managing COPD, we propose a personal, integrated prototype system for early detection and evolution

prediction of COPD. As such, we intend to build a sensor network that gathers multiple physiological signals, and a mobile application that extracts the multi-fractal spectra as mere signatures of these signals. Then, the mobile system will integrate the physiologic signatures with anthropometric and other individual clinical data. The overview of INCEPTION at individual level is presented in Figure 5.1.

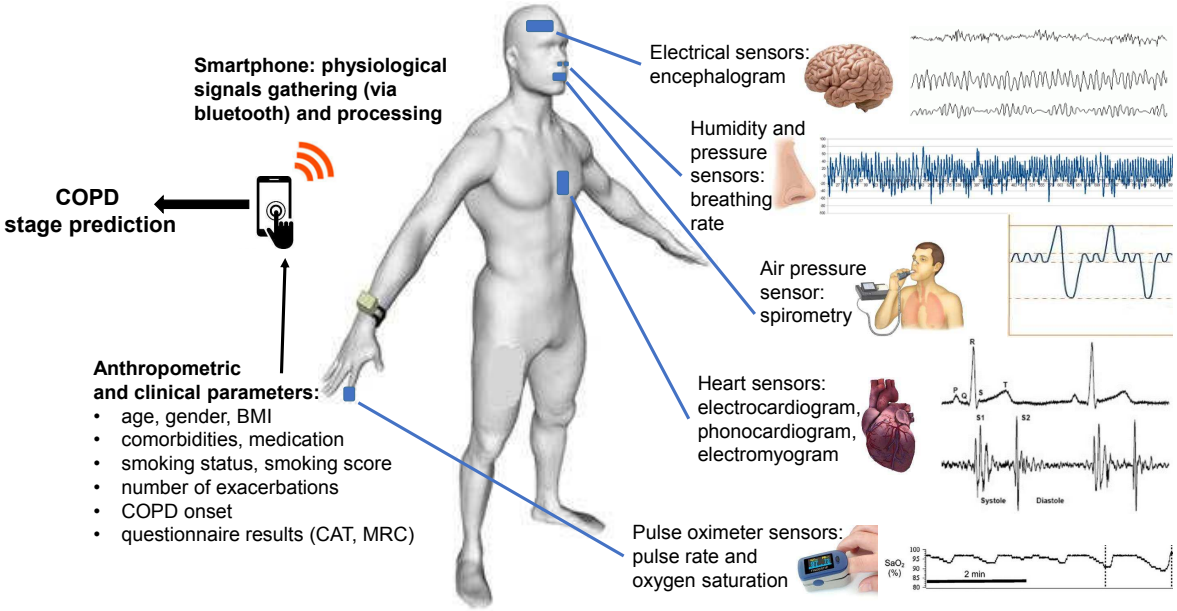

Figure 5.1: Overview of INCEPTION at individual-level. Physiological signals are gathered with a sensor network and transmitted via Bluetooth. Anthropometric and clinical parameters are recorded and integrated at smartphone-level and then sent over the cloud in a centralized database.

On the server side, we will collect the integrated data from a population of individuals, to build a complex network model of patients. Indeed, recent papers indicate the complex network model as very useful for generating COPD predictions. To this end, we will employ modularity clustering and network layout tools to build prediction models for both early detection and evolution prediction of COPD. he overview of INCEPTION at individual level is presented in Figure 5.2.

The prediction model will be instantiated as a smartphone application and tested in order to assess its predictive capacity. In order to undertake the objectives of our proposal, we assemble a multidisciplinary team, consisting of computer engineers (hardware and software) and specialized medical doctors from University of Medicine and Pharmacy "Victor Babeş" Timişoara. Our computer engineering teams will build the hardware and software parts of our demonstration model, whereas the medical team will provide the necessary medical expertise, in order to test and validate the model proposed by the computer engineers.

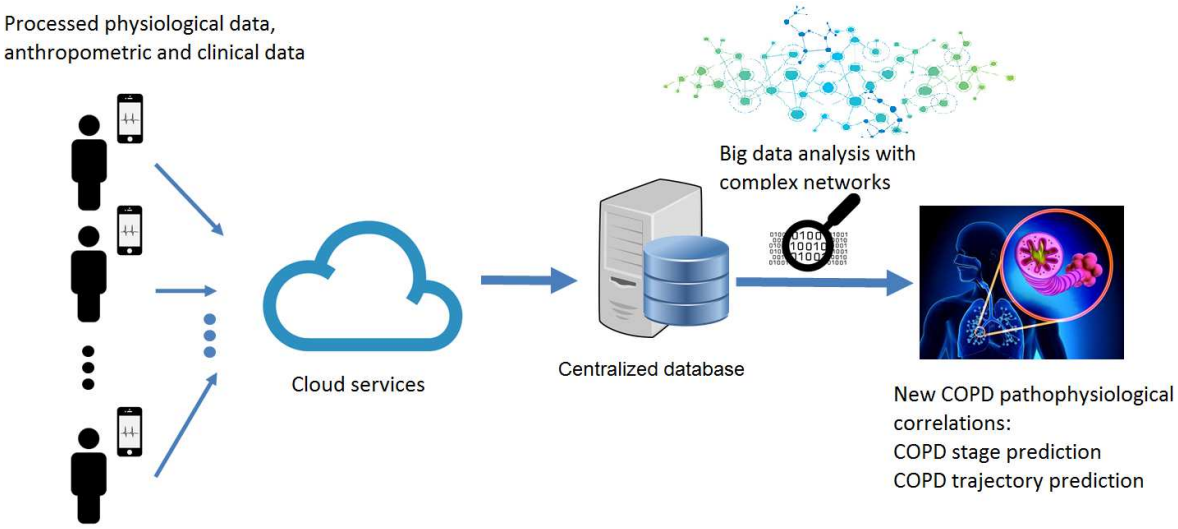

Figure 5.2: Overview of INCEPTION at server-level. Integrated individual data are collected from many individuals and stored in a centralized database. The gathered dataset is then processed according to the complex network paradigm, in order to find new pathophysiological correlations.

Part III

Relevant Bibliography

Bibliography

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

[2] D. Acemoglu, A. Ozdaglar, and E. Yildiz, *Diffusion of innovations in social networks*, 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC), pp. 2329-2334, 2011

[3] D. Acemoglu, G. Como, F. Fagnani, and A. Ozdaglar, *Opinion fluctuations and disagreement in social networks*, Mathematics of Operations Research 38, 1, 1–27, 2013

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

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

[6] C. Alt, O. Astrachan, J. Forbes, R. Lucic, and S. Rodger, *Social networks generate interest in computer science*, ACM SIGCSE Bulletin, 38, 1, 438-442, 2006

[7] L. Amaral, A. Scala, M. Barthélémy, and H.E. Stanley, *Classes of small-world networks*, Proceedings of the National Academy of Sciences, 97, 21, 11149-11152, 2000

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

[9] T. Arentze, and H. Timmermans, *Social networks, social interactions, and activity-travel behavior: A framework for microsimulation*, Environment and planning. B, Planning & design, 35, 6, 1012, 2008

[10] T. Arentze, P. van den Berg, and H. Timmermans, *Modeling social networks in geographic space: approach and empirical application*, Environment and Planning-Part A, 44, 5, 1101, 2012

[11] D. Auber, *Tulip: A huge graph visualization framework*, Graph Drawing Software, Springer, 105-126, 2004

[12] R. Axelrod, *The dissemination of culture a model with local convergence and global polarization, Journal of conflict resolution* 41, 2, 203-226, 1997

[13] F. Bagnoli, F. Franci, and R. Rechtman, *Opinion formation and phase transitions in a probabilistic cellular automaton with two absorbing states*, Cellular Automata, Springer, pp. 249-258, 2002

[14] N.T.J. Bailey *et al.*, *The mathematical theory of infectious diseases and its applications*, Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975

[15] S. Bandyopadhyay, A.R. Rao, B. Sinha, and B.K. Sinha, *Models for Social Networks with Statistical Applications*, Sage, Vol.13, 2011

[16] A.L. Barabási, *Network science*, Cambridge University Press, 2016

[17] A.L. Barabási, *The network takeover*, Nature Physics, 8, 1, 14, 2011

[18] A.L. Barabási, and R. Albert, *Emergence of scaling in random networks*, Science 286, 5439, 509-512, 1999

[19] M. Bastian, S. Heymann, and M. Jacomy, *Gephi: an open source software for exploring and manipulating networks.* ICWSM, 2009

[20] V. Batagelj, and A. Mrvar, *Pajek-program for large network analysis*, Connections 21, 2, 47-57, 1998

[21] M. A. Bender, D. P. Bunde, E. D. Demaine, S. P. Fekete, V. J. Leung, H. Meijer, C. A. Phillips, *Communication-Aware Processor Allocation for Supercomputers: Finding Point Sets of Small Average Distance*, C Algorithmica 50, 2, 279-298, 2008

[22] G. B. P. Bezerra, S. Forrest, M. Forrest, A. Davis, P. Zarkesh-Ha, *Modeling NoC traffic locality and energy consumption with rent's communication probability distribution*, In Proc. 12th ACM/IEEE Intl. Workshop on System Level Interconnect Prediction, New York, NY, USA, pp. 3-8, 2010

[23] A. Bigdeli, A. Tizghadam, and A. Leon-Garcia, *Comparison of network criticality, algebraic connectivity, and other graph metrics*, Proceedings of the 1st Annual ACM Workshop on Simplifying Complex Network for Practitioners, 4, 2009

[24] K.R. Bisset, J. Chen, X. Feng, V.S. Kumar, and M.V. Marathe, *EpiFast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems*, Proceedings of the 23rd ACM International Conference on Supercomputing, pp. 430-439, 2009

[25] S. Biswas, A.K. Chandra, A. Chatterjee, and B.K. Chakrabarti, *Phase transitions and non-equilibrium relaxation in kinetic models of opinion formation*, Journal of Physics: Conference Series 297, 1, 012004, 2011

[26] P. Bogdan, R. Marculescu, *Statistical physics approaches for network-on-chip traffic characterization*, In Proc. 7th IEEE/ACM Intl. Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS '09), New York, NY, USA, pp. 461-470, 2009

[27] P. Bogdan, M. Kas, R. Marculescu, O. Mutlu, *QuaLe: a quantum-leap inspired model for non-stationary analysis of NoC traffic in chip multi-processors*, In Proc. ACM/IEEE Intl. Symp. on Networks-on-Chip (NOCS), pp. 241-248, 2010

[28] B. Bolgár et al., *Drug repositioning for treatment of movement disorders: from serendipity to rational discovery strategies*, Current topics in medicinal chemistry 13, 2337–2363, 2013

[29] E. Bolotin, I. Cidon, R. Ginosar, A. Kolodny, *Routing table minimization for irregular mesh NoCs*, In Proc. Design Automation and Test in Europe (DATE), pp. 942-947, 2007

[30] J. Borondo, F. Borondo, C. Rodriguez-Sickert, C.A. Hidalgo, *To each according to its degree: The meritocracy and topocracy of embedded markets*, Scientific Reports, 4, 3784, 2014

[31] C.T. Butts, *Social network analysis with SNA*, Journal of Statistical Software 24, 6, 1-51, 2008

[32] A. Cairns, G. Poulos, and R. Bogan, *Sex differences in sleep apnea predictors and outcomes from home sleep apnea testing*, Nature and Science of Sleep, 8, 197, 2016

[33] N. Caseiro, and P. Trigo, *Comparing complex networks: An application to emergency managers mental models*, Third Brazilian Workshop on Social Simulation (BWSS), pp. 128-131, 2012

[34] D. Centola, *The spread of behavior in an online social network experiment*, Science 329, 5996, 1194–1197, 2010

[35] D. Centola, and M. Macy, *Complex contagions and the weakness of long ties1*, American Journal of Sociology 113, 3, 702-734, 2007

[36] K. Channakeshava, K. Bisset, V.S.A. Kumar, M. Marathe, and S. Yardi, *High performance scalable and expressive modeling environment to study mobile malware in large dynamic networks*, IEEE International Parallel & Distributed Processing Symposium (IPDPS), pp. 770–781, 2011

[37] H.F. Chau, C.Y. Wong, F.K. Chow, and C-H.F. Fung, *Social judgment theory based model on opinion formation, polarization and evolution*, Physica A: Statistical Mechanics and its Applications 415, 770-781, 2014

[38] X.-Q. Chen, M.D. Antman, C. Gesenberg, O.S. Gudmundsson, *Discovery pharmaceuticschallenges and opportunities*, The AAPS journal 8, E402–E408, 2006

[39] G. Chen, X. Wang, and X. Li, *Fundamentals of Complex Networks: Models, Structures and Dynamics*, John Wiley & Sons, 2014

[40] N.R. Clark, K. Hu, E.Y. Chen, Q. Duan, and A. Maayan, *Characteristic Direction Approach to Identify Differentially Expressed Genes*, arXiv preprint arXiv:1307.8366, 2013

[41] J.J. Clarkson, Z.L. Tormala, D.D. Rucker, and R.G. Dugan, *The malleable influence of social consensus on attitude certainty*, Journal of Experimental Social Psychology 49, 6, 1019–1022, 2013

[42] H. Claussen, W. Lichtner, L. Heres, P. Lahaije, and J. Siebold, *GDF, a proposed standard for digital road maps to be used in car navigation systems*, IEEE Vehicle Navigation and Information Systems Conference, pp. 324-330, 1989

[43] P. R. Cohen, *Darpa's big mechanism program*, Physical biology 12, 045008, 2015

[44] F. S. Collins, and H. Varmus, *A new initiative on precision medicine*, New England Journal of Medicine, 372, 9, 793–795, 2015

[45] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, *Introduction to Algorithms (Second ed.)Section 24.3: Dijkstra's algorithm*, MIT Press and McGraw-Hill, pp. 595-601, 2001

[46] J. Chang, E. Blazek, M. Skowronek, L. Marinari, and R. P. Carlson, *The antiinflammatory action of guanabenz is mediated through 5-lipoxygenase and cyclooxygenase inhibition*, European journal of pharmacology 142, 197–205, 1987

[47] E. Chautard, N. Thierry-Mieg, and S. Ricard-Blum, *Interaction networks: from protein functions to drug discovery. A review*, Pathologie Biologie 57, 324–333, 2009

[48] Y.W. Chen, L.F. Zhang, J.P Huang, *The Watts–Strogatz network model developed by including degree distribution: theory and computer simulation*, Journal of Physics A: Mathematical and Theoretical 40, 29, 8237, 2007

[49] C.-Y. Cheng, Y.-C. Lin, J.-S. Chen, C.-H. Chen, and S. Deng, *Cisplatin-induced acute hyponatremia leading to a seizure and coma: a case report*, Chang Gung Med J 34, 48–51, 2011

[50] F. Chung, B. Yegneswaran, P. Liao, S.A. Chung, S. Vairavanathan, S. Islam, A. Khajehdehi, and C.M. Shapiro, *STOP Questionnaire. A Tool to Screen Patients for Obstructive Sleep Apnea*, The Journal of the American Society of Anesthesiologists, 108, 5, 812–821, 2008

[51] F. Chung, B. Yegneswaran, P. Liao, S.A. Chung, S. Vairavanathan, S. Islam, A. Khajehdehi, and C.M. Shapiro, *Validation of the Berlin questionnaire and American Society of Anesthesiologists checklist as screening tools for obstructive sleep apnea in surgical patients*, The Journal of the American Society of Anesthesiologists, 108, 5, 822–830, 2008

[52] F. Chung, Y. Yang, R. Brown, and P. Liao, *Alternative scoring models of STOP-Bang questionnaire improve specificity to detect undiagnosed obstructive sleep apnea*, Journal of Clinical Sleep Medicine, 10, 9, 951–958, 2014

[53] B. M. Conti-Fine, M. Milani, and H. J. Kaminski, *Myasthenia gravis: past, present, and future*, The Journal of clinical investigation 116, 2843–2854, 2006

[54] L. Costa, F.A. Rodrigues, G. Travieso, and P.R. Villas Boas, *Characterization of complex networks: A survey of measurements*, Advances in Physics 56, 1, 167-242, 2007

[55] G. Csányi, and B. Szendrői, *Structure of a large social network*, Physical Review E 69, 3, 036131, 2004

[56] G. Csardi, and T. Nepusz, *The igraph software package for complex network research*, InterJournal, Complex Systems 5, 1695, 2006

[57] P. Csermely, T. Korcsmáros, H. J. Kiss, G. London, and R. Nussinov, *Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review*, Pharmacology & therapeutics 138, 333–408, 2013

[58] A. D'Andrea, F. Ferri, and P. Grifoni, *An overview of methods for virtual social networks analysis*, Computational Social Network Analysis, Springer, 3-25, 2010

[59] L. Daqing, K. Kosmidis, A. Bunde, S. Havlin, *Dimension of spatially embedded networks*, Nature Physics 7, 481484, 2011

[60] A. Das, S. Gollapudi, and K. Munagala, *Modeling opinion dynamics in social networks*, Proceedings of the 7th ACM international conference on Web search and data mining, pp. 403-412, 2014

[61] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, *Mixing beliefs among interacting agents*, Advances in Complex Systems 3, 01n04, 87-98, 2000

[62] D. De Martino, L. Dall'Asta, G. Bianconi, M. Marsili, *Congestion phenomena on complex networks*, Phys. Rev. E 79, 1, 015101, 2009

[63] L. Deng, Y. Liu, and F. Xiong, *An opinion diffusion model with clustered early adopters*, Physica A: Statistical Mechanics and its Applications 392, 17, 3546-3554, 2013

[64] A. Dharmarajan, and T. Velmurugan, *Lung cancer data analysis by k-means and farthest first clustering algorithms*, Indian Journal of Science and Technology, 8, 15, 2015

[65] M. Dickson, and J. P. Gagnon, *The cost of new drug discovery and development*, Discovery Medicine 4, 172–179, 2009

[66] D. Diez, A. Augusti, C.E. Wheelock, *Network analysis in the investigation of chronic respiratory diseases. From basics to application*, American Journal on Respiratory and Critical Care Medicine 190, 981988, 2014

[67] M.J. Divo, C. Casanova, J.M. Marin, V.M. Pinto-Plata, J.P. de-Torres, J.J. Zulueta, C. Cabrera, J. Zagaceta, P. Sanchez-Salcedo, J. Berto, R. Baz Davila, A.B. Alcaide, C. Cote, B.R. Celli, *Chronic obstructive pulmonary disease comorbidities network*, European Respiratory Journal, 46, 3, 640-650, 2015

[68] S. N. Dorogovtsev, J. F. F. Mendes, A. N. Samukhin, A. T. Zyuzin,*Organization of modular networks*, In Phys. Rev. E, 78, 5, 056106, 2008

[69] S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes, *Critical phenomena in complex networks*, Reviews of Modern Physics Vol. 80, pp.1275-1335, 2008

[70] A. Duma, and A. Topirceanu, *A network motif based approach for classifying online social networks*, 9th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI), pp. 311-315, 2014

[71] W. J. Egan, G. Zlokarnik, and P. D. Grootenhuis, *In silico prediction of drug safety: despite progress there is abundant room for improvement*, Drug Discovery Today: Technologies 1, 381–387, 2004

[72] D. Elkind, *Egocentrism in adolescence*, Child development, pp. 1025-1034, 1967

[73] J. Ellson, E. Gansner, L. Koutsofios, S.C. North, and G. Woodhull, *Graphviz: open source graph drawing tools*, Graph Drawing, Springer, pp. 483-484, 2002

[74] P. Erdös, and A. Rényi, *On the evolution of random graphs*, Publications de l'Institut des Mathematiques Appliquees de l'Academie des Sciences de Hongrie 5, 17-61, 1960

[75] E. Estrada, *The structure of complex networks*, Oxford University Press, 2011

[76] K. Falconer, *Fractal geometry: mathematical foundations and applications* Wiley, 2007

[77] W. Fan, and K.H. Yeung, *Online social networks: Paradise of computer viruses*, Physica A: Statistical Mechanics and its Applications 390, 2, 189-197, 2011

[78] R. Faner, T. Cruz, A. Lopez-Giraldo, A. Agust, *Network medicine, multimorbidity and the lung in the elderly*, European Respiratory Journal 44, 3, 775788, 2014

[79] R. Faner, A. Agust, *Multilevel, Dynamic Chronic Obstructive Pulmonary Disease Heterogeneity. A Challenge for Personalized Medicine*, Annals of the American Thoracic Society 13, 5, 2016

[80] H. Fang, J. Zhang, and N.M. Thalmann, *A trust model stemmed from the diffusion theory for opinion evaluation*, Proceedings of the 2013 international conference on autonomous agents and multi-agent systems, pp. 805-812, 2013

[81] R.J. Farney, B.S. Walker, R.M. Farney, G.L. Snow, and J.M. Walker, *The STOP-Bang equivalent model and prediction of severity of obstructive sleep apnea: relation to polysomnographic measurements of the apnea/hypopnea index*, Journal of Clinical Sleep Medicine, 7, 5, 459–465, 2011

[82] B. S. Feero and P. P. Pande, *Networks-on-Chip in a Three-Dimensional Environment: A Performance Evaluation*, IEEE Trans. Comput. 58, 1, pp.32-45, 2009

[83] L. Feng, Y. Hu, B. Li, H.E. Stanley, S. Havlin, and L.A. Braunstein, *Competing for Attention in Social Media under Information Overload Conditions*, arXiv preprint arXiv:1410.1668, 2014

[84] J. Fischer, Z. Dogas, C. Bassetti, S. Berg, L. Grote, P. Jennum, P. Levy, S. Mihaicuta, L. Nobili, D. Riemann, F.J. Puertas-Cuesta, F. Raschke, D.J. Skene, N. Stanley, D. Pevernagie, and Executive Committee (EC) of the Assembly of the National Sleep Societies (ANSS); Board of the European Sleep Research Society (ESRS), Regensburg, Germany, *Standard procedures for adults in accredited sleep medicine centres in Europe*, Journal of Sleep Research, 21(4), 357–368, 2012

[85] G. Fowler, *Facebook: One billion and counting*, The Wall Street Journal, B1, 2012

[86] A. Fronczak, P. Fronczak, J. A. Holyst, *Average path length in random networks*, Physical Review E - Statistical, Nonlinear and Soft Matter Physics, 70(5 Pt 2), 056110, 2004

[87] T.M.J. Fruchterman, and E.M. Reingold, *Graph drawing by force-directed placement*, Software: Practice and experience 21, 11, 1129–1164, 1991

[88] M. Gaertler, *Clustering. Network Analysis*, Lecture Notes in Computer Science LNCS 3418, pp. 178215, 2005

[89] A. Ganguly, K. Chang, S. Deb, P.P. Pande, B. Belzer, and C. Teuscher, *Scalable Hybrid Wireless Network-on-Chip Architectures for Multicore Systems*, IEEE Transactions on Computers, 60(10), 1485-1502, 2011

[90] I.E. Gabbay, and P. Lavie, *Age-and gender-related characteristics of obstructive sleep apnea*, Sleep and Breathing, 16, 2, 453–460, 2012

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

[92] S. D. Ghiassian, et al., *Endophenotype network models: Common core of complex diseases*, Scientific reports 6, 27414, 2016

[93] M. Girvan, M.E.J. Newman, *Community structure in social and biological networks*, Proceedings of the National Academy of Sciences, 99, 12, 7821-7826, 2002

[94] J. Golbeck, *Analyzing the Social Web*, Elsevier, 2013

[95] A. Graul, E. Cruces, and M. Stringer, *The year's new drugs & biologics, 2013: Part I*, Drugs Today (Barc) 50, 51–100, 2014

[96] D. Greenfield, S. W. Moore, *Brief announcement: fractal communication in software data dependency graphs*, Proc. Ann. Symp. on Parallelism in algorithms and architectures (SPAA), pp.116-118, 2008

[97] P. Guerrier, A. Greiner, *A generic architecture for on-chip packet-switched interconnections*, In Proc. Design Automation and Test in Europe (DATE), pp. 250-256, 2000

[98] A. Guille, H. Hacid, C. Favre, and D.A. Zighed, *Information diffusion in online social networks: a survey*, ACM SIGMOD Record 42, 2, 17-28, 2013

[99] R. Guimera, A. Diaz-Guilera, F. Vega-Redondo, A. Cabrales, A. Arenas, *Optimal Network Topologies for Local Search with Congestion*, In Phys. Rev. Lett., 89, 24, 248701, 2002

[100] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten, *The WEKA data mining software: an update*, ACM SIGKDD explorations newsletter, 11, 1, 10–18, 2009

[101] J. Hedner, L. Grote, M. Bonsignore, W. McNicholas, P. Lavie, G. Parati, P. Sliwinski, F. Barbé, W. De Backer, P. Escourrou, I. Fietze, J.A. Kvamme, C. Lombardi, O. Marrone, J.F. Masa, J.M. Montserrat, T. Penzel, M. Pretl, R. Riha, D. Rodenstein, T. Saaresranta, R. Schulz, R. Tkacova, G. Varoneckas, A. Vitols, H. Vrints, and J. Zielinski, *The European sleep apnoea database (ESADA): report from 22 European sleep laboratories*, European Respiratory Journal, 38, 3, 635–642, 2011

[102] R. Hegselmann, and U. Krause, *Opinion dynamics and bounded confidence models, analysis, and simulation*, Journal of Artificial Societies and Social Simulation 5, 3, 2002

[103] C. Hernandez, F. Silla, J. Duato, *A methodology for the characterization of process variation in NoC links*, In Proc. Design Automation and Test in Europe (DATE), Leuven, Belgium, pp.685-690, 2010

[104] M. Hirohashi, K. Takasuna, Y. Kasai, C. Usui, and H. Kojima, *Pharmacological studies with the alpha 2-adrenoceptor antagonist midaglizole. Part II: Central and peripheral nervous systems*, Arzneimittel-Forschung 41, 19–24, 1991

[105] J. Huang et al., *Systematic prediction of pharmacodynamic drug-drug interactions through protein-protein-interaction network*, PLoS Comput Biol 9, e1002998, 2013

[106] O. Hussain, Z. Anwar, S. Saleem, F. Zaidi, *et al.*, *Empirical Analysis of Seed Selection Criterion in Influence Mining for Different Classes of Networks*, ASONAM, pp. 1-8, 2013

[107] M. A. Javarone, and T. Squartini, *Conformism-driven phases of opinion formation on heterogeneous networks: the q-voter model case*, ArXiv 2014arXiv1410.7300J, 2014

[108] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, *ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software*, PloS one, 9, 6, e98679, 2014

[109] L. Jian-Guo, D. Yan-Zhong, W. Zhong-Tuo, *Multistage random growing small-world networks with power-law degree distribution*, Chinese Physics Letters, 23, 3, 746, 2006

[110] S.A. Joosten, K. Hamza, S. Sands, A. Turton, P. Berger, and G. Hamilton, *Phenotypes of patients with mild to moderate obstructive sleep apnoea as confirmed by cluster analysis*, Respirology, 17, 1, 99–107, 2012

[111] M. Karlgren, et al., *In vitro and in silico strategies to identify OATP1B1 inhibitors and predict clinical drug–drug interactions*, Pharmaceutical research 29, 411–426, 2012

[112] K. Kawaguchi, Y. Oribe, and H. Uzawa, *Tolbutamide effect on cultured human endothelial cells with special reference to platelet aggregation*, The Tohoku journal of experimental medicine 141, 563–568, 1983

[113] D. Kempe, J. Kleinberg, and E. Tardos, *Maximizing the spread of influence through a social network*, Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 137-146, 2003

[114] K.-T. Khaw, and E. Barrett-Connor, *Dietary potassium and stroke-associated mortality*, New England Journal of Medicine 316, 235–240, 1987

[115] R.G. Kim, W. Choi, G. Liu, E. Mohandesi, P.P. Pande, D. Marculescu, and R. Marculescu, *Wireless NoC for VFI-Enabled Multicore Chip Design: Performance Evaluation and Design Trade-Offs*, IEEE Transactions on Computers, 65(4), 1323-1336, 2016

[116] T. M. Laidlaw, et al., *Cysteinyl leukotriene overproduction in aspirin-exacerbated respiratory disease is driven by platelet-adherent leukocytes*, Blood 119, 3790–3798, 2012

[117] K. C. Lee, and D. C. Randall, *Potentiation of the pressor response to stress by tolbutamide in dogs*, Integrative physiological and behavioral science, 28, 22–28, 1993

[118] S. -J. Lee, K. Lee, H. -J. Yoo,*Analysis and Implementation of Practical, Cost-Effective Networks on Chips*, IEEE Des.Test 22, 5, 422-433, 2005

[119] C. E. Leiserson, *Fat-trees: universal networks for hardware-efficient supercomputing*, In Interconnection Networks for High-Performance Parallel Computers, IEEE Computer Society Press, pp. 170-179, 1994

[120] P. Lévy, M. Kohler, W.T. McNicholas, F. Barbé, R.D. McEvoy, V.K. Somers, L. Lavie, J.-L. Pépin, *Obstructive sleep apnoea syndrome*, Nature Reviews. Disease Primers 1, 15015–15015, 2014

[121] L. Lewis, *Drug–drug interactions: is there an optimal way to study them?*, British journal of clinical pharmacology 70, 781–783, 2010

[122] L. Li, A. Scaglione, A. Swami, and Q. Zhao, *Trust, opinion diffusion and radicalization in social networks*, Conference Record of the Forty Fifth IEEE Conference on Signals, Systems and Computers (ASILOMAR), pp. 691-695, 2011

[123] J. Li, S. Zheng, B. Chen, A.J. Butte, S.J. Swamidass, and Z. Lu, *A survey of current trends in computational drug repositioning*, Briefings in Bioinformatics 17(1):2-12, 2015

[124] L. Li, A. Scaglione, A. Swami, and Q. Zhao, *Phase transition in opinion diffusion in social networks*, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3073-3076, 2012

[125] Z. Liu et al., *In silico drug repositioning–what we need to know*, Drug discovery today 18, 110–115, 2013

[126] B. B. Mandelbrot, *The Fractal Geometry of Nature*, W.H. Freeman and Company, 1982

[127] R. Marculescu, P. Bogdan, *The chip is the network: toward a science of network-on-chip design*, Foundations and Trends in Electronic Design Automation Vol. 2, No. 4, pp.371461, 2009

[128] R. Marculescu, *Entering the Labyrinth of (Inexact) Computer Science: A Cyber-Physical Approach*, Distinguished Lecture @ ESE UPenn (April 28), 2015

[129] H. Marti-Soler, C. Hirotsu, P. Marques-Vidal, P. Vollenweider, G. Waeber, M. Preisig, M. Tafti, S.B. Tufik, L. Bittencourt, S. Tufik, J. Haba-Rubio, and R. Heinzer, *The NoSAS score for screening of sleep-disordered breathing: a derivation and validation study*, The Lancet Respiratory Medicine, 4, 9, 742–748, 2016

[130] J.C. Maxwell, *Developing the leader within you*, Thomas Nelson Publishers, 1993

[131] W.T. McNicholas, M.R. Bonsignore, and Management Committee of EU Cost Action B26, *Sleep apnoea as an independent risk factor for cardiovascular disease: current evidence, basic mechanisms and research priorities*, European Respiratory Journal, 29, 1, 156–178, 2007

[132] S.G. Memtsoudis, M.C. Besculides, and M. Mazumdar, *A rude awakeningthe perioperative sleep apnea epidemic*, New England Journal of Medicine, 368, 25, 2352–2353, 2013

[133] M. McDonald, and H. Wilson, *Marketing plans: How to prepare them, how to use them*, Wiley, 2011

[134] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, C. Diot. *Traffic matrix estimation: existing techniques and new directions*, In Proc. of Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM),pp. 161-174, 2002

[135] D. Mikhailidis, et al., *Platelet aggregation and thromboxane a2 release in primary biliary cirrhosis and effect of d-penicillamine treatment*, Prostaglandins, leukotrienes and essential fatty acids 31, 131–138, 1988

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

[137] M. Minsky, *The Society of Mind*, Appendix 4: Brain Connections, p. 314 Simon and Schuster, New York, 1988

[138] S. Moore, D. Greenfield, *The next resource war: computation vs. communication*, Proc. Intl. Workshop on System Level Interconnect Prediction (SLIP), pp.81-85, 2008

[139] L. Muchnik, S. Aral, and S.J. Taylor, *Social influence bias: A randomized experiment*, Science 341, 6146, 647-651, 2013

[140] B. Munos, *Lessons from 60 years of pharmaceutical innovation*, Nature Reviews Drug Discovery 8, 959–968, 2009

[141] P. G. Navarro, et al., *Penicillin degradation catalysed by Zn (II) ions in methanol*, *International journal of biological macromolecules* 33, 159–166, 2003

[142] N.C. Netzer, R.A. Stoohs, C.M. Netzer, K. Clark, and K.P. Strohl, *Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome*, Annals of Internal Medicine 131, 7, 485–491, 1999

[143] M. E. J. Newman, *The structure and function of complex networks*, SIAM Review 45, 167-256, 2003

[144] M.E.J. Newman, *Modularity and community structure in networks*, Proceedings of the National Academy of Sciences 103, 23, 8577–8582, 2006

[145] M.E.J. Newman, *Networks: An introduction*, Oxford University Press, 2010

[146] A. Noack, *An energy model for visual graph clustering*, International Symposium on Graph Drawing, pp. 425–436, 2003

[147] A. Noack, *Modularity clustering is force-directed layout*, Physical Review E 79, 026102, 2009

[148] T. Nugent, V. Plachouras, and J. L. Leidner, *Computational drug repositioning based on side-effects mined from social media*, PeerJ Computer Science 2, e46, 2016

[149] R. W. Numrich, M. A. Heroux, *Self-similarity of parallel machines*, Parallel Computing 37, 69-84, 2011

[150] U. Y. Ogras, R. Marculescu, *"Its a Small World After All": NoC Performance Optimization Via Long-Range Link Insertion*, IEEE Trans. VLSI Sys., 14, 7, 2006

[151] U. Y. Ogras, P. Bogdan, R. Marculescu,*An analytical approach for network-on-chip performance analysis*, Trans. Comp.-Aided Des. Integ. Cir. Sys. 29, 12, 2001-2013, 2010

[152] J. Palace, J. Newsom-Davis, B. Lecky, M. G. S. Group et al., *A randomized double-blind trial of prednisolone alone or with azathioprine in myasthenia gravis*, Neurology 50, 1778–1783, 1998

[153] F. Pammolli, L. Magazzini, and M. Riccaboni, *The productivity crisis in pharmaceutical r&d*, Nature Reviews Drug Discovery 10, 428–438, 2011

[154] P. P.Pande, C. Grecu, M. Jones, A. Ivanov, R. Saleh, *Performance Evaluation and Design Trade-Offs for Network-on-Chip Interconnect Architectures*, IEEE Trans. Comput. 54, 8, 1025-1040, 2005

[155] P.E. Peppard, T. Young, J.H. Barnet, M. Palta, E.W. Hagen, and K.M. Hla, *Increased prevalence of sleep-disordered breathing in adults*, American Journal of Epidemiology, 177, 9, 1006–1014, 2013

[156] T. M. Polasek, F. P. Lin, J. O. Miners, and M. P. Doogue, *Perpetrators of pharmacokinetic drug–drug interactions arising from altered cytochrome p450 activity: a criteria-based assessment*, British journal of clinical pharmacology 71, 727–736, 2011

[157] N.M. Punjabi, *The epidemiology of adult obstructive sleep apnea*, Proceedings of the American Thoracic Society, 5, 2, 136–143, 2008

[158] A. Rechtschaffen, A. Kales, *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*, 1968, US Government Printing Office, US Public Health Service

[159] V.A. Rossi, J.R. Stradling, and M. Kohler, *Effects of obstructive sleep apnoea on heart rhythm*, European Respiratory Journal, 41, 6, 1439–1451, 2013

[160] Z. Ruan, G. Iniguez, M. Karsai, and J. Kertesz, *Kinetics of Social Contagion*, arXiv preprint arXiv:1506.00251, 2015

[161] M. Sánchez-de-la-Torre, F. Campos-Rodriguez, and F. Barbé, *Obstructive sleep apnoea and cardiovascular disease*, The Lancet Respiratory Medicine, 1, 1, 61–72, 2013

[162] R. Santos-Silva, L.S. Castro, J.A. Taddei, S. Tufik, and L.R. Bittencourt, *Sleep disorders and demand for medical services: evidence from a population-based longitudinal study*, PLoS One, 7, 2, e30085, 2012

[163] D. Schlosshan, and M. Elliott, *Sleep 3: Clinical presentation and diagnosis of the obstructive sleep apnoea hypopnoea syndrome*, Thorax, 59, 4, 347, 2004

[164] K. Shameer, B. Readhead, and J.T. Dudley, *Computational and experimental advances in drug repositioning for accelerated therapeutic stratification*, Current topics in medicinal chemistry 15, 5–20, 2015

[165] C. E. Shardlow, et al., *Utilizing drug-drug interaction prediction tools during drug development: enhanced decision making based on clinical risk*, Drug Metabolism and Disposition 39, 2076–2084, 2011

[166] R. Sharma, M.A. Alam, and A. Rani, *K-means clustering in spatial data mining using weka interface*, International conference on advances in communication and computing technologies (ICACACT), 26, 30, 2012

[167] A. F. Shaughnessy, *Old drugs, new tricks*, BMJ 342, d741, 2011

[168] M. L. Shooman, *Reliability of computer systems and networks: fault tolerance, analysis, and design*, John Wiley & Sons, 2003

[169] G.E. Silva, K.D. Vana, J.L. Goodwin, D.L. Sherrill, and S.F. Quan, *Identification of patients with sleep disordered breathing: comparing the four-variable screening tool, STOP, STOP-Bang, and Epworth Sleepiness Scales*, Journal of Clinical Sleep Medicine, 7, 5, 467–472, 2011

[170] S. Simon, and N. Collop, *Latest advances in sleep medicine: obstructive sleep apnea*, CHEST, 142(6), 1645–1651, 2012

[171] S.H. Sleigh, C.L. Barton, *Repurposing strategies for therapeutics*, Pharmaceutical Medicine 24, 3, 151–159, 2010

[172] R. V. Sole, S. Valverde, *Information transfer and phase transitions in a model of internet traffic*, Physica A: Statistical Mechanics and its Applications, 289, 3-4, 15, 595-605, 2001

[173] A. Soule, A. Lakhina, N. Taft, K. Papagiannaki, K. Salamatian, A. Nucci, M. Crovella, C. Diot, *Traffic matrices: balancing measurements, inference and modeling*, In Proc. ACM Intl. Conf. on Measurement and Modeling of Computer Systems (SIGMETRICS), pp. 362-373, 2005

[174] A. Soule, A. Nucci, R. L. Cruz, E. Leonardi, N. Taft, *Estimating dynamic traffic matrices by using viable routing changes*, IEEE/ACM Trans. Netw. 15, 3, 485-498, 2007

[175] N. Steeghs, F. de Jongh, P. S. Smitt, and M. van den Bent, *Cisplatin-induced encephalopathy and seizures*, Anti-cancer drugs 14, 443–446, 2003

[176] L. Suciu, et al., *Evaluation of patients diagnosed with essential arterial hypertension through network analysis*, Irish Journal of Medical Science (1971-) 185, 443–451, 2016

[177] M. Takarabe, D. Shigemizu, M. Kotera, S. Goto, and M. Kanehisa, *Network-based analysis and characterization of adverse drug–drug interactions*, Journal of chemical information and modeling 51, 2977–298, 2011

[178] N. P. Tatonetti, P. Y. Patrick, R. Daneshjou, and R. B. Altman, *Data-driven prediction of drug effects and interactions*, Science translational medicine 4, 125ra31, 2012

[179] G. Tempesti, *Biological Inspiration in the Design of Computing Systems*, Proc. of the IEEE, 95, 3, 463 - 464, 2007

[180] S. K. Teo, D. I. Stirling, and J. B. Zeldis, *Thalidomide as a novel therapeutic agent: new uses for an old product*, Drug discovery today 10, 107–114, 2005

[181] C. Teuscher, *Nature-Inspired Interconnects for Self-Assembled Large-Scale Network-on-Chip Designs*, Chaos, 17(2):026106, 2007

[182] S. Theodoridou, et al., *Laboratory investigation of platelet function in patients with thalassaemia*, Acta haematologica 132, 45–48, 2014

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

[184] A. Topirceanu, M. Udrescu, A. Muresan, and S. Mihaicuta, *Network-based assessment of sleep apnea syndrome treatment response*, European Respiratory Journal, 44, Suppl 58, P2215, 2014

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

[186] A. Topirceanu, M. Udrescu, R. Avram, and S. Mihaicuta, *Data analysis for patients with sleep apnea syndrome: A complex network approach*, Soft Computing Applications, pp. 231–239, Springer, 2016

[187] B. Towles, W. J. Dally, S. Boyd, *Throughput-centric routing algorithm design*, In Proc. 15th ACM Symp. on Parallel Algorithms and Architectures (SPAA), 200-209, 2003

[188] B. Towles, W. J. Dally, *Worst-case Traffic for Oblivious Routing Functions*, IEEE Comput. Archit. Lett. 1, 1, 4-4, 2002

[189] W.-C. Tseng, J. Hu, Q. Zhuge, Y. He, E. H.-M. Sha, *Algorithms for optimally arranging multicore memory structures*, EURASIP J. Embedded Syst., Article 1, 2010

[190] L. Udrescu, L. Sbârcea, A. Topîrceanu, A. Iovanovici, L. Kurunczi, P. Bogdan, M. Udrescu, *Clustering drug-drug interaction networks with energy model layouts: community analysis and drug repurposing*, Scientific Reports, 6, 32745, 2016

[191] M. Udrescu and A. Topîrceanu, *Probabilistic modeling of tolerance-based social network interaction*, 3rd IEEE European Conference on Network Intelligence (ENIC), pp. 48–54, 2016

[192] M. Udrescu, A. Topirceanu, R. Avram, Ş. Mihăicuţă, *AER score: a social-network-inspired predictor for sleep apnea syndrome*, Chest 145, 3S, 609A–1, 2014

[193] K.T. Utriainen, J.K. Airaksinen, O. Polo, O.T. Raitakari, M.J. Pietilä, H. Scheinin, H.Y. Helenius, K.A. Leino, E.S. Kentala, J.R. Jalonen, H. Hakovirta, T.M. Salo, and T.T. Laitio, *Unrecognised obstructive sleep apnoea is common in severe peripheral arterial disease*, European Respiratory Journal, 41, 3, 616–620, 2013

[194] E. Vagiakis, F. Kapsimalis, I. Lagogianni, H. Perraki, A. Minaritzoglou, K. Alexandropoulou, C. Roussos, and M. Kryger, *Gender differences on polysomnographic findings in Greek subjects with obstructive sleep apnea syndrome*, Sleep Medicine, 7, 5, 424–430, 2006

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

[196] M. A. Valentovic, and W. C. Lubawy, *Impact of insulin or tolbutamide treatment on 14c-arachidonic acid conversion to prostacyclin and/or thromboxane in lungs, aortas, and platelets of streptozotocin-induced diabetic rats*, Diabetes 32, 846–851, 1983

[197] S. Valverde, R. V. Sole, *Self-organized critical traffic in parallel computer networks*, Physica A, 312(3-4), 636 - 648, 2002

[198] G.D. Vavougios, G. Natsios, C. Pastaka, S.G. Zarogiannis, and K.I. Gourgoulianis, *Phenotypes of comorbidity in OSAS patients: combining categorical principal component analysis with cluster analysis*, Journal of Sleep Research, 25, 1, 31–38, 2016

[199] D. J. Watts, and S. B. Strogatz, *Collective dynamics of small-world networks*, Nature 393, 6684, 440-442, 1998

[200] X. F. Wang, and G. Chen, *Complex Networks : Scale-Free and Beyond*, IEEE Circuits And Systems Magazine, 3(2), 6-20, 2003

[201] Z. Wang, J. Xu, X. Wu, Y. Ye, W. Zhang, M. Nikdast, X. Wang, and Z. Wang, *Floorplan optimization of fat-tree-based networks-on-chip for chip multiprocessors*, IEEE Transactions Computers, 63(6), 14461459, 2014

[202] Z. Wang, J. Xu, P. Yang, L.H. Kinh Duong, Z. Wang, X. Wang, Z. Wang, H. Li, R. Kioji, and V. Maeda, *A Holistic Modelling and Analysis of Optical-Electrical Interfaces for Inter/Intra-chip Interconnects*, IEEE Transactions on Very Large Scale Integration Systems, 24(7), 2462-2474, 2016

[203] W. Weidlich, *Sociodynamics-a systematic approach to mathematical modelling in the social sciences*, Nonlinear phenomena in complex systems - Minsk - 5, 4, 479–487, 2002

[204] P. D. Windschitl, J. P. Rose, M. T. Stalkfleet, and A. R. Smith, *Are people excessive or judicious in their egocentrism? A modeling approach to understanding bias and accuracy in people's optimism*, Journal of personality and social psychology 95, 2, 253, 2008

[205] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, *DrugBank: a comprehensive resource for in silico drug discovery and exploration*, Nucleic Acids Research, 34, S1, D668–D672, 2006

[206] L. Ye, G.W. Pien, S.J. Ratcliffe, E. Björnsdottir, E.S. Arnardottir, A.I. Pack, B. Benediktsdottir, and T. Gislason, *The different clinical faces of obstructive sleep apnoea: a cluster analysis*, European Respiratory Journal, 44, 6, 1600–1607, 2014

[207] E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, and A. Scaglione, *Binary opinion dynamics with stubborn agents*, ACM Transactions on Economics and Computation 1, 4, 19, 2013

[208] T. Young, P.E. Peppard, D.J. Gottlieb, *Epidemiology of obstructive sleep apnea: a population health perspective*, American Journal of Respiratory and Critical Care Medicine, 165, 9, 1217–1239, 2002

[209] G. Zahoránszky-Kőhalmi, C. G. Bologa, and T. I. Oprea, *Impact of similarity threshold on the topology of molecular similarity networks and clustering outcomes*, Journal of cheminformatics 8, 1, 2016

[210] B. Zhou, R. Wang, P. Wu, and D.-X. Kong, *Drug repurposing based on drug–drug interaction*, Chemical biology & drug design 85, 137–144, 2015

[211] ***, *www.drugs.com* [last accessed April 2016]

[212] ***, *www.rxlist.com* [last accessed April 2016]