Evolutionary Dynamics of Social Behaviours on Multilayer Networks

Ing. Alessandro Di Stefano
Dipartimento di Ingegneria Elettrica, Elettronica e Informatica
Università degli Studi di Catania
Tutor: Chiar.mo Prof. Ing. Aurelio La Corte
Coordinator: Chiar.mo Prof. Ing. Luigi Fortuna

A thesis submitted for the degree of
Ph.D. in Systems Engineering (XXVIII CYCLE)

December 2015

To my love Damiana and my Family

"By striving to do the impossible, man has always achieved what is possible. Those who have cautiously done no more than they believed possible have never taken a single step forward." (M. A. Bakunin)

Acknowledgements

First of all, I have no words to express my gratitude to my tutor, Prof. Ing. Aurelio La Corte, for his invaluable support and wise counsels during my Ph.D. experience and for giving me the opportunity to participate at stimulating international conferences and schools. He has been also more than a guide, a reference person for all these years. Then, I have to thank Prof. Ing. Luigi Fortuna for having given me insightful comments and pushed me to do my best during my Ph.D. studies. Also, I wish to express my thanks to my friend and collaborator Dott. Ing. Marialisa Scatà for having accompanied me along this wonderful journey and for collaborating with me through all my Ph.D. course. She has been a tireless guide and an enthusiastic collaborator, transmitting her love towards science and constantly encouraging me to push myself further. Furthermore, I must thank Prof. Pietro Liò, senior reader at the Computer Laboratory of the University of Cambridge, for his precious collaboration since 2012, with whom the most of the research presented in this Ph.D. thesis has been carried out. He has represented for me an incredible guide with a large amount of constructive feedbacks and precious advices at every stage of my research, making increase more and more my passion for research and suggesting me continously new ideas and topics to focus on. I would like to thank him also for having given me the opportunity to attend two visiting periods at the Computer Laboratory. I would like to thank all the members of the Pietro's group, who have warmly welcomed and supported me as a member of the group. Last but not least, I would like to thank my family: my parents, my incredible big brother Dario, and my grandparents, which has been supporting me in all my life, and my love, Damiana, able to give me love, joy, happiness and light every day of my life. Finally, I have to thank all my friends and the people that have been supporting me these last years.

Abstract

Nature shows as human beings live and grow inside social structures. This assumption allows explaining and exploring how it may shape most of our behaviours and choices, and why we are not just blindly driven by instincts: our decisions are based on more complex cognitive reasons, based on our connectedness on different spaces. Thus, human cooperation emerges from this complex nature of social network. One of the main aims of my Ph.D. dissertation is to explore how and why it happens, hence the work is mainly focused on studying the evolutionary dynamics of cooperation and social behaviours on a multilayer social network. Following a Bio-inspired approach, the social network analysis methodologies, and exploiting the mathematical framework of Evolutionary Game Theory (EGT), the target is to unveil the hidden dynamics, observe non-trivial patterns, finding out the hidden emergent behaviours in a population. The study of cooperation and its evolutionary dynamics on a social network, has raised up the need of a model able to explain the actual complexity of real-world networks, where individuals are connected through multiple types of relationships. For this reason, the mathematical framework of multilayer networks has been exploited, indeed it allows us to encompass these several interactions and relationships, exploring and unveiling how the different ties in the various layers can impact on the emergence of social behaviours in a population. Therefore, the presence of the same nodes in multiple layers of a system, known as multiplexity, is the key to understand emergent phenomena, adding an extra dimension of analysis which explains what is the role not only of the intralayer interactions, as in a single-layer framework, but also of interlayer interactions for the emergence of these phenomena. Furthermore, it is explored and quantified the role of some shaping factors, such as homophily, in this evolutionary process. Taking into account all these aspects, a novel

analytical model is proposed, together with a simulative investigation of the evolution of human cooperation using mutliplex evolutionary game theory, shedding light on the key role played by homophily and multiplexity in the evolution of cooperation.

Furthermore, the analysis of social behaviours in a multilayer network, along with the large amount of available data, the Big Data revolution, is also exploited to design a novel multilayer structure in healthcare context, giving rise to a smart healthcare system. In fact, together with the multidimensional approach to comorbidity, the inclusion of a social dimension of analysis allows finding out correlations and causality relations between diseases and patients also considering the connectedness and social contagion processes. In this way, we obtain an evolution from data to multi-agents through the introduction of personalised medicine in a multilayer architecture. The multilayer paradigm is built up on a health mining approach, which introduces methodologies for fusing, integrating and drawing inference from a plurality of heterogneous data, extracted from different sources, creating a complex data type, containing all the information enclosed in the various dimensions. The introduction of a social dimension in healthcare, the study of the evolution of behaviours, the social networking, Big Data and ICT strategies, considering a multilayer structure, allow extracting an "organised" knowledge, which connects, relates and computes all the facets of healthcare data. The Bio-inspired and the social network evolution, based on the organised knowledge, provides the basis for a new Bio-inspired Information and Communication Technology (ICT) paradigm, enclosing social, human, and cognitive aspects, other than the self-capabilities and context-awareness. Nodes and data enclose all these features, and the social multilayer network allows us to analyse the complex dynamic patterns involving these entities, highlighting the impact of social interactions and contagion processes, and increasing the overall level of awareness, from simple things to knowledge through the social objects. ICT interventions are the result of the multilayer analysis, the context and the system as a whole. This evolution process leads to a bio-inspired network-driven ICT, re-designing the ICT communication paradigm.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Evolution of Social Behaviours on Multilayer Networks: An Overview

One of the main properties of a complex system is that a large number of simple elements gives rise to collective phenomena impossible to predict or anticipate considering an individual unit. Ant colonies, biological systems, social networks are some examples of the emergent complexity. It raises up the following question: what is behind the emergent complexity? What are the mechanisms allowing to transform people into complex societies or simple cells into a complex entity? The answer is the connectedness of the simple units, forming a more complex entity, which is more than just the sum of their parts, known as *network*. In the last decades, we have witnessed with an incredible development of data-driven mathematical models, which have created a new discipline, named *'network science'* (1). Network theory has demonstrated to be the most suitable way to investigate the structural patterns of the interactions among the constituent elements of a variety of complex systems, also composed by a large number of elements, such as social networks, the brain and biological networks (2, 3, 4, 5). Analysing the spectral properties of the adjacency and Laplacian matrix of a network is possible to gain insight on the structure and dynamics occurring on the network (6). In recent years, a vast literature has witnessed with the development of tools and models to get a better understanding of how these networks may evolve. A lot of work has focused on social network properties and dynamics and, thanks to the technological advances with regards to real data acquisition both from real and online social networks, nowadays there is the availability of a huge quantity of data, called as *Big*

Data, which is expected even to grow in the next decades. The increasing resolution of data gathering techniques allows capturing new properties of the interaction patterns in complex systems, including the spatial, temporal and multiplex nature of interaction networks.

In fact, advances in the complex systems field have underlined that, in order to gain insight on the complexity of the large variety of systems, the description in terms of single networks is an oversimplification, which fails into capturing the dynamics of network patterns deriving from the simultaneous interactions of more than just the one network. For this reason, *multilayer networks* have been introduced, distinguishing the different kinds and channels of interactions between nodes through the layers. Depending on the interaction between the different systems or layers, we can discern different types of multilayer networks, e.g. multiplex networks are composed by the same nodes interacting trough different layers. Multilayer networks and their structural and dynamical features, have greatly attracted the interest in network science recently, resulting in a number of works about the structure and dynamics of multilayer and multiplex networks (7), demonstrating that the behaviour of interacting complex systems is very different from a simple combination of the isolated cases (6). An individual's behaviour can be different in each layer, even if it is conditioned by all of them (8).

Network science and multilayer networks allow describing interactions among non-trivial and complex entities, such as humans in a social network. But, what kind of behaviours may emerge from these interactions in a social network? Is it possible to solve the human conflict between the benefit of the single individual and that one of the population, such as risking one's life to save a stranger, or finding an equilibrium among the interests that enable them to cooperate toward a common good?

This pushed my motivation to study a standard approach used to solve such situations: *Game Theory.* Game Theory has been applied to various fields, from economy to biology, other than in computer science, such as communication networks, security, power control issues in wireless scenarios, distributed systems (e.g. peer-to-peer networks), artificial intelligence, where agents have to negotiate between them in order to coordinate a collective action. It allows modelling those situations where the individual behaviours and decisions affect the outcomes of others in the same environment, that is the conflictual scenarios where there is a strategic behavioural interdependence among individuals, affecting the dynamics of behaviours. The analysis of the dynamic patterns raises the challenge of the evolution of cooperation and how it can evolve and survive in various scenarios. *Cooperation* is the act where individuals can contribute something, at a cost to themselves, to provide a benefit for others. It is a widespread phenomenon in natural and social systems, but not fully-understood mainly due to its complexity. Cooperation, representing the most important challenge to Darwin's theory of evolution, is crucial to understand the evolutionary dynamics and transitions that lead from single-cell organisms to complex animal and human societies. Thus, understanding the evolution of cooperation remains a key challenge attracting the interest of research across the social and natural sciences.

To address the conundrum of human cooperation and understand the evolution of social behaviours within a population in networks, it is important to have a mathematical framework to capture these underlying mechanisms. *Evolutionary Game Theory (EGT)* provides a powerful theoretical framework to investigate strategic choices in a huge variety of complex systems (9). Despite the numerous application areas, however, the main fundamental problem that is studied in the realm of EGT is the exploration of the evolution of cooperation (10, 11) and cooperative behaviours in systems consisting of competitive individuals (12, 13, 14). EGT allows studying interactions of multiple nodes in a population, and find out the hidden dynamics, shedding light on how and why some behaviours emerge following a specific pattern, and which behaviours are able to persist in the population and those having a tendency to be driven out by others. To describe and study the problem of evolution of cooperation, the social dilemmas, such as Prisoner's Dilemma Game (PDG), are typically used (15). Social dilemmas represent situations where individuals face the conflictual situations between what is best for them and what is best for the society. In the Prisoner's dilemma, at each instance two players have to decide simultaneously whether they want to cooperate or defect. The dilemma is given by the fact that although mutual cooperation yields the highest collective payoff a defector will do better if the opponent decides to cooperate. A lot of research has been devoted to the identification of mechanisms that may lead to a cooperative resolution of social dilemmas, such as kin selection, direct and indirect reciprocity, network reciprocity, as well as group selection (16).

Humans are part of different social networks and live inside them. This could have important consequences for the evolutionary dynamics of social behaviours, mainly due to the multilayer nature of their adopted strategies. The presence of nodes in multiple layers of a system, together with their social connectedness, is the key to understand emergent phenomena and how the evolution of cooperation can work on top of this multilayer structure. One of the main targets becomes to try to answer some key open challenges, such as: how did the selfish process of natural selection gives rise to cooperation? how might social interactions can give a boost to cooperative behaviour? Despite some studies in EGT have already shed light on the promotion of cooperation (10, 11, 17, 18), the puzzle between the Darwinian fundamental assumptions of natural selection described in "The Origin of Species", for which "only the fittest survive" and the cooperation observed in human and animal societies remains unsolved (19).

1.2 Research Questions

This Ph.D. dissertation addresses the issue of evolutionary dynamics of social behaviours on multilayer networks. Thus, how evolutionary game theory and multilayer social networks, with its structural and dynamical features and considering the different kinds and channels of social interactions between nodes, can improve the investigation of the emergent dynamics of behaviours in a population. The novel approach is crucial to improve the analysis of evolutionary dynamics and addresses the following main research questions:

- Some of the main requirements of an Information and Communication Technology (ICT) system, such as the high complexity and connectivity, the reliability, the growing information load, risk management and energy-saving issues, under conditions of limited computational resources, limited time constraints and low overall knowledge, claim for a new modeling approach, able to face all these needs, getting a "satisficing" or sub-optimal solution. If, at a first glance, it might seem strange to look at biology as an inspiration for research related to networking and ICT fields, however, the two research fields show a deeper connection than one might expect. In fact, the Internet and, more in general ICT, has a lot of features very close to those of biological systems (20). **To what extent can biological processes and models inspire the analysis and management of the complexity of the ICT systems, increasing knowledge and trying to solve also the associated computational problems?**

- Real-world complex systems, such as social networks, suggest how nodes or agents usually interact using different ways and channels, thus a single layer network is not enough to represent and describe the complex set of multiple interactions between different networks. For instance, a social network can be described as a set of people who interact through some patterns. In general, the connections, which identify relationships between nodes on a network, have been considered at the same level. However, in a real context, this assumption is not true. In fact, in a real social network, different and not mutually exclusive relationships can be considered between the same two people (e.g. friends, relatives, colleagues, etc.), therefore the actual interactions and relationships cover different levels or layers. This concept introduces a new point of view of the social network analysis, and shows how the nodes and their relationships must be considered and weighed on different contexts (or layers) and, at the same time, the weight of their decisions and their behaviours has an impact on so many different levels. In order to gain a better understanding of the complexity and how these complex systems function and evolve, multilayer networks become the more popular and natural paradigm. **What are the main properties and features of multilayer networks? How can multilayer networks help to represent and investigate the dynamics and patterns of real-world networks?**

- ICT systems are continuously changing as a result of innovation, the increasing connectedness and the large amount of data. The Big Data revolution requires novel ICT models and paradigms, and poses new issues and challenges, in terms of heterogeneity and computational analysis. This is affecting not only the technologies, but also human, social and economic aspects. Furthermore, the introduction of the Internet of Things (IoT) paradigm has transformed simple things in interconnected smart objects, able to measure, monitor, detect events and human activities, understand and cooperate in order to solve a problem. In the era of high connectedness, users are becoming more active in interacting, sharing and collaborating, through social networks. In this way, this process produces a collective intelligence, spread out in many different areas and related to networks' phenomena. **Despite the huge amount of heterogeneous data to be analysed, the increased connectedness and complexity of social networks,**

how can ICT systems and models be able to manage them, capturing the complex dynamic patterns, and increasing both performance and context-awareness?

- The issue of model and explore individual and social behaviours in a context or setting where the outcomes depend on the behaviour of others, is fundamental and become also more interesting in network contexts. Furthermore, the increased number of nodes in networks and the multilayer description of interactions through many different layers, make the setting even more complex and difficult to analyse and model analytically. The study of the dynamic evolutionary patterns emerging from these complex systems represents an even more challenging question. **What is the more suitable mathematical formalisation to analyse social interactions and behaviours? What are its main properties and features? What are the behaviours having the ability to persist in the population, and which ones instead have a tendency to be driven out by others?**
- Humans tend to cooperate building complex societies, as well as predators hunt in groups to catch more preys as possible (8). More in general, cooperation is an act where individuals can contribute something, at a cost to themselves, to provide a benefit for others. We can find situations where this kind of act can arise at almost every layer of human societies and also in the animal world. Many models and mechanisms have been proposed to explain the emergence and evolution of cooperation by studying interactions in a population. Nevertheless, the evolution of cooperation among individuals remains an unsolved puzzle: it has being observed since ancient times but, only in the recent years, a lot of research efforts have been done trying to understand and deepen the origin inside social networks. A vast literature on the evolution of cooperation on complex networks (19, 21, 22, 23) highlights many aspects which offer insights on how cooperation can evolve and survive in different scenarios (24, 25, 26). To study cooperation and its evolutionary dynamics, we need to understand the impact of the structure and the nature of social relationships among individuals. **Therefore, how did the selfish process of natural selection, described by Charles Darwin, gives rise to cooperation? How might social interactions can give a**

boost to cooperative behaviour? And what may be the role of a linkage polariser, such as homophily, in this evolutionary process?

1.3 Methodology

I focused my research on an interdisciplinary approach, since I believe that the best ideas should address topics and issues from different fields, combining two or more academic fields in an innovative way, overcoming the traditional boundaries between apparently disconnected research areas. This methodology gave me the advantage of covering many research topics and obviously facing with many interesting issues, whose meaning can be revealed only breaking down barriers and thinking them as one largest and most complex issue.

In particular, I started my Ph.D. in Systems Engineering pushed by curiosity and desire to deepen the bio-inspired approach and algorithms, which were issues I dealt with during my MSc thesis. I studied and proposed a novel bio-inspired model for converged networks (e.g. NGN, NWGN, based on self-organisation and biodiversity. The idea was to exploit biodiversity to design a multilayer approach to security, introducing a self-protecting module in each node and an architecture consisting of three security layers (node, community and ecosystem), suitable to increase the security degree of the emerging paradigm of networks. The MSc thesis experience allowed me to investigate a huge variety of bio-inspired models and I was really wondered and fascinated by how such microscopic interactions could affect macroscopically the system with good performance. In particular, how the interactions of many simple self-organised agents, for instance imitating the behaviour of bees or ants (e.g. swarm intelligence), were able to push towards a sub-optimal solution, thanks to the emerging intelligence from these interactions of many simple agents.

Then, I began asking myself a number of questions about studying and analysing these interactions among entities: one of the main targets became to study and focus on what could it happen when the interaction is among non-trivial and complex entities, such as humans within a social network. What kind of behaviours can emerge from these interactions in a social network? Is it possible to solve the human conflict between the benefit of the single individual and that one of the population, such as risking one's life to save a stranger, or finding an equilibrium among the interests that enable them to

cooperate toward a common good? This raised up my interest in studying a standard approach used to solve such situations: Game Theory. On the other hand, the idea to better understand the dynamics of interactions in human societies led me to focus on exploring properties and dynamics of social networks. Social network analytics (27, 28, 29, 30) and the availability of a huge amount of data and simultaneously the technological advances with regards to data acquisition both from real and online social networks, enabled to investigate dynamic patterns and the evolution of social behaviours, such as human cooperation, in various scenarios. To better understand the evolution of social behaviours and, in particular, human cooperation in a population in a social network, I decided to concentrate upon the mathematical framework of Evolutionary Game Theory, able to capture the underlying mechanisms and the hidden dynamics, and shedding light on how and why some behaviours emerge following a specific pattern. To deal with the complexity of social interactions, I started exploiting the paradigm of multilayer networks, since the presence of nodes in multiple layers of a system is the key to understand emergent phenomena, adding an extra dimension explaining what is the role not only of the intralayer interactions, as in a monoplex framework, but also of interlayer interactions for the emergence of these phenomena.
The final target of my research is to study social behaviours, analyse the emergence and their evolution on a multilayer network. To do this, first I searched, studied and evaluated numerous scientific contributions, related to the different issues involved in this field. The following is a list of my research keywords, corresponding to the topics of interest that I have studied during my Ph.D. and on which I'm going to focus also in the next future. The following is a list of the main macro-areas of this Ph.D. dissertation:

- Bio-Inspired ICT
- Multilayer networks
- Social networks
- Game Theory and Evolutionary Game Theory (EGT)
- Human cooperation
- Multiplex EGT

My research work has been articulated over the years in different and logically linked steps. I started my Ph.D. using a bio-inspired approach applied to ICT, then I exploited multilayer networks, social networks methodologies and game-theoretic approach to describe the complexity of the real-world scenarions, and study interactions among individuals. In particular, I focused on human cooperation issue and, through multiplex EGT, I have quantified the role of multiplexity and of homophily, as a shaping factor of social interactions, in the emergence of cooperation in a population.

1.4 Dissertation Outline

The structure of this Ph.D. dissertation is as follows:

1. **Chapter 2** discusses the main concepts of bio-inspired approach and algorithms, and shows how the interplay between bio-inspired approach and Information and Communication Technology (ICT) gives the opportunity to use algorithms, tools and analytical models able to optimise and improve the design and management of methodologies proper of traditional vision of ICT. A bio-inspired approach allows solving certain problems and meet specific requirements, such as reliability, information load, risk management and energy saving, under conditions of limited computational resources, time constraints and low overall knowledge. To prove the importance of a bio-inspired approach to ICT, two algorithms for Wireless Sensor Networks (WSNs), both based on heuristics, are proposed, respectively an IoTs clustering algorithm (IMLM) and a energy-aware routing algorithm.

2. **Chapter 3** reviews the main structural notions, models, properties and measures of the mathematical framework of multilayer networks. At the beginning, the motivations behind the transition from Single- to Multi-layer Networks are explained. In fact, multilayer networks are a recent mathematical framework which has been introduced to explain and understand the complexity of a huge variety of real-world systems, since single networks are not enough to describe the complex set of interactions between different networks. Some representative examples of the major constraints of the traditional single networks are presented, stressing the importance to use a multilayer approach. Finally, in the second part is focused on the structural properties and measures in multilayer and multiplex networks.

3. **Chapter 4**, leveraging the bio-inspired approach to ICT, deals with the Big Data and the data-intensive computing issues in the future vision of a smart healthcare. To this aim, a multidimensional approach to comorbidity and the introduction of a social dimension of analysis, allow finding out correlations and causality relations between diseases and patients also considering the connectedness and social contagion processes. The proposed multilayer architecture represents an evolution from data to multi-agents through the introduction of personalised medicine, giving rise to a smart healthcare paradigm. The last part is devoted to present the bio-inspired and social evolution of nodes and data in a multilayer network, with an increasing level of awareness, and the development of an organised knowledge.

4. **Chapter 5** focuses on game theory, a mathematical tool able to describe and analyse the strategically interdependent interactions among individuals. In particular, the first part of the chapter is intended to define some of the main notions underlying the classical game theory, while the second part deals with the framework of Evolutionary Game Theory, allowing to explore the evolutionary dynamics of behaviours in a population.

5. **Chapter 6** represents the main focus of this dissertation, joining the mathematical frameworks of multiplex networks and EGT, the bio-inspired approach and the social network analysis methodologies, in order to find out the hidden emergent behaviours within a population across network layers. The investigation of evolutionary dynamics through Evolutionary Game Theory on multiplex networks allows unveiling and studying the existing social conflicts and dilemmas among the interests of the single nodes and groups, their counterparts in various layers, not neglecting what is captured from homophily, the patterns of similarity and dissimilarity. The simulations, conducted both macroscopically and microscopically across the network layers in the multiplex, show quantitatively the role of homophily in human cooperation.

6. **Chapter 7** concludes by revisiting the research questions posed in Sect. 1.2, and summing up the main contributions of this dissertation, other than highlighting some key aspects to be investigated in the future research.

1.5 Publications

- **Di Stefano, A., La Corte, A., Leotta, M., Liò, P., Scatà, M. (2013)**. *"It measures like me: An IoTs algorithm in WSNs based on heuristics behaviour and clustering methods"*. Ad Hoc Networks, 11(8), 2637-2647.
- **La Corte, A., Di Stefano, A., Scatà, M., Leotta, M. (2013)**. *"A Energy-Preserving Model for Wireless Sensors Networks based on Heuristic Self-Organized Routing"*. In: IEEE International Conference on Systems, Man and Cybernetics (IEEE SMC), 2013.
- **Scatà, M., Di Stefano, A., Giacchi, E., La Corte, A., Liò P. (2014)**. *"The Bio-Inspired and Social Evolution of Node and Data in a Multilayer Network"*. In: 5th International Conference on Data Communication Networking (DCNet), 2014.
- **Di Stefano, A., La Corte, A., Scatà, M. (2014)**. *"Health Mining: a new data fusion and integration paradigm"*. In: 11th International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics (CIBB), 26-28 June, 2014.
- **Giacchi, E., Di Stefano, A., La Corte, A., Scatà, M. (2014)**. *"A dynamic context-aware multiple criteria decision making model in social network"*. In: Information Society (i-Society), 2014 International Conference on, pp. 157-162, 10-12 Nov. 2014 doi: 10.1109/i-Society.2014.7009032.
- **Di Stefano, A., La Corte, A., Scatà, M. (2015)**. *"A Novel Multi-agent Social Multilayer Framework for Improving Health Information Exchange and Management"*. Poster presentation, NetSci 2015, Zaragoza (Spain), 1-5 June, 2015.
- **Catania, E., Di Stefano, A., Guardo, E., La Corte, A., Pagano, S., Scatà, M. (2015)**. *"Energy Awareness and the Role of "Critical Mass" In Smart Cities"*. International Refereed Journal of Engineering and Science (IRJES) ISSN (Online) 2319-183X, (Print) 2319-1821 Volume 4, Issue 7 (July 2015), pp. 38-43.

- **Di Stefano, A., Scatà, M., La Corte, A., Liò, P., Catania E., Guardo E., Pagano S. (2015)**. *"Quantifying the Role of Homophily in Human Cooperation Using Multiplex Evolutionary Game Theory"*. PLoS ONE 10(10): e0140646. doi: 10.1371/journal.pone.0140646
- **Di Stefano, A., La Corte, A., Liò, P., Scatà, M. (2016)**. *"Bio-Inspired ICT for Big Data Management in Healthcare"*. Springer Book Chapter in: "Intelligent Agents in Data Intensive Computing" - "Studies in Big Data", Vol. 14, Springer book series in Computer Science. ISSN: 2197-6503.
- **Giacchi, E., Corrente, S., Di Stefano, A., Greco S., La Corte A., Scatà, M. (2016)**. *"A context-aware and social model of preferences dynamics through a multiple criteria decision making framework"*. Decision Analytics - Springer (submitted).
- **Scatà, M., Di Stefano, A., La Corte, A., Liò, P., Catania, E., Guardo, E., Pagano, S., (2016)**. *"Combining Evolutionary Game Theory and Network Theory to analyze Human Cooperation patterns"*. Chaos, Solitons and Fractals, Elsevier (submitted).

Chapter 2

Bio-Inspired ICT

2.1 Introduction

The potential consequences that may arise as a result of innovation and widespread of future Information and Communication Technologies (ICTs) are raising up the interest in this research field. ICT should be planned and designed to conduct almost any activity which involves logical strategies and operations in a variety of application areas. ICTs, represented in services, logical strategies, infrastructures, methodologies, interventions and platforms, are expected to contribute to the realisation of a sustainable and smart society, because of its strong influence and impact on the environmental, social, economic and technology aspects of our lives. Biological systems act as an inspiration for research related to networking and ICT fields, in fact the two research fields show a deeper connection than one might expect. In fact, the Internet, and more in general ICT, has a lot of features very close to those of biological systems (20). The natural world is enormous, dynamic, heterogeneous and highly complex. Nevertheless, biological organisms are able to survive, self-organise, and evolve in a such a complex and challenging world, exploiting only the local knowledge, and not with a centralised control. A continuously increasing and similar complexity is experienced by computer networks, since they are becoming larger and more interconnected, even if the same extent of robustness, adaptability, and scalability is reached. A lot of research efforts have been made on studying these analogies between the two worlds, and evaluated if there are some concepts and approaches to be derived from biological systems. Future ICT requires an innovation in terms of novel architectures, capable of integrating high computational capabilities, extracting knowledge from data and from smart communications, social-based networking, and developing context-aware platforms. Information Systems (IS) engineering technologies are increasingly becoming embedded into our job and home lives, creating a complex interdependence between people and technology since individuals and groups communicate, collaborate and exchange knowledge through a variety of ICT systems (31). ICT may be seen as the technical aspect of the socio-technical systems which are used in society and organisations (31).

Research and innovation in some topics such as Big Data, data-intensive computing, context-awareness, social networking, provide the basis for the future challenges of ICT. Traditionally the research, design, and applications in the ICT field involve the development of methodologies and tools, and it is characterised by features meant to support specific tasks, linked to specific contexts of technologies. The complexity beyond the new issues, for example, the heterogeneity of data and objects, the multitude of information sources, the high connectedness, complexity of systems and the related intensive computing, requires to transform the basis of ICT following the real computational nature of world around us. The plurality of resources in terms of data, nodes, communication paradigms and smart infrastructures, should coexist in the same space and merges coherently in order to produce knowledge, as the real representation of world needs and behaviours. ICT solutions, if wisely applied, may increase the success of organisations, the efficiency and transparency of governments (32).

The impact of ICT on enhancing the quality of services and decreasing the overall costs has been the focus of numerous studies in the last two decades (32). This confirms the importance which covers the ICT on following the evolutionary trends on social, economic and technological aspects. Through its influence and the resulting impact, ICT should be able to mine and convey knowledge into practices and methods to improve the performance of tasks related to growing scientific and socio-economic interests. The challenge of ICT is to extract knowledge as best as possible from information sources, by using the actual technologies and analysing the real world scenarios, translating everything into actions, interventions, strategies and innovative platforms, addressing to the complex systems as well as users.

The bio-inspired approach, when applied to ICT, gives the opportunity to use algorithms, tools and analytical models to optimise and improve the design and management of methodologies proper of traditional vision of ICT. This could introduce a new assessment perspective, considering innovative approaches and, at the same time, new challenges linked to the new issue of finding suitable models to evaluate and translate into computational terms the need to extract as much knowledge as possible from the systems we know. The target is to make this knowledge available to those who have to design ICT interventions and services, considering the multitude of resources, in terms of data and sources, computational limits and social dynamics. ICT systems are evolving toward innovative approaches by investing more on developing new fields such as context-aware, social networking, multilayer networks, evolutionary game theory, smart mobility, dynamic complex systems, personalisation of models and services, smart platforms and services, with the aim of providing, not only good user experiences and user expectations, but for the future, also the profiling of all his features. The goal of the research for the future ICT, by drawing inspiration from bio-inspired models, is to rewrite and redesign networks nodes, as information sources, and data, as packet information of global knowledge, finding consistency starting from the heterogeneity, ubiquity, dynamicity of the new social-based complex systems.

2.2 Bio-inspired and ICT: analogies and interplay

Surprisingly, ICT and biological systems have a lot of common features (20). One of the analogies, for instance, is related to the similar architecture, as the "hourglass" model, typical of Internet protocol stack, has a structure close to many biological systems. In the hourglass model, one layer exploits the set of lower layer protocols but, at the same time, protocols that run on top provide new additional information and functions exploited by the above layers in the stack. In a biological system, for instance, bacteria eat severe different nutrients; all of these nutrients contain some or all of the raw building blocks needed to power a bacterial cell, even if a bacterium must first metabolize these nutrients before using them, reassembling the building blocks into the multitude of complex macromolecules required for survival.

As well as it is almost impossible to build a different version of application layer adapted to a different physical layer technology, it would be unthinkable to use different metabolic processes to convert nutrients to the macromolecules it needs. In analogy with the Internet model, all nutrients are converted into a small number of common currencies. These currencies are then used to build the large number of complex macromolecules required to power the cell. Considering the hourglass metaphor, the so-called "bow tie" structure is a nearly universal feature of complex systems. The acknowledgement mechanism ensembles a form of integral feedback: it allows regulating the transmission speed of packets through the Internet. Biological systems use an analogous type of feedback to rule various processes between cells, e.g. bacteria rule the speed and direction of movement when tracking the concentration of certain chemicals in their environment (20).

Other significant similarities emerge taking into consideration both social networks and biological systems, as they represent two examples of complex systems. We find a lot of entities (nodes, proteins, etc.), connected to each other in several ways and also for very different reasons. They interact using both weak and strong ties (33), inducing some other actions inside the communities they belong to and, more in general, also in whole network. Furthermore, both the Internet and biological systems are large scale networks, and they show a complex, strictly organised internal structure. The human body has many different organs and physiological systems, each of which serves a specific target. The Internet also contains a number of specialised devices: in the core network the high speed routers forward data in a highly optimised manner, while at the edges of the network there are application-oriented devices, such as laptop computers and cellular phones. As well as an high-speed router is very important to forward an important message through the network, the kidney is fundamental in oxygenating your blood. In addition, complex systems are robust against perturbations or expected failures but not if we consider an unexpected one. In particular, interdependent systems increase their vulnerability and become increasingly prone to cascade failures due to the coupling factor between the different blocks or layers they consist of (34). A social network tends to emphasize the significance of a particular event when it has a great spreading factor; similarly, a biological system shows an extremely high vulnerability when, for example, a disease is able to travel quickly corrupting the body cells, involving different organs.

The analogies between the biological and networking systems has been exploited in so many different research fields, such as social insects (35), artificial immune systems, the fireflies synchronisation, the transport networks inspired by *physarum polycephalum*,

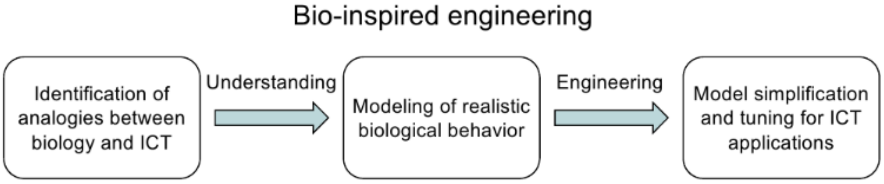


Figure 2.1: Bio-inspired Engineering. Steps needed to adapt biological mechanisms to technical engineered solutions.

epidemic spreading models, etc. Therefore, examining some of the most common structures and algorithms used in telecommunications networks, it is easy to find out striking analogies with the biological systems. In fact, the evidence suggests that Nature and engineers not only have to solve similar problems but regardless they come to surprisingly similar solutions. Thus, it seems completely reasonable that networking issues may have much in common with those that biology has already encountered and solved a long time ago. The idea is to study deeply the biological systems since they may inspire very interesting solutions to networking and ICT problems.

To develop bio-inspired models and algorithms, we need to shed light on the general modelling approaches. The first modelling approaches date back to the early 1970s. Since then a lot of technical solutions mimicking biological counterparts have been developed. Fig. 2.1 illustrates the bio-inspired methodology.

The first step is to identify the analogies between biological and networking systems, that is which structures and methods seem to be similar. The second step consists of creating detailed models for the biological behaviour which will later be used to develop the technical solution. The third step is the translation from biological models into a model describing the bio-inspired technical system, and the engineering of the biological models, which includes the model's simplification and tuning for the specific application.

Furthermore, we can distinguish the following categories of bio-inspired solutions to issues related to computation and communication:

- *Bio-inspired computing*: it represents a class of algorithms focused on efficient computing techniques, for instance for optimisation processes and pattern recognition.

- *Bio-inspired systems*: they constitute a class of system architectures for distributed systems in massive and collaborative way, e.g. distributed sensing and exploration.
- *Bio-inspired networking*: it consists of a class of strategies to obtain a scalable networking under uncertain conditions, e.g. for the autonomous organisation in massively distributed systems.

Exploiting the biological principles, several application domains in networking can be distinguished. The following is a list some of the main biological domains and the correspondent networking applications and algorithms:

- Swarm intelligence - distributed search and optimisation; routing in computer networks,(e.g. MANETs, WSNs, etc.); task and resource allocation.
- Firefly synchronisation - Robust and fully distributed clock synchronisation;
- Activatorinhibitor systems - Self-organisation of autonomous systems;
- Artificial immune system - Network security and anomaly detection;
- Epidemic spreading - Analysis of worm and virus spreading in the Internet;
- Cellular signaling networks - Coordination and control in massively distributed systems;

Another important property of the bio-inspired algorithms related to communication and coordination is that we may identify similarities between techniques studied at a microscopic layer, such as cells and pathways between cells, and the techniques observed at a macroscopic layer, such as the coordination among people in a group or all over the world (35). In other words, many models show similar features both at microscopic and macroscopic layers, and exploit analogous coordination and communication mechanisms (see Fig. 2.2). This further degree of similarity allows using existing communication models in other domains, e.g. the mathematical models of proteins' spreading between cells may be considered equal to the virus spreading throughout people, so the problem is to choose the proper biological model as inspiration to

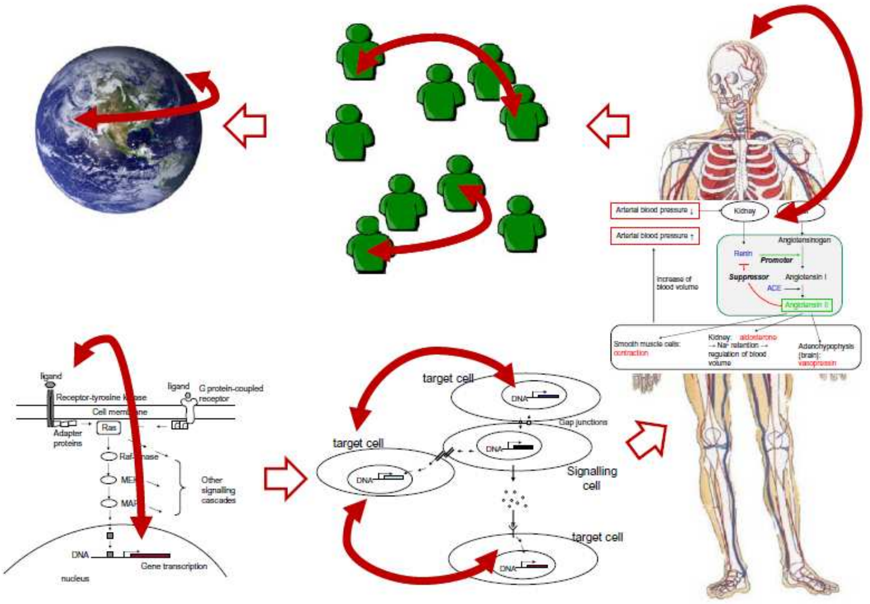


Figure 2.2: Bio-inspired approach. Figure illustrates communication and coordination at microscopic and macroscopic levels, showing the information exchange within a cell, between cells, within the human body, among people, and around the globe.

solve a technical network or social problem, otherwise, if the matching is not correctly identified, the solution may result limited in terms of functionality or effectiveness.

As stated in the previous subsection, ICT represents the project, the development, the implementation, the support and the management of information systems through telecommunications systems, therefore ICT constitutes a fundamental resource in the modern organisations, where it becomes increasingly more important to manage and use quickly and efficiently the large amount of data and the growing volume of information. Although the complexity, dynamicity and the diversity of the Nature, and the great problems related to the survival in the world, the biological organisms are able to evolve, self-organize, self-repair, and self-protect. To realise all these aspects, they leverage their knowledge and they have not any centralised control entity. Similarly, the entities in a network are growing in number and are becoming more and more connected, so they may mimic the robustness and the adaptiveness of the biological systems. Many research works have underlined how we can learn from the Nature. For this reason, bio-inspired research in ICT field is a growing research field.

Some of the main requirements of an ICT system, such as the high complexity and connectivity, the reliability, the growing information load, risk management and energy-saving issues, under conditions of limited computational resources, time constraints and a low overall knowledge, claim for a new modelling approach, able to face all these needs, getting a satisficing or sub-optimal solution. A bio-inspired approach allows solving certain problems and meet these specific ICT requirements. As we will see in the next sections, such kind of approach has been used as a model that relates the cooperation of multi-agent systems, the intelligence of the node, according to the paradigm of Internet of Things (IoT), and also the satisficing concept of heuristic decisions. What are the analogies between our system and a biological scenario? A biological system is characterised by the following features: high complexity; high connectivity; communication, cooperation and coordination; relation with other systems of the same nature; relation and communication with external environment.

Therefore, it is clear how a networked system (e.g. social network, energy-aware WSN, etc.), is a complex system similar to a biological one. Following the Dressler's approach, the first step is to focus on identification of mechanisms and models applicable to biological technical solutions for ICT systems. The biological approach makes it possible to give methodologies, algorithms and models useful to optimise and improve the design and management of traditional issues related to ICT systems, but also to evaluate new points of view which consider issues that previously have never been taken into account. The biological approach has the advantage of being found in Nature, and the Nature shows us the operation and the performance of certain processes (35). ICT systems have been evolved following innovative approaches by investing more on developing new fields such as context-awareness, social networks and mobile, dynamic adaptive complex systems, smart platforms, models and services for personalisation, providing the best user experience and user expectation. The user could be viewed as a transfiguration of knowledge, communication skills and cognitive ability. Also the behaviour of the nodes of a network, representing human beings, follows its human features, socio-psychological traits, involving cognitivity, awareness, contagiousness etc. One of the aims of future ICT networks is to redesigning network nodes, since a node is becomes a sort of "bio-shell" of the user, trying to reduce the gap existing between the network node and the human user. The node acquires cognitive skills, user's habits, understands the context in which it is located, and predicts behaviours and reactions, acquires the ability to make inferences and decides based on contexts, perceived risks and manages the interactions and social relations, from workplace to everyday life. The node acquires more human features and, locating at the center of a truly pervasive network, it becomes a really “smart subject”. A collection of nodes, connected with strong and weak ties (33), forms a network of smart subjects able to interact within a community or among different communities. The structure of the social network will determine the dynamics of diffusion and relationships. The bio-inspired approach represents the way to guide and inspire strategy for rethinking and redesigning ICT. In the next sections, we will see in detail the importance of joining the bio-inspired approach with ICT. In particular, we will focus on two bio-inspired algorithms, applied to IoT and WSNs, based on the concept of heuristics.

2.2.1 Heuristics

One of the most interesting and promising area in computer science is the design of algorithms and computer architectures based on our reasoning process and on how the brain works. Human neural circuits receive, encode and analyse the “available information” from the environment in a fast, reliable and economical way.

The evolution of human cognition could be seen as the result of a continuous improvement of neural structures which drive the decision-making processes from the inputs to the final behaviours, cognitions and emotions. Heuristics are simple, efficient rules, hard-coded by evolutionary processes or learned, which have been proposed to explain how people make decisions and solve also complex problems under limited knowledge or incomplete information. It is common experience that the majority of human reasoning and decision making can be modelled by fast and frugal heuristics that make inferences with limited time and knowledge. For example, Darwin’s deliberation over whether to marry represents an interesting example of such heuristic process. In other words, heuristics suggests the inability to achieve the complexity of the traditional models of rationality and a heuristic approach is a solution to the problems, even if complex, that do not rely on a clear path but rely on intuition upon temporary circumstances in order to generate a decision or a reasoning.

The concept of heuristics is the result of a lot of research works on the mechanisms which rule decisions about uncertainty (36). The basic idea in these works is that “people rely on a limited number of heuristic principles which reduce the complex tasks of

assessing probabilities and predicting values to simpler judgmental operations". Although representing a valid rule, D. Kahneman and A. Twersky have underlined also how it is prone to sistematic errors, following a certain statistical distribution, which makes it possible to describe and predict them. Heuristics have impacted in different research fields, among them human-machine interactions, in particular it has been proved useful for understanding human logic (37), since it allows studying human behaviour in human-computer interactions, empowering the analysis and helping in designing related architectures and algorithms. As stated in (36), the following are three main mechanisms underlying heuristics are: availability, representativeness and anchoring.
Availability refers to the probability of relying more upon knowledge readily available rather than examining other alternatives. For instance, people evaluate the probability of an event according to the occurrences among their acquaintances, thus it constitutes a useful clue, even though it could be affected by some factors, such as kinship, frienship, etc., producing a bias in exalting small risks or underestimating more dangerous risks. The *representativeness* derives from the assessment of similarities between an outcome and a model. It consists of categorisation and generalisation, as to predict the behaviour of an unknown subject, ii is needed to identify the group to which it belongs, but at the same time we can associate its behaviour to the typical behaviour of the group. The classical example is when you need to guess if a person is a computer scientist or a clerk employed in the public administration. His description as a shy person with passion for details leads to think that he is more likely a computer scientist, but the error may be generated from not considering the base-rate, that is the fact that there are more clerk employed in public administration than computer scientists. From this example it is clear that this kind of heuristics may produce assessment errors tending to ignore the base-rate. D.Kahneman and A. Twersky have also suggested that people usually make inferences starting from an initial value, or *anchoring*, thus the choice is ruled by this starting value which, though randomly, influences people's answers. In (37, 38), authors shed light on the dual process which characterizes human thinking: from one side people make decisions in a rapid and intuitive way, but obviously error-prone; from the other side, there is a slower, reflective and more statistical decision mechanism. The idea underlying the models used in the two following subsections represent a trade-off between these two approaches to decisions, since heuristics

produces intuitive and error-prone decisions, while the other system, based on statistics and thoughtful allows making corrections. The coexistence of the two mechanisms ruling decisions is linked with the experimental evidences of the presence of areas for emotions in the brain, which may be triggered or activated before than the cognitive areas.

2.3 It Measures Like Me: An IoTs algorithm in WSNs based on heuristics behaviour and clustering methods

In (39), we stem from the consideration that nodes of a WSN, deployed on a general topology, should follow a bio-inspired approach to respect the trustability, information load, risk and energy-saving requirements, upon bounded conditions of time, knowledge and computational power. This allows introducing a multi-agent model related to Internet of Things and heuristics models, in order to obtain a smart organised network where nodes have a social and human cognition. Our model is based on hierarchical clustering method and aggregation/rejection mechanism, following sociological and heuristics theories. The model follows the principle of sense of community and the logic of tie for similarity. The key target is to integrate the concept of cooperation of a multi-agent system with the node's intelligence of Internet of Things and the "Satisficing" of heuristics decision, in order to get a "Social Smart Behaviour" of the overall network.

2.3.1 Introduction

Wireless sensor networks (WSNs) are large networks made up of many autonomous low-power, low-cost, and small-sized sensor nodes. WSNs use sensors to co-operatively monitor complex physical or environmental conditions, such as motion, temperature, sound etc. Such sensors are generally equipped with data processing and communication capabilities to collect data and route information back to a sink. The network must possess self-organising capabilities since positions of individual nodes are not predetermined. Cooperation among nodes is the dominant feature of this type of network because sensor nodes use their processing abilities to locally carry out simple computations and transmit only the required and partially processed data (40). Sensor nodes can be either thrown in mass or placed one by one in the sensor filed, hence the deployment may be deterministic or self-organising. The future of WSNs is the integration of bio-inspired ideas, hierarchical clustering methods, and sociological models and concepts, such as sense of community and the “satisficing” theory, to form a social network model (41, 42). This will be possible using the node intelligence to allow network to self-organise itself into communities deciding how to join, through an aggregation/rejection mechanism, trying to keep the key requirements regarding the quality of service (QoS), efficiency, security, trustability and computational power. For this reason we have based our algorithm on multi-agent system model, where a single agent is an intelligent node, exploiting the Internet of Things approach (39). Then, we introduce the heuristic model to give to the node the ability to decide about the interactions with other nodes obtaining a social smart behaviour of the network. This approach is characterised by the assessment of the trustability value and the risk perception value for each node; this will rule the formation of the community and the aggregation/rejection mechanism of the nodes. Proposing an algorithm based on the models mentioned above, the idea has been to emphasise the importance of the concept of cooperation and sense of aggregation to group or community. The model accepts and follows the natural tendency to aggregate and reject each other according to a bio-inspired and self-organised approach, following a model of aggregation/rejection, applying a clustering method to a multi-agent model, based on heuristic decisions, in order to get eventually a “satisficing” model. It allows increasing the global knowledge in a WSN with nodes characterised by bounded conditions like limited time, limited knowledge and limited computational power.

The next sections are organised as follows: in sect. 2.3.2 we specify what are the main reason and features for using a bio-inspired approach for the model, in sect. 2.3.3 first we focus on heuristics and Internet of Things, then it is explained how we use these concepts in our model. In sect. 2.3.4 we present and describe our propsed algorithm “It measures like me” (IMLM). Finally, sect. 2.3.5 is dedicated to conclusion.

2.3.2 Why using a Bio-inspired approach?

As underlined before, a bio-inspired approach allows solving certain problems and meet specific requirements, such as reliability, information load, risk management and energy saving, under conditions of limited computational resources, time constraints and low overall knowledge. In (39), such kind of approach has been used as a model that relates the cooperation of multi-agent systems, the intelligence of the node, according to the

IoT paradigm, and also the "satisficing" concept of heuristic decisions. What are the analogies between our system and a biological scenario? Starting from the features of a biological system, it is clear that a energy-aware WSN, that has to send aggregated information related to single clusters, is a complex system similar to a biological one. We follow the Dresslers approach (35) described in Fig. 2.1, and the identification of analogies step is summarised in the following scheme:

- High complexity: IoT node intelligence;
- High connectivity: sense of community and social behaviour, other than the aggregation model;
- Communication, cooperation and coordination: multi-agent system, heuristics and trustability model;
- Relation with other systems of the same nature: logic of similarity, heuristics and information load;
- Relation and communication with external environment: social and human cognition.

The proposed approach tends to solve decisional issues through heuristics, cognitive aspects, using the proposed trustability model, security problems exploiting risk perception model, and shared knowledge management, using a controlled information load. The understanding and engineering steps will be treated in the following subsections.

2.3.3 Heuristics that makes WSNs "Smart and Things"

2.3.3.1 Inference, Heuristics and Satisficing

How do nodes deployed in a topology make inference about unknown aspects of a context? The possible approaches could be three (43): one follows the Laplacian demon theory that considers the mind as a supercomputer, with unlimited time, unlimited knowledge and unlimited computational power. This follows the classical view that human inferences rules are those of probability and statistics. Another approach is fully heuristics so that inference is systematically subject to human error: this perspective is diametrically opposed to the classical rationality principle. The issue is much more complex because it would identify the conditions under which the human mind seems to be more rational or more irrational. The heuristic would suggest the inability to achieve the complexity of the classical canons of the models of rationality. The third approach achieves a balance of compromise between both, and it is the approach of a controlled heuristics, on which we have based our proposed model of (39). This follows the theory of H. Simon (41), which is based on the concepts of “bounded rationality” and “Satisficing”. Simon starts from hypothesis that information systems of processing should have the need to satisfy rather than optimise. Hence, the term “Satisficing”, that is the union of “sufficing” and “satysfing”, is suitable with our model and with models that generally deal with conditions of limited time, limited knowledge and limited computational power. The theory that follows the “bounded rationality” considers human minds appropriate in the environment in which they live, only if they have the right perception of their limits, according to a cognitive, ecological and saving logic, but still meeting the target. Therefore, this approach remains heuristic but not at all, and finds the right trade-off between the heuristic decisions and the sense of community, control strategy and suitable criteria. The heuristic approach is a solution to the problems, that do not rely on a clear path, but rely on intuition upon temporary circumstances in order to generate new knowledge. We overcome the simple heuristics in the model due to the bounded rationality of Simon, as we also rely on the good sense of the community in decision-making, and we also add on trustability and risk perception (39). The heuristic models that generally rely on bounded rationality, follow the two sides defined by H. Simon, that is, cognitive mode and ecological mode (38, 41). In models such as “Two Alternative Choice Tasks” in general, there are two types of inference: inference from memory, decisions are taken considering declared knowledge, studies, memory and history; inference from given, decisions are made considering data and information extracted from a calculation or data extracted from an experiment. Following the process suggested by Simon, we should involve only the first type of inference. The initial process, and probably the most natural one, is to base its decisions only from those we have acquired in the past. In our proposal the component “inference from memory” is represented by an array that keeps track of our past contacts. This allows us to make inductive inference during aggregation to a community. Obviously, the inductive inference needs to be investigated in relation to the surrounding environment, topology and context of the communities created.

Aggregation/Rejection	**CHx**
Inference from Memory Cue Value 1: Trustability	+/-
Inference from Memory Cue Value 2: Risk Perception	+/-
Inference from Givens Cue Value 3: Measure(Temperature)	+/-
Inference from Givens Cue Value 4: Variance	+/-

Table 2.1: Cue Values for Inference on Aggregation/Rejection.

These models are characterised by cognitive algorithms that allow the creation of a framework for modelling the inference from memory. Such kind of tool allows the use of limited knowledge to make fast inference, using intelligent insights about unknown properties, based on indicators uncertainty. A subject must know the “cue values” that can be linked to the target variable to make inference, in a positive or in a negative way. Each “cue” also has a validity which indicates the frequency with which the cue correctly predicts the target defined with regards to the environment. The “cue values” are criteria and suggestions for assessment in order to achieve the targets. In Table 2.1 we show the cue values for our algorithm. Each cue will be characterised by a validity and a discrimination rate. In our proposal, after an initial self-organised sensing phase, the node aggregate and form communities, considering similarity measurements of temperature, trustability, risk perception and variance values.

2.3.3.2 Trustability and Risk Perception

Our model follows the principles of multi-agent systems, indeed the set of nodes will be deployed in a certain environment and will interact with each other using organisational rules that follow the hierarchical clustering, and exploiting the two main principles of multi-agent system: organisation and cooperation. Cooperation is related to the interactions among agents. It is the fundamental feature of multi-agent system where the overall system exhibits significantly greater functionality than the individual component. Cooperation allows to reach the target through coordination and conflict that regulate the community and which result in the aggregation and rejection processes. In this way we have an autonomous, multi-agent and self-organised system. Nodes become smart objects which have different communication, information and processing capabilites.

Starting from a WSN, our design choice of the proposed algorithm consists of the

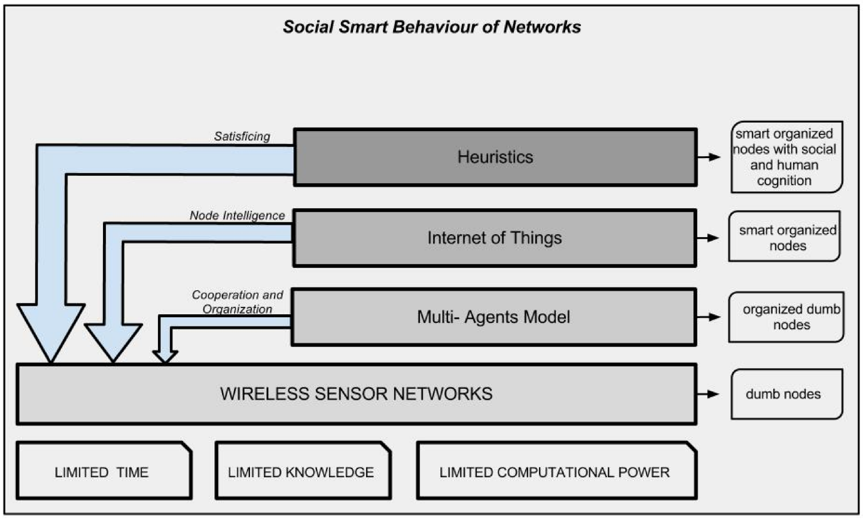


Figure 2.3: Steps to get a Social Smart Behaviour.

introduction of a heuristic model, allowing us to reach the perfect compromise by "satisficing" (41) and the compliance by smart objects in bounded conditions. The heuristic will allow us to explain how the nodes make decisions, come to judgments and solve complex problems with incomplete information (44). The purpose of the proposed algorithm is to use fast and frugal heuristics to make inferences. The main advantage is that using heuristics we reduce the complexity of the tasks in much more simple and immediate operations. People have two systems for making decisions in rapid intuitive error prone and slower reflective statistical mode. When an agent interacts with other agents, it can or not trust on their neighbors. The trustability, in our model (39), has been implemented as a personal recording one by one that measures the trustability level of the node with which it must interact. We take as a reference model the one presented in (44), by adding the empathy mechanism and contextualising it in accordance with our problems. Thus, we consider in the model, as well as α_{ij}, that is the dynamical memory for the trustability of i on its partner i, also the parameter A_i, that is the risk perception, which regulates the value of trust in the nodes of the community. Furthermore, we consider the oblivion mechanism to update the network in terms of knowledge. This results into the need of the network to forget periodically, and update dynamically. Summing up, if in (44), heuristics is used to balance between the costs and the risk of being infected, in our model it is used to balance between the bounded condition and sending reliable parameters, not neglecting risk and energy saving, as showed in Fig. 2.3. Therefore, to achieve this target, introducing all the features mentioned above, we also make the nodes more "human", as well as smart and self-organised. Nodes are able to decide, then the limited conditions allow obtaining resolute decisions and, considering communities and interactions between them, we have a *Social Smart Behaviour.*

2.3.4 IMLM Algorithm

2.3.4.1 Introduction

"It Measures Like Me" (IMLM) algorithm is applied to WSNs, in which a large number of sensor nodes is deployed in a extended region to monitor and measure some parameter, such as temperature. IMLM aims at reducing the power-consumption and introducing a social smart behaviour of the network. IMLM fuses an aggregation/rejection model, in terms of clustering, with a heuristic multi-agent model related to the single node. IMLM uses heuristics to mitigate the speed of node rejection with a decision taken in a short lap of time (limited time), using a reduced amount of information (limited knowledge) and consuming as low battery as possible (limited power consumption) (38, 43). The main assumption of the clustering process takes advantage from the first law of geography: "everything is related to everything else, but near things are more related than distant things" (45). The basic idea is that we can aggregate a large amount of known nodes in a WSN. The aggregation mechanism concerns with radio visibility between the couples of nodes. The algorithmic approach is self-organised and consists of nodes' "instinct" to aggregate themselves to other communities, while the rejection policy is hierarchically managed by Cluster Heads (CHs). The proposed model follows rules similar to those ones of cohesive attraction or cohesive force, that is the action or property of like molecules sticking together, being mutually attractive. The cluster aggregation is similar to the molecular aggregation based on the instinct to follow its own nature. The node is attracted by neighbourhood inside its radio range and it will aggregate "naturally" with one of them. The same thing happens in the case of oil in a glass of water: the two liquids split each other to form two different clusters, then they mix again cause an external force that is represented in the algorithm by the CH decision to reject one or more CH. IMLM is based on a multi-agent model that considers abstract entities, called "agents", that work autonomously in the algorithm in different ways according to their states and roles. These roles depend on hierarchical levels and

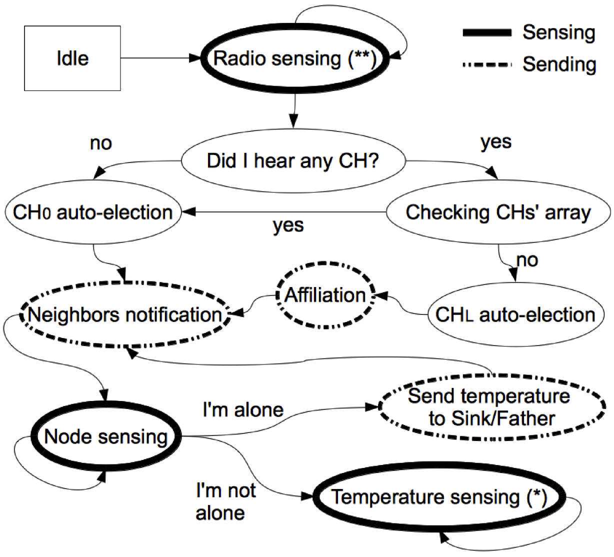


Figure 2.4: Algorithm Description - Aggregation mechanism.

on the internal state: idle, Cluster Head (CH), that could be a CH_L or a CH_0, going up the hierarchical ladder, and still climbing to the sink node.

2.3.4.2 Description

Before focusing on the detailed operation of the algorithm, the followings are the different types of messages exchanged between the nodes in the various steps with a brief description for each of them:

- *Cluster Head Notification Message (CHNM)*: notification message sent by a neighbouring CH;
- *Node Affiliation Message (NAM)*: node affiliation to a CH;
- *Measurement Message (MM)*: it allows nodes to communicate a single measurement or a mean value;
- *Variance Request Message (VRM)*: it allows CHs to ask "children" for sub-community variance values: it is set "true" when it is needed to forward the message, otherwise it is "false";
- *VRM Response (VRMR)*: the sub-community sends variance value;

- *Rejection message (REJM)*: it allows CHs to reject a child, in particular "true" is used to reject it, while "false" is used to mantain the child.

The IMLM operation is described as follows and figures are used to outline graphically the various steps (see Figs. 2.4 and 2.5). At the beginning the node stays in the idle state and listens to CHs via radio sensing for a random period of time. The node listens to CH Notification Messages (CHNM) to know if there are CHs in the neighbourhood. Both in the case in which an idle node does not recognise and in the case in which recognises the presence of a CH that rejected it in the recent past, it will self-elect itself as a CH_0. Otherwise, if the node finds an available CH, it will become a CH with a lower hierarchical level (CH_L) and it will send a Node Affiliation Message (NAM) to the "father" (i.e., the node of higher hierarchical level). Hence, the node notifies to the neighbourhood its actual state in both cases using CHNM messages. After "Neighbours notification", the node will wait for NAM messages from its children and it will register their identities (IDs). CH will have to associate a random trustability value, in the interval between 0 and A_i for the empathy mechanism (which explains the process for which we trust in a different way of one rather than another, without an apparently reasonable logic: first, aggregation mechanism is ruled by the logic of the first encounter, then it will be ruled by the trustability and risk perception values following the hierarchy). If the CH is alone and if it is a CH_0, it will send its measured temperature to a sink node, otherwise if it is a alone CH_L, it will send it to the father. Instead if the CH is not alone, it will wait for Measurement Messages (MM) from children; MM can be either single measurements or mean values of sub-communities. The IMLM algorithm uses a heuristic mechanism based on trustability estimation directed from CH to its children. For this reason, the CHs evaluate the trustability among all children and relate sub-communities. In the trustable case, if the CH is the root of the hierarchical tree (CH_0), it will send a Variance Request Message (VRM) set to "false" to children, and the mean value of the whole community to the sink. The next step is to return in the "temperature sensing" state. If the CH has a lower hierarchical level, it will send the mean value of its community to the father and it will wait for a VRM. A received VRM, set to "false", allows the node to come back to a temperature sensing of its sub-community, while VRM set to "true" forces CH to forward the request (VRM) to its children. In the latter case CH has to wait for a VRM Response (VRMR) to collect variance values from sub-communities. Then it calculates its local variance value to be sent to the father.

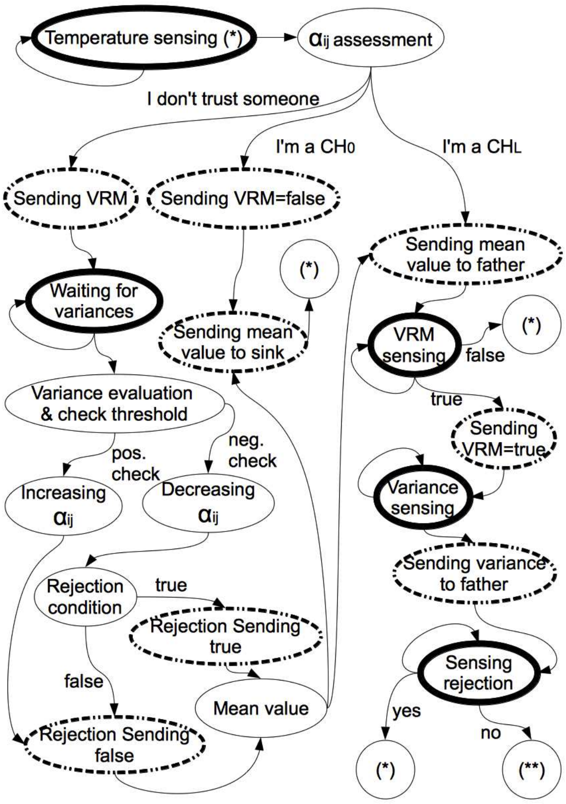


Figure 2.5: Algorithm Description - Heuristics and Rejection mechanism.

It will listen to the REJection Message (REJM) to see if it still belongs or not to the community. The α_{ij} assessment allows identifying untrusted children. This condition occurs when the related α_{ij} is less than the risk perception, A_i (44). In this case, the autonomous agent will be "scared" of specific sub-communities, so it will ask them for updated variance values that result in a local new variance value. It is needed to evaluate also variances related to trusted sub-communities; these values will be estimated weighing them with a coefficient that is inversely proportional to the trustability value and directly proportional to the last variance value related to the sub-community. The variance calculation is based on (46). The Ward's method aims to minimise the inner-cluster variance. The variance of a community is calculated as follows:

$$S = S_w + S_b \tag{2.1}$$

where S is the matrix of total variances and co-variances, S_w the matrix of internal variances and co-variances, S_b the matrix of external variances and co-variances. If we consider a uni-variate measurement and two clusters, “1” and “2”, the global variance will be calculated as follows:

$$\sigma_{tot} = \sigma_1 n_1 + \sigma_2 n_2 + (\mu_1 - \mu_{tot})^2 + (\mu_2 - \mu_{tot})^2 / n_1 + n_2 \tag{2.2}$$

where σ_1, σ_2 are variance values of the two communities; μ_1, μ_2 are the corresponding mean values; n_1, n_2 represent the number of nodes in each cluster. The new community variance value will be compared with a fixed threshold. If the check is positive, the specified trustability, related to the sub-community, will be increased of a fixed quantity V_α, otherwise, it will be decreased of the same quantity. In the latter case, the CH will have to see if the sub-community is suitable yet in order to send a Rejection Message (REJM), “true” or “false”, according to the new trustability and A_i values. If the trustability value is less than $-A_i$, the corresponding sub-community will be thrown away, otherwise it will be maintained. The rejected node will register the last CH in a specified scheduling queue, not to allow the association to a “old” community for a certain period of time. Each CH in the queue is affected by a oblivion factor, following a negative exponential function $(1-\lambda)^\tau$. If the oblivion factor reaches a fixed threshold, the associated CH will be thrown away from the queue. The last step consists of the mean value calculation, considering all the “alive” sub-communities, and finally the sending of it to the father or to the sink. Each CH_0 communicates a mean value to the sink, that forwards information to an elaboration center, integrated with GPS positions of the community. The elaboration center will reconstruct a measurement map, using interpolation algorithms like Kriging (47).

2.3.5 Conclusion

The aim of our clustering algorithm is to reduce power consumption of nodes in WSNs, through the aggregation of them, based on the geographic position and a common range measurement. This feature also allows reducing the waste of energy related to sink nodes, especially in communications to the satellite. A challenging issue of the aggregation process is the waste of overhead related to the cooperation among nodes. Heuristic behaviour addresses to solve this question, mitigating the flow of information exchanged between nodes in a "satisficing" way. Furthermore, the self-organisation of nodes in communities is similar to the principles ruling human society. Thus, IMLM creates a "Social Smart behaviour", adding a social feature to the IoT principles.

2.4 A Energy-Preserving Model for Wireless Sensor Networks based on Heuristic Self-Organised Routing

One of the main targets related to WSNs is to reduce power consumption of nodes and of the whole network. An ideal WSN should be networked, scalable, fault-tolerant, energy-aware, and also smart and efficient. Unfortunately, however this is not always true. The basic idea of the work (48) is that senders use a heuristic approach to select the sub-optimal next hop in order to reach just one sink, considering some key requirements such as general performance (QoS and security), efficiency, trustability, high computational power and energy-aware behaviour. The model presented in (48) tries to satisfy the need for reaching the nearest sink node, considering a trade-off between the shortest path and heuristic decisions, in a top-level strategy based on a heuristics approach in order to reduce the overall power consumption of the network.

2.4.1 Introduction

A WSN consists of spatially distributed autonomous sensors to monitor physical or environmental conditions. Each node is an autonomous agent able to decide how to route information (40). There are many types of routing and forwarding algorithms for WSNs, characterised by different kinds of actions about how to reach the destinations. Overall, an efficient routing protocol should perform aggregation, clustering, self-organisation and cooperation for power saving and to increase network lifetime; it should also consider a threshold for sensor nodes in data transmission in order to get energy-saving, and a multi-path dissemination to improve fault-tolerance.

In (48), we propose an energy-preserving model for WSNs based on a heuristic and self-organised routing. We consider a network with specific nodes which need to send information to just one of a set of sinks. In order to make simpler and reduce the communication flow that would produce a large amount of energy consumption, each node routes information exploiting a cognitive behaviour of its own neighbourhood, without using traditional routing protocols that could produce large routing tables, for a huge amount of sensors. The hierarchy follows a trend ranging from the sink to the probable sources. From the latter, following the hierarchy created previously, each node will choose the next hop according to an heuristic mechanism, based on trustability, goodness' perception of the path and battery level of node. The heuristic approach considered within the model will allow us to explain how the nodes take decisions to solve complex problems with incomplete information using trustability level and perception, following a top-level strategy that leads to solve problems exploiting this underlying heuristics.

The work (48) is organised as follows: sect. 2.4.2 deals with some energy-aware protocols, also introducing what is required to achieve an efficient energy-aware routing protocol. In sections 2.4.3, we analyse the heuristic theory underlying the proposed model, and how we use and propose this strategy, along with trustability and perception, in multi-agent systems. In sect. 2.4.4, we present our model, and finally in sect. 2.4.5 we sum up the presented model with conclusions.

2.4.2 Energy-Aware Routing Protocols

Routing or forwarding of data packets in WSNs can be divided into three categories: flat-based routing, hierarchical-based routing, location-based routing (49). In flat-based routing, nodes have the same role and responsibility in forwarding/routing data, while in hierarchical-based routing, decisions are influenced by the hierarchical rank of sensor nodes. Furthermore, another classification of routing protocols is based on how routes are created (50). In a proactive approach all routes are created in advance and updated regularly; in a reactive approach, routes are computed only when they are required and a hybrid approach is a combination of these two ideas. In general an efficient routing protocol should perform data aggregation for power saving, dynamic clustering to increase network lifetime, a threshold for sensor nodes on data transmission and dissemination, in order to help energy-saving, multi-path selection dissemination to improve fault-tolerance, self-configuration and adaptation of the sensors nodes to changes in network topology and finally time synchronisation.

In (48) we propose an energy-efficient model based on a heuristic approach. The key idea applied to a WSN's scenario, is that of giving to sinks the power to create a dynamic hierarchy. The root of the hierarchical tree is represented by the sink and leafs follow iteratively the hierarchy in order to cover the entire network spatially. Using the previously created hierarchical tree, senders will be able to reach one of the avaliable sinks, exploiting only the knowledge of their neighbourhood. Heuristics will help single nodes to choose the next hop until to the sink.

2.4.3 Inference and Heuristics in the model

In this section, we are going to explain inference mechanisms and focus on the overall heuristic approach adopted in the model. The question is how nodes deployed in a topology do make inferences about unknown aspect of the context? There could be three possible approaches (43): Laplacian demon theory, fully heuristics and controlled heuristics. Many decision problems may be considered as optimisation problems. These problems are tipically too difficult to be solved exactly within a reasonable amount of time and heuristics become the best methods of solve them. Furthermore, when the quality of solution is critical and an optimal solution does not exist, it becomes important to investigate efficient procedures to obtain the best possible solutions, considering key factors to be minimised. Heuristics are criteria, methods, or principles for deciding which, among several alternative courses of action, promises to be the most effective, in order to achieve some goal (51). The model follows the main principles of multi-agent systems: cooperation and self-organisation. Cooperation is related to the interactions among agents. It is a fundamental feature of multi-agent system where the overall system exhibits significantly greater functionality than the individual component (52). Exploiting hierarchical methods within an heuristic approach to detect a good path, we achieve an autonomous, multi-agent and self-organised system. Therefore, nodes become smart objects which have different communication, information and processing capabilities to make decisions. Our design choice consists of the implementation of a heuristic model. The heuristic approach inside the model will allow us to explain how the nodes make decisions, come to judgments and solve complex problems with incomplete information. The main advantage is that using a heuristics model with a dinamic hierarchical phase, from sink to source, and a heuristic approach, from source to sink, we reduce the complexity of tasks, obtaining a self-organised energy-aware system. The

ID FRAME	MAC ADDR	NODE'S LEVEL

Table 2.2: Identification frame.

trustability and perception are used when a node/agent interacts with other agents to decide the next-hop in order to reach the sink. We refer to the model presented in (44), adding some mechanisms to maximise the overall lifeness of the network.

2.4.4 Model

Sink nodes in WSNs are able to receive messages from other nodes and collect various kind of data. Usage of multiple sinks is related to power consumption reduction in WSNs, as shown in (53), for this reason our topology follows a hierarchical organisation in which the highest role is delegated to sink nodes. We identify three phases: Topology Discovery, Data Sending and Heuristic Approach. Our model tries to satisfy the need of reaching the nearest sink node, considering heuristic decisions to reduce the overall power consumption of the network and network's errors. The aim of the proposal is to use simple network signaling and light logic to maximise network lifetime.

2.4.4.1 Topology discovery

The *Topology discovery* phase is initialised by sink nodes, which represent 0-level hierarchical nodes (HN_0), sending a identification frame to their neighbours, at a certain time interval, $T_{identification}$. Each HN_0's neighbour will become HN_1, storing node IDs of previously identified HN_0s and in turn they will send an identification frame to their neighbourhood. The *Identification frame* is defined as follows:

Network's nodes, without a specific role, will listen to identification frames and they will become L-level hierarchical node, choosing own level L as the minimum of the levels of nodes which have sent it identification frames:

$$L = min\ (nodes\ levels) + 1 \tag{2.3}$$

Consequently the HN_L will store IDs of the HN_{L-1} nodes so that it will be able to use one of them to reach a sink as shown in Fig. 2.6. The HN_{L-1} black circled node is the “delegated node”, chosen by HN_L node to send data to the sink, using the heuristics we are going to describe in sect. 2.4.4.3.

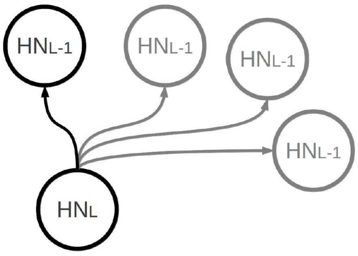


Figure 2.6: Multiple routes. Each node could reach multiple upper level nodes.

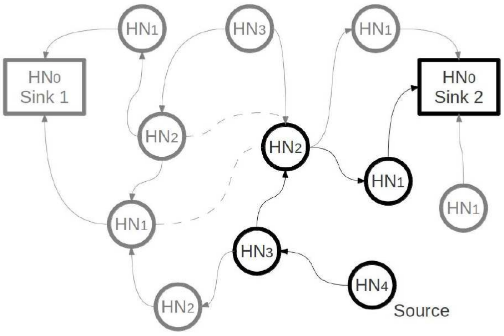


Figure 2.7: HN_3 routes on the right.

2.4.4.2 Data Sending

When a HN_L node needs to send data to a sink, it will use the delegated node, using a *Data Frame*, defined as showed in Table 2.3.

At each Data Frame (DF), the node will store, the Sender (Se), the Receiver (Re), the last Data Frame Counter (DFC), the Data packet, and it will calculate the related Cyclic Redundancy Check (CRC) of the frame. Figs. 2.7 and 2.8 represent a network with two sinks and two different paths.

2.4.4.3 Heuristic Approach

If a source node needs to send data to a sink, it will refer to one of the hierarchical upper level neighbours. It will choose the specific neighbour, using a heuristic approach,

DF	Se	Re	DFC	DATA

Table 2.3: Data Frame.

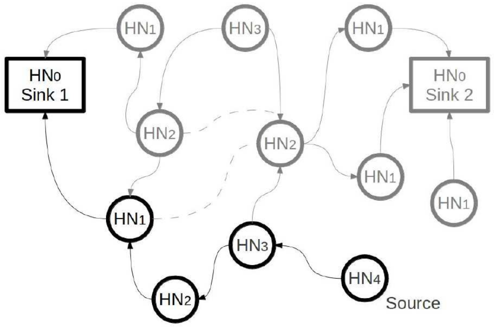


Figure 2.8: **HN_3 routes on the left.**

similar to (44), taking in account both battery level and communication quality. A generic node i stores a risk perception level, A_i, that intuitively represents a mean "trouble" level of the neighbouring nodes. Each node will store a trustability level, α_{ij}, related to each upper level hierarchical node and they will use the most "trustable" node in the neighbourhood to reach the sink. The choosen node will be called "delegated node". Considering a generic node i, and its delegated node j, if the trustability level, α_{ij}, related to j is greater than the risk perception perceived by i, A_i, it will send data to node j without considering j's battery level, or connection quality between them.

$$Condition: \ if \ \alpha_{ij} > A_i => i \ sends \ data \ to \ j \tag{2.4}$$

The model considers nodes with a finite memory, so they "forget" their history, following an *Oblivion Mechanism.* Hence, risk perception and trustability will be updated at certain time steps, as follows:

$$\alpha_{ij} = \alpha_{ij}(1 - r_\alpha)^\tau; \quad A_i = A_i(1 - r_A)^\tau \tag{2.5}$$

In this way the model will force nodes to check model's parameters against neighbours and refresh their perceived knowledge. The r_α parameter represents the rate of decrease of the Trustability Level, likewise r_A is the rate of decrease of Risk Perception. If node i does not trust nobody due to the Oblivion Mechanism or Negative Checks (as explained below), then it will start a "check phase". The check phase consists of asking to delegated node the CRC of the last Data Frame and to each upper level nodes battey levels, using *CRC Frame* and *Battery Frame* (see Tables 2.4 and 2.5).

CRC Frame	Se	Re

Table 2.4: CRC Frame.

Battery Frame	Se	Re

Table 2.5: Battery Frame.

Battery levels and connection quality parameters will be used to recalculate trustability levels related to each upper level neighbour. We propose to evaluate the correction of trustability level related to the delegated node, as follows:

$$\alpha_{ij_{\text{next}}} = \alpha_{ij_{previous}} + C'_{weight} C_{ij} + B'_{weight} b_j \tag{2.6}$$

where C_{ij} is the parameter related to connection quality between i and j, C'_{weight} is the weight related to the importance of considering the connection quality. b_j is equal to 1 if the battery level of node is lower than the quantity: $Mean_{battery\ levels} - Standard\ Deviation_{battery\ levels}$, otherwise it is equal to zero. We choose such kind of threshold ($Standard\ Deviation_{battery\ levels}$), to consider the global amount of energy in the neighbourhood and its distribution among nodes. B'_{weight} is the weight related to the importance of considering battery levels in our model. B'_{weight} and C'_{weight} are design values that will be choosed according to the specific target to obtain. The A_i factor is a key parameter in our model because it rules the behaviour of a node, according to the amount of α_{ij}, following a heuristic approach. For this reason, periodically, at a certain time step $T_{perception}$, nodes will recalculate the risk perception perceived, as follows:

$$A_i = A_i + v_A \left(\frac{B'_{weight} b_{negative\ checks} + C'_{weight} C_{negative\ checks}}{Total\ checks} \right) \tag{2.7}$$

where $b_{negative\ checks}$ represent the total number of negative checks related to battery levels and $C_{negative\ checks}$ represents the total amount of network negative checks. $Total\ checks$ is the amount of total checks, both for battery and communication "rightness". v_A is a key parameter because it rules how the risk perception should grow up.

2.4.5 Conclusion

In a WSNs context, nodes are able to collect data and send only the required information to the sink. In a scenario with multiple sinks, an ideal model should satisfy the need to reach the nearest sink node. In order to reduce power consumption of the network and, at the same time, maintain the same communication performance, in (48) we have proposed a novel strategy based on a heuristics approach that rules the underlying heuristic decision process under certain parameters, such as battery level, communication performance, trustability and risk perception. In this way, we have proposed a model based on heuristic assumptions, dynamic hierarchy and decision processes to obtain a cognitive energy-preserving behaviour for the sensor nodes, exploiting a self-organised routing.

Chapter 3

Multilayer Networks

3.1 From Single- to Multi-layer Networks

Since the turn of the millennium, networks have become a universal paradigm for simplifying large-scale complex systems, for describing their structure and for studying their systems-wide functionalities. Network theory has constituted a way to characterise the connectedness among single entities or components which interact in a huge variety of complex systems, ranging from biological to social systems. 'Complex Systems science' represents an emergent field of knowledge dealing with collective and large scale phenomena arising from the interaction of many components. One of the main features is that a complex system cannot be understood and predicted from individual behaviour, but instead the system should be considered as a whole. Some examples are the self-organised behaviour of neurons in the brain or the complex cooperative phenomena in ant colonies or human cities. Thus, together with phenomena such as self-organisation or pattern formation, the complexity is essentially the result of network, connecting the single parts in a system which does not corresponds only to the sum of its parts. The striking advances of data-driven models has allowed empowering and deepening the analysis of such complex systems, creating a new kind of science, known as "network science" (1). Recently, a lot of research efforts have been devoted to study networked systems, e.g. social networks, made up by nodes, which are the components or the single entities, while the links among them describe the interactions (4). Interestingly, some pioneering works in this field have discovered some similarities between networks, describing quite different phenomena, such as heavy tailed degree distributions of nodes or the relatively high clustering of node triples. These regularities along with its interdisciplinarity have even raised up the attention for network science, making it one of the hottest research topics in the 21st century (2, 3, 4). The "engine" behind this rapid development, increasing investigation and growing understanding of network science has been represented by the vast amount of data, that are now routinely being collected, modelled and analysed, regarding these complex systems (10, 16, 17, 18).

In 1960, Erdos and Rényi developed the first model of networks, known as random graph theory, where the 'random' hypothesis means that every pair of nodes is randomly connected with the same probability, giving rise to a Poisson degree distribution (54). Simultaneously the lattice networks have been exploited to model physical systems, where instead the feature is that each node has the same number of links. Since then, many types of modelling approaches have been proposed and studied in the various disciplines based on different features of the different networks, making the graph theory a well-estalished mathematic tool to study networked systems. Despite its ability to desribe the structure of a variety of networks, graph theory is not able to include the real-world networks which, as Barabàsi observed in 1999 in his pioneering work (55), do not follow the Erdos-Renyi model, but rather in most systems the organisational principles arise in natural way, pushing towards new models and to the emergence of a new science, that of complex networks. Barabàsi networks, known as scale-free networks, are non-homogenous structures whose degree distribution, which indicates the number of links per node, in most cases follows a power-law form (55). The most important and interesting aspect of these scale-free networks is the ability to approximate real networks, such as the Internet (4), the World Wide Web (WWW) (56), social networks, infrastructures networks, biological networks, Protein-Protein Interactions (PPI), gene regulation and biochemical pathways, and networks in physics (13, 16, 18, 57). Scale-free networks have also highlighted different properties of networks, for instance, in term of robustness, in contrast with Erdos-Renyi networks, this kind of networks is extremly robust to random failures due to its heterogeneity (55, 56). Network concepts and notions have been exploited to study different aspects and properties, such as robustness, epidemic spreading and the social interactions and behaviours in frienship (7, 58) or scientific collaborations (59). Evidently, a great part of our current knowledge on networks is based on ideas and concepts borrowed from statistical physics, such as percolation theory, fractals and scaling analysis. Over the years, new measures and methods have been introduced to characterise network properties, such as measures of node clustering, that is the formation of triads in network, degree-degree correlations between neighbouring nodes, centrality measures to weigh the node importance in the network, other than measures to detect community structures. All these measures not only have been proven useful to explore and better understand networks as a whole, but also to identify network motifs or patterns occurring repeatedly and providing information about their functionality (13).

If, from one hand, network theory has successfully allowed characterising the interaction among the constituents of various types of complex systems, however, from the other hand, up until recently, attention was almost exclusively given to single-layer networks, in which all components were treated on a single equivalent layer, neglecting all the extra information about the temporal- or context-related properties of the interactions among nodes or entities. Only in the last years, also exploiting the enhanced multi-scale resolution in real data sets (60), network scientists have directed their interest and research efforts to the multiplex nature of real-world systems, explicitly considering the presence of a node and its interactions in different layers. In fact, these complex systems are not single isolated networks, but continuously interact with other networks, and this creates an interdependence.

The concept of interdependence means that a failure of a node in a network can cause a failure of dependent nodes lying in another network, and so on, leading to cascading failures and catastrophic phenomena (34, 61). Blackouts represent a classical example of interdependent networks, since most often they are the result of cascading effects between commmunication and power grid systems (34, 61, 62, 63). Furthermore, different kinds of critical infrastructures are also coupled together, such as financial systems, power transmission or communications systems only to name a few. Therefore, the deeper understanding of complex networks has showed as they are organised as a *Network of Networks (NoN)*, producing a shift in paradigm from single layer networks to multiplex, multilayer and interdependent networks (61). Thus, providing a suitable representation and modelling of these interconnected systems, identifying the rules and mechanisms, control them, are some of the major efforts of recent research in complex systems. ‘Complex networks theory’, born in the last fifteen years, has been exploited the interdisciplinary effort of a lot of scientists, in order to extract meaning, information and knowledge from the vast amount of big Data and the underlying complex systems and mechanisms. Among the main targets, from one side the analysis and identification of the main unifying principles describing the structure of network, and from the other side the understanding of the emergent dynamics able to explain and capture what actually observed in such systems. In particular, in (4) the authors present a complete compendium of the ideas and concepts involved in both structural and dynamical properties of complex networks.

The traditional complex network approach has mostly been concentrated on the single-layer case, in which each single system's unity is represented as a network node, and each link or connection among units is represented as a real number indicating the link's weight. Thus, every link of the network belongs to the same equivalent layer and it constitutes a big constraint because it is not able to capture the real-world systems, where instead phenomena occur simultaneously in different layers. This is the reason why a multilayer approach, where a multitude of entities interact with each other through somehow complicated patterns, results by far more suitable to describe such systems. The idea is to consider a complex system including multiple subsystems, or layers of connectivity, focusing on the mechanisms of evolution, adaptation, transformation due to internal and external dynamics regarding locally the single components or globally the whole multilayer structure. The understanding of these multiscale and multicomponent dynamics is challenging, but a lot of work has been done during the last years to describe and understand the structure and dynamics of this kind of systems (7, 58, 64, 65). The introduction of some concepts, such as network of networks (61, 63), multidimensional networks, multilevel networks, multiplex networks, interacting networks, interdependent networks, other than the related and even different mathematical approaches, based on tensorial representation (58, 64) or otherwise (65), represent a demonstration of the efforts in this field.

The main target of this chapter is to discuss the framework of multilayer networks, and the main structural notions, models and measures introduced in literature to study and describe it.

3.2 Real-world examples of Multilayer networks

As seen in the previous section, the framework of multilayer networks is a recent mathematical tool introduced to explain and understand the complexity of a huge variety of real-world systems, since single networks are not enough to describe the complex set of interactions between different networks. The following are some representative examples of the major constraints of the traditional single networks, where it becomes essential to use a multilayer approach.

The first example is related to social networks. A social network can be described as a set of nodes, corresponding to people or groups (communities), and a set of links or connections or edges, representing the relationships or interactions between them (7). The traditional network science assumes that all the connections or social interactions lie in the same layer or level, but the real scenarios are extremely different from this contrained description. In fact, in general social relationships occur among nodes or groups belonging not only to one layer, but rather social interactions involve nodes on different layers. Thus, considering only a single layer perspective, used in classic complex network models, does not allow modelling properly the real-world situations. For instance, if we take into account the spreading of information or rumours on top of a social network, such as Facebook or Twitter. Exploiting the graph theory, we can see users as the nodes of the graph and all the relationships as network's links or connections, but the connection between two nodes could be the result of various and several direct or indirect interactions which occur in another context, for example due to the fact that they are both fan of the same football team and they occasionally met at the stadium, or maybe they could be co-workers, sharing the same working environment, or even live in the same part of the city, going to the same pub. Thus, the problem is that considering only a layer in the diffusion process, we cannot capture the real dynamics of information spreading, since a node lives, interacts and evolve in the various layers, being influenced from each single interaction in every layer. In other words, the spreading process will involve each layer and only through a multilayer approach we can study its dynamics.

Another clear example of the multilayer or multiplex nature of real-world systems is the transportation network. For instance, if we deal with the Air Transportation Network (ATN) or subway networks, using the traditional single-layer modelling approach, where nodes represent the various airports and links the direct flights between them, we will lose a lot of information about dynamics (e.g. delay propagation in the ATN), as there are different types of flights, e.g. commercial airlines include different flights and routes, so it may be represented better as a distinct layer, containing all the connections operated by the same company. Also in the case we want to model the transportation networks in the city, the single parts of the city may be connected via different kinds of public transports, ranging from bus to underground, suburban rail, or riverboat networks. Even if we consider the connections among different cities, they may include not only roads, but also railways, and air transports.

Another interesting example, taken from biology, is the description of C. elegans neural networks, which consist of 281 neurons and nearly 2000 connections: these connections involve not only a layer, because two neurons may be linked through a chemical link or by an ionic channel with gap junctions' interactions, and accordingly the dynamics will completely change.

To sum up, all the previous examples are only some of the cases where using the traditional single-layer modelling approach would be a strong oversimplification, since the different nodes interact simultaneously in more than just the one network. As observed, this key concept could be applied to social as well as technological, air transportation and biological systems. Therefore, we need to replace single-layer networks with another mathematical framework which provides a better description of such systems. For this reason, multilayer network, that is a combination of networks that are interrelated in a nontrivial way, has recently emerged as a key concept to quantitatively describe the interactions not just within, but also among different networks. Another mathematical description typically used in the examples above, such as in transportation or social networks, is in terms of multiplex networks, i.e. networks where each node appears in a set of different layers, and each layer describes all the edges of a given type.

3.3 The mathematical formulation of multilayer networks

Network science investigates the structural properties and patterns of the interactions among the elements of a variety of complex systems, ranging from social groups to infrastructure and technological systems, and also including the brain and biological networks (2, 3, 4, 5). In the last decade, a vast amount of literature has attempted to disentangle noise and stochasticity from non-random patterns and mechanisms, in order to get a better understanding of how these systems function and evolve. As we explained in the previous sections, the evidence suggests that to understand the complexity and unveil the real properties of these systems, we need to switch from single layer networks to a multi-layer framework, able to characterise the complex set of interactions among different networks. The intrinsic interdependence of systems, with many components interacting with each other through different channels allows explaining the self-organisation and emergent phenomena, from which we can extract the knowledge. Network science and the science of complexity have revolutionised our understanding of several phenomena and processes, considering a multilevel, multichannel, multicomponent, a system-of-systems analysis. The multiple subsystems and layers of connectivity allow exploring the internal and external dynamics affecting the various components, both at local and global scale, which make the system evolve, adapt and transform accordingly. The multiscale and multilayer dynamics poses severe challenges regarding the analysis, observation, investigation and understanding, thus a lot of work in literature has been done in the last years to understand and describe adequately the structure and dynamics of the multilayer neytworks (58, 64, 65). In particular, some notions corresponding to the different mathematical approaches have been introduced, such as Network of Networks (NoN), multilevel networks, interacting networks, interdependent networks, multilayer networks and multiplex networks, only to name a few (61, 63). Most of these approaches are based on a tensorial representation (58, 64), even though there also other descriptions based more on topological formalism (65). In this section, the aim is to briefly present and discuss the mathematical framework of multilayer networks, and also to extend some notions and network measures from single layer to multilayer networks.

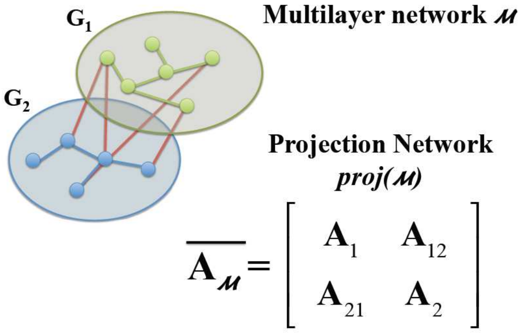


Figure 3.1: Schematic represenation of a multilayer network. $\mathbb{M} = (G, C)$ is composed by two graphs: G_1 and G_2. The interlayer connections are in red, while the intralayer connections are in green for graph G_1 and in blue for graph G_2.

Multilayer networks are formed by a set M of layers constituted by single networks, and by interlinks connecting the nodes in the different layers. Formally, a multilayer graph $\mathbb{M}$ is described by a pair (G, C), where $G = \{G_\alpha; \alpha \in \{1, \ldots M,\}\}$ is a set of (directed or undirected, weighted or unweighted) graphs $G_\alpha = (X_\alpha, E_\alpha)$ (called *layers*) and $C = \{E_{\alpha\beta} \subseteq X \times X; \alpha, \beta \in 1, \ldots, M, \alpha \neq \beta\}$ is the set of interconnections between nodes of different layers G_α and G_β, with $\alpha \neq \beta$. The elements of $E_{\alpha\beta}$ are called *interlayer* connections or *crossed layers* (see red edges in Fig. 3.1) while the elements of each E_α are called *intralayer* connections (see Fig. 3.1, green edges for graph G_1 and blue edges for graph G_2). The set of nodes of the layer G_α will be denoted by $X_\alpha = \left\{x_1^\alpha, \ldots, x_{N_\alpha}^\alpha\right\}$ and the adjacency matrix of each layer G_α will be denoted by $A^\alpha = \left\{a_{ij}^\alpha\right\} \in \mathbb{R}^{N_\alpha \times N_\beta}$, where:

$$a_{ij}^\alpha = \begin{cases} 1 & \text{if } (x_i^\alpha, x_j^\alpha) \in E_\alpha, \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

for $1 \leq i, j \leq N_\alpha$, $1 \leq \alpha \leq M$. Furthermore, associated with $E_{\alpha\beta}$ we define a similar adjacency matrix, called *interlayer adjacency matrix*, $A^{\alpha\beta} = (a_{ij}^{\alpha\beta}) \in \mathbb{R}^{N_\alpha \times N_\beta}$, given by:

$$a_{ij}^\alpha = \begin{cases} 1 & \text{if } (x_i^\alpha, x_j^\beta) \in E_{\alpha\beta}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

The projection network related to multilayer $\mathbb{M}$, indicated with $proj(\mathbb{M}) = (X_{\mathbb{M}}, E_{\mathbb{M}})$, is given by:

$$X_{\mathbb{M}} = \bigcup_{\alpha=1}^{M} X_\alpha \qquad E_{\mathbb{M}} = \left(\bigcup_{\alpha=1}^{M} E_\alpha \right) \bigcup \left(\bigcup_{\alpha,\beta=1,\alpha\neq\beta}^{M} E_{\alpha\beta} \right) \tag{3.3}$$

We will denote the adjacency matrix of *proj*($\mathbb{M}$) by $A(\mathbb{M})$.
This mathematical model is well suited to describe phenomena in social systems (66), air transportation networks (65) and brain networks (67), as well as many other complex systems, such as the so-called *systems medicine*, whose definition is born with the introduction of complex network methodology in biomedicine, which involves a systemic view of the organism, where there is an interplay among the various elements building living beings. Systems medicine exploits multilayer networks as a tool for combining the characterisation of the main constituents of the cell: genes, proteins and metabolites. Until now, many different complex networks have been studied, e.g. gene-gene coexpression networks, protein-protein interaction networks, metabolite-metabolite coexpression network. But each one has been considered separately, not including the strong correlations and interdependencies with the other complex networks. The interdependent representation of the cell, other than of the living being, may give a new insight about the exhibition of systemic pathological conditions. Moreover, focusing on the interdependencies among genes and proteins, we can build a multilayer network encoding both experimental setup (coexpression matrices from experimental profiles) and annotated reactions (protein-protein interaction network). This structure naturally pictures the gene control upon the production of proteins, turning into catalysers of the metabolic reactions. The multilayer representation allows also performing a multivariate statistics and an integrated clustering.
In general, by exploiting this multilayer representation, we simultaneously consider:

- the links within the different groups,
- the nature of the links and the relationships between elements that may also belong to different layers,
- the specific nodes belonging to each layer involved.

Multilayer networks can be distinguished, among others, in *multiplex networks* (64, 65, 66) and *interacting networks of networks* (61, 63). In interacting networks of networks, the nodes in the different layers represent different elements of the system. For example, in the cell, metabolites, proteins and transcription factors remain distinct biological entities. In a multiplex network, instead, the same set of nodes forms M networks, one in each layer corresponding to different types of interactions. Mathematically, a multiplex network is a particular case of multilayer network consisting of a fixed set of nodes in the various layers, that is $X_1 = X_2 = \cdots = X_M = X$, and where the only possible type of interlayer connections are those in which a given node is only connected to its counterpart (or replica) nodes in the other layers of the network, i.e., $E_{\alpha\beta} = \{(x;x); x \in X\}$, for each $\alpha, \beta \in \{1, \ldots, M\}, \alpha \neq \beta$. The paradigm of multiplex networks is suitable for describing social systems which can be seen as a superposition of a multitude of complex social networks, where nodes represent individuals and links capture a variety of different social relations. A given multiplex network $\mathbb{M}$, can be associated with several (monolayer) networks providing valuable information about it. A specific example is the projection network $proj(M) = (X_{\mathbb{M}}; E_{\mathbb{M}})$, where the elements of its adjacency matrix $A_{\mathbb{M}}$ are:

$$\overline{a_{ij}^{\alpha}} = \begin{cases} 1 & \text{if } a_{ij} = 1, \text{ for some } 1 \leq \alpha \leq M \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

Multiplex networks can be seen as monolayer networks with a certain modular structure in the mesoscale, and thus network a multiplex structure $\mathbb{M}$ can be represented as a monolayer network $\tilde{\mathbb{M}} = (\tilde{X}, \tilde{E})$, where $\tilde{X}$ is the disjoint union of all the nodes of $G_1, G_2, ..., G_M$, that is formally expressed as follows:

$$\tilde{X} = \bigsqcup_{1 \leq \alpha \leq M} X_\alpha = \{x^\alpha; x \in X_\alpha\} \tag{3.5}$$

and $\tilde{E}$ is formally given by the following expression:

$$\left(\bigcup_{\alpha=1}^{M} \left\{ (x_i^\alpha, x_j^\alpha); (x_i^\alpha, x_j^\alpha) \in E_\alpha \right\} \right) \bigcup \left(\bigcup_{\alpha,\beta=1,\alpha\neq\beta}^{M} \left\{ (x_i^\alpha, x_j^\beta); x_i \in X \right\} \right) \tag{3.6}$$

Note that $\tilde{\mathbb{M}}$ is a (monolayer) graph with $N \times M$ nodes whose *adjacency matrix*, called *supra-adjacency matrix* of $\mathbb{M}$, can be written as a block matrix:

$$\tilde{A} = \begin{bmatrix} A_1 & I_N & \cdots & I_N \\ I_N & A_2 & \cdots & I_N \\ \vdots & \vdots & \ddots & \vdots \\ I_N & I_N & \cdots & A_M \end{bmatrix} \in \mathbb{R}^{(NM)\times(NM)} \tag{3.7}$$

where I_N is the identity matrix with N dimensions. Therefore, the multilayer structure can be represented using a tensorial representation, and this procedure is called *flattening*, *unfolding* or *matricisation.* Thus, "flattening" the tensor allows also reducing the rank or the order of such a tensor constraining the space of possible multilayer networks. Therefore, this flattening process yields "supra-adjacency matrices" (or "super-adjacency matrices") (58), with the advantage over tensors of being able to represent missing nodes in a convenient way. In other words, with regards to computational issues and methods, most people are also much more familiar with working with matrices rather than with tensors.

It is important to note that the behaviours of $\mathbb{M}$ and $\tilde{\mathbb{M}}$ are related but different, since a single node of $\mathbb{M}$ corresponds to different nodes in $\tilde{\mathbb{M}}$. Therefore, the properties and behaviour of a multiplex $\mathbb{M}$ can be understood as a type of non-linear quotient of the properties of the corresponding (monolayer) network $\tilde{\mathbb{M}}$.

The framework of multilayer network extends that of other mathematical objects, that is every other mathematical object which exhibits a multilayer network structure, such as multiplex networks, networks of networks, multidimensional networks, etc. used in literature, can be represented exploiting the mathematical formulation of multilayer networks, by only introducing some constraints. We can summarise the mathematical properties of these mathematical objects as follows:

- *Multiplex networks* - as defined before, a multiplex network $\mathbb{M}$, with M layers is made up by a set of layers $\{G_\alpha; \alpha \in \{1, \ldots, M\}\}$, where each layer is a graph $G_\alpha = (X_\alpha, E_\alpha)$, with a set of nodes $X_\alpha = \{x_1, \ldots, x_N\}$, which is the same in all the layers, and this constraint formally means that $X_1 = X_2 = \cdots = X_M = X$ and interlayer connections are only possible between a node and its counterpart (or replica) nodes in the other layers of the network, which means that: $E_{\alpha\beta} = \{(x; x); x \in X\}$, for each $\alpha, \beta \in \{1, \ldots, M\}, \alpha \neq \beta$.

- *Interdependent networks* - a collection of different networks, corresponding to the various layers, whose nodes are interdependent to each other. In other words, there is a dependence of the nodes of one layer from another node, which is a control node, belonging to a different layer (61). These dependencies constitute constraints, and are represented by additional edges connecting the different layers. This structure is known as *mesostructure.* We can consider an interdependent (or layered) network as a multilayer network by identifying each network with a layer.

- *Interacting or interconnected networks* - if we consider a set of interacting networks $\{G_1, \ldots, G_L\}$, they can be modelled as a multilayer network of the L layers and whose crossed layers $E_{\alpha\beta}$ correspond to the interactions between networks G_α and G_β (see Fig. 3.2).

- *Multidimensional networks* - Formally, an edge-labeled multigraph (or multidimensional network) is a triple $G = (V, E, D)$, where respectively V is a set of nodes, D is a set of labels representing the different dimensions, and E is a set of labeled edges, that is a set of triples $E = \{(u, v, d)\,; u, v \in V; d \in D\}$. The rule is that, considered a pair of nodes $u, v \in V$ and a label $d \in D$, there could be only one edge (u, v, d). In the particular case of a directed graph, the edges $(u, v, d) \neq (v, u, d)$. Fixed the cardinality of D equals to m, each pair of nodes in G can be connected by at most m possible edges. If we also consider the weights, the edges become quadruplets (u, v, d, w), where $w \in \mathbb{R}$ is the weight of the relation between nodes $u, v \in V$ and labeled with $d \in D$. Furthermore, a multidimensional network $G = (V, E, D)$ can be modelled as a multiplex network and, hence, as a multilayer network by mapping each label to a layer (7).

- *Multilevel networks* - considering a network, whose graph is $G = (X, E)$, a multilevel network is a triple (X, E, S) (7), where $S = (S_1, \ldots, S_p)$ is a family of subgraphs (or *slices*) $S_j \in S$, with $S_j = (X_j, E_j)\,, j = 1, \ldots, p$ of the network G, which is the projection network of $\mathbb{M}$, such that:

$$X = \bigcup_{j=1}^{p} X_j \qquad E = \bigcup_{j=1}^{p} E_j \tag{3.8}$$

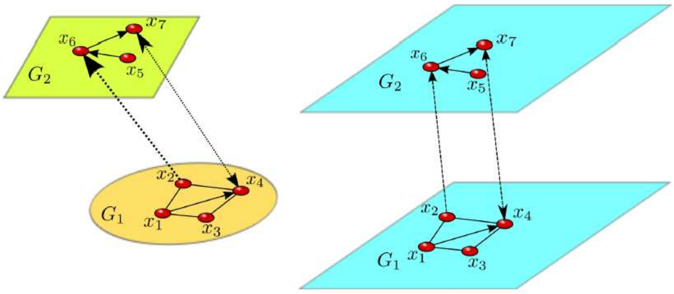


Figure 3.2: Schematic illustration of interacting networks and its mapping into a multilayer network. Each different colored network on the left side corresponds to a different blue layer on the right side.

Clearly, a multilevel network $M = (X, E, S)$ can be seen as a multilayer network, with layers $\{S_1, \dots, S_p\}$ and crossed layers $E_{\alpha\beta} = \{(x; x)\,; x \in X_\alpha \cap X_\beta\}$, and also as a multiplex network if $X_\alpha = X_\beta$ for all $1 \leq \alpha, \beta \leq p$.

- *Temporal networks* - A temporal network $(G(t))_{t=1}^{T}$ can be represented as a multilayer network with a set of layers $\{G_1, \dots, G_T\}$, where $G_t = G(t), E_{\alpha\beta} = \varnothing$, if $\beta \neq \alpha + 1$ (t is an integer, and not a continuous parameter) (68), while crossed layers are given by (see Fig. 3.3):

$$E_{\alpha,\alpha+1} = \{(x, x)\,; x \in X_\alpha \cap X_{\alpha+1}\} \tag{3.9}$$

- *Hypernetworks (or hypergraphs)* - A hypergraph is a pair $H = (X, H)$, where X is the (non-empty) set of nodes and $H = \{H_1, \dots, H_p\}$ includes (non-empty) subsets of X, known as *hyperlinks* of H. Therefore, considering a graph $G = (X, E)$, an hyperstructure S is defined as a triple (X, E, H) constituted of the vertex set X, the edge set E, and the hyper-edge set H. A hypernetwork (or hypergraph) can be represented as a multilayer network, defining a layer with G_h, a complete graph of nodes $(x_1, \dots, x_k)$ for each hyperlink $h = (x_1, \dots, x_k) \in H$, and the interlayer connections are $E_{\alpha\beta} = \{(x, x)\,; x \in X_\alpha \cap X_\beta\}$ (see Fig. 3.4).

Analysing the different multilayer mathematical models and definitions, what emerges clearly is that each model has its features and it results suitable to represent some situations, but not able to describe others. For example, hypergraphs are not able to combine all the scales of the analysis of a system, from local to global to mesoscale.

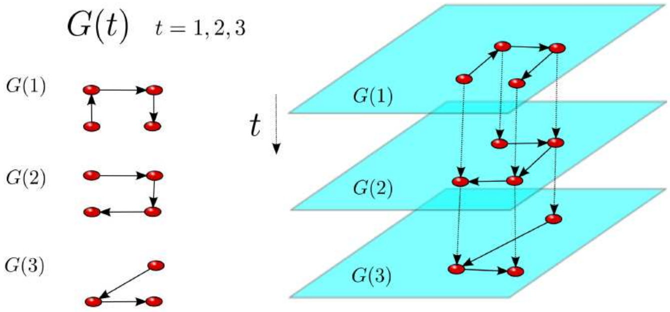


Figure 3.3: Schematic illustration of a temporal network and its mapping into a multilayer network. On the left side, it is showed that at each time instant $t = 1, 2, 3$, a different graph characterises the structure of interactions between the system's constituents. On the right side, it is illustrated the corresponding multilayer network representation, where each time instant is mapped into a different layer.

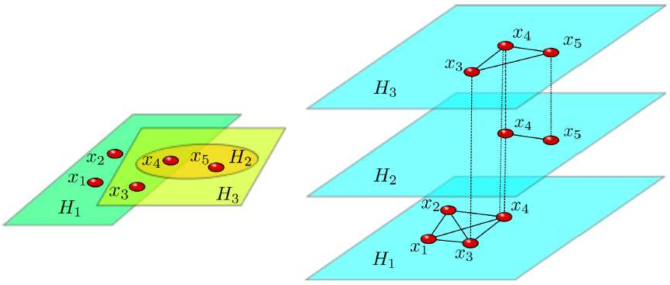


Figure 3.4: Schematic illustration of a hypergraph and its transformation into a multilayer network. Red nodes on the left side define three hyperlinks ($H_1, H_2,$ and H_3), each of which corresponds to a layer consisting of a complete graph of its nodes.

The classical example is that of a social network, where the different groups can be linked through some of their members, but also two people who know the same person do not necessarily know each other, since they may belong to entirely different groups or levels. Modelling this situation with hypernetworks, the main problem regarding systems with mesoscale structures is that they are node-based models, while many real systems combine a node-based point of view with a link-based perspective. On the other hand, as underlined in (64), multiplex networks have intrinsically a mesostructure, called interslice or interlayer coupling, which connects a node of a specific slice or layer to its copy in another layer. The mathematical formulation of multiplex networks has been recently developed through many works (65, 69, 70, 71). A considerable amount of effort has been devoted to the characterisation and modelling of multiplex networks, proposing a number of measures in various contexts of real-world multiplex networks, such as air transportation systems (65). Some other works are pointing towards a statistical mechanics formulation of multiplex networks (69)(see 3.4.6), to the extension of classical network metrics to the case of multiplexes (65) (see 3.4.1, 3.4.2, 3.4.3, 3.4.5, 3.4.4), and to model the growth of such systems (72). Finally, other works aim at characterising the dynamics and the emergent properties of multi-layer systems, especially with respect to epidemic and information spreading (70), cooperation (8, 71, 73), synchronisation (74), diffusion processes (75) and random walks on multiplex networks (64). In (65) a comprehensive formalism and many structural measures on multiplex systems are proposed, thus allowing to characterise multiplex systems with respect to node degree, edge overlap, node participation to different layers, clustering coefficient, reachability, betweenness, closeness and eigenvector centralities. A review in this field can be found in (58) and in (7).

In the following section, we are going to describe some of these key measures, starting from the traditional properties of monoplex structures and extending their definition to the context of multilayer networks and multiplex networks.

3.4 Structural measures and properties in multilayer and multiplex networks

3.4.1 Spectral properties

In general, the analysis of the spectral properties of the adjacency and Laplacian matrices of a network allows gaining insights into its structure and dynamics (4). Therefore, also in the case of multilayer networks, the introduction of suitable matrix representations enables the deep understanding of dynamics on top of complex networks and capturing the specific role of multilayer and multiplex topologies. Given a multilayer network $\mathbb{M}$, to analyse its dynamics and spectral properties, we can use the adjacency matrix A^α of each layer G_α, the adjacency matrix of the projection network $\overline{A_{\mathbb{M}}}$ and, most of all, the supra-adjacency matrix $A_{\mathbb{M}}$, whose spectrum is directly related to several dynamical processes occurring on the multilayer network.

In (6), the authors introduce the mathematical concept of *quotient graph*, which underpins the notion of multilayer network and gives crucial insights into the structure and spectral properties of the network. In fact, applying interlacing of eigenvalues of quotients of matrices, they show how the pattern of connections between layers constraints the dynamics on the whole system. A network quotient is a reduction or a simplification of the original network, and their spectral analysis aims also at quantifying the information loss in terms of eigenvalue spectrum due to this reduction process. The authors prove that if $\lambda_1 \leq \cdots \leq \lambda_N$ is the spectrum of the supra-adjacency matrix $A_{\mathbb{M}}$ of an undirected multilayer network and $\mu_1 \leq \cdots \leq \mu_{n_\alpha}$ is the spectrum of the adjacency matrix A^α of the layer G_α, then for every $1 \leq k \leq n_\alpha$, we have: $\lambda_k \leq \mu_k \leq \lambda_{k+N-n_\alpha}$. Therefore, considering a network quotient as a partition or identification of its node set, a multilayer network can be indeed recovered from its underlying network and these two quotients, the network of layers (which represents the connection pattern between layers), and the aggregate network (which results from the projection of all layers onto an aggregated single-layer network).

In (75), the authors focus on analysing the spectral properties of multiplex networks, considering the Laplacian matrix, and studying the time scales associated with diffusion processes that take place on multiplex networks. The proposed mathematical setting not only allows studying diffusion dynamics, but it also represents a good approximation for other dynamical processes, whose dynamics can be captured by the behaviour of the eigenvalues of the Laplacian matrix. The Laplacian matrix of a multiplex $\mathbb{M}$, known also as *supra-Laplacian matrix*, denoted by $L_{\mathbb{M}} = L$ is a $MN \times MN$ matrix, having the following form:

$$L = \begin{bmatrix} D_1\mathbf{L^1} & 0 & \cdots & 0 \\ 0 & D_2\mathbf{L^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_M\mathbf{L^M} \end{bmatrix} + \begin{bmatrix} \sum_\beta D_{1\beta}\mathbf{I} & -D_{12}\mathbf{I} & \cdots & -D_{1M}\mathbf{I} \\ -D_{21}\mathbf{I} & \sum_\beta D_2 L^2 & \cdots & -D_{2M}\mathbf{I} \\ \vdots & \vdots & \ddots & \vdots \\ -D_{M1}\mathbf{I} & -D_{M2}\mathbf{I} & \cdots & \sum_\beta D_{M\beta}\mathbf{I} \end{bmatrix} \tag{3.10}$$

In (3.10), $\mathbf{I}$ is the $N \times N$ identity matrix and $\mathbf{L}^\alpha$ is the $N \times N$ Laplacian matrix of the network layer α, whose elements are $L^\alpha_{ij} = s^\alpha_i \delta_{ij} - w^\alpha_{ij}$, where s^α_i is the strength of the node i in layer α, given by: $s^\alpha_i = \sum_j w^\alpha_{ij}$. The authors in (75) prove as the diffusion dynamics depends strongly on the spectral properties of L. More in general, the behaviour of any linearised dynamical process on a complex system is related to the Laplacian matrix of the underlying network and particularly to its second smallest eigenvalue, also called algebraic connectivity λ_2, e.g. it is essential to evaluate the time required to synchronise phase oscillators (76), or to converge to the maximum entropy state in a diffusion process (75). Futhermore, the largest eigenvalue of the Laplacian matrix plays a determinant role in the assessment of the stability of the synchronisation manifold in networks of coupled oscillators. In (77), based on the features of a multiplex network, they model its Laplacian matrix in terms of a decomposition between intra- and interlayer structure. The target is to characterise the spectrum of the Laplacian, using perturbation theory, and hence the behaviour of several dynamic processes, assessing the diffusion time scales in any multiplex structure, and inferring the optimal value of the synchronisation ratio in terms of the master stability function. One of the most interesting findings of (75) is that, in some particular cases, the coupling of networks shows a super-diffusive behaviour, meaning that diffusive processes in the multiplex are faster than in any of the networks that form it separately. Even if it is worth to specify that in general, the superdiffusive behaviour is not guaranteed and depends on the specific structures coupled together.

In this field of studying spectral properties, also irreducibility has been analysed in fact, as we know, in network theory not only the eigenvalues, but also the eigenvectors play a key role in determining the spectral properties (77). Studying the spectral irreducibility of a multiplex network means the study of the existence of the Perron vector, that is the existence of a positive and normalised eigenvector, guaranteed by the Perron-Frobenius theorem (as it allows verifying if the corresponding matrix is irreducible). Then, we can relate this irreducibility with that one in each layer and on the projection network. For example, in (77), the authors demonstrate that if $w_{ij} > 0$ and the adjacency matrix of the projection network $\overline{A_{\mathbb{M}}}$ is irreducible, then the matrix corresponding to the global heterogeneous-like centrality is irreducible. In (78), it is proved the irreducibility also considering random walkers in multiplex networks, thus random walkers have a unique stationary state.

3.4.2 Centrality measures

The importance or ranking of nodes in the network, that is the identification of nodes that play a central structural role is traditionally a key aspect when analysing complex networks. In single-layer networks, the classical measures used to assess the structural centrality of a node are the node degree, the closeness, the betweenness, eigenvector centrality. In this section, we are going to extend these measures to multilayer networks. The *degree* of a node measures the relevance of a node according to the number of links incident upon a node. We define the degree in a multiplex network as follows:

$$\mathbf{k}_i = \left(k_i^{[1]}, \ldots, k_i^{[M]}\right), i = 1, \ldots, N \tag{3.11}$$

where $k_i^{[\alpha]}$ is the degree of the node i on a given layer α, defined as: $k_i^{[\alpha]} = \sum_j a_{ij}^{[\alpha]}$.

One of the major issues related to centrality measures is to rank structurally the nodes, creating an ordered list of vertices according it their importance in the network. As the node degree is a vector and we can define many complete orders in $\mathbb{R}^M$, we need to clarify which of these nodes are relevant. To this aim, we can aggregate the degree of nodes and define the so-called *overlapping degree* of the node i, as:

$$o_i = \sum_j o_{ij} = \sum_\alpha a_{ij}^{[\alpha]} = \sum_{\alpha=1}^{M} k_i^{[\alpha]} \tag{3.12}$$

where $o_{ij} = \sum_\alpha a_{ij}^{[\alpha]}$ is the *edge overlap* of edge $i - j$, and by definition: $0 \leq o_{ij} \leq M, \forall i, j$. We can note as overlapping degree o_i represents the correct factor to normalise the components of the degree vector k_i.

Closeness and betweenness centralities are strictly connected with the metric structure of the network so, defining the geodesic and metric structure, it is easy to extend these measures to multilayer networks. *Eigenvector centrality* is instead based on the spectral properties of the adjacency matrix, and its definition depends not only on the number of links of each node but also on the quality of these connections. In literature there are a lot of definitions of this centrality measure, but probably the simplest way to calculate it in multiplex networks is to consider separately the eigenvector centrality in each layer $1 \leq \alpha \leq M$, given by: $c_\alpha = (c_i^\alpha, \cdots, c_N^\alpha)$. Thus, eigenvector centrality is another vector:

$$\mathbf{c}_i = (c_i^1, \cdots, c_i^M) \in \mathbb{R}^M \tag{3.13}$$

where each element is the centrality in the corresponding layer. Starting from (3.13), the following (3.14) gives the *indipendent layer eigenvector-like centrality* of $\mathbb{M}$:

$$C = (\mathbf{c}_{[1]}^T \mid \mathbf{c}_{[2]}^T \mid \cdots \mid \mathbf{c}_{[M]}^T) \in \mathbb{R}^{N \times M} \tag{3.14}$$

The main drawback of this this parameter is that it does not fully consider the multilayer interactions between layers and the subsequent influence in the centrality of each node.

Since the centrality measure of a node also depends on the centrality of its neighbours, distributed among all the layers, we need to define the so-called *uniform eigenvector-like centrality*, defined as a positive and normalised eigenvector $\tilde{c}$ (if it exists) of the matrix $\tilde{A}$ given by the following (3.15):

$$\tilde{A} = \sum_{\alpha=1}^{M} (A^{[\alpha]})^T \tag{3.15}$$

where $(A^{[\alpha]})^T$ is the transpose of the adjacency matrix A of the layer α. This type of centrality measure does not consider neither the different importance or influence of a node through the various network layers, nor the mutual influence between layers. In fact, in a multilayer or multiplex structure, the relevance of a node within a specific layer changes according to the other layers, therefore it becomes fundamental to take into account also the situation where there is a heterogeneous influence among layers. To include also this influence in the centrality definition, the influence matrix $W = (w_{\alpha\beta}) \in \mathbb{R}^{N \times M}$, with $W \geq 0$, is introduced, where $w_{\alpha\beta}$ measures the influence on the layer G_α given by the layer G_β. Having fixed G and W, we define the *local heterogeneous eigenvector-like centrality* of G on each layer G_α as a positive and normalised eigenvector $c^*_\alpha \in \mathbb{R}^N$ (if it exists) of the following matrix (3.16):

$$A^*_\alpha = \sum_{\beta=1}^{M} w_{\alpha\beta}(A^{[\beta]})^T \tag{3.16}$$

Thus, C^* represents the local heterogeneous eigenvector-like centrality matrix, defined as:

$$C^* = (\mathbf{c}^*_1 \mid \mathbf{c}^*_2 \mid \cdots \mid \mathbf{c}^*_M) \in \mathbb{R}^{N\times M} \tag{3.17}$$

It is important to observe how the centrality of the node x_i^α belonging to the layer α depends not only locally but also globally on the neighbours linked to it belonging to the other layers. This leads to another definition of centrality, known as *global heterogeneous eigenvector-like centrality* of $\mathbb{M}$, defined, starting from the influence matrix W, as a positive and normalised eigenvector $c^\otimes \in \mathbb{R}^{NM}$ of the following matrix ((3.18)):

$$A^\otimes = \begin{bmatrix} w_{11}(A^{[1]})^T & w_{12}(A^{[2]})^T & \cdots & w_{1M}(A^{[M]})^T \\ w_{21}(A^{[1]})^T & w_{22}(A^{[2]})^T & \cdots & w_{2M}(A^{[M]})^T \\ \vdots & \vdots & \ddots & \vdots \\ w_{M1}(A^{[1]})^T & w_{M2}(A^{[2]})^T & \cdots & w_{MM}(A^{[M]})^T \end{bmatrix} \in \mathbb{R}^{NM\times NM} \tag{3.18}$$

which is the result of the Khatri-Rao product of the matrices W and $((A^{[1]})^T \mid (A^{[2]})^T \mid \cdots \mid (A^{[M]})^T)$, where:

$$W = \begin{bmatrix} w_{11} & \cdots & w_{1M} \\ \vdots & \ddots & \vdots \\ w_{M1} & \cdots & w_{MM} \end{bmatrix} \tag{3.19}$$

Introducing the following notation:

$$c^\otimes = \begin{bmatrix} c_1^\otimes \\ c_2^\otimes \\ \vdots \\ c_M^\otimes \end{bmatrix} \in \mathbb{R}^{NM} \tag{3.20}$$

where $c_1^{\otimes},\ c_2^{\otimes},\ \cdots,\ c_\alpha^{\otimes},\ \cdots,\ c_M^{\otimes} \in \mathbb{R}^N$, we can define the *global heterogeneous eigenvector-like matrix* $C^{\otimes}$ of $\mathbb{M}$, as follows:

$$C^{\otimes} = \left(c_1^{\otimes} \mid c_2^{\otimes} \mid \cdots \mid c_M^{\otimes}\right) \in \mathbb{R}^{N \times M} \tag{3.21}$$

3.4.3 Clustering

It is well-known as one of the most significant characteristics of complex real-world single-layer networks, e.g. in collaboration networks, is the tendency of nodes to form triangles, i.e. simple cycles involving three nodes. This tendency is expressed by the famous popular saying "*the friend of your friend is my friend*". To quantify this tendency, Watts and Strogatz (56) introduced ths so-called *node clustering coefficient.* This notion can be easily extended to multilayer networks in different ways. Given a network $G = (X, E)$, the clustering coefficient of a given node i is given by:

$$c_G(i) = \frac{\text{number of links between the neighbours of } i}{\text{largest possible number of links between the neighbours of } i} \tag{3.22}$$

If i, j and k are three people with mutual relations between i and j, as well as between i and k, the clustering coefficient measures the likelihood that also j and k are related to each other. The *global clustering coefficient* of G is given by the average of the clustering coefficients of all the nodes in the network. Instead, the *local clustering coefficient* is the density of the local node's neighbourhood, and it represents a measure of transitivity. In some works (27), the global clustering coefficient is defined in relation with the network features, as a network transitivity, which means the capability of nodes in the network to form triangles, thus formally we have:

$$T = \frac{\text{number of triangles in the network}}{\text{number of triads in the network}} \tag{3.23}$$

To extend the clustering coefficient defintion to the multilayer networks, we need to consider not only intralayer connections but also interlayer links. It is important to notice how in general, different layers may show similar or dissimilar patterns of clustering, but it is interesting to study to which extent the multiplexity affects the formation of triangles, i.e. how the presence of different layers can give rise to triangles which cannot exist at the level of single layers. Therefore, we need to extend the notion of triangle to take into account the richness added by the presence of more than one layer. Before defining the clustering coefficient for multilayer networks, we need to introduce some basic notions about triangles and triads. The notion of triangle is extended in multilayer networks due to the presence of more than one layer, thus we define a 2-triangle as a triangle formed by an edge belonging to one layer and two edges belonging to a second layer. Similarly, we define a 3-triangle as a triangle composed by three edges all lying in different layers. Furthermore, we define the notion of 1-triad, when the two edges of triangles are on the same layer, while in the case of a 2-triad, the two links belong to different layers of the system. In order to quantify the added value provided by the multiplex structure in terms of clustering, the authors of (65) consider two parameters of clustering interdependence, I_1 and I_2. I_1 (I_2) is the ratio between the number of triangles in the multiplex which can be obtained only as 2-triangles (3-triangles), and the number of triangles in the aggregated system. Then, $I = I_1 + I_2$ is the total fraction of triangles of the aggregated network which cannot be found entirely in one of the layers. Given these notions, we can give two further definitions of clustering coefficient for multiplex networks. In terms of multilayer adjacency matrix, we have:

$$C_{i,1} = \frac{\sum_{\alpha}\sum_{\alpha'\neq\alpha}\sum_{j\neq i,m\neq i}\left(a_{ij}^{[\alpha]}a_{jm}^{[\alpha']}a_{mi}^{[\alpha]}\right)}{(M-1)\sum_{\alpha}\sum_{j\neq i,m\neq i}\left(a_{ij}^{[\alpha]}a_{mi}^{[\alpha]}\right)} = \frac{\sum_{\alpha}\sum_{\alpha'\neq\alpha}\sum_{j\neq i,m\neq i}\left(a_{ij}^{[\alpha]}a_{jm}^{[\alpha']}a_{mi}^{[\alpha]}\right)}{(M-1)\sum_{\alpha}k_i^{[\alpha]}\left(k_i^{[\alpha]}-1\right)} \tag{3.24}$$

Since each 1-triad can theoretically be closed as a 2-triangle on each of the M layers of the multiplex excluding the layer to which its edges belong, in order to have a normalised coefficient, we have to divide the term by $M-1$. In addition to this, we define a second clustering coefficient for multiplex networks as the ratio between the number of 3-triangles with node i as a vertex, and the number of 2-triads centered in i. Therefore, we have:

$$C_{i,2} = \frac{\sum_{\alpha}\sum_{\alpha'\neq\alpha}\sum_{\alpha''\neq\alpha,\alpha'}\sum_{j\neq i,m\neq i}\left(a_{ij}^{[\alpha]}a_{jm}^{[\alpha'']}a_{mi}^{[\alpha']}\right)}{(M-2)\sum_{\alpha}\sum_{\alpha'\neq\alpha}\sum_{j\neq i,m\neq i}\left(a_{ij}^{[\alpha]}a_{mi}^{[\alpha']}\right)} \tag{3.25}$$

where $M-2$ is a normalisation coefficient. While $C_{i,1}$ (see (3.24)) is a suitable definition for multiplexes with $M \geq 2$, $C_{i,2}$ can only be defined for systems composed of at least three layers. Averaging over all the nodes of the system, we obtain the network clustering coeffcients C_1 and C_2.

In (65) the authors also extend the definition of transitivity. To this purpose, they propose two measures of transitivity: T_1, as the ratio between the number of 2-triangles and the number of 1-triads, and T_2 as the ratio between the number of 3-triangles and the number of 2-triads. We can observe as clustering interdependencies I_1 and I_2, average multiplex clustering coefficients C_1 and C_2, and multiplex transitivities T_1 and T_2 are all global network variables which give a different view on the multilayer patterns of clustering and triadic closure with respect to the clustering coefficient and the transitivity computed for each layer of the network.

3.4.4 Entropy, Participation Coefficient, Social Reinforcement and Interdependence

In Sect. 3.4.2 we have defined the degree centrality of a node. As we know, the degree distribution is the probability distribution of these degrees over the single layer or the whole network. Most often degree distribution is heterogeneous between layers and there is a weak correlation between the degrees of the same node at different layers. This suggests the need to introduce a measure able to quantify the ability of a node to be connected through a pattern in all the layers of the multiplex structure. Indeed, two nodes i and j could have the same overlapping degree, $o_i = o_j$ (see (3.12)), but for instance node i could be a hub on a generic layer α and only an isolated node on the other layers, because all its connections lie in that layer, while the node j has the same number of links on each layer, so that we have: $o_i = k_i^{\alpha}$ and $o_j = Mk_j^{\alpha}, \forall \alpha$. Thus, even though the two nodes have the same overlapping degree, if we look at them from a multiplex perspective, they have totally different roles. By contrast, it may occur also the opposite situation, where the overlapping degree are different, but the two nodes are very similar looking at the contribution of each layer to the overall overlapping degree of the two nodes. In order to describe the degree distribution over the various layers of the multiplex, we define the so-called *entropy* of the multiplex degree, as follows:

$$H_i = -\sum_{\alpha=1}^{M} \frac{k_i^{[\alpha]}}{o_i} \ln\left(\frac{k_i^{[\alpha]}}{o_i}\right) \tag{3.26}$$

From the above definition, it follows that entropy $H_i = 0$ if all the connections lie in a single layer, while it will reach the maximum value when the degree distribution over the various layers is uniform. Similarly, we can define the multiplex participation coefficient P_i of the node i as follows:

$$P_i = \frac{M}{M-1}\left[1 - \sum_{\alpha=1}^{M}\left(\frac{k_i^{[\alpha]}}{o_i}\right)^2\right] \tag{3.27}$$

Then, P_i measures how the links of node i are distributed among the M layers of the multiplex, and its values ranges in the interval $[0,1]$, where values close to 0 mean that links are concentrated in just one or few layers (in the limit case $P_i = 0$ all the edges of the node lie in one layer), while values close to 1 mean that the distribution is nearly uniform among layers (in the limit case $P_i = 1$, the node has the same links on each of the M layers). The overall participation coefficient P of the multiplex is calculated as the average of P_i over all the nodes of multiplex, i.e. $P = 1/N \sum_i P_i$. It is clear how the two quantities P_i and H_i give very similar information.

Up to now we have focused on measures related to the single node in the multiplex, our purpose is now to quantify the importance of each layer as a whole. In Sect. 3.4.2 we have defined the edge overlap and the overlapping degree (see (3.12)), which allows evaluating the presence of correlations across the layers of the multiplex. Although the edge overlap distribution gives information about inter-layer correlations, it does not allow assessing whether one layer is more relevant than others in the multiplex. Hence, we can define the following quantity, defined in (65) as *social reinforcement*, which measures the conditional probability to find a link between two nodes at a layer α' given the presence of a link between the same nodes at layer α:

$$P\left(a_{ij}^{[\alpha']} \mid a_{ij}^{[\alpha]}\right) = \frac{\sum_{ij} a_{ij}^{[\alpha']} a_{ij}^{[\alpha]}}{\sum_{ij} a_{ij}^{[\alpha]}} \tag{3.28}$$

where the denominator is the total number of links at layer α, or the size of the layer, K^α, obtained summing all the degrees k_i of the nodes in the layer, while the numerator quantifies the number of the edges which are present also in the layer α'. In (65), the authors consider the multi-layer network of Indonesian terrorists, where social reinforcement is able to quantify if one of the layers (in this case the Trust layer) acts as a dominant layer in the multiplex, or in other words if there is a layer representing the strongest connection between two nodes, whose existence can lead to an edge creation also in the other layers. Thus, collaboration and communication in this example of multiplex network is ruled by this mechanism of social reinforcement, and we can think to extend it also in other social contexts.

Reachability, as well as in the single-layer networks, is related to the existence and length of a shortest path between a pair of nodes in the multiplex network. It is clear that in presence of more than one layer, the shortest path can change across layers, thus to capture how the multiplex affects the reachability of each node in the network, it is possible to define the *node interdependence*, as follows:

$$\lambda_i = \sum_{j \neq i} \frac{\psi_{ij}}{\sigma_{ij}} \tag{3.29}$$

where σ_{ij} is the total number of shortest paths between the two nodes i and j, while ψ_{ij}, among all of these shortest paths, only considers those exploiting links in two or more than two layers. Therefore, λ_i ranges in the interval $[0, 1]$ and, in particular, $\lambda_i = 1$ when all the shortest paths make use at least two layers, while $\lambda_i = 0$ if each of the shortest paths exploits only one of the M layers of the multiplex structure. Overall, the *network interdependence* is the average of λ_i over all nodes: $\lambda = 1/N \sum_i \lambda_i$. It is easy to observe how the interdependence is anti-correlated to the overlapping degree, as if a node has a high overlapping degree, it will have a lot of opportunities to choose the first edge to reach another node, while if it has a low overlapping degree, its shortest paths are more constrained in terms of edges and layers.

3.4.5 Correlations

Multiplex networks contain a lot of information, more than considering the single layers isolatedly. This is one of the reasons why it is interesting to model and study a system as a multiplex network, instead of simply aggregating together its component networks trated as single-layer networks. Multiplex network allows including measures of correlation between both nodes and layers. Finding out the correlations is one of the most important aspects in multilayer networks. Degree correlation in a single layer network indicates how much the degree of a node can be correlated, positively or negatively, or not correlated with the degree of its neighbours. In particular, a positive degree correlation indicates that the hubs are preferentially connected to each other through the network, while a negative degree correlation means that they are preferentially connected to low-degree nodes. In the case of multiplex networks, the notion of correlation is by far richer than in the single-layer case, as we can still investigate the classical degree-degree correlation in each layer, but evaluating and measuring correlations between a property of a node in one layer and the same or other properties of the same node in another layer is surely more interesting and gives more information. Before distinguishing between the most important types of correlations in multiplex networks, listing the most important properties of each of them, it is worth to note that multiplexity introduces a novel complexity due to the non-trivial patterns of node involvement across layers (79). We can outline the various types of correlations in multiplex networks as follows:

- *interlayer degree correlations*: it indicates wheter a hub in one layer is also a hub in another layer, or a low degree node. Thus, it allows quantifying if there are assortative or disassortative correlations: in the former case, nodes having a certain degree are preferentially connected to other nodes having similar degree, in the latter case they are connected to dissimilar degree nodes. Social networks represent the most clear example of assortative networks, in contrast with other types of networks, such as technological and biological networks, showing disassortative degree correlations, where hubs are preferentially linked to poorly-connected nodes. The interlayer degree-degree correlation allows exploring whether there are correlations in the degrees of a node across different layers. To quantify it, we can use the Pearson, Spearman and Kendall's correlation coefficients (79).
- *overlap* and *multidegree*: as we defined in (3.12), the edge overlap gives information about patterns correlation through the network. For instance, mobile communications occur at different layers, such as phone, email, instant messaging applications (e.g. Whatsapp, Viber, Hangouts, Skype, etc.), hence the edge overlap is high and it is able to quantify this type of correlation. Multidegree of the node specifies the overlapping pattern.
- *multistrengths* and *inverse multiparticipation ratio of weighted multiplex*: if we represent the multiplex as a weighted multiplex network (80), the weights of the links in the different layers can be correlated with other structural properties of

the multiplex. The classical example is that of citation networks, where authors tend to cite collaborators differently from other scientists. The definition of these two correlation measures derives from the concept of *multilink*, which is the set of links connecting two nodes in the different layers of multiplex. The multistrength of a node is the distribution of weights in the multiplex, while the inverse multiparticipation ratio is the distribution of weights incident upon nodes having a certain degree K in the multiplex. Then, these two parameters represent two correlation measures able to capture the correlation between the weights distribution and topology of the multiplex (80).

- *node pairwise multiplexity*: the activity of a node B_i is the number of layers in which the node is present. The evaluation of activity across the different layers can give information about correlation in fact, if two nodes are not active in all the layers, it is more likely that their activity patterns are correlated, as for example they are active on the same layers. To capture these correlations, we use the node pairwise multiplexity (79).
- *layer pairwise multiplexity*: analogously to the previous measure, also this measure is based on layer activity, so that two layers show correlated activity patterns if, for instance, the active nodes are the same. To capture these correlations between layers, we use the layer pairwise multiplexity.

3.4.6 Multiplex networks models, network ensembles and entropy

As we know, the issue of generative models of a network has been addressed in the case of monolayer networks, unveiling some of the most important mechanisms, such as Barabàsi-Albert model based on preferential attachment, explaining the emergence of power-law distributions (55), and the static models, as a generalisation of random graphs to the case of more complex structures, able to describe network ensembles with certain properties (e.g. exponential random graphs) (2, 3). Also in multilayer networks, the generative models can be split into two main classes: *growing multilayer network models*, where the basic mechanism is the generalised preferential attachment, and *multilayer network ensembles*, which are ensembles of N nodes with specified structural constraints (72). In statistical mechanics the concept of ensemble is defined as a large number of copies of a system (which can be also infinite), considered all at once, each of which represents a possible state in which the real system might be in (that is a microstate). Also in network theory, ensemble models consider not only a single network, but a probability distribution over many possible networks. Apart from constraints, the other features of the network ensemble are completely random (randomised network ensembles), so that for a give real network, it is possible to generate different randomised network ensembles, which depend on the fixed structural constraints (69). The ensembles allow identifying non-random patterns and properties in real networks, thus an observed real network represents a single realisation of a larger statistical ensemble including all the possible realisations consistent with the defined constraints. Furthermore, given a set of constraints, we can distinguish between the microcanonical and the canonical network ensembles: in the first case, network ensemble satisfies the hard constraints, while in the second case, constraints are satisfied in average. It is important to observe how the distinction between microcanonical and canonical network ensembles has to be take into account when analysing dynamical processes on network, as it can strongly influence their critical behaviour. From a mathematical point of view, a statistical ensemble of networks can be defined as a set of graphs where, for each graph $\vec{G} = \left(G^1, G^2, \ldots, G^M\right)$, we can define a probability $P(\vec{G})$. Now, we define the concept of entropy in multiplex ensembles. It is well known as the concept of entropy appears in many different theories, starting from the first probabilistic interpretation of thermodynamic entropy given by Ludwig Boltzmann, and becoming a crucial concept of information theory. In statistical mechanics, defined a set of macroscopic variables, entropy constitutes a measure of the spreading out of probability over different possible microscopic states. Therefore, given the probability $P(\vec{G})$, the entropy of the multiplex S is given by:

$$S = -\sum_{\vec{G}} P(\vec{G}) \log P(\vec{G}) \tag{3.30}$$

It measures the logarithm of the typical number of multiplex networks in the ensemble. Moreover, as described comprehensively in (69), other than distinguishing between microcanonical and canonical multiplex ensembles, there is also a further distinction between uncorrelated and correlated multiplex ensembles, since in ensembles of networks, links in the different layers can be uncorrelated or correlated, and this produces different definitions of $P(\vec{G})$ and entropy as a consequence.

Chapter 4

Social Networks, Big Data and ICT in Healthcare

In the future, life and death decisions will depend on having more data and more organised knowledge. These data will overcome traditional scale and dimensions, thus we will need to think about new kind of strategies which involve ICT. Collect, organise and compute every aspect will be crucial for survival of patients and for healthcare management. Following a bio-inspired approach to ICT, in (81) we relate big data and the data-intensive computing issues in the future vision of a smart healthcare. The multidimensional approach to comorbidity and the introduction of a social dimension of analysis allow finding out correlations and causality relations between diseases and patients also considering the connectedness and social contagion processes. In this way, we obtain an evolution from data to multi-agents through the introduction of personalised medicine in a multilayer architecture.

4.1 Introduction

Innovation and widespread of Information and Communication Technologies (ICTs) are pushing forward the interest in this research field. ICT should be planned and designed to conduct almost any activity which involves logical strategies and operations in a variety of application areas. In chap. 2 we have discussed about the similarities and connections between biology and ICT (20), and about some of the main challenges of bio-inspired ICT in such a complex, heterogeneous, and increasingly highly connected world. In the era of high connectedness, users are becoming more active in interacting, sharing and collaborating, through social networks. The latest trends in network expansion and data provisioning have introduced new challenges in network science and engineering, such as: availability, efficiency, evolvability, computational intelligence, data intelligence, effectiveness intelligence. In this way, this process produces collective intelligence, spread out in many different areas and related to networks phenomena. Nodes, data and interactions represent both subject and object of the knowledge distillation in social networks. This produces a large amount of unstructured information, and a multitude of social subjects and objects as information sources and users. Heterogeneity and plurality, based on bio-inspired approach, are the representation of the bio-diversity: in terms of knowledge, it is represented by health, in terms of data, it is represented by complexity and, in terms of social networking, it is represented by collective intelligence.

In this scenario, as well as biodiversity in biological phenomena, big data represent an excellent resource to improve the understanding of the networks phenomena, optimising ICT interventions, pushing towards greater knowledge and collective intelligence. The healthcare industry historically generates large amounts of data, related to patient care (82). The current trend is toward digitalisation of these large amounts of data. The basic idea is to improve the quality of healthcare delivery and reduce the costs. Moreover, big data in healthcare hold the promise of supporting a wide range of medical and healthcare functions, including clinical decision support, disease surveillance, and population health management (82). Big data in healthcare refers to electronic health data sets, such as Electronic Health Records (EHRs) so large (for instance, big data for U.S. healthcare will soon reach the zettabyte scale, and not long after, yottabyte) and complex that they are almost impossible to manage with traditional data management tools and methods. For all these reasons, Public Health and Clinical Intervention management is a future challenges of next future research. Medical statistics, clinical epidemiology, decision analysis and health economics, are different disciplines that, in the smart health context, are becoming more increasingly connected in a fruitful and compelling way (83).

Comorbidity refers to the co-occurrence of different medical conditions or diseases in the same patient. In the comorbidity perspective, the patient becomes the centerpiece of a system characterised by multiple, complex and correlated conditions or diseases (84). In the biomedical research field, for example, one of the main problems is to find out the correlations between human diseases and the subtle mechanisms, at the molecular level, shared by apparently dissimilar diseases. By exploring genes-diseases relations, the pathogenesis may emerge, and it could lead to better diagnosis and treatment (85). In this field, the study of comorbidity relations becomes crucial, as it allows finding out dynamic patterns, correlations and causality mechanisms between diseases and patients and, even more interesting, it becomes crucial to study the social dimension of comorbidity, focusing on how the strength of the ties, the connectedness, and communications lead to an evolutionary dynamics of the social network and represent a further degree of analysis in the comorbidity assessment. Medicine is subject to many challenges, due to many kinds of innovations in different fields. The social networking, the ubiquitous networks, big data and ICT strategies can support the future innovation in healthcare design and management, making the medicine able to face the different opportunities in important challenges. As an added value in the future development of medicine, the involvement of patients in the healthcare process is increasingly becoming an important factor. The social dimension will change radically the approach for the future medicine, involving the precision medicine (86), by exploiting the dynamic analysis, the study of the dynamic evolution of behaviours and needs, the disease contagion processes, and also through the sharing, cooperation and collaborative approach thanks to the new technologies.

In this chapter at the beginning, we will give a brief scientific background of the main issues and challenges of big data, and in particular big data storage and analytics, focusing more on data-intensive computing (see sect. 4.2). In sect. 4.3, we will shed light on the challenges of big data in the healthcare context. After having discussed a health mining approach (see sect. 4.4), we will raise up the issue of comorbidity relations, bringing out the importance of considering a social dimension (see sect. 4.5). In sect. 4.6, we will focus on the future challenges of medicine and innovation in healthcare through the introduction of P5 medicine. In sect. 4.6.1, we propose a novel multiagent system in a multilayer structure, which allows extracting a smart organised knowledge from big data. In sect. 4.7, we summarise some of the main proposed aspects of bio-inspired ICT in healthcare. Finally, in sect. 4.8, we will discuss about a social evolution of the concepts of node and data, developing a novel social multilayer paradigm in order to investigate the complex dynamic patterns in the analysis of ICT systems.

4.2 Big Data: An Overview

4.2.1 Introduction and background

Big Data Analysis is an interdisciplinary topic which combines many different disciplines, such as social networking (87), computational social science (88), data integration and fusion (89), data mining (90), trend discovery, multimedia and business management, opinion analysis and sentiment analysis (91), machine learning (92), graph mining, knowledge-based systems (91), decision support system (93), healthcare and biomedical applications (82), cloud computing.

Big data is defined as large amount of data which requires new technologies and architectures so that it becomes possible to extract value by capturing and analysing process. In the near future there will be a high availability of digital data difficult to be managed and analysed by using only the traditional techniques and conventional software tools and technologies. This kind of data, in all shapes and sizes, is growing at very fast rates. The actual interest of research is to shed light on suitable methodologies to perform an effective analysis on this huge quantity of data, exploiting the benefits brought by it. As stated by (94), big data can be defined using the following properties associated with variety, velocity, volume, variability, value (the so-called “5V model”) as illustrated in Fig. 4.1. These features represent the advantages, the main issues and the basis for the future challenges.

ICT should attend and provide support to the designing of innovative systems able to handle large amount of data efficiently and effectively. The challenge is also to insert this process in a context of social networking linked to application areas, such as the healthcare, in which nodes and information become subject and object of interactions, information sources and users of the services. Through the mechanisms of data fusion and integration, another challenge may be to select, among all the collected data, the most important data related to time, space and, more in general, to the context-aware process. The challenges consist not only of storing, collecting, processing and managing the vast volume of data, but also to analyse and extract meaningful value from it. If this process of extraction is inserted in a social context, the obtained knowledge may become object of social sharing and part of the social contagion process (30), promoting cooperation and spreading collective actions in the population.

The concept of big data has permeated social sciences and humanities (60). The complexity lies in collecting data combining datasets from multiple sources, matching coherently incomplete data, and managing the data paying attention on ownership questions and privacy issues, since many datasets contain sensitive data regarding users (60). Issues and challenges are focused on the following topics: big data, big data analytics and storage and data-intensive computing. Data-intensive computing has become a research problem in science, industry and computer academia. The data explosion has lead to the development of various data-intensive computing applications, including for example text data processing, large scale graph computing. Data-intensive computing research faces challenges of scalability of massive data management and processing of integrated or unstructured data, fault tolerance and high availability issues. Therefore, data-intensive computing models need to be suitable for large scale data sets parallel computing, multiple virtual machine task scheduling and constructing new data-intensive computing applications. Although industry and academia has proposed some approaches for data-intensive applications, issues related to both data-intensive and computation intensive features are still not solved at present.

4.2.2 Big Data Analytics and Storage

The Internet's network expansion and the resource provisioning have introduced new challenges in network science and engineering. The first challenge is the availability, both in terms of network infrastructure and services; it becomes really challenging facing the scaling and the large amount of network applications which exceed the speed of network upgrades (e.g. video streaming). The problem becomes even stronger due to the mobility and heterogeneity of devices. Therefore, this means that the Internet should become more flexible, adaptable, robust and resilient, maintaining its availability. Another challenge is the efficiency in delivering information, in particular in social networks because of the large amount of applications daily installed from users. The third feature of the future Internet should be evolvability, so the Internet architecture has to be able to face the emerging trends in data-intensive computing and the continuous infrastructure and applications changes. The future Internet architecture has to address the increasing computational complexity. Other challenges are the data intelligence, in order to sample, analyse and find out correlations and causality relations among data, increasing the quality metrics, e.g. using the users' quality of experience. Furthermore, to manage the network resources, data-driven approaches are needed, simplifying the problem formulation and speeding up the decision-making processes, reducing cost and improving network performance. Big data opens a new era for science discovery through data-driven computing. This new paradigm applies to the design of the future Internet, facing issues such as supporting new applications, the efficient resource utilisation, and continuous evolution.

Following the 5V model, explained in (95), major interest for the bio-inspired approach is addressed to the heterogeneity of data, identified as “Variety” in the 5V model. This feature has a dual nature, in fact from one side it represents a problem, but it is also an opportunity. The achievable benefits are for end-users, enterprise consumers, services providers, and prosumers which can extract knowledge and a lot of information from this huge data variety. The concept of “variety”, which can be identified as the biodiversity of our system, using a bio-inspired approach, is the result of unstructured data. Unstructured data refers to information without a pre-defined data model or does not fit well into relational tables. The nature of the unstructured leads to consider that we have a variety of formats which can collect. Moreover, there is no standardisation on content extraction and storage of it and, finally, the data are from multiple and different sources (social media, data providers, mobile apps, sensors, etc.) and also in different layers (7). We can collect data in different and long time intervals, introducing a large-scale resolution problem (60). Thus, big data introduces a need to manage efficiently the measures of data being stored and manipulated. Data storage is, in fact, the basis for big data networking. From the analytical point of view, there are also high challenges to be addressed. Thus, the research is interested in investigating all the new techniques and paradigms being developed and implemented. The challenge related to the problem and, at the same time, the opportunity of exploiting a large amount of data, is not only to store and manage the great amount of data, but also to analyse and extract meaningful value from it. There are several approaches to collect, store, process, and analyse big data. The emerging volume of big data types needs for capabilities of innovative technologies for the unstructured data. For this reason, looking for future generation technologies for data analytics is becoming a growing research interest. Big data analytics are practical methodologies applied to very big data sets, composed by structured and unstructured data. Big data analytics is where advanced analytics techniques operate on big data sets (96). The analytic process, starting from multiple sources, such as mobile devices, network nodes, Internet of Things, datasets, sensors, etc., is a collection of practical skills and techniques which include data mining, data visualisation, predictive analysis and artificial intelligence. The nature of big data involves the collection of many samples of data. The traditional analytical tools must therefore be optimised to support a large set of data. In statistical terms, the larger is the set of data, the more accurate is the analysis. The evolution process of big data analytics is related to the rapid growth of applications in this field. This produces a great opportunity growing in area and topics diversity.

The principal features and challenges linked to technologies identified in Big data analytics are the following: (a) innovation in data warehousing, because traditional processes are slow and limited in scalability; (b) ability to converge data from multiple structured and unstructured sources; (c) innovation and challenge in time of value extraction from multiple sources.

Big data analytics can benefit customers, relations, business intelligence, and many analytic applications (96). In this research field, there are three major advances: 1) direct analytic over massively parallel processing data warehouses, 2) indirect analytic over Hadoop and 3) direct analytic over Hadoop. There are various developmental topics within Big Data Analytics: MapReduce, scalable database, real-time stream processing and Big Data appliance. MapReduce was introduced by Google in order to process and store large datasets, providing a programming paradigm which allows useable and manageable distribution of many computationally tasks. It simplifies the programming work, improving data processing efficiency. Hadoop is a highly popular free Map-Reduce implementation by the Apache Foundation, and it is widely used for example in Facebook, Yahoo, etc. Microsoft Dryad model is a parallel computing model based on pipeline computation. It uses a Directed Acyclic Graph (DAG) to represent computational task decomposition. Each node in DAG represents a scheduled task. The stored big data are made available by analytical functions. After that, the data mining and knowledge discovery are parts of the features of the process handling this large amount of data, relating to the other questions which go around the intensive computing.

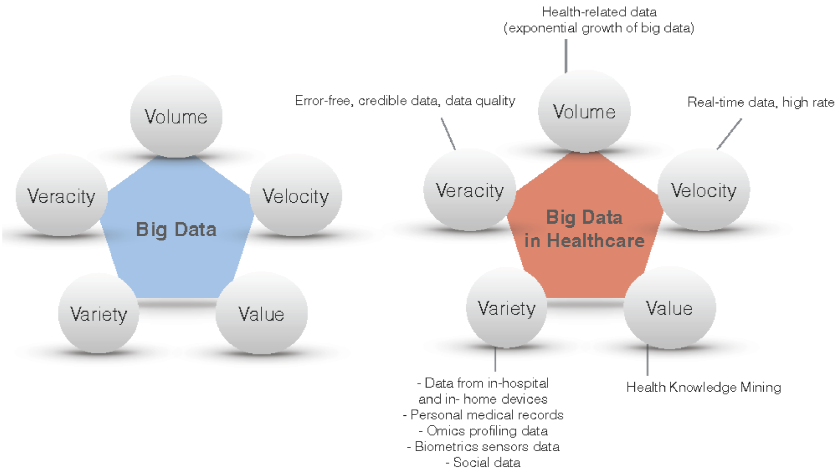


Figure 4.1: Big Data. Traditional 5V Model and Healthcare 5V Model.

4.2.3 Data-Intensive Computing

The data-intensive computing problem involves different fields from medicine to biology and also to environmental sciences and other aspects. The problem is the large volume of data to be managed and processed. Along with the incredible growth of the data volume, we are witnessing also with the data dependent applications, and thus data-intensive computing is becoming the key technology to meet these applications. Data-intensive computing applications include Internet text data processing, in fact the Internet is the world's largest database and contains a lot of different types of data and applications, moreover some web pages have behind a huge database (called "deep web"), whose size is enormously greater than the webpage itself.

The data management and data mining problems, in this fourth era (97), enriched by new application developments and information, needs for more data intelligence, and the data become more complex, asking for massive data processing services. As the Internet provides various applications, such as web search, research assistant, enterprise search, customer relationship management, Internet applications represent the most important driving force for data-intensive computing technology development. Scientific research generates a large amount of data, which includes experimental data, observational data, papers, technical reports, project reports, and scientific literature. As described by (97), the advances in data acquisition technology and the growing data production and data collection are becoming a key factor in scientific research. Simultaneously, the cloud computing and virtualisation technologies are revolutionising the scientific research, transforming it in a data-intensive computing application.

Data-intensive computing is different from traditional high performance computing, as it not only requires high storage capabilities and high speed data transmission, but also computationally complex issues and the analysis and visualisation of the results. The main difference is related to application environment, data size and application requirements. Its features are embodied in the data, processing technology, complex application development and application mode. Data-intensive computing deals with massive, changing, distributed and heterogeneous data. In addition to the traditional structured data, there are semi-structured and unstructured data. As data dynamically change features, data processing must be real-time and the traditional static database management technology is not able to do this type of dynamic processing. Furthermore, the new paradigm of data-intensive computing revolutionises the meaning of computing, as it includes search, query and other traditional data processing, but also includes smart processing, such as analysis and understanding. Therefore, data-intensive computing is associated with the storage and management platform, and combined with a high flexibility and customisation ability, it has to exploit easy search, query and analysis tools. By using these tools, users can construct complex data analysis or understanding application.

Data-intensive computing needs the massive storage, high performance computing platform, which usually cannot be achieved locally. It may include data acquisition, pre-processing and data analysis process. In this complex procedure, data-intensive computing service interface must provide full description function and favourable web service interaction between client and server. Data-intensive computing is a parallel computing technology that processes large scale intensive data sets. The end users do not need to care about parallel processing details. To allow program developers to fully exploit the data-intensive computing convenience and availability, a proper data-intensive computing programming model needs to meet the following requirements: (a) suitability for large scale data sets parallel computing, (b) the possibility of multiple virtual machine task scheduling, and finally (c) the program developers should be able to construct new data-intensive computing applications for end users on the network. Some examples of programming models are MapReduce and Dryad, showing different features as discussed above. Data-intensive computing research prospect is very broad and the sensation is that a lot of work has to be done in this research field to meet the different requirements of the various applications.

4.3 Big Data in Healthcare

Big data in healthcare is a very important issue not only because of its volume, but also because of the heterogeneity of data types and the speed at which it must be managed. In fact, the data related to patient healthcare and wellbeing in the healthcare industry include clinical data from clinical decision support systems; patient data in electronic patient records (EPRs); sensor data (e.g. body monitoring); social media posts; and less patient-specific information (e.g. pubMed) (82). Big data available in healthcare industry represent an opportunity, as discovering associations and understanding patterns within the data through big data analytics, has the potential to improve care, and analytical and computational processes. Organised data allows improving the diagnosis process, treatments, producing, as a consequence, a higher quality at an overall reduced cost (82).

Furthermore, the analysis of data in healthcare helps to identify the most effective treatments, offering analytics tools to make predictions and inferences about patients (predictive modelling), developing a proactive medicine, e.g. suggesting lifestyle changes. Other advantages are the possibility to collect and publish data on medical procedures, assisting patients in determining the care protocols, identifying and minimising fraud by implementing advanced analytic systems for fraud detection and checking the accuracy and consistency of claims. Through analytics, it becomes possible to monitor adherence to drug and treatments and detect trends that lead to individual and population wellness benefits.

Individual and population data would inform each physician and her patient during the decision-making process and help determine the most appropriate treatment option for that particular patient. Using big data, potential benefits include not only the detection of diseases at earlier stages when they can be treated more easily and effectively, but also the management of both specific individual and population health. It becomes possible to predict patients at risk for advancement in disease states; causality between diseases and possible comorbid conditions.

Big data could help reduce economic wastes in all the healthcare aspects. In clinical operations, it will help to detect more clinically relevant and cost-effective ways to diagnose and treat patients. In the research field, predictive modelling will produce a faster and more targeted processes. Statistical tools and algorithms will improve clinical trial design and patient recruitment to better match treatments to individual patients, thus reducing trial failures. In Public Health, the analysis of disease patterns and tracking disease outbreaks will improve public health surveillance and interventions, in addition the large amounts of data can be used to identify needs, provide services, and predict and prevent crises, especially for the benefit of populations (82). Big data analytics in healthcare can contribute to evidence-based medicine, thus combining and analysing a variety of structured and unstructured data-EMRs, clinical data, and genomic data.

In terms of genomic analytics, big data allow executing gene sequencing more efficiently and cost effectively and make genomic analysis a part of the regular medical care decision process and the growing patient medical record (82). The device monitoring, that is the real-time analysis of large volumes of fast-moving data from in-hospital and inhome devices, allow monitoring and predicting events. Applying analytics to patients profiles, using predictive modelling, we can identify individuals who would benefit from proactive care, lifestyle changes, or preventive care because they are at risk of developing a specific disease. Health-related data generate an incredible volume of data. Data includes personal medical records, radiology images, human genetics and population data genomic sequences, etc., and new forms of big data, such as 3D imaging, genomics and biometric sensor readings, are also emerging, increasing this exponential growth. Data are accumulated in real-time and at a rapid pace, or velocity, such as more continuous control by insulin pumps, blood pressure readings, etc. Moreover, it is important to note that in many medical situations, constant real-time data can result in the survival of the patient. The early detection of infections through real-time data analytics, and the application of the suitable treatments could reduce the patient morbidity and mortality and even prevent hospital outbreaks.

The enormous variety of data-structured, unstructured and semi-structured is a dimension that makes healthcare data both interesting and challenging. In healthcare, structured and semi-structured data include instruments readings and data generated by the ongoing conversion of paper records to electronic health and medical records. Unstructured data is the office medical records, doctor notes, hospital admission, paper prescriptions, radiograph films, and other images. Genetics, genomics, social media research are creating new forms of structured and unstructured data, the problem is to capture, store and organise in order to manipulated and analyse them for useful information. Healthcare applications in particular need more efficient ways to combine and convert varieties of data including automating conversion from structured to unstructured data. The potential of big data in healthcare lies in combining traditional data with new forms of data, both individually and on a population level. The veracity, as we discuss previously, is related to the error-free and credible data, so the problem is to get veracity increasing the data quality. Improving coordination of care, avoiding errors and reducing costs depend on high-quality data, as well as the advances in drug safety and efficacy, diagnostic accuracy and more precise targeting of disease processes by treatments. On the other hand, the increased variety and high velocity hinder the ability to cleanse data before analysing it and making decisions, magnifying the issue of “trust” data (82). There are many other issues to consider, such as the number of architectures and platforms, the dominance of the open source paradigm in the availability of tools, the challenge of developing methodologies and the need for user-friendly interfaces.

It is clear that big data analytics has the potential to transform healthcare context, to gain insight from their clinical and other data repositories and make informed decisions. At this stage, the sensation is that the healthcare organisations and industry still have to face a lot of issues and challenges to speed up the use of big data analytics and data-intensive computing techniques in this field (82).

4.4 A new Data Mining in Healthcare

In healthcare context, there are many opportunities and issues related to data, because we have a multitude of heterogeneous sources from which we can extract awareness, related to the patient health. From the social, economic and political aspect, there are a lot of perspectives to analyse these processes, and the mining process, as a collection of different highly increased data, will become hard. Nowadays, the problem is that healthcare should be rationed from different points of view.

First of all, the required innovation is in classifying and analysing clinical and biomedical observations and studies, sensor measures, social behavioural dynamics which represent the new dimension of collected data in the health context. The plurality of the available informative resources does not allow fast processing, and therefore it would be useful to define a single paradigm able to simplify the procedures optimising the efficiency of the subsequent inference analysis, decision-making and management processes (83). With the advance of ICT in healthcare context, following a bio-inspired approach, exploiting processes and methodologies of social and biological world, it is possible to collect many different kinds of data, in time and space. In time, we can consider the need of high resolution process, to evaluate dynamically every changes in a long period (60). In space, we consider many layers of knowledge, related to the type of interactions among patients on different dimensions (7). Multiple types of data provide many complementary perspectives of a single or different correlated aspects, and highlight the need for algorithms able to unify these heterogeneous resources. On the other hand, multiple layers consider different relationships which can influence and change (30) the dynamics of population and personal health, in small and large scale. The final target is to combine relevant information in the most efficient way and obtain a complex and smart mining process, extracting awareness and knowledge, as a semi-complete understanding of the healthcare issues, social dynamics which include influences and contagion processes, clinical and biomedical observations and experiments. From multiple sources and various data, we obtain a large amount of smart complex big data. Clinical data, statistics, classifications and databases (e.g. ICD9) provide and maintain the traditional approach to the collection of bio-medical information. Instead, the Internet of Things (IoTs) perspective, linked to body area sensors, home and social monitoring, allows us to gather context-aware data about patients, thereby reducing hospitalisation costs through early and fast interventions and treatments (83). The social dimension introduces a new point of view. Each patient is a node, each community is a group of nodes or a population of nodes, and the ties, strong or weak, are the relationships and interactions between them. There is a stream of sharing in which we can convey, control and exploit the diffusion dynamics, the contagion of some behaviours (e.g. cooperation, good behaviours related to health, public good). In particular, this influence rules the decision process of each node or of the entire community. Each node is a social subject, and what it shares is a social object. Each of them has a role in the knowledge extraction from these processes and to predict health phenomena (30, 98).

The traditional data mining process represents techniques used to explore database and detect unknown patterns useful to build predictive models (99). In (83) we have

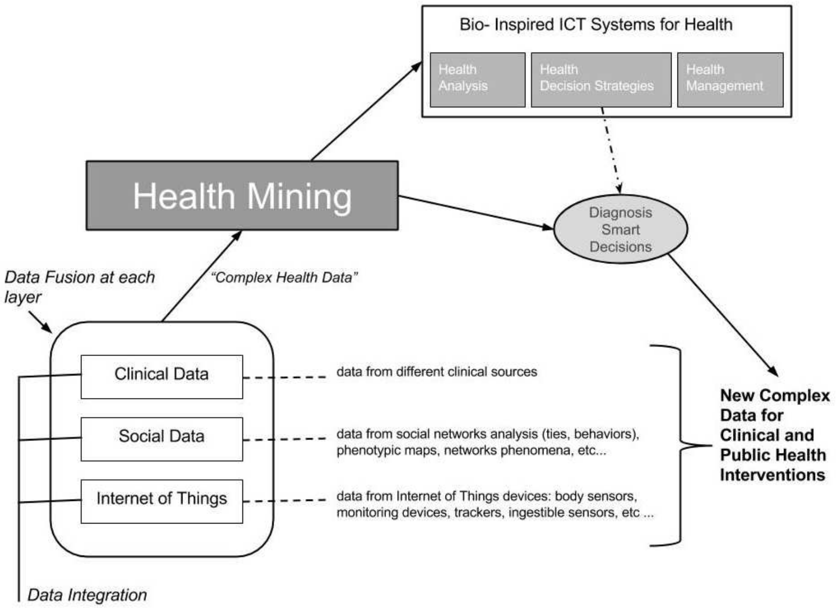


Figure 4.2: Health Mining Architecture. A novel data fusion and integration paradigm.

proposed a new point of view of the data mining process, introducing a multi-layered approach, capturing different aspects of the collected data from several sources. The three data layers, defined in the architecture, create a reality mining in health context, considering not only the clinical data, but also social and IoT information. Finally, the Bio-Inspired approach (35) provides methodologies and algorithms to redefine and optimise the analytic process and improve ICT design and management systems. In (81), we consider and propose some other dimensions, such as comorbidity, linked to the social dimension, following bio-inspired and multilayer approach, obtaining an efficient patient profiling and also new kinds of data to re-populate our big data collection.

4.5 Comorbidity and Complex Health Data

The Health Mining paradigm (83) presented in the previous section (see Fig. 4.2), represents a new vision of data, a complex health data, obtained through an intra-layer fusion and an inter-later integration of data of different nature from multiple heterogeneous sources. This new complex form of data represents a first step towards the resolution of data heterogeneity issues related to patients and diseases. Furthermore, the complex data allow deepening the associations between diseases, and their co-occurrence in the same patient, referred as 'comorbidity'. The new type of complex data has intrinsically also a social dimension which, as we are going to describe in the next subsections, is the key aspect to consider if we want to address the comorbidity aspects of diseases. We will stress the importance of taking into account also the social dimension along with the other dimensions of analysis.

4.5.1 A Multidimensional Approach

Comorbidity refers to the co-occurrence of different medical conditions or diseases in the same patient. A comorbidity relation between two diseases occurs when the two diseases appear simultaneously in a patient, not by chance alone (84). It is worth to underline that some infections or diseases can coexist in the same patient only by coincidence, but there is not any pathological association between them (85). In most cases, however, multiple diseases (chronicle or acute events) occur together in the same patient because of the associations between them. These associations can be direct or indirect causal relationships and can share some risk factors (some diseases share almost the same risk factors). In the comorbidity perspective, the patient becomes the centerpiece of a system characterised by multiple, complex and correlated conditions or diseases (84).

To evaluate comorbidity, in the age of big data, with a large amount of available data, it becomes necessary to use statistical parameters to make predictions or inferences, learning from large data sets. This underlines the increasingly need for statistics and uncertainty. In fact, even if, from one hand, the uncertainty increases the complexity of the analysis in terms of comorbidity, on the other hand, uncertainty and statistics allow gaining knowledge from ever larger data sets (100). Building probability models and sophisticated simulation-based statistical methods, such as filters, it becomes possible to evaluate comorbidity risk. For instance, statisticians are developing models for complex interactions between genes, learning the structure of a network describing gene relationships or phylogenetic trees. The growing of medical databases creates more and more difficulties in managing them, for this reason there is an increasing demand for making causal inferences from observational data. The problem is the difficulty to turn correlation into causality, even with large amount of data.

Big data help in increasing sample sizes, and it allows doing better estimates but this apparent precision goes down if we do not consider the selection bias, regression to the mean, causation properties, and all the complex issues related to the different dimensions of analysis. To produce knowledge from big data, it becomes necessary to analyse and understand the various aspects using a multidimensional approach. In the biomedical research field, for example, one of the main problems is to find out the correlations between human diseases and the subtle mechanisms, at the molecular level, shared by apparently dissimilar diseases. By exploring genes-diseases relations, the pathogenesis may emerge, and it could lead to better diagnosis and treatment. Diseases are more likely comorbid if they share associated genes (85). What becomes crucial in the comorbidity assessment is to discover the positive and negative associations between diseases, and therefore the analysis has to be also focused on pathways-genes associations. Ashley et al. in (101) have mapped the personal genome, gene-environment interactions and the dependent risks for the clinical assessment. In addition to the molecular and genetic data, the population-based disease associations are also important in the analysis of comorbidity. From the co-occurrence of diseases we can estimate the correlation degree which is proportional to the strength of comorbidity risk (85).

The research of recent years has shed light how many apparently different diseases, in reality from a molecular perspective show similar mechanisms. A part of the research is exploring the relationships between genes and diseases at a molecular level, trying to understand the pathogenesis. It may lead to more powerful diagnosis and treatments. Therefore, from a biological point of view, it is important to analyse both pathways-diseases associations and gene-diseases associations, to understand the molecular mechanism of a particular disease. The authors of (101) have analysed the personal genome, gene-environment interactions and risks of conditional dependence for the clinical assessment. To find out the molecular origins and the comorbidity of a disease, it is useful to study, along with molecular and genetic data, also the associations between diseases based on populations. The problem is to quantify the strength of the comorbidity risk, in this direction the disease ontology helps to promote the study of diseases and risk factors.

Comorbidity, in a “smart” health perspective, is a key factor to improve the analysis related to stratification of the patients risk and to adapt the treatments planes. Predictions made taking into account comorbidity risk may allow improving the patient’s management. Comorbidity may also influence the patient survival in fact, if we want to predict the possible consequences of diseases in the survival analysis, it is fundamental to examine not the single disease in isolation, but the comorbidity, the environment, the patient's age and the treatment plane. Kan et al. (102) have observed how the life expectancy decreases in function of the number of comorbid diseases. The authors of (84) have underlined multiple dimensions of analysis of comorbidity. The *disease space* allows inferring diseases relationships. It consists of a coordinate system with clinical and molecular data; it leads to a patient disease network, with multiple interactions. Taking into account these links, it is possible to evaluate the associations between the diseases and also the "exogenous factors", such as the large number of variables, the lack of accuracy in measurements (due to the limited sample size) and the technological limitations in generating data. These noise factors produce uncertainty and increase the complexity of the system and, as a consequence, it becomes increasingly difficult to discriminate between causality and correlation.

Another macro-dimension is time, in fact the events represent the turning points in remodulating the 'comorbidity map', which is a way to illustrate the links between diseases (nodes) and quantify the correlations mechanism of human diseases (101) (see sect. 4.5.2). The pathological episodes have to be contextualised in time, in fact consequences are linked with the context and the involved environmental factors. Furthermore, time allows distinguishing between transient and persistent contributions to the comorbidity map, as some events are able to generate persistent changes in relationships (e.g. acute or chronic events), while others are re-absorbed in a relatively short time. The time has two components: the sequence of events and the time span which indicates the horizon of the contributions, transient or persistent. The interdependence and the perturbation of systems (deviation from stationarity of the system) can produce both local and global scale dynamics: locally, they can influence only a limited part of the comorbidity map, but in some cases the effects may determine a diffuse impact.

Trying to disentangle the complexity of the analysis, we need a multidimensional approach considering, beyond the temporal dimension, also the following dimensions: (a) clinical, which involves diagnostics and intervention strategies in response to changing conditions of the system. The treatment plane has to follow three different phases: perturbation (acute events), transition (uncertainty and change of interdependencies) and return to stationarity, assessing changes and remodulating the map. (b) The genetic dimension that affects dynamically stationarity since disease mechanisms, risk factors often alter the expression levels, pathway activation/inhibition and epigenetic influences. The common pathways may help in diagnosis and therapeutic process, detecting the molecular causes of several associated diseases. Functional analysis based on gene sets can explain crossover influences between different conditions. Omics layers, such as genome, epigenome, proteome, interactome, and so on, are crucial to detect molecular relationships. (c) The therapeutic dimension aims at restoring stationarity, but it may add further complexities according to the positive or negative effects or interventions (e.g. side-effects of some drugs). Finally, (d) the computational dimension considers all the complexities trying to make inferences using a variety of approaches, topology-driven, time space driven or risk-driven, and so the clustering networks of diseases, computing distance between them. The multidimensional approach allows shedding light on the dynamic dimension in the analysis of comorbidites. The common patterns of influences shared by diseases, grouped together according to their distance in the comorbidity map, are subject to events, perturbations that may induce cascade effects or disruptions of the clusters. In the analysis of comorbidity, there is always an index condition, which is the subject of the state, becoming the dominant disease or the attractor state driving the other linked diseases. Therefore, the dynamic dimension is crucial in this kind of analysis, since it can be helpful in detecting these 'attractors', that is the stable points to which the system would return after small perturbations. Building network-based predictive models could allow predicting the occurrence of perturbations and could be a very important support system for clinicians, suggesting the definition of protocols aimed at personalisation of therapy: the calibration of therapy should be simplified by using some comorbidity measures and by predictive evidence obtained by dynamically analysing and revealing the stationarity patterns of the system.

4.5.2 Social and Comorbidity Map: a new Social Dimension

In the previous subsection, we have underlined the importance of dealing with different dimensions in the analysis of comorbidity. In this subsection, we want to stress the importance of another dimension, the social one. In fact, among the conditions that induce comorbidity/multimorbidity, we need to address also the social contagion effect (30), because it acts in time as another remodulator of the comorbidity map. The occurrence of certain events on the social network with a strong impact, such as an epidemics, may influence locally the evolution of the single diseases or risk factors, but above all these events can change globally the links/connections in the map. Therefore, we can distinguish between some persistent components, such as aging, the genetic/inherited predisposition to contract a specific disease, and other components which impact the transient phase, inducing temporary changes in the map, such as the influence of behaviours of people around us, the psychological stress related to a specific event, e.g. the death of a family member or spouse, the infections, some therapeutic interventions and the adverse interactions between drugs and diseases, the outbreak of an epidemic, etc..

The problem related to modelling the comorbidities, the social behaviours and all the possible consequences linked with some conditions, can be addressed only by exploiting a paradigm which fuses and integrates a large amount of different kinds of data derived from different contexts, ranging from microscopic sources, such as in body sensors, to macroscopic sources, linked with social relationships, communities, and globally social environment, more than clinical and medical data. These data fusion and integration mechanisms, from different context and sources, allow weighing the contributions of various factors in the map and to evaluate the overall components both transient and persistent. Among them, the social aspects can modify notably the weights and links in the social map, resulting in a rewiring and a variation of these weights.

In defining a new measure, it is crucial to consider also the social dimension, along with the above mentioned dimensions (clinical, molecular, etc.): the strength of the ties, the connectedness, and communications lead to an evolutionary dynamics of the social network and represent a further degree of analysis in the comorbidity assessment. The exogenous and environmental factors, called 'confounding factors', act as a noise in the system, amplifying some phenomena or creating new spurious perturbations. These perturbations may generate effects at a local level, with consequences limited in space and time, or at a global level, triggering resonances with important implications on the entire map. As Christakis and Fowler have underlined in (30), the social contagion mechanism produces a spread of some factors or behaviours, such as obesity, smoking or happiness. Moreover, they have demonstrated that this contagion, even if it decreases with social distance, is relevant until *the friends of friends of friends.*

Some of these factors are risk factors in the map, such as obesity, smoking and the alcohol consumption, which may generate new comorbid diseases and cause some social behaviours. On the other hand, some social behaviours may bring out some genotypic-phenotypic traits leading to some comorbidities. Through exploiting the social and comorbidity map, the target is to study this reciprocal influence. Social dimension and social relationships can act as a selection bias in choosing one treatment plane or drug rather than others, or still shedding light on some side-effects of some drugs inhibiting the usage, or mimicking "good" behaviours of some friends in the social network.

To quantify the comorbidity, the authors of (85) have used two measures: the relative risk, that is the fraction between the number of patients diagnosed with both diseases and random expectation based on disease prevalence as the quantified measure of comorbidity tendency of two disease pairs; the ϕ-correlation (Pearson correlation for binary variables) to measure the robustness of the comorbidity association. Beyond these measures, we need to use also a social measure of contagion, considering a dynamic evolution of comorbidities which depends on a different response of the patient linked to a genotypic-phenotypic diversity, which may induce different social behaviours and, as a consequence, to different maps.

4.6 Personalised Medicine and Social Dimension

According to the theory of "expert patient" (103), the individual-patient has a unique knowledge and awareness of own health status. Thus, this knowledge would be translated in data, useful to design the strategies of decision-making process about clinical trials or treatments. This kind of data could be shared and may influence other social contacts (30), in different layers of the network (7), triggering a dynamic contagion of "good behaviours" (30), with the aim of increasing the rate of success in terms of personal and community wellness. The individual knowledge may become social knowledge and, consequently, collective intelligence, as well as the individual participation may become collective and social cooperation. The healthcare process would benefit from these dynamics.

Leroy Hood, molecular biologist and oncologist, in (104) proposed the evolution of medicine in P4, passing from P0 and P3. The transition was also due to changes in the diagnostic process, that became complex and interdisciplinary, and in technologies

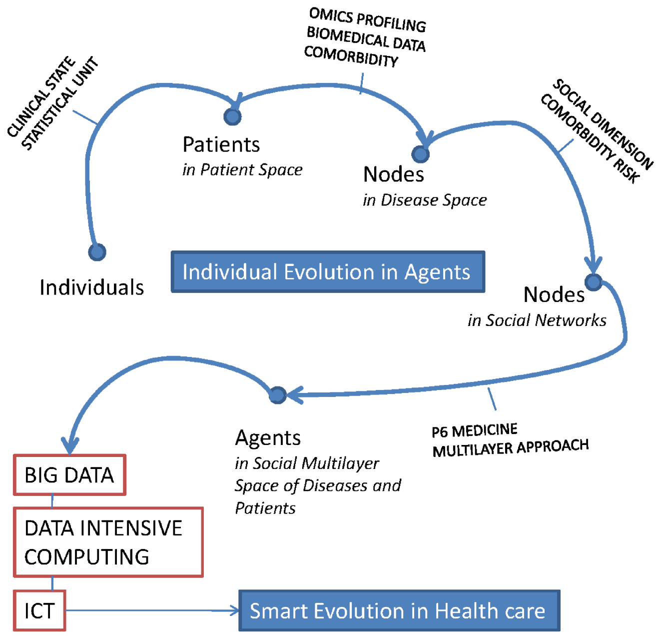


Figure 4.3: A social smart evolution in healthcare.

supporting health processes. The P4 medicine (104) is an approach proposed as an extension of what was simply defined as "personalised" and "genomic" medicine. Introduced in the field of oncology, the P4 medicine is a model of treatments of patients based on the following features: predictive, personalised, preventive and participatory. This kind of approach creates a predictive model for the treatment of patients, combining different kinds of data, obtained from different layers of biological information. The *predictive* aspects include the theory of evidence-based medicine (103) and the aim is to use models and tools to predict the future health status of a patient or community. Big data about omics profile of patients, the data-intensive computing techniques, and the multilayered approach allow treating different kinds of sources, moreover high-throughput based information can be important to make an optimal predictive analysis. The *personalised* approach enables to design the pharmacological trials, health strategies and therapies, based on the individual patients, improving the clinical practice, making the patients more active in the experimental trials. This is possible considering the evolution in nano-genomics, nano-proteomics, and related on the studies about the molecular and cellular pictures of each patient (103). Treatments could be more

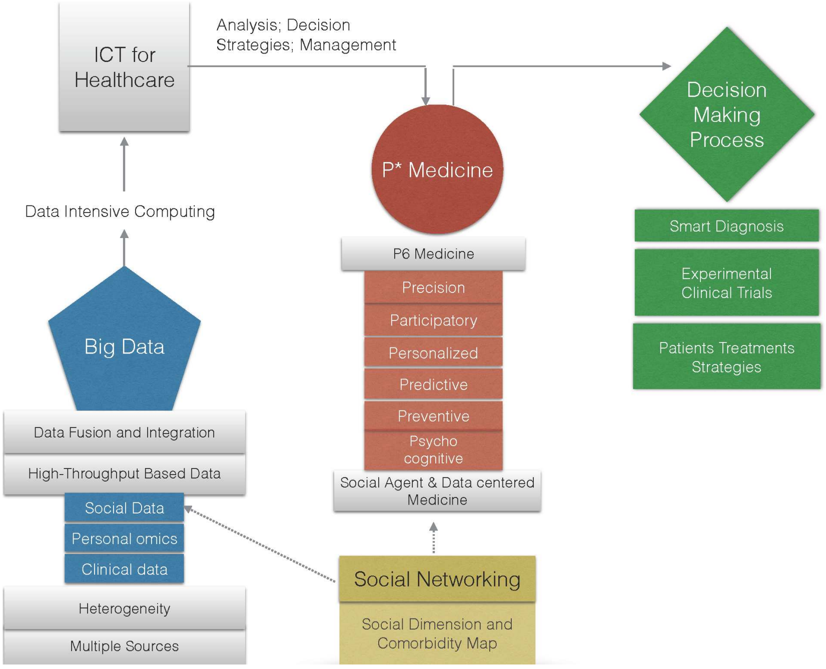


Figure 4.4: Personalised medicine. A social evolution in healthcare.

effective if the healthcare process faces the complexity of disease. In fact, it is a cluster of symptoms, multifaceted and affected by multifactorial pathologies. As a third factor of complexity, along with the biological information associated with diseases and their interactions, the future medicine is facing with the introduction of strategies for the assessment of comorbidities related to possible risk factors and underlying causes. Recently, the most important application of personalised medicine has been discussed in the research studies of A. L. Barabàsi (105), known as "network medicine".

Considering a *preventive* perspective, the medicine switches from reactive to proactive. Taking into account the shift from an illness-centered approach to an holistic one, the doctors, with the support of computational models, big data (aggregation and integration of clinical and social data), molecular-biological profiles, and everything is related to environmental factors which could influence risk factors, should be able to advice patients to prevent particular diseases.

The social dimension could be important to convey the dynamics of social contagion about good behaviours, for improving the preventive process, following specific strategies linked to the pursuing of the common wellness or public good. The *participatory* aspect allows the patients to be more involved in healthcare process, becoming gradually more self-aware (103) of their own wellness and lifestyle, participating actively to the decision process about their own health. In this context, following this kind of approach and introducing the social networking platform, the community helps to develop self-organisation and self-awareness capabilities, improving the efficiency of strategies related to the population's health. It is worth to note that no single one of these dimension can be managed in isolation from the rest (103).

Recently, the research in healthcare has been taking into consideration another key aspect of human life. In fact, considering the individual-patient as a whole (104), thus a single unity, with a clinical and biological profile, in a global perspective, in order to improve the personalisation and the efficiency of the model, the evolution of P4 medicine has introduced a new component linked to the psycho-cognitive features, which characterises and distinguishes each single individual. This approach leads to the P5 medicine (103, 104). This new challenge in medicine, introducing the psycho-cognitive skills, consists of considering a personal profile of the patient, who is not only a biological-genetic entity, but the profile must be able to reflect efficiently his needs, habits, behaviours and cognition. Through this novel approach, the rate of success will increase in terms of predictions, since it will be able to combine different types of data (heterogeneity), a large collection of data in time (big data), obtained from different levels of research and fields of study (multiple sources). Furthermore, emphasising psychological and cognitive aspects, the purpose is also to improve the general state of the patient's wellness.

The P5 model is an approach for person-centered and relationship-based care model; it allows the design of a new profile related to the psycho-cognitive and behavioural aspects of each patients. In fact "every human being of adult years and sound mind has a right to determine what shall be done with his own body". This means that patients should participate more actively to the medical care process, for example by weighing benefits of the proposed alternatives, based on risks assessment. The fifth "P", referring to these aspects, is important for the assessment of the wellness of the individual-patient, making him, inside the healthcare process, a decision-maker who actively participates in the process. The P5 model allows considering the individual-patient as a person and not only as a collection of data extracted from clinical case studies and statistical analysis of the single unit. This approach by far increases the individual's satisfaction level and, as a consequence, the overall quality of life, the relationship with physician, and the efficiency in the searching process of more suitable and/or optimal therapy or therapeutic plan. The entire health system benefits from this model, making the prevention more effective through interventions in promoting strategies for reducing the unhealthy behaviours. Acting on psycho-cognitive aspects, it could influence individuals to adopt or modify their inappropriate beliefs, bringing out the underlying risks. The P5 model introduces the psychological, cognitive, emotional and social aspect levels to improve the development of health personalisation. The final goals are to improve design techniques of the new pharmacological trials, clinical practices and make patients more actively involved in experimental trials, changing the medicine paradigm, from reactive to proactive. The social, cultural, biological, psycho-cognitive and behavioural features and needs of a single individual-patient or community, will help the physicians in choosing the treatment or in deciding about experimental trials.

The social dimension will change radically the approach for the future medicine, as the *precision medicine* (86) will exploit the dynamic analysis, the study of the dynamic evolution of behaviours and needs, the disease contagion processes, and also the sharing, cooperation and collaborative approach through the new technologies. This evolution could be identified in P6 medicine as a social evolution of the healthcare process, making medicine, clinical cases and data (genetic, molecular, phenotypic) public and available, transforming them in collective knowledge, objects for future studies, in order to exploit and benefit from the enormous potential gained by open, public and social form of medicine. Big data could help to solve the aggregation problem of genetic, molecular, environmental and social data, and the analysis of the comorbidity dimension associated to risk factors of diseases (as underlined in sect. 4.5.1). Moreover, it is known that similarity in lifestyle, along with the genetic predisposition, cause us to be susceptible to similar diseases (106). Thus, using analytical tools for prediction of the co-occurring risk factors, and taking into account the similarity aspects, the complexity of the problem increases.

4.6.1 From Data to Multiagent: A Multilayer Architecture

To describe the healthcare processes, following a bio-inspired approach, we exploit the concept of Multi-Agent System (MAS). A multi-agent system is composed of multiple agents which interact each other to obtain a specific purpose which may be, for instance, the cooperative behaviour to get a specific target. Multi-agent systems are often used to solve problems that are difficult or impossible for an individual agent, in fact the underlying concept is that the overall system exhibits significantly greater functionality than the individual component. The agents are intelligent, including some processing and cognitive abilities. The agents are autonomous, self-organised entities and only with a local view of the system (limited knowledge). The agent need to embody the knowledge of other similar agents with which it will interact, including their own preferences and knowledge.

In our system, we exploit MAS in order to model social entities, represented by nodes or cluster of nodes, therefore we consider complex agents, passive or active sources of data and with a specific role in the network. The introduction of context-aware and cognitive abilities gives agents the possibility to convey the extracted knowledge, decide, apply strategies based on heuristics in order to reduce the computational complexity (44). Big data analytics and storage techniques and the data-intensive computing applications and programming models, allow addressing the complexities, improving the performance of a future smart data-intensive computing process in healthcare.

The multilayer structure represents a way to describe the interactions among nodes on more layers. A set of entities, which can be agents, communities or clusters, diseases, patients or data, based on the layer we take into account, interacts through some patterns. The connections and the relationships between entities in a complex system, are considered on different layers, with corresponding different types of interactions. Therefore, the entities and the relationships among nodes must be considered and weighed on different layers and, at the same time, the weight of their decisions and their behaviours has an impact on so many different levels. It is only recently that multilayer structures have become a popular paradigm for the modelling of interrelated subsystems and entire systems (7).

In Fig. 4.5 we illustrate the multilayer structure in healthcare. Starting from the bottom, the first layer is feeded by the large amount of data in healthcare and it corresponds to patients-data associations. The second layer is focused on patients-diseases associations, hence the patients space allows shedding light on which patients are associated with different diseases. This layer transforms the associations, considering not only data, but the diseases connections: from a genetic point of view, two diseases are connected if they have been associated with the same gene (107), from a metabolic perspective, two disease are connected if they are associated with adjacent metabolites in the metabolic network (108) and, based on phenotypic traits, two diseases are connected if they are coexpressed in a significant number of patients in a population (109). The third layer is that of comorbidity relationships, so we use clinical and molecular data to build the disease space, and we consider the multidimensional approach described above in sect. 4.5.1. The fourth layer consists of adding the social dimension in comorbidity analysis: the habits and social behaviours, shared with similar friends, and the social contagion process, obtained studying the connectedness of the network, lead to clustering of the social nodes which influence each other. The comorbidity and social dimension allow us to capture the subtle connections among nodes, and help us to make this knowledge more connected and deeper. Finally, the fifth layer is made up of complex agents which enclose all the addressed "traits" of the entities, hence the multiagent system allows extracting an "organised" knowledge, furthermore applying all the strategies targeting at reducing the computational complexity (e.g. heuristics).

In (110), starting from the previous multilayer structure, we have considered an additional layer above all the other layers, called *Health Information Exchange (HIE) and Management Layer*, consisting of control entities, such as ICT (Information and Communication Technology), physicians, and the same patients that interact with the various layers, through the social network and the various kinds of associations between layers. These entities exploit the tools provided by ICT in order to observe and estimate the various medical, social and economic aspects. The HIE and Management layer solves the interoperability issues among the different healthcare information facilitating the exchange, moving clinical information among disparate healthcare information systems without losing the meaning of the information being exchanged. Furthermore, it is also useful to public health authorities and ICT to assist in analyses of the health of the population and of the single patients.

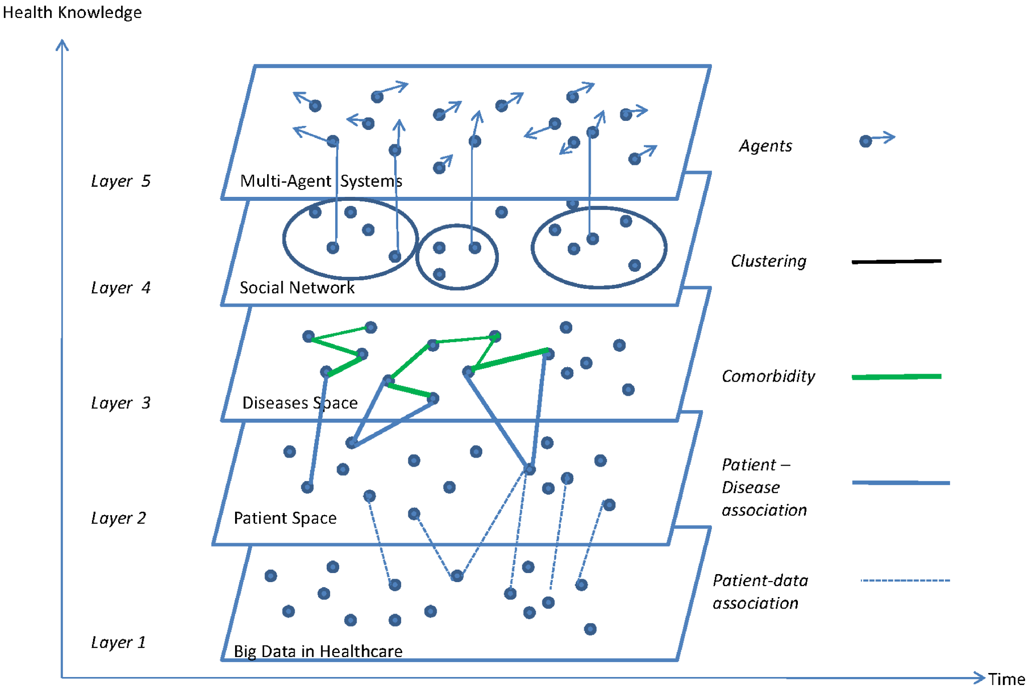


Figure 4.5: Multilayer in Healthcare.

4.7 Conclusions and Future Work

ICTs, represented in services, logical strategies, infrastructures, methodologies, interventions and platforms, are expected to contribute to the realisation of a sustainable and smart society, because of its strong influence and impact on the environmental, social, economic and technological aspects of our lives. The bio-inspired approach, when applied to ICT, starting from the striking similarities between biological and ICT systems, gives the opportunity to use algorithms, tools and analytical models to optimise and improve the design and management of methodologies proper of traditional vision of ICT. These bio-inspired systems help to extract knowledge, considering social dynamics and biodiversity, and make it available to those who have to design ICT interventions and services, considering the multitude of resources, in terms of data and sources, computational limits and social dynamics.

In the era of big data and high connectedness, we are witnessing with a large amount of data available from different sources, e.g. social networks, sensors, etc., creating new challenges, such as those related to the '5V model' but, at the same time, these data are an opportunity to increase the collective intelligence and extract knowledge through interactions among social agents. Nodes, data and interactions become both subject and object of the spread of knowledge in social networks. Therefore, big data analytics, data-intensive computing, data mining and machine learning tools, allow improving the understanding of the networks phenomena, optimising ICT interventions, pushing towards greater knowledge and collective intelligence.

In Healthcare, big data analytics and storage techniques have become crucial because of the large amounts of data, related to patient care. The Health Mining paradigm allows analysing these data, the emerging patterns, and extract complex health data, an organised knowledge useful to deepen the comorbidity relations between diseases in the patient. The study of comorbidity relations allows finding out dynamic patterns, correlations and causality mechanisms between diseases and patients. We have underlined how the social dimension of comorbidity, focusing on how the strength of the ties, the connectedness, and communications lead to an evolutionary dynamics of the social network and represent a further degree of analysis in the comorbidity assessment.

The future medicine will change towards the precision medicine, overcoming the p5 medicine, exploiting the complex dynamic analysis, the study of the evolution of behaviours, the social networking, the ubiquitous networks, big data and ICT strategies. Patients are becoming the centerpiece in the healthcare process (patient-centered), agents able to interact through different layers, such as patient-data and patient-disease associations, comorbidity relations, clustering in the social network. This multilayer structure allows extracting an organised knowledge which connects, relates and computes all the facets of healthcare data. In the near future, these topics will change radically the basic idea of data-intensive computing applications, data mining approach, big data analytics and storage techniques, re-shaping all these tools and techniques in the new social dimension and personalised medicine approaches.

4.8 The Bio-Inspired and Social Evolution of Nodes and Data in a Multilayer Network

In (111), following a bio-inspired approach, applied to multilayer social networks, the idea has been to build a novel paradigm aimed to improve methodologies and analysis in the field of ICT. The social network and the multilayer structure allow carrying out an analysis of the complex patterns, in terms of the dynamics involving the main entities, nodes and data. The nodes represent the basic kernel from which generating ties, interactions, flow of information, influences and action strategies that affect the communities. The data, gathered from multiple sources, after their integration, will become complex objects, enclosing different kinds of information.

The proposed approach introduces a level of abstraction that originates from the evolution of nodes and data transformed in "social objects". This new paradigm consists of a multilayer social network, divided into three layers, generating an increasing awareness, from "things" to "knowledge", extracting as much "knowledge" as possible. This paradigm allows redesigning the ICT in a bio-networks driven approach. The new ICT paradigm is expected to contribute to the process of improvement in the realisation of a knowledge-based networking, characterised by innovation, making the networks sustainable with processes based on a strategic bio-inspired approach, considering also the social, human and cognitive aspects.

The future network needs to meet some requirements, such as ubiquity, mobility, dynamicity, reliability. The ubiquitous nature leads to a logical fusion and integration of different aspects of real and online social network platforms. The network nodes acquire a common representation through identity features. These features, following a bio-inspired approach, enclose genotypic and phenotypic traits. In addition to these traits, it is important to consider also context-aware capabilities, self-organisation, self-protection, perception, decision-making processes and cognitive behaviour. Considering these features, the nodes interact through social networks and they are able to self-organise dynamically in communities and groups, based on aggregation metrics.

In (111), the node is an abstraction, an object which collects bio-inspired features as well as human and social capabilities. Similarly, data shared inside the network are a complex object, like a box, which travels across the network through interactions

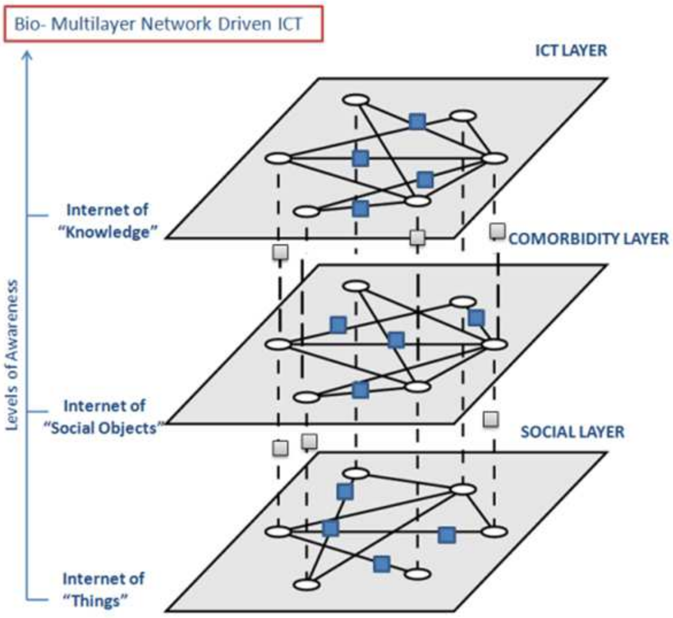


Figure 4.6: Bio-Inspired Multilayer Network for ICT.

between nodes. Data and nodes represent the objects that trigger influence, interaction, contagion and decision criteria. These two entities enclose many different social aspects. In the future of ICT we will expect to be able to obtain context-aware services, which stem from a bio-inspired approach able to drive methodologies for analysing social complex networks. Then, we propose an evolutionary and innovative multilayer perspective of nodes and data. Furthermore, to solve the heterogeneity issue of these entities, in (111) we have considered a social object oriented approach, following the bio-inspired principles in a multilayer social network.

The proposed paradigm is the result of a biologically-inspired approach applied to complex social networks. Exploiting a multilayer architecture, it consists of an evolution process that involves both node and data. This approach starts from the consideration that all techniques and models related to information and communication have to be based on what governs the network dynamics. These processes in a social-based context are determined from nodes and data. This requires an evolution of these entities inside a multilayer network which shows the patterns of the different relationships and interactions among nodes and communities, through the sharing of data. The main goal is to solve all the issues related to heterogeneity, which can be an obstacle to more

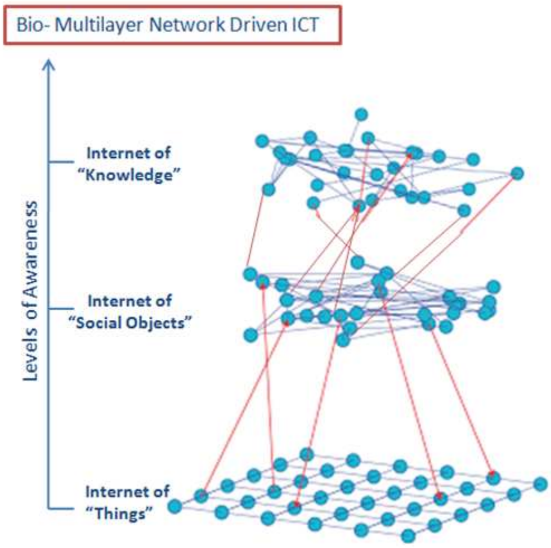


Figure 4.7: Dynamic Complex Patterns of the Multilayer Structure.

complex and deep analysis of networks. The network is characterised by a multitude of nodes of different nature, and by a multi-channel data collection. The aim is to obtain an evolution of the networks considering bio-inspired social objects. This evolution will become useful to enable the ICT procedures to be in line with the bio-inspired processes that rule the complex social network. The future ICT will be driven from these objects creating strategies and applications with an innovative and dynamic approach.

Fig. 4.6 illustrates a multilayer social network, divided in three different layers. The three layers show the same topology but actually it could be different across layers. The first layer is the social layer, characterised by interactions between nodes through the sharing of data. In the second layer we introduce the comorbidity perspective, which is a medical concept referring to the co-existing of different diseases in the same subject, it creates other different relationships between nodes, that represent patients/individuals who share the same diseases or morbidities. In the third layer we consider the ICT in terms of interventions which involves the single entities and the system as a whole. In this perspective, the multilayer organisation enables us to analyse the complex patterns of analysis. Starting from a simple node which interacts with other nodes, we can obtain an evolution and a growing awareness.

Fig. 4.7 shows the dynamic patterns as the result of the coupling effect of interdependent layers. Only by studying the inter-layer interactions between nodes, it is possible to detect the emergent behaviours and focus on the key features related to data and nodes, from which these patterns are generated. One perturbation in one layer could drive changes in the other layers through interactions.

The evolution process from node and data to “social objects”, is indicated in Fig. 4.8. The “social objects” pave the way to the higher level of awareness, referred as “knowledge”, the abstraction of the outcome of the flow of information and social objects. The social objects merge together all the different cognitive, social and human aspects and the various contexts.

The node, in the proposed paradigm, becomes an abstract object which contains any kind of presence and/or participation in the social networks. This can encompass simple network nodes, both hardware and software, IoT sensors, human nodes, etc. The node’s presence is defined as a set of bio-inspired features, such as genotype and phenotype. The genotype is represented by the immutable traits of that object. The phenotype is a combination of observable features, behavioural manifestations of genotype, and the result of interactions between genes, environment, and random factors. The multitude of heterogeneous nodes, with capabilities of self-organisation, through mechanisms of aggregation and clustering techniques, becomes an organised structure of communities and groups. Enabling context-awareness and cognitive capabilities, the nodes become smart, able to decide their strategies inside and outside the communities. Adding abilities extracted from complex social networks analysis, in terms of emerging behaviours, we will obtain the abstraction, which is the social object node.

The data, the other entity of the network, are any kind of collected information, useful to network analysis. The data could consist of statistical data, data gathered from sensors, social data, derived from online and real social network platforms. Collecting data may be relatively easy, but the complexity arises in combining and integrating datasets from multiple sources and different contexts, in order to extract the real knowledge about networks.

This is the reason why we need a complex mining, as that described in (83), able to fuse and integrate in a unique structure these heterogeneous data, collected from different sources and of different nature. Furthermore, we have to integrate data considering the different contexts and environmental conditions in which these data are generated, considering who created them and for what purpose, so we have to consider a context-

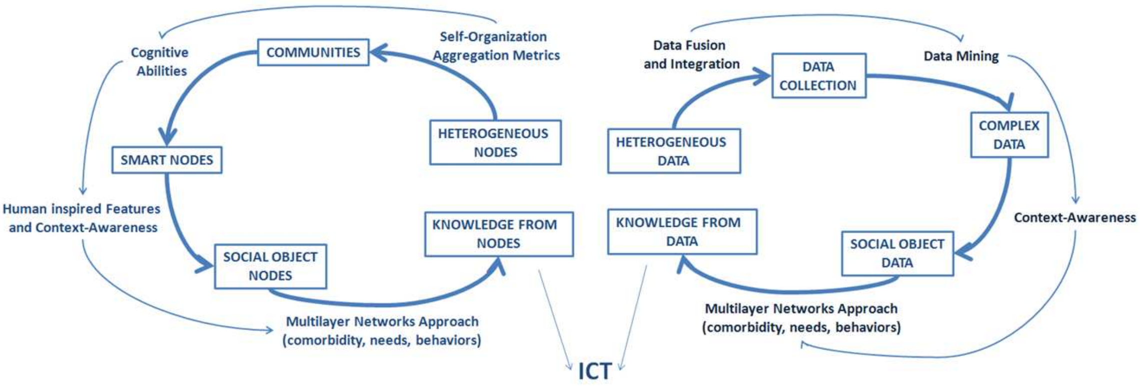


Figure 4.8: Evolution of Node and Data. The evolution, involving data and nodes, is a process which starts from disaggregated and heterogeneous things, and gets the social objects and finally, the knowledge for the ICT.

aware data mining, related to how attributes should be interpreted according to the different contexts.

The evolution process, that involves nodes and data, enables us to give a new definition, which tries, first of all, to solve the issue related to the management of heterogeneous data and different nodes. In Fig. 4.9 we show the future directions of social-object oriented paradigm. The social multilayer network allowed us to analyse the complex dynamic patterns involving these entities, shedding light on the different types of interactions at various layers. The multiplex structure, consisting of three layers, the social layer, the comorbidity layer and the ICT layer, allows considering, respectively, the social interactions and the social contagion between nodes through the sharing of data, the comorbidity relations between diseases, and the ICT interventions as a result of the analysis of the complex patterns involving entities, the context and the system as a whole.

The multilayer social network paradigm describes the evolution of data and nodes, considering an increasing level of awareness, from things to knowledge between social objects nodes through the social objects data. This evolution process leads to a bio-inspired network-driven ICT, redesigning the ICT communication paradigm. There is still much work to be done on the complex analysis, considering different aspects simultaneously, as layers of a multilayer or multiplex structure, and data, in order to get a knowledge-based mining, supported by the collected data in healthcare. The game-

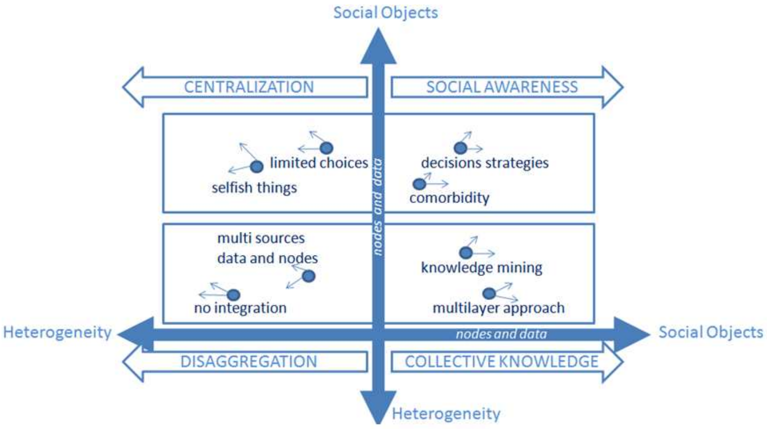


Figure 4.9: Future directions of social-object oriented paradigm.

theoretic modelling of social interactions, along with multiplex, can allow investigating and study from a social perspective the issues of comorbidity and the influence of social behaviours.

Chapter 5

Game Theory and Evolutionary Dynamics

One of the main targets of this Ph. D. dissertation is to study the evolutionary processes and the dynamics of cooperative behaviours of people within a social network, the latter modelled exploiting the mathematical framework of multilayer networks (see chap. 3). In chap. 3, we have explored the multi-layer connectedness of a complex system, in terms of structure, describing mathematically the connections among nodes in the network, extending the concepts and main properties of single-layer networks, derived from graph theory, to the case of multilayer networks. Beyond the structural interdependence of nodes in a network, there is also a behavioural interdependence of the individuals who live in the system, since the outcome for each individual depends not just on his chosen action or behaviour, but also on the actions or behaviors of its neighbours with whom he interacts.

To describe and analyse such interdependent interactions and the dynamics of behaviours, we exploit the mathematical framework of *Game Theory*. Game-theoretic ideas arise in many contexts, ranging from economy, e.g. the pricing of a new product in comparison with similar products of other firms, or how to bid in an auction, to politics, such as the voting system or international relationships with other countries, to finacial and strategical-military, other than sociology and transportation networks. Finally in biology, game theory has been used as a model to understand many different phenomena: animals behaviours, sex ratio, biological altruism, etc. John Maynard Smith, in the preface to “Evolution and the Theory of Games” assesss how “paradoxically, it has turned out that game theory is more readily applied to biology than to the field of economic behaviour for which it was originally designed". As Jon Elster stated in 1982, "..if one accepts that interaction is the essence of social life, then game theory provides solid micro-foundations for the study of social structure and social change", thus game theory is everywhere these days. The common and basic idea, in all the possible applications, is that there is a strategic interdependence that game theory is aimed at analysing.

The chapter is organised as follows: in sect. 5.1 we will briefly discuss the main principles and concepts of the classical game theory; sect. 5.2 will be focused on evolutionary game theory as a way to study evolutionary dynamics of behaviours in a population, such as the emergence of cooperation in social dilemmas.

5.1 Classical Game Theory

Classical Game theory was born in the 1920s, when John von Neumann thought to develop a scientific approach to bluffing in the game of poker. The subsequent study applied to these conflictual situations, involving a decision-making problem, conducted him to write, together with the economist Oskar Morgensten the first prototypical framework of game theory in 1944 (112). Von Neumann and Morgensten defined a game "as any interaction between agents that is governed by a set of rules specifying the possible moves for each participant and a set of outcomes for each possible combination of moves.". After, in 1950, a lot of research in classical game theory was carried out in Universities, such as Princeton University, and in particular one of the most important contributions came from the American mathematician John Nash, who invented a simple but powerful concept, the so-called "Nash equilibrium", which is used to analyse the outcome of strategic interactions among decision makers.

After having discussed the basic principle and underlined the importance of using game theory to describe social interactions, now we are going to define some of the main notions underlying the classical game theory, which is called "classical" in order to distinguish it from the "evolutionary" game theory (see sect. 5.2).

To describe a game and its basic elements, we need to define the following three main aspects:

1. *players* are the set of participants;

2. *strategies* represent all the possible choices or strategies available for each player;
3. *payoff* is number associated to each choice or strategy selected by a player. Every player tends to maximise his payoff as much as possible. The payoff values are generally inserted into a *payoff matrix*, which describes the *rules* of the game.

Another important distinction is between *single stage games* or *one-shot games*, where players play simultaneously, choose their actions independently and make their choice only once, and *repeated* or *iterated games*, where instead there are more rounds, representing a certain number of repetitions of some one-shot game. In the latter case, the presence of more than one round means that a player has to consider not only its action seen temporally isolated within a single round, but also the impact of his current action on the future actions of other players (sometimes called "reputation"). Furthermore, an iterated game is by definition dynamic, as a player varies his strategy according to the payoff obtained in the previous rounds.

Once described the basic properties of a game, which consists of players, strategies, payoffs and rules, our interest shifts to understand and reasoning about how players will behave, or will choose their strategies in a certain game. In the following description, we will focus on two-player game, but it is clear can be applied in the same way to games with any number of players. First, we need to do a few assumptions. The first assumption of classical game theory is that the interacting decision makers are *rational*, thus they act in order to maximise their payoffs according to his predictions on the strategies adopted by the other player. It is important to notice how the framework of game theory does not require this assumption to care only about personal payoff as, for instance, a player may also care about both his payoff and that of the other players; in this case, payoff will reflect this situation. In general, payoff matrix should describe completely the payoff associated with all possible outcomes of the game. The assumption of rational behaviour combines the idea of maximising the payoff, and that each player is able to select the optimal strategy. These two ideas are both reasonable if we consider experienced player playing a simple game, but in the case of complex games, this assumption of choosing the optimal strategy may fail as players will make mistakes in the strategies adoption, and instead it becomes interesting to see how players make mistakes and learn from playing the game.

Furthermore, another assumption is that players know all the rules and the structure of the game, that is all the possible strategies and related payoffs. This assumption is also reasonable if we consider games with complete information, nonetheless some research works, e.g. in economy, have analysed games with incomplete information (28), where this assumption is no longer valid.

5.1.1 Best Response, Dominant Strategy and Nash Equilibrium

One of the key concepts of game theory is that of *best response*, which is the best choice for a player, given his belief about the choice of the other player in a two-player game. Mathematically, if S is a strategy chosen by player 1 (P_1), and T is a strategy chosen by the player 2 (P_2), in the payoff matrix there is an entry which corresponds to the pair (S,T). As a result, we indicate respectively with $P_1(S,T)$ and $P_2(S,T)$, the payoffs obtained by the two players. Therefore, we can say that a strategy S for P_1 is a best response to a strategy t for P_2 if the payoff is at least equal to the payoff obtained by choosing another strategy S' paired with T:

$$P_1(S,T) \geq P_1\left(S',T\right), \forall S' \text{ of Player 1.} \tag{5.1}$$

We can give a symmetric definition for the other player P_2, and we will do the same also for the other following definitions. We can note that there may exist more than one strategy corresponding to the best response to strategy T, and this makes more difficult to predict which of these strategies P_1 will choose. For this reason, we can also define strictly the best response, considering the strict inequality, as follows:

$$P_1(S,T) > P_1\left(S',T\right), \forall S' \text{ of Player 1.} \tag{5.2}$$

In this case, it is clear how the strategy should be played by P_1 when facing the strategy T of P_2 is S.

Other fundamental concepts, related to the definition of best response, are that of *dominant strategy*, defined as the best response to every strategy chosen by P_2, and *strictly dominant strategy*, defined as a strict best response to every strategy chosen by P_2. The notion of a dominant strategy is slightly weaker than the strictly dominant one, since it can be equal to another strategy as the best choice against some opposing strategies, then for a player there could be more than one dominant strategy, so there is no certainty on the strategy to be adopted. In the case of *Prisoner's Dilemma* (see sect. 5.2.4), the analysis is facilitated since both players have strictly dominant strategies, and so it is easy to reason about what will be the strategies of each player.

When none of the players has not a strictly dominant strategy, we need for another way to predict what is likely to happen. To this purpose, in 1950 John Nash proposed a simple but extremely effective and powerful equilibrium concept, a principle to reason about behaviours in generic games. Nash's starting point was that, in the absence of dominant strategies, players are likely to choose strategies which represent best responses to each other or, in other words, supposing that each player chooses a strategy, the pair of strategies (S, T) is a Nash equilibrium (NE) if S is a best response to T, and vice versa T is the best response to S. We can formulate the concept of an equilibrium for a two-player game with respective payoff matrices $\mathcal{P}_A$ and $\mathcal{P}_B$, where A and B denote the two players. Denoting by $\mathcal{P}_A(S, S')$ the payoff for player A when A plays S and B plays S', this is simply the (S, S')-entry of the matrix $\mathcal{P}_A$. We can say that a pair of strategies $\left(\tilde{S}_A, \tilde{S}_B\right)$ is a Nash equilibrium for a two-player game if no player can improve his payoff by changing his strategy from his equilibrium strategy to another strategy considering that his opponent keeps his equilibrium strategy. In terms of the payoffs matrices this means that:

$$\mathcal{P}_A\left(S_A, \tilde{S}_B\right) \leq \mathcal{P}\left(\tilde{S}_A, \tilde{S}_B\right), \forall S_A \quad \text{and} \quad \mathcal{P}_B\left(\tilde{S}_A, S_B\right) \leq \mathcal{P}\left(\tilde{S}_A, \tilde{S}_B\right), \forall S_B \tag{5.3}$$

Thus, a strategy $\tilde{S}_A$ is a best response to a strategy S_B if:

$$\mathcal{P}_A\left(S_A, S_B\right) \leq \mathcal{P}\left(\tilde{S}_A, S_B\right), \forall S_A. \tag{5.4}$$

The concept of equilibrium introduced by John Nash (who shared the Nobel Prize in Economics in 1994 with game theorists Reinhard Selten and John Harsanyi) allows finding out the optimal outcome of a game, indeed a player has not any reason or incentive to deviate from his strategy after considering the other player's choice, because he will not receive a higher payoff, if the other player will not change his strategy. A game may have multiple Nash equilibria or none at all. Nash equilibrium can be seen as an equilibrium in beliefs, as if each player believes that the other player will play a strategy which belongs to a Nash equilibrium, then also the other will choose the other part of the equilibrium.

There are two ways to find the Nash equilibria: the first consists of simply checking all the possible pairs of strategies and see if each one represents a pair of best responses to each other; the second is to calculate the best response of each player (or the best responses, if there are more than one) to every strategy chosen from the other player, and then find the strategies which constitute mutual best responses to each other.

5.1.2 Pareto-Optimality and Social Optimality

We can observe how in the Nash equilibrium condition, the two players are optimising their strategies reasoning at individual level, but this does not necessarily mean that they will get a whatever good outcome. Therefore, it is interesting to classify the outcomes of a game not only according to the strategies or at equilibrium, but also considering whether the chosen strategies are "good for society" as a whole. To this purpose, we define two concepts.

The first is the so-called *Pareto-optimality* (whose name derives from the Italian economist Vilfredo Pareto), according to which the choice of strategies (one by each player) is *Pareto-optimal* if there is no other choice of strategies in which all players receive payoffs at least as high, and at least one player receives a strictly higher payoff. To better understand this concept, we can consider a choice of strategies that is not Pareto-optimal, which means that it does exist an alternate choice of strategies with a higher payoff at least for one player without degrading any player, or in other words there exists a pair of strategies corresponding to a Nash equilibrium. This alternate choice of strategies is better than the one is currently being played, then if the players could jointly select the strategies, then they would wish to switch to this better solution. The basic concept of Pareto-optimality is that, even though both the players realise that there is a better solution, it is not possible to maintain it without a binding agreement between the two players.

The second key concept is that of *social optimality*, which represent a even stronger condition, defined as follows: a choice of strategies (one by each player) is a *socially optimal* if it maximises the sum of the players' payoffs. This definition results appropriate only when is is possible to add the payoff of the different players (in fact, not always it makes sense to add up payoff experimented by players). It is important to note how socially-optimal outcomes must be Pareto-optimal, so the Pareto-optimality is a necessary but not sufficient condition to have a social-optimality. In fact, an outcome which is not Pareto-optimal means that there would be a different outcome where all payoffs were at least the same and and one was larger, and this would be an outcome with a larger sum of payoffs. It is not a sufficient condition, since a Pareto-optimal outcome need not be also socially optimal.

5.2 Evolutionary Game Theory

In sect. 5.1 we have presented the basic ideas of classical game theory, now in this section we will focus on *Evolutionary Game Theory (EGT)*, which applies the basic ideas of game theory also in those situations where nobody makes explicit decisions. If from one hand, classical game theory assumes that the interacting decision makers are rational, as they act so as to maximise their payoffs, from the other hand if we aim at applying the game-thoretic concepts to organisms (different from human beings), this rationality assumption must be relaxed. Therefore, EGT deviates from this assumption and make the model applicable to biology.

EGT is applied to settings in which individuals can show different forms of behaviour (including those that may not be the result of conscious choices), and we will consider which forms of behaviour have the ability to persist in the population, and which ones instead have a tendency to be driven out by others (28). Incorporating the ideas from Darwinian evolution, EGT has been applied most widely in the area of evolutionary biology, the domain in which the idea was first articulated by J. M. Smith and G. R. Price (12), and more recently also in social contexts. Evolutionary biology is based on the idea that an organism's genes largely determine its observable characteristics, and hence its *fitness* in a certain environment (121). Organisms with fitness will have more possibility to reproduce, and thus it will lead their genes to increase the presence in the population. Hence, the higher is the fitness of an organism, the higher will be the rates of reproduction, winning over time.

The key concept of EGT is that many behaviours involve the interaction of multiple organisms in a population, and the success of any one of these organisms depends on how its behaviour interacts with that of others. As a consequence, the fitness of an individual organism cannot be measured in isolation but, rather, it has to be evaluated in the context of the full population in which it lives.

Although classical and evolutionary game theory differ quite radically in how they consider strategic interactions and in the basic assumptions, there are some analogies between the two theories. Indeed, the genetic features of an organism and the corresponding behaviours represent its strategy in a game, its fitness corresponds to payoff, and payoff will depend on the strategies (characteristics) of the interacting organisms. Also, other concepts prove the deep connection between classical and evolutionary game theory, such as the idea of equilibrium formulated by Nash, matches with the the concept of evolutionarily stable strategy (ESS) in EGT (see sect. 5.2.2), which is crucial to make predictions about the results of evolution on a population.

5.2.1 Fitness, Diversity, Selection and Replication

What is important to highlight is the importance of analysing evolutionary processes, as most biological, social and economic systems are governed by evolutionary pressures. These evolutionary systems are made up of entities of different nature, such as animals, genes, cells, behaviours, etc., but they share some common features represented by diversity, selection, and replication. Diversity means that entities in the system show dissimilarities affecting their so-called *individual fitness*. As we said before, fitness is just a measurable indicator that determines how a population of entities evolves, so entities with higher fitness will tend to persist in the population. In particular, the linkage between fitness and the future population composition is the *selection mechanism*, which reduces the diversity of the system, favouring the higher fitness entities. Meanwhile, there are some other mechanisms which generate and preserve a certain amount of diversity and heterogeneity, otherwise the system would lock. In biology, diversity is generated by genetic mutations, while in economic systems, this role is played by innovations and in social and ICT systems by the flow of information and data through the network, which modifies behaviours.

In EGT, strategies (which may be seen as behavioural phenotypes) are selected on the basis of the payoff they obtain, that is the relative frequency of strategies which obtained higher payoffs in the population will increase at the expense of those which obtained relatively lower payoffs. *Replication* (or *Inheritance*) can be considered as a preservation mechanism for the properties of the entities in the system (or the entities themselves), in fact replication means the inheritance from one generation to the next at least to some extent. In biological systems, replication is constituted by genetic transmission, while in social systems, replication is the imitation in the social learning processes.

5.2.2 Evolutionarily Stable Strategy

In the classical game theory, we have defined the key concept of Nash equilibrium for a two-player game, which represents a situation of equilibrium in which neither player has an incentive to deviate from the strategy they are currently using, so the equilibrium is a choice of strategies that tends to persist once the players are using it. In general, the study of dynamic systems often begins with the identification of their stable states, the so-called *static analysis*, which does not consider the dynamics of the system explicitly, but only its rest points.

An analogous notion of Nash equilibrium does exist also in EGT and it is that of an *Evolutionarily Stable Strategy (ESS)*, proposed by J. M. Smith and R. Price in 1973. It allows determining these stable states and represents the most important concept in the static analysis. ESS can be defined as a genetically-determined strategy that tends to persist once it is prevalent in a population. We can assume a certain strategy evolutionarily stable if, when the population is using this strategy, any small group of invaders using a different strategy will eventually die off over multiple generations (28). In biological terms, these invaders could be represented as mutants born into the population with a different behaviour, or migrants joining the population. We can say that a population playing an ESS is uninvadable by any other strategy (113).

Looking closely at the given definition of ESS, it is important to observe the assumptions underpinning its theoretical framework: the ESS is derived for a system composed of a single infinite population of individuals who are repeatedly randomly drawn to play a 2-player symmetric game; furthermore, it only considers monomorphic populations (where all individuals are playing the same strategy) which can be invaded by only one type of mutant strategy at a time. Indeed, this assumption of one single infinite population represents a mean-field approximation used to match the average payoff actually obtained by a population with the expected value of a probability distribution of payoffs (which would be obtained by explicitly modelling players' interactions). Futhermore, this assumption effectively eliminates the impact of arbitrarily small invasions on the incumbent population. To clear this last observation, we can consider the simple case of a 2-player population, where a player i can punish the other player j with a certain magnitude P at a cost $C < P$: obviously, the punishment of the player j represents an advantage for i, and it will produce a favour in the evolution process. Different is the result when we consider a larger population where the punishable players are more than one, in fact in this case the punishment's effect is divided by the size of the population n, wich means that player i will incur the same cost C, but the average payoff of the incumbent population will be decreased only P/n so that, if n is infinite, the effect of punishment will be almost zero. This observation reasoning is important is the foundation of the argument that the concept of ESS is a refinement of Nash equilibrium (see below).

We can capture the concept of ESS also in terms of payoffs, considering a population using a strategy S, a small group of invaders, using another strategy T, should have strictly lower fitness than the users of the majority strategy. As fitness means reproductive success, the sub-population of invaders will shrink over time and evenually die off in the evolutionary process. To clear and summarise the concept of ESS, we give the following fundamental notions of EGT:

1. The *fitness* of an organism in a population is the expected payoff it receives from an interaction with a random member of the population.
2. A strategy T *invades* a strategy S at a certain level δ, for some small positive number δ, if a *delta* fraction of the underlying population uses T and a $1 \ldots \delta$ fraction of the underlying population uses S.
3. A strategy S is evolutionarily stable if there is a (small) positive number γ, such that when any other strategy T invades S at any level $\delta < \gamma$, the fitness of an organism playing S is strictly greater than the fitness of an organism playing T.

Once defined and discussed about ESS, we can now understand the relationship with Nash equilibria. In particular, we can say that "If a strategy S is evolutionarily stable, then (S, S) is a Nash equilibrium". The opposite direction does not hold: it is possible to have a game where (S, S) is a Nash equilibrium, but S is not evolutionarily stable.

Therefore, the concept of an evolutionarily stable strategy can be viewed as a refinement of the concept of a Nash equilibrium: the set of evolutionarily stable strategies S is a subset of the set of strategies S for which (S, S) is a Nash equilibrium. Similarly, the concept of a *strict Nash equilibrium* (when the players use the same strategy which is the unique best response) is a refinement of evolutionary stability: if (S, S) is a strict Nash equilibrium, then S is evolutionarily stable. In other words, these two different notions of equilibrium naturally refine each other. Furthermore, it is interesting to note how, despite the extremely close similarities between the conclusions of ESS and Nash equilibrium, they are built on completely different underlying concepts. From one hand, in a Nash equilibrium, we consider players choosing mutual best responses to each other's strategy. So it demands on the ability of the players to choose optimally and to coordinate on strategies that are best responses to each other. Evolutionary stability, on the other hand, supposes no intelligence or coordination on the part of the players. Instead, strategies are viewed as being naturally placed into the players, because their behaviour is encoded in their genes. ESS increases the fitness of the more successful strategies with a higher reproductive success in the selection process.

5.2.3 Evolutionary Dynamics

M. Nowak in (18) defines *evolutionary dynamics* as the mathematical formalisation of the process of evolution whereby a population changes over time. Natural selection operates such that genotypes (or strategies) with higher fitness tend to become more common, whereas lower-fitness genotypes tend to die out. Mutation (re)introduces variation into the population. This process can also represent cultural evolution and social learning, in which people imitate those with higher payoffs and sometimes experiment with novel strategies. Therefore, EGT is a combination of game theory (mathematical formalisation of social interactions and strategic behaviours), and evolutionary dynamics. There is a population of agents, each of them has a strategy; they interact with each other and earn payoffs. Payoff is translated into fitness, and the frequency of strategies in the population changes over time accordingly: higher-payoff strategies tend to become more common, whereas lower-payoff strategies tend to die out.

To explain the evolutionary dynamics, we exploit the so-called *replicator equation*, introduced for the first time by P. D. Taylor e L. Jonker in 1978 (114). Denoting by x the state of the population, that is the distribution of strategy frequencies, we assume that the x_i are differentiable functions of time t, which means to assume that the population is infinitely large (or that x_i are expected values for an ensemble of populations). Within this hypothesis, we can now postulate a law of motion for $x(t)$. Furthermore, we assume that individuals, meeting randomly, engage in a game with payoff matrix W, so that $(Wx)_i$ is the expected payoff for an individual using strategy S_i, and x^TWx is the average payoff in the population state x. Interpreting the payoff as fitness and considering that postulate that the growth rate for each individual of the part of the population using strategy s_i is proportional to its payoff, we can define the replicator equation as follows (13):

$$\dot{x_i} = x_i\left[(Wx)_i - x^TWx\right] \tag{5.5}$$

where $\dot{x_i}$ is the derivative of x_i, and the term x^TWx exists in relation to the constraint: $\sum_i x_i = 1$. The (5.5) represents mathematically the principle of natural selection, for which strategies with higher fitness will spread more efficiently in a population. The states $x_i = 1, x_j = 0, \forall j \neq j$ are solutions of (5.5) and, in particular, they are the absorbing states, which play a key role in the system dynamics in absence of mutations. Regarding the solutions of the replicator dynamics, we are interested in equilibrium or rest points, or the frequency distributions that make the second member of the equation equal to zero, verifying either $x_i = 0$ or $(Wx)_i = x^TWx = 0, \forall i = 1, \ldots, n$. The solutions are all the mixed strategy Nash equilibria of the game (115), and Nash equilibria are the stable rest points. Henece, the replicator equation describes the evolutionary mechanism by which the players, or the population, can reach the Nash equilibrium, or an ESS. Moreover, the different basins of attraction of the different equilibria allow finding out which of them is selected when there are more than one equilibrium. Thus, the following are the hypotheses underlying the replicator equation:

1. The population in infinitely large;
2. Individuals meet randomly or play against every other one, such that the payoff of strategy S_i is proportional to the payoff averaged over the current population state x.
3. There are no mutations (or migrations among physically distinct groups), thus the increasing or decreasing in frequency of strategies depends only on reproduction mechanism.
4. The variation of the population is linear according to the payoff difference.

The first two hypotheses are crucial to derive the replicator equation, replacing the fitness of a strategy using its mean value, with the population described in terms of frequencies. Some works has been focused on investigating what happens when the population is rather finite and they have identified an adjusted replicator equation, showing a different dynamics and according to population size, it may also invert the direction of evolutionary process. Other works have proved how in case of small populations, ESS may lose their stable feature (116). In particular, the second hypothesis is related to temporal and spatial constraints of interactions, which requires another approximation for representing the fitness of a strategy in the population, as the expected value of fitness is no longer valid. With regards to the third assumption, we may also include mutation, producing the so-called *replicator-mutator equation* (117), in which appears also a term which explictly involves the covariance of fitness and strategies. The fourt assumption is the definition of replicator dynamics. Some works which do not include the hypothesis of linearity consider the *generalised replicator equation*, having the following form:

$$\dot{x_i} = x_i \left[(W_i(x) - x^T W(x) \right] \tag{5.6}$$

where the functions $W_i(x)$ are chosen according to the particular case to be modelled (14).

5.2.4 The Emergence of Cooperation and Social Dilemmas

As we will see in the chap. 6, one of the most important challenges to which EGT is being applied is the understanding of the emergence of cooperation in different contexts, among these the human societies (17). Why should I help a competitor? Cooperation in a competitive world is a conundrum. In general, natural selection is a mechanism in contrast with the evolution of cooperation. In a *social dilemma*, there is the struggle between what is good for the individual and what is good for the population. The population does best if individuals cooperate, but for each individual there is a temptation to defect. Cooperation is an act where individuals can contribute something, at a cost to themselves, to provide a benefit for others. In EGT, cost and benefit are measured in terms of reproductive success, where reproduction can be cultural or genetic.
To better understand the challenge of cooperation we are going to present the most challenging and stringent social dilemma in terms of cooperation, called *Prisoners dilemma*

(or PDG), and the *Snowdrift game* (or SG) as a more relaxed dilemma. In both these pairwise social dilemmas, a player has a binary choice: cooperate (C) or defect (D) (15). Cooperation allows maximising the social welfare, while defection maximises one's own payoff regardless of the other's choice. The payoff matrices of the two games are illustrated in the following tables:

Table 5.1: Payoff Matrix of the Prisoner's Dilemma Game.

	Cooperate	Defect
Payoff to Cooperation	$b-c$	$-c$
Payoff to Defection	b	0

Table 5.2: Payoff Matrix of the Snowdrift Game.

	Cooperate	Defect
Payoff to Cooperation	$b-c/2$	$b-c$
Payoff to Defection	b	0

In the PDG, cooperation results in a benefit b to the opposing player, but incurs a cost c to the cooperator (where $b > c > 0$); defection has no costs or benefits. In both cases, independently of whether the opponent plays C or D, it is, therefore, better to play D. In evolutionary settings, payoffs determine reproductive fitness, and it follows that D is the Evolutionarily Stable Strategy (ESS). Thus, defectors tend always to overcome cooperators, and in a population containing both cooperators and defectors, defectors will have a higher fitness. Considering a longer timespan, there will be an increasing number of cooperators until the population will consist completely of defectors. This can be formalised using replicator dynamics (14), which admits pure defection as the only stable equilibrium. The social dilemma is thus established, since mutual cooperation yields both an individual and total benefit higher than that of mutual defection.

The social dilemma of the PDG can be relaxed by assuming that cooperation yields a benefit that is accessible to both interacting players, and that costs are shared between cooperators. This results in the so-called *Snowdrift Game* (also known as the *Hawk-Dove Game*, or the *Chicken Game*). In the SG, the order of P and S is exchanged, such that $T > R > S > P$, therefore cooperation yields a benefit b to the cooperator as well as to the opposing player, and incurs a cost c if the opponent defects, but only a cost $c/2$ if the opponent cooperates. Thus, both strategies can invade when rare, resulting in a mixed evolutionarily stable state at which the proportion of cooperators is $1 - c/(2b - c)$. It is important to note that in this state the population payoff is smaller than it would be if everybody played C, hence the SG still represents a social dilemma (15). Its essential ingredient is that, in contrast to the PDG, cooperation has an advantage when rare, which implies that the replicator dynamics (14) of the SG converges to a mixed stable equilibrium at which both C and D strategies are present. The PDG is in fact the most stringent cooperative dilemma, where for cooperation to arise a mechanism for the evolution of cooperation is needed (118).

Other than the classical PDG, it is more interesting to study its repeated or iterated form, known as *Iterated Prisoner's Dilemma* (or IPD), a single game consists of a number of rounds of PD, which allows individuals to react to an opponent's past behaviour. Repeated interactions between players open up a whole new world of possible strategies determining whether to cooperate or defect in the next round based on the outcome of the previous rounds. In exploring the evolution of cooperation in a population exploiting the framework of EGT on multiplex networks (see chap. 6), we will consider an IPD game with a memory-one game, where the player has only a finite memory of the previous play. This assumption derives from (119), in fact the authors have proved that, giving only a finite memory of previous play, the payoff obtained is exactly the same as if we would consider a player with a longer memory.

Chapter 6

The Evolutionary Dynamics of Human Cooperation through Multiplex Evolutionary Game Theory

In previous chapters, we have described the bio-inspired ICT, the role of Big Data and the introduction of a social dimension for analysis, which have been revolutionising the bio-inspired ICT in the healthcare context. We have presented the mathematical framework of multilayer networks, able to capture the complexity of the real-world scenarios, allowing us to observe also the non-trivial patterns between entities belonging simultaneously to different layers. Furthermore, Evolutionary Game Theory (EGT) constitutes the fundamental framework for investigating the dynamics of social behaviours in a population, allowing to gain a better understanding of the emergence of cooperation in different contexts, among these the human societies. In this chapter, representing the main focus of this Ph.D. dissertation, joining the mathematical frameworks of multiplex networks and EGT, and following a bio-inspired approach and the social network analysis methodologies, the target is to find out the hidden emergent behaviours within a population across network layers. Furthermore, we will explore and quantify the role of some shaping factors, such as *homophily*, in this evolutionary process. To this aim, in the next sections we will conduct an analytical and simulative investigation of the evolution of human cooperation using multiplex EGT (73).

6.1 Quantifying the Role of Homophily in Human Cooperation Using Multiplex Evolutionary Game Theory

6.1.1 Introduction

Charles Darwin observed how animals, from ants to people, interacting each other, are able to create social groups in which most of them work together for common good. By the way, it was in contrast with his idea of individual fitness surviving over the long term (17, 120). This altruistic behaviour could be justified among kin in the natural selection mechanism. In (121), the authors explain that kin selection is conditioned by "kin recognition", as an individual recognises kin and behaves accordingly. Much research effort has been done in exploring this behaviour, but the understanding of how and why it may work out and evolve among people, linked by every type of relation, remains an open and major challenge.

It raises the conundrum of cooperation, a widespread phenomenon in natural and social systems, but not fully-understood mainly due to its complexity. Cooperation produces a human conflict between the benefit of the single individual and that one of the population, such as risking one's life to save a stranger. The reason why people do something for someone else, cooperating or helping, even though there is often a low probability for direct reciprocity or socially reward, is that actions are contagious (122). Humans tend to cooperate building complex societies, as well as predators hunt in groups to catch more preys as possible (8). More in general, cooperating means to contribute towards a common good at a cost to themselves, providing a benefit for others. Many models and mechanisms have been proposed to explain the emergence of cooperation, nevertheless, only by studying interactions inside population, we are able to explain the hidden patterns leading to cooperation (123, 124). Cooperation may induce assortative interactions among individuals, transforming it into the most profitable strategy (125). Previous works on evolution of cooperation have identified some mechanisms, other than kin selection, related to interactions among individuals which can favour it, such as direct reciprocity, indirect reciprocity, spatial selection, and multilevel selection (18). *Direct reciprocity* is related to a cost of cooperating in order to obtain a gain in the near future. *Indirect reciprocity* involves the dependence of an individual's actions from the previous behaviours of the others. *Spatial selection* is linked with interactions and clusters of cooperators. *Multilevel selection* refers to competition existing between groups and between individuals. Rand and Nowak in (121) underline the importance to distinguish between interaction patterns that are mechanisms for the evolution of cooperation and behaviours that require an evolutionary explanation such as strong reciprocity, upstream reciprocity, and parochial altruism.

Therefore, how did the selfish process of natural selection give rise to cooperation? how might social interactions can give a boost to cooperative behaviour? And what may be the role of a linkage polariser, such as *homophily*, in this evolutionary process? The evolution of cooperation among individuals is an unsolved puzzle: it has being observed since ancient times but, only in the recent years, a lot of research efforts have been done trying to understand and deepening the origin inside social networks. A vast literature on the evolution of cooperation on complex networks (19, 21, 22, 23) highlights many aspects which offer insights on how cooperation can evolve and survive in different scenarios (24, 25, 26, 115). To study cooperation and its evolutionary dynamics, we need to understand the impact of the structure and the nature of social relationships among individuals. The study of network properties and dynamics is the result of a growing research interest in all the aspects related to social networks, from extracted data to emerging behaviours (3, 27, 56). Therefore, both the structural and behavioural dimensions are fundamental to analyse what is the origin of the observed social dynamics within a population (28). Social network analysis is intended to deepen the nature of nodes and ties (126), the actions and interactions between them and all the features and behaviours emerging from the combination of both aspects. These structural and behavioural dimensions allow unveiling the social contagion dynamics (30, 70, 127, 128), showing how the influence runs through the ties connecting nodes, with regards to several phenomena at a population scale, such as diseases, smoking, happiness, etc. (29). Then, network thinking is central in the analysis of contagion processes (129).

Social ties are crucial for collective action. In (130) the authors have formalised the problem of collective action of large groups towards cooperative and uncooperative behaviours, considering how the role of a single actor or a group of people, community or coalition, could trigger a dynamic action within a population, which could represent a social contagion process. The most well-known theory of *critical mass* in the social sciences, is that by Granovetter (131), by considering people that have to make a binary choice, for instance, whether to join a protest or not, or whether to cooperate or not.

In particular, it is argued that the large group problem can be solved by introducing the concept of critical mass, intended as the minimum number of initial contributors, whose efforts can produce a bandwagon effect, which has the power to involve the rest of population, for example, persuading the remaining members of the population towards the adoption of a specific behaviour (130). Therefore, starting from a minority (132), the question is how many people should be involved in a collective action such that a single individual, interacting with them, becomes more likely to join the action? The answer is that each individual has his own threshold in terms of how many other people connected with him should join the action before he will do the same.

The actions of the nodes could be affected by a huge number of factors, among them one of the most important is the role played by *homophily*. The concept of homophily, that is the principle that similarity breeds connection, can explain how social connections are forged and severed over time (133). In (134), the authors define homophily, in terms of information consumed and rumors spreading, as the tendency to interact with users and have similar consumption patterns. Homophily has been introduced and investigated in several works and across various domains (135, 136) from friendship to information transfer. This concept generates some interesting behaviours observed in nature, shaping social relationships with a significant impact on information sharing, influence dynamics, and all the interactions people form and experience. Following this tendency to associate with others who are similar to them, we observe that a contact between similar people occurs at a higher rate that among dissimilar; in terms of social networks, this simply means that the attributes of vertices correlate across edges and it is known as *assortative mixing*.

Among the various aspects of homophily, cognitive homophily is referred to the similarity in interests, beliefs, which can represent a reason towards a choice. People select each other because they share a similar representation of reality, strengthening some contacts rather than others. Nevertheless, other studies in the social sciences have pointed in the opposite direction, e.g. organisational ecologists have suggested that similarity can lead to competition for scarce resources (137), therefore competition among organisations using similar strategies, of similar size, and in geographical proximity with one another tends to be stronger than competition among dissimilar organisations (138). Furthermore, it is crucial to distinguish between homophily, social dependence and social influence. Homophily means that similar nodes are more likely to contact. Social dependence means that nodes exchange resources in order to satisfy their goals. Social influence means that nodes which interact become more similar. Homophily or assortative mixing, however, is only a statement of pattern, and does not say much about the underlying mechanism. For example, if we observe a pattern of homophily in a social network, e.g., on political beliefs or socioeconomic status, we generally cannot distinguish between the edge forming as a result of the similar attributes, or the attributes becoming more similar as a result of the edge. The concept of homophily is important in the dynamics of collective action and critical mass mobilisation (130). Therefore, despite a lot of research efforts in studying the role of homophily in different fields (139), there is still much work to be done in studying its real effect on the evolution of social behaviours within a population.

From the other hand, homophily alone cannot explain why we connect or choose a strategy when interact with others, so it becomes essential considering the multiple types of relationships between nodes, known as multiplexity. In fact, the constituents of a huge variety of real-world complex systems, such as social networks, interact with each other following complicated patterns. Therefore multiplexity allows us to encompass these several interactions and relationships, exploring and unveiling how the different ties in the various layers can impact on the diffusion of social behaviours within a population. The presence of nodes in multiple layers of a system is the key to understand emergent phenomena, adding an extra dimension explaining what is the role not only of the intralayer interactions, as in a monoplex framework, but also of interlayer interactions for the emergence of these phenomena. Multiplex networks consist of multiple channels of connectivity, and they provide the more natural description for systems in which entities have a different set of neighbours in each layer (140) (see Fig. 6.1). In social networks, these layers may correspond to different types of relationships: kin, co-workers, friends, etc. A fundamental aspect of describing multiplex networks is to quantify the interconnectivity between the different types of connections. In fact, interlayer connections can generate new structural and dynamical correlations between components of a system, and it is important to take them into account (141). Multiplex networks are not just a particular case of interdependent networks (34), in fact, as in multiplex systems, many or even all of the nodes have a counterpart in each layer, so one can associate a vector of states to each node. In the multiplex case, the presence of nodes in multiple layers of a system also entails the possibility of self-interactions. This

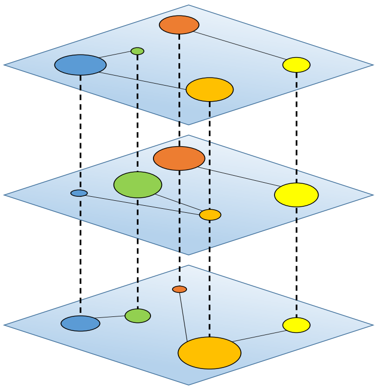


Figure 6.1: Schematic example of a multiplex network. The multiplex is made of N = 5 nodes embedded within M = 3 layers, each one containing 3 links. The size of nodes is proportional to centrality measure. The dashed lines represent interlayer connections, while the continuous lined represent the intra-layer connections.

feature is absent in interdependent networks, which were generated as interconnected communities within a single, larger network (61, 62, 63). In multiplex framework, being the same node at different layers has deep dynamical consequences and give rise to unexpected emergent phenomena (64).

To understand the evolution of social behaviours and, in particular, the emergence of human cooperation within a population in networks, it is important to have a mathematical framework to capture these underlying mechanisms. Fortunately, *Evolutionary Game Theory* (EGT) has provided a powerful framework to investigate cooperative behaviour in systems consisting of competitive individuals (12, 13, 14). EGT allows studying interactions of multiple nodes in a population, and find out the hidden dynamics, shedding light on how and why some behaviours emerge following a specific pattern. Among the classical games, we consider the Prisoner's Dilemma Game (PDG), that is one of the most common paradigms used to describe and study the problem of evolution (15). The investigation of evolutionary dynamics through EGT on multiplex networks allows unveiling and studying the existing social conflicts and dilemmas among the interests of the single nodes and groups, their counterparts in various layers, not neglecting what is captured from homophily, the patterns of similarity and dissimilarity (8, 71).

In (73), we targeted at investigating the evolutionary dynamics of human cooperation dilemma considering the multiplexity of interactions between nodes. To explore the nature of human cooperation, we take into account a Critical Mass (130), able to pop up a new behaviour and trigger a collective action within a population. To analyse the contagiousness of the action (30), we investigate the social connectedness, using a multiplex evolutionary game theory framework (71) and bringing out the real reason why similarity breeds connection (133). Therefore, we focus on both the role of homophily and multiplexity (58, 65), stressing also the importance of the coupling between layers using the communicability function inside the multiplex network (142). Taking into account all these aspects, we propose a novel analytical model and simulate the evolution of human cooperation using evolutionary game theory. In (73), coherently with (8, 71), we analyse the problem of the emergence of cooperation in multiplex networks using EGT. Our findings highlight the key role played by homophily and multiplexity in the evolution of cooperation. In fact, despite the apparently constrained nature of homophily in reducing the boundaries of connectedness, homophily allows observing a new nature of the interaction patterns people experience, looking at these patterns through multiplexity.

6.1.2 Materials and Methods

6.1.2.1 Critical Mass, Centrality and Homophily in Multiplex Networks

Critical mass is defined as the minimum coalition $min(n)$, such that if actors organise into coalitions of size n, at least n people will prefer mutual cooperation to unilateral defection, and it is calculated as follows (130):

$$min(n)\ s.t. \left\{ \sum_{i=1}^{N} H(R_i - T_i) \right\} \geq n \tag{6.1}$$

where n is the overall population and $min(n)$ is the minimum coalition size. The latter depends on the Heaviside function of the difference between Reward and Temptation payoffs, R_i and T_i respectively, evaluated considering different types of games (130). In (73) we extend the concept of Critical Mass introducing a social network approach, considering a scale-free network (55), and taking into account centrality and homophily measures in a multiplex structure. Furthermore, we aim not only to evaluate the minimum coalition size, but also to define a new kind of "Critical Mass" (CM), as the minimum information enclosed in one or more nodes configurations able to trigger

a diffusion process of a behaviour within a population. This represents the role of CM for investigating the human cooperation. The idea is to observe and track the diffusion of behaviours between connected nodes using a multiplex approach. We take into account a scale-free network, thus CM is transformed into a set of nodes that depends on the network structure taking into account centrality. In particular, we choose the eigenvector-like centrality measure, which is defined in (143). The eigenvector-like centrality allows including the concept of influence in our analysis; starting from the spectral properties of the adjacency matrix, considers not only the number of links of each node, but also the quality of such connections. Central nodes are the most influential nodes which can condition the behaviours of their neighbouring nodes.
In Fig. 6.2 we show the centrality measures and its distribution in the multiplex network. The analysis of CM is then further extended considering a multiplex structure

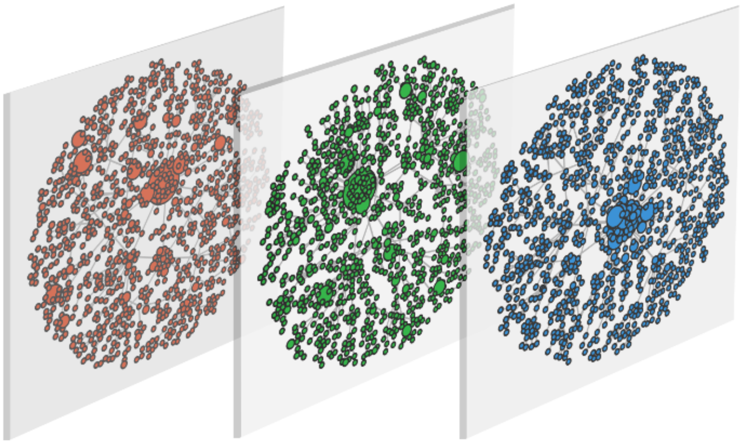


Figure 6.2: Centrality distribution in the multiplex network. The multiplex is made of $N = 1000$ nodes embedded within $M = 3$ layers, each one modelled by a different scale-free network. The size of nodes is proportional to centrality measure.

$\mathbb{M}$, taking a different eigenvector-like centrality measure in each layer α, in order to consider different degrees of importance (or influence) in different layers of the network, and to include this information in the definition of a matrix of mutual influence between layers. Thus, to calculate the centrality of a node within a specific layer, one must take into account all the other layers, as it may depends not only on the neighbours that are linked to x^α within that layer, but also on all other neighbours of x^β that belong to the other layers. In other words, one needs to consider the situation where the influence amongst layers is heterogeneous. To this aim, one can introduce an influence matrix W, defined as a non-negative matrix, such that $W^{\alpha\beta}$ measures the influence on a layer α given by the layer β. Given a multiplex network $\mathbb{M}$ and an influence matrix $W = (w^{\alpha\beta})$, we define the global heterogeneous eigenvector-like centrality of $\mathbb{M}$ as in (7). For each layer α, we introduce the adjacency matrix, denoted by $A^{\alpha} \in \mathbb{R}^{N\times N}$, where each element is:

$$a^{\alpha}_{xy} = a^{\alpha}_{yx} = \begin{cases} 1, & \text{if } x \text{ and } y \text{ are connected} \\ 0, & \text{otherwise} \end{cases} \text{, for } 1 \leq \alpha \leq M \tag{6.2}$$

Now we extend the homophily measure considering a multiplex structure. In each layer α, we define an Homophily matrix H^{α}, where each element h^{α}_{xy} represents the homophily measure between two nodes x and y in the layer α, calculated as:

$$h^{\alpha}_{xy} = \frac{1}{1+\delta^{\alpha}_{xy}}$$

where δ_{xy} measures the homophily difference between two nodes x and y. Then, the Homophily matrix is defined as follows:

$$H^{\alpha} = \begin{bmatrix} 1 & \cdots & \frac{1}{1+\delta^{\alpha}_{1,N}} \\ \vdots & \ddots & \vdots \\ \frac{1}{1+\delta^{\alpha}_{N,1}} & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{N\times N} \tag{6.3}$$

For each layer α, we define the matrix Z^{α}, as the Hadamard product between the homophily matrix H^{α} and the adjacency matrix A^{α}, as follows:

$$Z^{\alpha} = H^{\alpha} \circ A^{\alpha} = \begin{bmatrix} 0 & \cdots & \frac{a^{\alpha}_{1,N}}{1+\delta^{\alpha}_{1,N}} \\ \vdots & \ddots & \vdots \\ \frac{a^{\alpha}_{N,1}}{1+\delta^{\alpha}_{N,1}} & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{N\times N} \tag{6.4}$$

where each element is given by:

$$z^{\alpha}_{xy} = \frac{a^{\alpha}_{xy}}{1+\delta^{\alpha}_{xy}}$$

Note that Z^{α} degenerates in the adjacency matrix A^{α} if, for each pair of nodes, we have $\delta^{\alpha}_{xy} = 0$, that is a network with no homophily difference between nodes. In order to obtain an overall measure that includes both the concepts of centrality and homophily in the multiplex structure, in a first step we need to evaluate the global heterogeneous eigenvector-like centrality and homophily of the multiplex $\mathbb{M}$, defined as a positive and normalised eigenvector $o^{\otimes} \in \mathbb{R}^{NM}$ (if it exists) of the matrix:

$$Z^{\otimes} = \begin{bmatrix} w^{11}\left(Z^1\right)^T & w^{12}\left(Z^2\right)^T & \cdots & w^{1M}\left(Z^M\right)^T \\ w^{21}\left(Z^1\right)^T & w^{22}\left(Z^2\right)^T & \cdots & w^{2M}\left(Z^M\right)^T \\ \vdots & \vdots & \ddots & \vdots \\ w^{M1}\left(Z^1\right)^T & w^{M2}\left(Z^2\right)^T & \cdots & w^{MM}\left(Z^M\right)^T \end{bmatrix} \in \mathbb{R}^{(NM)\times(NM)} \tag{6.5}$$

where $Z^{\otimes}$ is the KhatriRao product of the matrices:

$$W = \begin{bmatrix} w^{11} & w^{12} & \cdots & w^{1M} \\ w^{21} & w^{22} & \cdots & w^{2M} \\ \vdots & \vdots & \ddots & \vdots \\ w^{M1} & w^{M2} & \cdots & w^{MM} \end{bmatrix} \tag{6.6}$$

and

$$Z^T = \left[\left(Z^1\right)^T \quad \left(Z^2\right)^T \quad \cdots \quad \left(Z^M\right)^T\right] \tag{6.7}$$

Note that we consider a symmetric homophily measure between two different nodes, that is $\delta_{xy} = \delta_{yx}$. In other words, we consider realistically that, in terms of similarity, two connected nodes present a symmetric measure, so that: $\left(Z^{\alpha}\right)^T = Z^{\alpha}$. Introducing the following notation:

$$o^{\otimes} = \begin{bmatrix} o^{1\otimes} \\ o^{2\otimes} \\ \vdots \\ o^{M\otimes} \end{bmatrix} \in \mathbb{R}^{NM} \tag{6.8}$$

where $o^{1\otimes}, o^{2\otimes}, \cdots, o^{\alpha\otimes}, \cdots, o^{M\otimes} \in \mathbb{R}^N$, we can define the global heterogeneous eigenvector-like matrix $O^{\otimes}$ of $\mathbb{M}$, as follows:

$$O^{\otimes} = \left[o^{1\otimes} \quad o^{2\otimes} \quad \cdots \quad o^{M\otimes}\right] \in \mathbb{R}^{N\times M} \tag{6.9}$$

Once defined (6.9), in a second step, for each node x, we define an overall measure of its centrality and homophily, denoted by λ_x, in the multiplex network $\mathbb{M}$. Λ is a column vector of size N, which includes all the measures λ_x. It allows quantifying the overall weight, in terms of centrality and homophily, of each node in the multiplex $\mathbb{M}$, as follows:

$$\Lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{M} \left(o^{i\otimes}\right)_1 \\ \sum_{i=1}^{M} \left(o^{i\otimes}\right)_2 \\ \vdots \\ \sum_{i=1}^{M} \left(o^{i\otimes}\right)_N \end{bmatrix} \in \mathbb{R}^N \tag{6.10}$$

Note that:

$$\sum_{x=1}^{N} \lambda_x = 1$$

Now we want to define the CM in the multiplex structure both in a quantitative and qualitatively way. To this purpose, on one hand, we evaluate the minimum coalition size $\bar{n}$ and, on the other hand, we also identify the nodes which maximise the diffusion process of a behaviour within a population of size N. First we consider the multiplex as a single layer of $N \times M$ nodes, and we calculate the CM size $\bar{n}$, as follows:

$$min\left(\bar{n}\right)\ s.t.\ \left(\sum_{i=1}^{NM} H\left(R_i - T_i\right)\right) \geq \bar{n} \tag{6.11}$$

We identify a node as "critical" when it triggers a certain behaviour in all the layers in which it is involved. Therefore, the CM results in a set of "critical" nodes able to give a boost to a certain behaviour in a more effective and faster way, due to its high centrality and homophily weight. In fact, more a node is central in the network structure and more it is similar to the other nodes in the different layers of the multiplex structure, more it becomes relevant in triggering a behaviour. Therefore, assumed that in the multiplex network a "critical" node adopts the same behaviour in all the layers, starting from (6.11) and considering a multiplex network, the CM size $\bar{\bar{n}}$ of $\mathbb{M}$ is given by:

$$\bar{\bar{n}} = \frac{\bar{n}}{M} \leq n \tag{6.12}$$

as we can exclude the replicas of a node that belongs to the set of CM nodes. In other words, $\bar{n}$ is the CM size of the "aggregate layer", obtained considering in a single layer all the connections in the different layers and the nodes in the multiplex (including its counterparts in the different layers of the multiplex). Therefore, the definition of $\bar{\bar{n}}$ allows us to leave out the critical nodes' counterparts from the CM size. We define the Critical Mass density, denoted by τ_{CM}, as follows:

$$\tau_{CM} = \frac{\bar{\bar{n}}}{N} \tag{6.13}$$

In Fig. 6.3 we show the CM density τ_{CM} according to the population of size N and the number of layers M. The plots are generated considering a population N ranging from 5000 nodes to 30000 nodes, and a number of layers M ranging from one layer to seven layers.

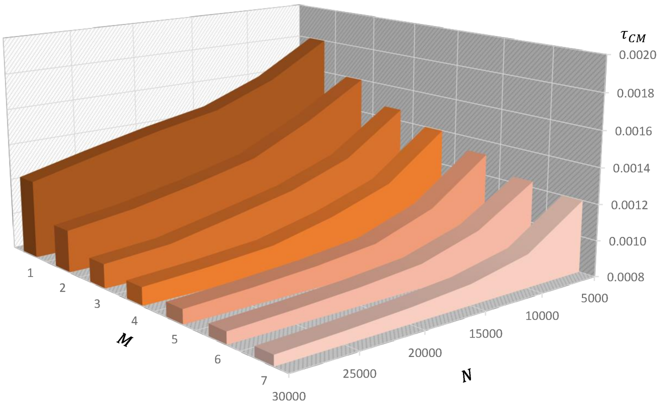


Figure 6.3: Critical mass density. The Critical Mass Density as a function of the population's size N and the number of layers M.

The boundaries of the variation ranges related to the variables N and M are due to the convergence process of the τ_{CM} to a limit value denoted by $\overline{\tau}_{CM}$. In fact, although increasing the size of the "aggregate layer", τ_{CM} maintains nearly the same value:

$$\lim_{N \times M \to +\infty} \tau_{CM} = \overline{\tau}_{CM} \tag{6.14}$$

The plot highlights how the τ_{CM} decreases with the population and the number of layers in the considered variation intervals. Thus, these findings shed light on the significance of the inequality.

To identify qualitatively the set of CM nodes, we take into account the set R, representing all the permutations S_k of the $\bar{\bar{n}}$ nodes in the population N, so that we overall have $N!/\bar{\bar{n}}!$ permutations.

The set R is defined as follows:

$$R = \left\{S_1, S_2, ...S_{N!/\bar{\bar{n}}!}\right\} \tag{6.15}$$

For each subset S_k of R, we can define an overall centrality and homophily measure, as follows:

$$\Lambda_{S_k} = \sum_{x \in S_k} \lambda_x \tag{6.16}$$

The CM in the multiplex structure $\mathbb{M}$, indicated with $\bar{S}$, is the subset which maximises the overall centrality and homophily measure, that is $\Lambda_{\bar{S}}$, as follows:

$$\overline{S} \in R \; s.t. \; \Lambda_{\overline{S}} = \arg\max_{S \in R} \left\{ \sum_{x \in S} \lambda_x \right\} \tag{6.17}$$

Note that it may be more than one subset that satisfy the (6.17). Furthermore, from a computational point of view, calculating $\bar{S}$ is simple, since we have only to consider the $\bar{\bar{n}}$ nodes with the higher values of λ_x.

6.1.2.2 Communicability in Multiplex

In (73), we want to stress the importance of the coupling between layers in exploring the evolution of behaviours in the multiplex structure. To this aim we exploit the communicability function defined in (142), which quantifies the number of possible routes that two nodes have to communicate with each other. Therefore, considering a multiplex formed by M layers, denoted by $L_1, L_2, , L_M$, and their respective matrices $Z_1, Z_2, , Z_M$, representing the Hadamard product between the homophily matrices and the adjacency matrices of the multiplex $\mathbb{M}$, its matrix is then given by $\mathfrak{M} = Z_L + C_{LL}$, where Z_L is:

$$Z_L = \oplus_{a=1}^{M} Z_\alpha \tag{6.18}$$

and C_{LL} is a matrix describing the interlayer interaction, defined as follows:

$$C_{LL} = \begin{bmatrix} 0 & C_{12} & ... & C_{1M} \\ C_{21} & 0 & ... & C_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ C_{M1} & C_{M2} & ... & 0 \end{bmatrix} \in R^{NM \times NM} \tag{6.19}$$

where each element $C_{\alpha\beta} \in \mathbb{R}^{N \times N}$ represents the interaction of layer α with layer β. Here it is assumed that: $C_{\alpha\beta} = C_{\beta\alpha} = C = \omega_{\alpha\beta} I = \omega_{\beta\alpha} I$, for all layers α and β, as we consider a symmetric interaction between layers. ω is the parameter describing the strength of the interlayer interaction, and $I \in \mathbb{R}^{N \times N}$ is the corresponding identity matrix. So we can now explain the multiplex matrix as follow:

$$\mathfrak{M} = \begin{bmatrix} Z_1 & \omega_{12} I & ... & \omega_{1M} I \\ \omega_{21} I & Z_2 & ... & \omega_{2M} I \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{M1} I & \omega_{M2} I & ... & Z_M \end{bmatrix} \in \mathbb{R}^{NM \times NM} \tag{6.20}$$

Since we are interested in accounting for all the walks between any pair of nodes in the multiplex, we consider the number of walks of length k between two generic nodes x and y in the multiplex, which is given by the α, β-entry of the K-th power of the adjacency matrix of the network. Consequently, the walks of k length in the multiplex are given by the different entries of $\mathfrak{M}^K$. As underlined in (142), the walks can include hops of two different kinds, e.g., intra-layer and interlayer hops, and we are interested in giving more weight to the shortest walks than to the longer ones. The communicability between two nodes x and y in the multiplex is given by a weighted sum of all walks from x to y as follows:

$$G_{xy} = I + \mathfrak{M} + \frac{\mathfrak{M}^2}{2!} + ... = \sum_{k=0}^{\infty} \frac{\mathfrak{M}^k}{k!} = [exp(Z_L + C_{LL})]_{xy} \tag{6.21}$$

Now we introduce the communicability matrix G, where each element $G_{\alpha\beta} \in \mathbb{R}^{N \times N}$ is the matrix representing the communicability between every pair of nodes belonging to two different layers α and β, of the multiplex $\mathbb{M}$. It is defined as follows:

$$G = exp(Z_L + C_{LL}) = \begin{bmatrix} G_{11} & G_{12} & ... & G_{1M} \\ G_{21} & G_{22} & ... & G_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ G_{M1} & G_{M2} & ... & G_{MM} \end{bmatrix} \in \mathbb{R}^{NM \times NM} \tag{6.22}$$

In particular, $[G_{\alpha\beta}]_{xy}$ represents the communicability between the node x in the layer α and the node y in the layer β.

6.1.2.3 Evolutionary Dynamics

We use the Prisoners Dilemma game (PDG) as a general metaphor for studying the evolution of cooperation. In this classical social dilemma, two players simultaneously decide whether to cooperate (C) or to defect (D): cooperation results in a benefit b to the opposing player, but incurs a cost c to the cooperator (where $b > c > 0$); defection has no costs or benefits. In both cases, it is best to defect for rational individuals in a single round of the PDG, regardless of the opponent strategy. However, mutual cooperation leads to a higher payoff than mutual defection, but cooperation is irrational. The social dilemma is thus established, since mutual cooperation yields both an individual and total benefit higher than that of mutual defection. The payoff matrix of the PDG is illustrated in Table 6.1:

Table 6.1: Payoff Matrix of the Prisoner's Dilemma Game.

	Cooperate	Defect
Payoff to Cooperation	$b - c$	$-c$
Payoff to Defection	b	0

In evolutionary settings, payoffs determine reproductive fitness, and it follows that D is the *Evolutionarily Stable Strategy* (ESS). This can be formalised using replicator dynamics (14), which admits pure defection as the only stable equilibrium. The PDG is in fact the most stringent cooperative dilemma where, for cooperation to arise, a mechanism for the evolution of cooperation is needed (118). The pairwise nature of the game is translated to a population scale by making the nodes playing with each other, and accumulating the payoff obtained from each interaction. After each round of the game, the strategies of the nodes are updated so that those nodes with less payoff are tempted to imitate the strategy of those fittest individuals. We focus on memory-one game since in (119) the authors have proved that, giving only a finite memory of previous play, the payoff obtained is exactly the same as if we would consider a player with a longer memory. In unstructured populations, in which players are well-mixed, evolutionary dynamics leads all the individuals to defection (7). However, the existence of a network of interactions, so that each node can only play with those directly connected to it, the population can sometimes promote the emergence of cooperation. This mechanism promoting cooperation, known as network reciprocity (16), was observed to be substantially enhanced when the network substrate is a scale-free network (21, 118), a real-world network, with a power law dependence of the degree distribution $P(k) \sim k^{\gamma}$, with the exponent γ typically satisfying $2 < \gamma < 3$. For this reason, we decide to adopt a scale-free as network substrate (55).

We simulate the evolutionary process in accordance with the standard Monte Carlo simulation procedure, composed of elementary steps; including the distribution of competing strategies, which is an elementary step entails randomly selecting a player and one of its neighbours, calculating the payoffs of both players, and finally attempting a strategy adoption. First, a randomly selected player x acquires its payoff P_x by playing the game with all its neighbours on the layer α. Next, player x randomly chooses one neighbour y on the layer β, who then also acquires its payoff P_y on the layer β in the same way as previously did player x. Lastly, player x adopts the strategy S_y from player y with a probability determined by the Fermi function (144):

$$W(S_y \to S_x) = \eta_x \frac{1}{1 + exp[\frac{P_x - P_y}{\delta_{xy} K}]} \tag{6.23}$$

One player x on the layer α of the multiplex $\mathbb{M}$ adopts the strategy S_y of another node playing on the layer β, taking into account the payoff difference, the homophily measure δ_{xy} and a communicability measure η_x in the multiplex network. We take into account a degree of uncertainty in the decision making process given by the factor K. In fact, the temperature K represents a noise level (or selection intensity) and quantifies the uncertainty related to the strategy adoption process; it can vary in the range $]0, +\infty[$. The selected value of K is a traditional and frequently employed choice that does not qualitatively affect the evolutionary outcomes, as shown in many preceding works and reviewed comprehensively in (57). In the $K \to 0$ limit, the adoption of a successful strategy is deterministic, while in the $K \to +\infty$ limit, the strategy learning is blind. The factor δ_{xy}, related to the homophily measure, means that the more the players have a high value, the more one player tends to imitate the strategy of the other one at each round. δ_{xy} can vary in the range $]0, +\infty[$; in particular for $\delta_{xy} \to 0$, the two nodes present the highest homophily value, while in the $\delta_{xy} \to +\infty$ limit, there is no homophily. Furthermore, we introduce the scaling factor η_x of player x which depends on the strategies of related players from the other layers, and it is the key quantity that takes into account the communicability function between layers (23). If we consider only the strategy of the counterpart x' on another layer β, we are in the simplest case and we can assume that η_x is minimal if $S_x = S_{x'}$, otherwise it assumes the maximal value. To avoid frozen states the scaling factor ranges in the interval $[0.1,\ 1]$, assuming $\eta_{x_{min}} = 0.1$ as the minimal scaling factor and $\eta_{x_{max}} = 1$ as the maximal value. In our definition, we consider a more general case where not only the counterpart node x' but also its neighbours on the other layer β determine η_x; in other words, the counterpart and its neighbours can influence the strategy adoption due to the communicability, that includes the interlayer interaction and the number of possible walks from node x to y, where y are all the neighbouring nodes connected with the counterpart node x' on the layer β of the node x on the layer α.

The scaling factor η_x changes linearly between $\eta_{x_{min}}$ and $\eta_{x_{max}}$ in accordance with:

$$\eta_x = 1 - (\eta_{x_{max}} - \eta_{x_{min}}) \frac{\sum_{y\in\beta, S_y = S_x} [G_{\alpha\beta}]_{xy}}{\sum_{y\in\beta} [G_{\alpha\beta}]_{xy}} \tag{6.24}$$

where the numerator is the sum of the communicability functions calculated between the node x on the layer α and all the neighbouring nodes y belonging to the layer β, adopting the same strategy as player x. While the denominator represents the sum of the communicability functions calculated between the node x on the layer α and all the neighbouring nodes y belonging to the layer β. Therefore, the ratio quantifies the influence, in terms of communicability, on the strategy adoption of the player x on the layer α, due to the strategies adopted by the counterpart node and its neighbours on the layer β. In particular, more are players on the layer β with a high communicability with the node x adopting the same strategy as player x, more likely x will adopt the same strategy in the next round. On the other hand, if there are nodes on the layer β with a high communicability, but adopting a different strategy, the player x will be most likely pushed to change its strategy. Thus, this ratio depends on the communicability function and it may result in a bias regarding the strategy adoption of the player x in the next round of the game. Each Monte Carlo step gives a chance for every player to change its strategy once on average.

6.1.3 Results and Discussion

The simulations have been conducted choosing a scale-free network with $N = 1000$ nodes. We take into account different values of homophily randomly chosen following a normal distribution around a mean value, with standard deviation σ. Furthermore, we have considered two different values of interlayer interaction strength $\omega_{\alpha\beta}$; in particular, $\omega_{\alpha\beta} = 0.3$ indicates a low interlayer interaction strength between the layers 1 and 3, while $\omega_{\alpha\beta} = 0.6$ represents a high interlayer interaction strength between the layers 2 and the others. The reasons behind our choice of the interlayer interaction strength between layers are explained in Fig. 6.4.

Fig. 6.5 shows the fraction (or density) of cooperative nodes against the rounds or time steps. ρ varies in the range [0, 1], where 0 corresponds to the global defection, while 1 means a global cooperation of population. We have simulated the evolutionary dynamics for a fixed number of simulations, and the colour corresponds to the population's density, so 'red' indicates the highest density, while 'blue' means the lowest density. In Fig. 6.5, the PDG is played between the interacting nodes in a multiplex network with $M = 3$ layers. We have considered two different values of σ, where $\sigma = 8$ means a low homophily value (Fig. 5A), while $\sigma = 1$ means a higher homophily value (Fig. 5B), fixed a CM value. We show the evolution of cooperation until 200 rounds as, in correspondence of that value, the convergence has already been reached.

This macroscopic evolution highlights how the higher is the homophily value, more quickly nodes converge to cooperation and the density of cooperative nodes tends to the maximum value. In other words, increasing the homophily value of the multiplex network $\mathbb{M}$, we note a faster emergence of cooperation. Instead, considering lower homophily values, we find a lower density of cooperative nodes, that means a slower convergence to cooperation.

The results are coherent with our theoretical expectations: in fact the more the homophily, the more the nodes tend to choose the same cooperative strategy, solving the social dilemma towards the most profitable strategy with the highest payoff for the evolutionary fitness of population. The switching from the pure rational strategy to the most profitable one is due to the interaction between nodes through the different layers

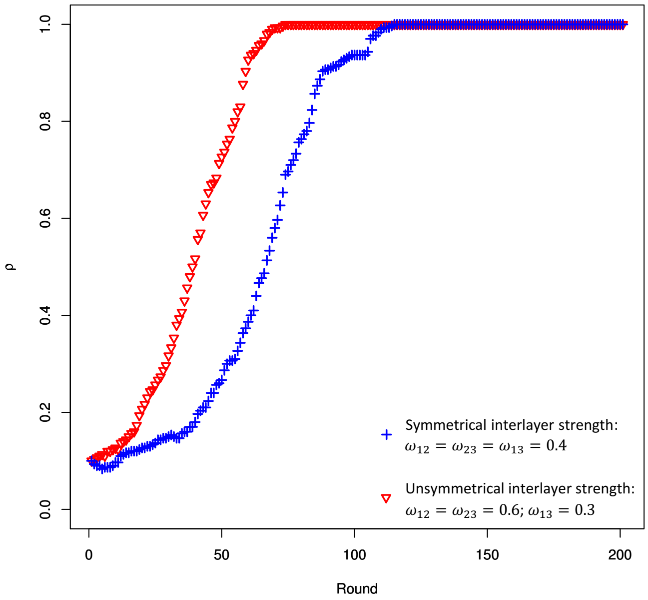
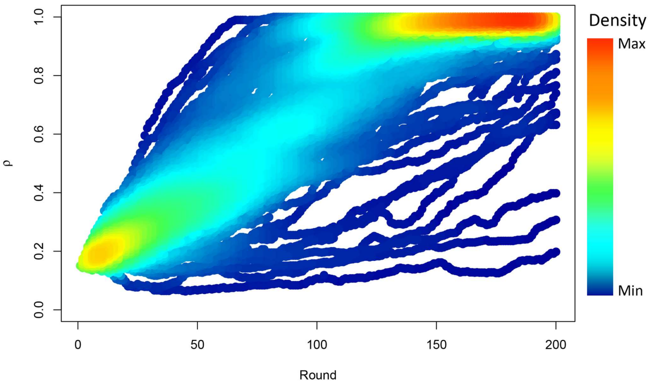


Figure 6.4: Evolution of cooperation considering different interlayer interaction strength. The evolution of cooperation against the round as a function of interlayer interaction strength. The 'blue' plot represents the case of constant interlayer strength: $\omega_{\alpha\beta} = 0.4$. The 'red' plot represents the case of variable interlayer strength (one dominant layer): $\omega_{\alpha\beta} = 0.3$ between layers 1 and 3; $\omega_{\alpha\beta} = 0.6$ between the layer 2 and the other layers of the multiplex. We show the evolution of cooperation until 200 rounds as, in correspondence of that value, the convergence has already been reached. It can be observed that the emergence of cooperation is quicker considering a variable interlayer strength (one dominant layer), than the constant case. The dominant layer acts as a behaviour's polariser of the nodes in the other layers.

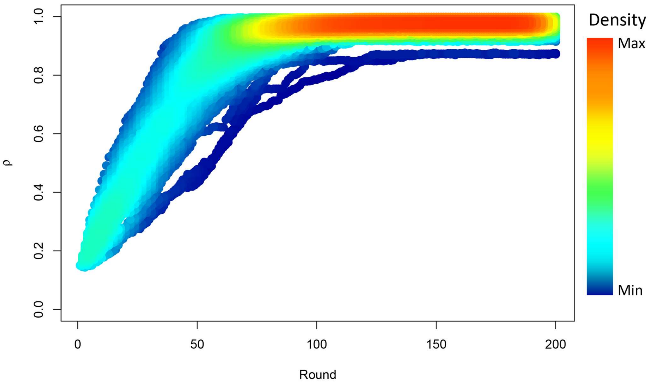
of the multiplex network, creating a sort of 'learning process' driven by homophily, which acts as a catalyst towards cooperation.

In Fig. 6.6, we illustrate the microscopic evolution of cooperation, considering both the cases, respectively with low homophily ($\sigma = 8$) and high homophily value ($\sigma = 1$). We have simulated the evolution of cooperation in the multiplex network, showing the evolutionary dynamics of one of the layers, as the evolution in one layer is representative of the overall one in all the layers of the multiplex. During the steps of evolutionary dynamics, the nodes becomes coloured when they cooperate, otherwise they are 'white'. In particular, we coloured in 'blue' the cooperative nodes in the case of low homophily $\sigma = 8$, while we indicated with 'red' the case of high homophily $\sigma = 1$. The size of nodes are log-proportional to the values of Λ (see (6.10)), so it depends on both the centrality and homophily measures of the multiplex network (see sect. 6.1.2).

As in Fig. 6.5, the Fig. 6.6 highlights the different speed in the emergence of cooperation of the evolutionary process. The formation of cooperative groups in the different parts of the network and also the group size depend on the homophily value. When we consider a low homophily value, nodes tend not to interact with the others in the multiplex network, then the defective behaviour tends to persist more in the population, not favouring the formation of cooperative groups and globally slowing the emergence of cooperation. As a consequence, the group size will be small in this case of low homophily (see Figs. 6A, 6B and 6C). Instead, when we consider a high homophily value, nodes are pushed to interact with each other, so the convergence towards cooperation becomes quicker, and there is a natural formation of larger cooperative groups than in low homophily case (see Figs. 6D, 6E and 6F). Analysing concurrently the corresponding figures of microscopic evolution, we see clearly this difference, both in speed and size, in the formation of cooperative groups. In both cases of respectively low and high homophily, we illustrate the evolutionary process until the convergence has already been reached.

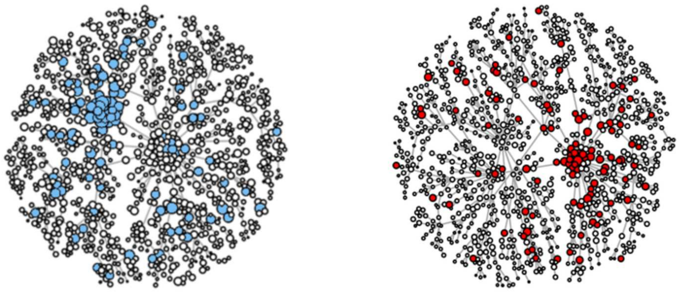


(A) Evolution of cooperation in the low homophily case ($\sigma = 8$).

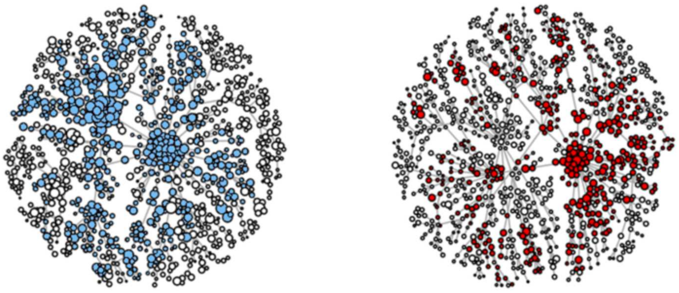


(B) Evolution of cooperation in the high homophily case ($\sigma = 1$).

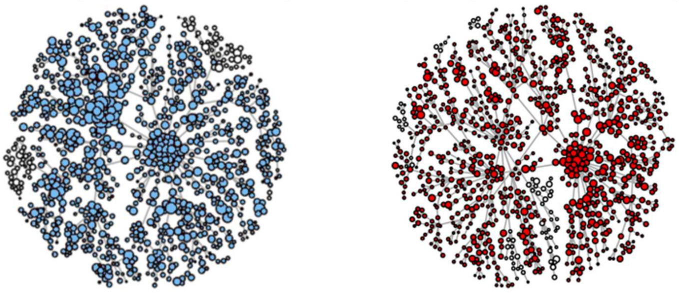
Figure 6.5: Emergence of cooperation over time. The figure illustrates the fraction of cooperative nodes against the rounds or time steps: low homophily (**A**) and high homophily (**B**). The figure shows the evolutionary dynamics of the PDG played between the interacting nodes in a multiplex network with $M = 3$ layers. In both cases $N = 1000$ nodes. The results are obtained choosing a fixed number of simulations and the colour corresponds to the density: 'red' indicates the highest density (that is the maximum number of overlapping points), while 'blue' means the lowest density. As can be observed, increasing the homophily value of the multiplex network $\mathbb{M}$, we note a faster emergence of cooperation.



(A) Low homophily case ($\sigma = 8$): time step 1. **(D)** High homophily case ($\sigma = 1$): time step 1.



(B) Low homophily case ($\sigma = 8$): time step 50. **(E)** High homophily case ($\sigma = 1$): time step 20.



(C) Low homophily case ($\sigma = 8$): time step 120. **(F)** High homophily case ($\sigma = 1$): time step 70.

Figure 6.6: Temporal evolution of cooperation. The figure highlights the microscopic emergence of cooperation in the evolutionary process. The formation of cooperative groups in the network and also the group size depend on the homophily value. Figs. **A**, **B**, **C** - in the low homophily case ($\sigma = 8$), the defective behaviour tends to persist more in the population, not favouring the formation of cooperative groups and globally slowing the emergence of cooperation. Yet, the group size will be smaller in this case of low homophily. Figs. **D**, **E**, **F** - in the high homophily case ($\sigma = 1$), the convergence towards cooperation becomes quicker, and there is a natural formation of larger cooperative groups than in low homophily case. Analysing the corresponding figures of the evolution, we see clearly this difference, both in speed and size, in the formation of cooperative groups.

To sum up, starting from (8, 71), we analysed the emergence of cooperation on multiplex networks. To this aim, we defined a novel analytical model able to analyse the problem of human cooperation in multiplex networks using evolutionary game theory, exploring the role played by multiplexity and homophily in the evolution of cooperation. Therefore, first we have introduced the critical mass in a multiplex network, proposing also the selection criterion to detect nodes to trigger the evolution. To capture the effect of multiplexity and stress the importance of the coupling between the network layers, we have exploited the communicability function defined in (142). We observed how the emergence of cooperation is quicker considering a variable interlayer strength in the different layers, with one dominant layer, than the constant case with the same interlayer strength. This have suggested us that the dominant layer acts as a behaviour's polariser of the nodes in the other layers. We have redesigned the study of evolution considering the homophily as a shaping factor. In particular, we have studied its crucial role in breeding connections and rules interactions within a population, and then influence the strategies of players in multiplex. After having included these concepts of mutiplexity, communicability and homophily in our model, we have investigated the evolutionary dynamics both at macroscopic and microscopic scales. From one hand, the macroscopic evolution have highlighted the crucial role of homophily in solving the social dilemma, moving the population from the pure rational strategy (defection) towards the most profitable strategy with the highest evolutionary fitness (cooperation). From the other hand, the microscopic evolution has pointed out the impact of homophily on the formation of cooperative groups in the network and on groups' size. The results have shown as homophily significantly affects the formation of cooperative groups, both in speed and size. Then, the introduction of multiplexity and homophily not only is a more realistic representation of social systems but, as shown before, it has a key effect on the evolutionary dynamics of cooperation.

Chapter 7

Discussion

In this chapter we will revisit the research questions posed in the first chapter, summarising the main results, contributions and the underlying key aspects we have addressed and presented in this Ph.D. dissertation.

7.1 Research Contributions and Questions Revisited

We can sum up the main reserach contributions presented in this dissertation, answering the research questions posed in sect. 1.2:

- **To what extent can biological processes and models inspire the analysis and management of the complexity of the ICT systems, increasing knowledge and trying to solve also the associated computational problems?** In Chapter 2, we have discussed and showed what are main challenges of ICT, the analogies between biology and ICT, and how using the bio-inspired approach and algorithms, applied to ICT, it becomes possible to optimise and improve the design and methodologies proper of the traditional vision of ICT. The increasing connectedness, the most stringent computational requirements, and the overall enhanced complexity of ICT systems can be faced by exploiting novel bio-inspired approaches, reducing complexity through heuristic approach, re-thinking network considering nodes as humans and evaluating their behaviours in a "Social Smart behaviour", extracting knowledge from network..
- **What are the main properties and features of multilayer networks? How can multilayer networks help to represent and investigate the dy-**

namics and patterns of real-world networks? As explained in Chapter 3, multilayer networks and its mathematical formalisation, meet the need for understanding and gaining insight on the complexity of the large variety of systems and its real properties. Multilayer networks allow to describe the multiple set of interactions among nodes through the different network layers, unveiling also the non-trivial patterns and resulting dynamics. The structural measures presented in Chap. 3 represent the way to characterise the main concepts and aspects to be investigated when dealing with multilayer networks.

- **Despite the huge amount of heterogeneous data to be analysed, the increased connectedness and complexity of social networks, how can ICT systems and models be able to manage them, capturing the complex dynamic patterns, and increasing both performance and context-awareness?** In Chapter 4, we have described the role of bio-inspired ICT in dealing with the Big Data in healthcare context, considering the social connectedness and using the methologies of social network analysis to analyse the dynamics of networked systems, such as healthcare system. Bio-inspired ICT, along with the introduction of a social dimension of analysis, allows to re-think the health system, described as a complex interdependent network, where the single entities interact through the various layers. Exploiting the multilayer paradigm, the novel Health mining introduced and the concept of multi-agent system, we have proposed a framework able to model the social entities, considering complex agents, with context-aware and cognitive abilities which makes nodes able to transmit the extracted knowledge, decide and apply ICT strategies and procedures. This produces an improved and empowered analysis and an overall better management of the healthcare system, and an increased awareness and adaptability of the system according to different targets.

- **What is the more suitable mathematical formalisation to analyse social interactions and behaviours? What are its main properties and features? What are the behaviours have the ability to persist in the population, and which ones instead have a tendency to be driven out by others?** In Chapter 5, we have described the classical Game Theory and

Evolutionary Game Theory methodologies used to describe and analyse the interdependent interactions and the dynamics of behaviours. Indeed, game-theoretic approach allows to model those situations where the individual behaviours and decisions affect the outcomes of others in the same environment, that is the conflictual scenarios where there is a strategic behavioural interdependence among individuals, affecting the dynamics of behaviours. We have discussed about Evolutionary Game Theory (EGT), representing the mathematical framework for investigating the dynamics of social behaviours in a population, gaining a better understanding of the emergence of cooperation in different contexts, among these the human societies. In fact, EGT allows to evaluate which are the behaviours able to persist in the population, using the replicator dynamics and calculating the fitness function, which determines reproductive success of a behaviour or a strategy in a population.

- **Therefore, how did the selfish process of natural selection, described by Charles Darwin, gives rise to cooperation? How might social interactions can give a boost to cooperative behaviour? And what may be the role of a linkage polarizer, such as homophily, in this evolutionary process?** In Chapter 6, we have explored one of the most important challenges to which EGT is being applied, that is the emergence of human cooperation in a social network. Modelling the social network as a multiplex network, we have defined a novel analytical model able to analyse the problem of human cooperation on multiplex networks using EGT, exploring the joint role played by multiplexity and homophily in the evolution of cooperation. Combining multiplex network and EGT frameworks, other than considering homophily, as a shaping factor of relationships in the social network, we have investigated the evolutionary dynamics both at macroscopic and microscopic scales. The results have shown the striking role of homophily in the emergence of cooperation and how the introduction of multiplexity and homophily not only is a more realistic representation of social systems but it has a key effect on the evolutionary dynamics of cooperation.

7.2 Concluding Remarks and Future Works

This Ph.D. dissertation has been focused on evolutionary dynamics of social behaviours, in particular human behaviours, on multilayer networks. In particular, starting from the bio-inspired approach to ICT and introducing some social network measures, and exploiting the mathematical framework of evolutionary game theory, we have investigated the evolution of human cooperation on a multiplex network. Multiplexity, as an extra dimension of analysis, has allowed us to give a more realistic representation of social systems, describing the complexity of social systems, but also to observe how the non-trivial patterns and the different social interactions and relationships on the various layers can impact on the emergence of social behaviours in a population. From the other hand, it has been also quantified the role of homophily as a linkage polariser or shaping factor of the social connections, in this evolutionary process. Although the apparently constraining nature of homophily in boundering connections, due to the patterns of similarity and dissimilarity, homophily has enabled us to shed light on a new nature of the interaction patterns people experience, looking at these patterns through multiplexity. Moreover, we have observed how the introduction of multiplexity and homophily has a key effect on the evolutionary dynamics of cooperation. In particular, we have found out the striking role of homophily in the emergence of human cooperation in the social dilemma represented by the PD game, both at macroscopic and microscopic scales. Macroscopically, homophily is able to solve the social dilemma, moving the population from the pure rational strategy (defection) towards the most profitable strategy with the highest evolutionary fitness (cooperation), while the microscopic evolution has pointed out the impact of homophily on the formation of cooperative groups in the network and on groups' size.

The analysis of social interactions and human behaviours and the investigation of its evolutionary dynamics has been the core part of my research during my Ph.D. If, from one hand, the mathematical frameworks of game theory and EGT have constituted the way to model such interactions and the strategic interedependence among people when interacting, also in evolutionary perspective, from the other hand, multilayer networks has been the mathematical tool to describe more realistically such complex dynamics in the social network. Social networking and the bio-inspired approach has been the way to characterise nodes and their features in the network, considering also the single traits of the individuals and how these can affect their relationships.

The fascinating development of network science in the last decades has been creating new growing challenges in various fields, from neuroscience to biology, and social sciences. Directions for future research are many, and the idea of joining all these aspects is thus promising. In terms of evolutionary games, perhaps the most obvious path to take is considering other types of games on mutlilayer networks, such as the ultimatum game, rock-paper-scissors games or other social dilemmas, studying the emergence of dynamic patterns and collective behaviours. As underlined in Chap. 4, this evolutionary perspective of nodes and data, creating an organised and increasing knowledge, due to interactions on the different layers, paves the way to explore collective awareness of the single individuals or groups of individuals, e.g. developing platforms for sustainability and social innovation. Reasoning in terms of epidemics or disease spreading processes, we can investigate them from a multilayer perspective, studying the contagions on the different layers and how the social connections may help in building prevention measures, inhibiting a pandemic diffusion. Therefore, joining the multilayer framework and EGT provides the basis to further research in epidemiology.

Another aspect to be investigated is to look at of the concept of homophily and the patterns of similarity/dissimilarity from other perspectives. One of these is how the structural dynamics of emotionally complex collaboration and in depth moral commitments has generated regular expulsion events of founding populations, generating a rapid dispersal of human populations, which has pushed them to take new routes and cross significant environmental and behavioural barriers, facing also the risks and difficulties. Therefore, along with cognitive and cultural complexity, we may recognise the influence of emerging emotional complexity on significant behavioural changes.

Another issue involving homophily is the spreading of misinformations or rumours on online social networks as one of the main risks for our society, thus characterising the similarity of the consumption patterns in online social networks may help to measure the degree of misinformation in the network and detect the most rumorous clusters.

References

[1] Albert-László Barabási. **The network takeover**. *Nature Physics*, **8**(1):14, 2011. 1, 42

[2] Réka Albert and Albert-László Barabási. **Statistical mechanics of complex networks**. *Reviews of modern physics*, **74**(1):47, 2002. 1, 43, 48, 68

[3] Mark EJ Newman. **The structure and function of complex networks**. *SIAM review*, **45**(2):167–256, 2003. 1, 43, 48, 68, 121

[4] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. **Complex networks: Structure and dynamics**. *Physics reports*, **424**(4):175–308, 2006. 1, 42, 43, 45, 48, 57

[5] Santo Fortunato. **Community detection in graphs**. *Physics Reports*, **486**(3):75–174, 2010. 1, 48

[6] Rubén J Sánchez-García, Emanuele Cozzo, and Yamir Moreno. **Dimensionality reduction and spectral properties of multilayer networks**. *Physical Review E*, **89**(5):052815, 2014. 1, 2, 57

[7] Stefano Boccaletti, G Bianconi, R Criado, Charo I Del Genio, J Gómez-Gardeñes, M Romance, I Sendina-Nadal, Z Wang, and M Zanin. **The structure and dynamics of multilayer networks**. *Physics Reports*, **544**(1):1–122, 2014. 2, 43, 45, 46, 53, 56, 75, 82, 89, 94, 127, 134

[8] Joan T Matamalas, Julia Poncela-Casasnovas, Sergio Gómez, and Alex Arenas. **Strategical incoherence regulates cooperation in social dilemmas on multiplex networks**. *Scientific reports*, **5**, 2015. 2, 6, 56, 120, 124, 125, 141

[9] Zhihai Rong, Zhi-Xi Wu, and Wen-Xu Wang. **Emergence of cooperation through coevolving time scale in spatial prisoners dilemma**. *Physical Review E*, **82**(2):026101, 2010. 3

[10] Robert Axelrod. **The Evolution of Cooperation**. *Basic Books, New York. Econornetrica*, **39**:383–96, 1984. 3, 4, 43

[11] Martin Nowak and Roger Highfield. *SuperCooperators: Altruism, evolution, and why we need each other to succeed.* Simon and Schuster, 2011. 3, 4

[12] John Maynard Smith. *Evolution and the Theory of Games.* Cambridge university press, 1982. 3, 110, 124

[13] Josef Hofbauer and Karl Sigmund. *Evolutionary games and population dynamics.* Cambridge University Press, 1998. 3, 43, 44, 115, 124

[14] Josef Hofbauer and Karl Sigmund. **Evolutionary game dynamics**. *Bulletin of the American Mathematical Society*, **40**(4):479–519, 2003. 3, 116, 117, 118, 124, 133

[15] Michael Doebeli and Christoph Hauert. **Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game**. *Ecology Letters*, **8**(7):748–766, 2005. 3, 117, 118, 124

[16] Martin A Nowak and Robert M May. **Evolutionary games and spatial chaos**. *Nature*, **359**(6398):826–829, 1992. 3, 43, 134

[17] Elizabeth Pennisi. **How did cooperative behavior evolve?** *Science*, **309**(5731):93–93, 2005. 4, 43, 116, 120

[18] Martin A Nowak. **Five rules for the evolution of cooperation**. *science*, **314**(5805):1560–1563, 2006. 4, 43, 114, 120

[19] Zhen Wang, Lin Wang, Attila Szolnoki, and Matjaž Perc. **Evolutionary games on multilayer networks: a colloquium**. *The European Physical Journal B*, **88**(5):1–15, 2015. 4, 6, 121

[20] Michael Meisel, Vasileios Pappas, and Lixia Zhang. **A taxonomy of biologically inspired research in computer networking**. *Computer Networks*, **54**(6):901–916, 2010. 4, 13, 15, 16, 70

[21] Francisco C Santos and Jorge M Pacheco. **Scale-free networks provide a unifying framework for the emergence of cooperation**. *Physical Review Letters*, **95**(9):098104, 2005. 6, 121, 134

[22] Matjaž Perc and Attila Szolnoki. **Coevolutionary gamesa mini review**. *BioSystems*, **99**(2):109–125, 2010. 6, 121

[23] Matjaž Perc, Jesús Gómez-Gardeñes, Attila Szolnoki, Luis M Floría, and Yamir Moreno. **Evolutionary dynamics of group interactions on structured populations: a review**. *Journal of The Royal Society Interface*, **10**(80):20120997, 2013. 6, 121, 135

[24] Julia Poncela, Jesús Gómez-Gardenes, Luis M Floria, Angel Sánchez, and Yamir Moreno. **Complex cooperative networks from evolutionary preferential attachment**. *PLoS one*, **3**(6):e2449, 2008. 6, 121

[25] Keke Huang, Tao Wang, Yuan Cheng, and Xiaoping Zheng. **Effect of Heterogeneous Investments on the Evolution of Cooperation in Spatial Public Goods Game**. *PloS one*, **10**(3), 2015. 6, 121

[26] Matjaz Perc and Zhen Wang. **Heterogeneous aspirations promote cooperation in the prisoners dilemma game**. *PLoS One*, **5**(12):e15117, 2010. 6, 121

[27] Stanley Wasserman. *Social network analysis: Methods and applications*, **8**. Cambridge university press, 1994. 8, 62, 121

[28] David Easley and Jon Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world.* Cambridge University Press, 2010. 8, 107, 110, 112, 121

[29] Nicholas A Christakis and James H Fowler. **The spread of obesity in a large social network over 32 years**. *New England journal of medicine*, **357**(4):370–379, 2007. 8, 121

[30] Nicholas A Christakis and James H Fowler. **Social contagion theory: examining dynamic social networks and human behavior**. *Statistics in medicine*, **32**(4):556–577, 2013. 8, 73, 82, 88, 89, 121, 125

[31] Grace Eden, Marina Jirotka, and Bernd Stahl. **Responsible Research and Innovation: Critical reflection into the potential social consequences of ICT**. In *Research Challenges in Information Science (RCIS), 2013 IEEE Seventh International Conference on*, pages 1–12. IEEE, 2013. 14

[32] Pujan Ziaie. **Challenges and issues of ICT industry in developing countries based on a case study of the barriers and the potential solutions for ICT deployment in Iran**. In *Computer Applications Technology (ICCAT), 2013 International Conference on*, pages 1–6. IEEE, 2013. 14

[33] Mark S Granovetter. **The strength of weak ties**. *American journal of sociology*, pages 1360–1380, 1973. 16, 21

[34] Amir Bashan, Yehiel Berezin, Sergey V Buldyrev, and Shlomo Havlin. **The extreme vulnerability of interdependent spatially embedded networks**. *Nature Physics*, **9**(10):667–672, 2013. 16, 44, 123

[35] Falko Dressler and Ozgur B Akan. **A survey on bio-inspired networking**. *Computer Networks*, **54**(6):881–900, 2010. 16, 18, 20, 25, 83

[36] Amos Tversky and Daniel Kahneman. **Availability: A heuristic for judging frequency and probability**. *Cognitive psychology*, **5**(2):207–232, 1973. 21, 22

[37] P Slovic, B Fischhoff, and S Lichtenstein. **Facts versus fears: Understanding perceived risks. D. Kahneman, P. Slovic, and A. Tversky (Eds), Judgment under uncertainty: Heuristics and biases**, 1982. 22

[38] Daniel G Goldstein and Gerd Gigerenzer. **Models of ecological rationality: the recognition heuristic.** *Psychological review*, **109**(1):75, 2002. 22, 26, 29

[39] Alessandro Di Stefano, Aurelio La Corte, Marco Leotta, Pietro Lió, and Marialisa Scatá. **It measures like me: An IoTs algorithm in WSNs based on heuristics behavior and clustering methods**. *Ad Hoc Networks*, **11**(8):2637–2647, 2013. 23, 24, 26, 28

[40] Ian F Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci. **Wireless sensor networks: a survey**. *Computer networks*, **38**(4):393–422, 2002. 23, 34

[41] Herbert A Simon. **Rational choice and the structure of the environment.** *Psychological review*, **63**(2):129, 1956. 24, 26, 28

[42] David W McMillan and David M Chavis. **Sense of community: A definition and theory**. *Journal of community psychology*, **14**(1):6–23, 1986. 24

[43] Gerd Gigerenzer and Daniel G Goldstein. **Reasoning the fast and frugal way: models of bounded rationality.** *Psychological review*, **103**(4):650, 1996. 25, 29, 36

[44] Franco Bagnoli, Andrea Guazzini, and Pietro Liò. **Human heuristics for autonomous agents**. In *Bio-Inspired Computing and Communication*, pages 340–351. Springer, 2008. 28, 32, 37, 39, 94

[45] Waldo R Tobler. **A computer movie simulating urban growth in the Detroit region**. *Economic geography*, pages 234–240, 1970. 29

[46] Joe H Ward Jr. **Hierarchical grouping to optimize an objective function**. *Journal of the American statistical association*, **58**(301):236–244, 1963. 32

[47] Wim Van Beers. **Kriging metamodeling in discrete-event simulation: an overview**. In *Proceedings of the 37th conference on Winter simulation*, pages 202–208. Winter Simulation Conference, 2005. 33

[48] Aurelio La Corte, Alessandro Di Stefano, Marialisa Scatà, and Marco Leotta. **A Energy-Preserving Model for Wireless Sensors Networks based on Heuristic Self-Organized Routing**. In *IEEE International Conference on Systems, Man and Cybernetics (IEEE SMC 2013)*, 2013. 34, 35, 36, 41

[49] Jamal N Al-Karaki and Ahmed E Kamal. **Routing techniques in wireless sensor networks: a survey**. *Wireless communications, IEEE*, **11**(6):6–28, 2004. 35

[50] Shio Kumar Singh, MP Singh, DK Singh, et al. **Routing protocols in wireless sensor networks–A survey**. *International Journal of Computer Science & Engineering Survey (IJCSES) Vol*, **1**:63–83, 2010. 35

[51] Judea Pearl. *Heuristics: intelligent search strategies for computer problem solving*. Addison-Wesley Pub. Co., Inc., Reading, MA, 1984. 36

[52] Gerd Gigerenzer and Peter M Todd. *Simple heuristics that make us smart*. Oxford University Press, USA, 1999. 36

[53] Mujdat Soyturk and Turgay Altilar. **A routing algorithm for mobile multiple sinks in large-scale wireless sensor networks**. In *Wireless Pervasive Computing, 2007. ISWPC'07. 2nd International Symposium on*. IEEE, 2007. 37

[54] Paul Erdös and Alfréd Rényi. **On the evolution of random graphs**. *Publ. Math. Inst. Hung. Acad. Sci*, **5**:17–61, 1960. 43

[55] Albert-László Barabási and Réka Albert. **Emergence of scaling in random networks**. *science*, **286**(5439):509–512, 1999. 43, 68, 125, 134

[56] Duncan J Watts and Steven H Strogatz. **Collective dynamics of small-worldnetworks**. *nature*, **393**(6684):440–442, 1998. 43, 62, 121

[57] György Szabó and Gabor Fath. **Evolutionary games on graphs**. *Physics Reports*, **446**(4):97–216, 2007. 43, 134

[58] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. **Multilayer networks**. *Journal of Complex Networks*, **2**(3):203–271, 2014. 43, 45, 48, 52, 56, 125

[59] Gregorio D'Agostino and Antonio Scala. *Networks of networks: the last frontier of complexity*, **340**. Springer, 2014. 43

[60] Arkadiusz Stopczynski, Vedran Sekara, Piotr Sapiezynski, Andrea Cuttone, Mette My Madsen, Jakob Eg Larsen, and Sune Lehmann. **Measuring large-scale social networks with high resolution**. *PloS one*, **9**(4):e95978, 2014. 44, 74, 75, 82

[61] Sergey V Buldyrev, Roni Parshani, Gerald Paul, H Eugene Stanley, and Shlomo Havlin. **Catastrophic cascade of failures in interdependent networks**. *Nature*, **464**(7291):1025–1028, 2010. 44, 45, 48, 50, 53, 124

[62] Mark Dickison, Shlomo Havlin, and H Eugene Stanley. **Epidemics on interconnected networks**. *Physical Review E*, **85**(6):066109, 2012. 44, 124

[63] Jianxi Gao, Sergey V Buldyrev, Shlomo Havlin, and H Eugene Stanley. **Robustness of a network of networks**. *Physical Review Letters*, **107**(19):195701, 2011. 44, 45, 48, 50, 124

[64] Manlio De Domenico, Albert Solé-Ribalta, Emanuele Cozzo, Mikko Kivelä, Yamir Moreno, Mason A Porter, Sergio Gómez, and Alex Arenas. **Mathematical formulation of multilayer networks**. *Physical Review X*, **3**(4):041022, 2013. 45, 48, 50, 56, 124

[65] Federico Battiston, Vincenzo Nicosia, and Vito Latora. **Structural measures for multiplex networks**. *Physical Review E*, **89**(3):032804, 2014. 45, 48, 50, 56, 63, 64, 65, 125

[66] Michael Szell, Renaud Lambiotte, and Stefan Thurner. **Multirelational organization of large-scale social networks in an online world**. *Proceedings of the National Academy of Sciences*, **107**(31):13636–13641, 2010. 50

[67] Ed Bullmore and Olaf Sporns. **Complex brain networks: graph theoretical analysis of structural and functional systems**. *Nature Reviews Neuroscience*, **10**(3):186–198, 2009. 50

[68] Petter Holme and Jari Saramäki. **Temporal networks**. *Physics reports*, **519**(3):97–125, 2012. 54

[69] Ginestra Bianconi. **Statistical mechanics of multiplex networks: Entropy and overlap**. *Physical Review E*, **87**(6):062806, 2013. 56, 69

[70] Emanuele Cozzo, Raquel A Banos, Sandro Meloni, and Yamir Moreno. **Contact-based social contagion in multiplex networks**. *Physical Review E*, **88**(5):050801, 2013. 56, 121

[71] Jesús Gómez-Gardenes, Irene Reinares, Alex Arenas, and Luis Mario Floría. **Evolution of cooperation in multiplex networks**. *Scientific reports*, **2**, 2012. 56, 124, 125, 141

[72] Vincenzo Nicosia, Ginestra Bianconi, Vito Latora, and Marc Barthelemy. **Growing multiplex networks**. *Physical review letters*, **111**(5):058701, 2013. 56, 68

[73] Alessandro Di Stefano, Marialisa Scatà, Aurelio La Corte, Pietro Liò, Emanuele Catania, Ermanno Guardo, and Salvatore Pagano. **Quantifying the Role of Homophily in Human Cooperation Using Multiplex Evolutionary Game Theory**. *PLoS ONE*, **10**(10):e0140646, 10 2015. 56, 119, 124, 125, 131

[74] Vincenzo Nicosia, Miguel Valencia, Mario Chavez, Albert Díaz-Guilera, and Vito Latora. **Remote synchronization reveals network symmetries and functional modules**. *Physical review letters*, **110**(17), 2013. 56

[75] Sergio Gomez, Albert Diaz-Guilera, Jesus Gomez-Gardeñes, Conrad J Perez-Vicente, Yamir Moreno, and Alex Arenas. **Diffusion dynamics on multiplex networks**. *Physical review letters*, **110**(2):028701, 2013. 56, 57, 58

[76] Juan A Almendral and Albert Díaz-Guilera. **Dynamical and spectral properties of complex networks**. *New Journal of Physics*, **9**(6):187, 2007. 58

[77] Albert Sole-Ribalta, Manlio De Domenico, Nikos E Kouvaris, Albert Diaz-Guilera, Sergio Gomez, and Alex Arenas. **Spectral properties of the Laplacian of multiplex networks**. *Physical Review E*, **88**(3):032807, 2013. 58, 59

[78] Manlio De Domenico, Albert Solé, Sergio Gómez, and Alex Arenas. **Random walks on multiplex networks**. *arXiv preprint arXiv:1306.0519*, 2013. 59

[79] Vincenzo Nicosia and Vito Latora. **Measuring and modeling correlations in multiplex networks**. *Phys. Rev. E*, **92**:032805, 2015. 67, 68

[80] Giulia Menichetti, Daniel Remondini, Pietro Panzarasa, Ral J. Mondragn, and Ginestra Bianconi. **Weighted Multiplex Networks**. *PLoS ONE*, **9**(6):e97857, 06 2014. 67, 68

[81] Alessandro Di Stefano, Aurelio La Corte, Pietro Liò, and Marialisa Scatà. **Bio-Inspired ICT for Big Data Management in Healthcare**. In *Intelligent Agents in Data-intensive Computing*, **14**, pages 1–26. Springer, 2016. 70, 83

[82] Wullianallur Raghupathi and Viju Raghupathi. **Big data analytics in healthcare: promise and potential**. *Health Information Science and Systems*, **2**(1):3, 2014. 71, 73, 79, 80, 81

[83] Alessandro Di Stefano, Aurelio La Corte, and Marialisa Scatà. **Health Mining: a new data fusion and integration paradigm**. In *Proceedings of CIBB 2014*, 2014. 71, 82, 83, 101

[84] Enrico Capobianco et al. **Comorbidity: a multidimensional approach**. *Trends in molecular medicine*, **19**(9):515–521, 2013. 71, 84, 86

[85] Mohammad Ali Moni and Pietro Liò. **comoR: a software for disease comorbidity risk assessment**. *Journal of clinical bioinformatics*, **4**(1):1–11, 2014. 72, 84, 85, 89

[86] Thomas A Peterson, Emily Doughty, and Maricel G Kann. **Towards precision medicine: advances in computational approaches for the analysis of human variants**. *Journal of molecular biology*, **425**(21):4047–4063, 2013. 72, 93

[87] Wei Tan, M Brian Blake, Iman Saleh, and Schahram Dustdar. **Social network-sourced big data analytics**. *IEEE Internet Computing*, **17**(5):62–69, 2013. 73

[88] Ray M Chang, Robert J Kauffman, and YoungOk Kwon. **Understanding the paradigm shift to computational social science in the presence of big data**. *Decision Support Systems*, **63**:67–80, 2014. 73

[89] Douglas Henrique Teodoro, Rémy Choquet, Daniel Schober, Giovanni Mels, Emilie Pasche, Patrick Ruch, and Christian Lovis. **Interoperability driven integration of biomedical data sources**. *Studies in health technology and informatics*, **169**:185–9, 2011. 73

[90] Petra Perner. **Mining sparse and big data by case-based reasoning**. *Procedia Computer Science*, **35**:19–33, 2014. 73

[91] Albert Weichselbraun, Stefan Gindl, and Arno Scharl. **Enriching semantic knowledge bases for opinion mining in big data applications**. *Knowledge-Based Systems*, **69**:78–85, 2014. 73

[92] Xue-Wen Chen and Xiaotong Lin. **Big data deep learning: Challenges and perspectives**. *Access, IEEE*, **2**:514–525, 2014. 73

[93] Rahul Sharan Renu, Gregory Mocko, and Abhiram Koneru. **Use of big data and knowledge discovery to create data backbones for decision support systems**. *Procedia Computer Science*, **20**:446–453, 2013. 73

[94] Avita Katal, Mohammad Wazid, and RH Goudar. **Big data: Issues, challenges, tools and Good practices**. In *Contemporary Computing (IC3), 2013 Sixth International Conference on*, pages 404–409. IEEE, 2013. 73

[95] Richard K Lomotey and Ralph Deters. **Towards Knowledge Discovery in Big Data**. In *Service Oriented System Engineering (SOSE), 2014 IEEE 8th International Symposium on*, pages 181–191. IEEE, 2014. 75

[96] Philip Russom et al. **Big data analytics**. *TDWI Best Practices Report, Fourth Quarter*, 2011. 75, 76

[97] Anthony JG Hey, Stewart Tansley, Kristin Michele Tolle, et al. *The fourth paradigm: data-intensive scientific discovery*, **1**. Microsoft Research Redmond, WA, 2009. 77

[98] Stephan Kitchovitch and Pietro Liò. **Community Structure in Social Networks: Applications for Epidemiological Modelling**. *PLoS ONE*, **6**(7):e22220, 07 2011. 82

[99] Hian Chye Koh, Gerald Tan, et al. **Data mining applications in healthcare**. *Journal of healthcare information management*, **19**(2):65, 2011. 82

[100] D Spiegelhalter. **The future lies in uncertainty**. *Science*, **345**(6194):264–265, 2014. 84

[101] Euan A Ashley, Atul J Butte, Matthew T Wheeler, Rong Chen, Teri E Klein, Frederick E Dewey, Joel T Dudley, Kelly E Ormond, Aleksandra Pavlovic, Alexander A Morgan, et al. **Clinical assessment incorporating a personal genome**. *The Lancet*, **375**(9725):1525–1535, 2010. 85, 86

[102] Wei-Chih Kan, Jhi-Joung Wang, Shuo-Yu Wang, Yih-Min Sun, Chien-Ya Hung, Chin-Chen Chu, Chin-Li Lu, Shih-Feng Weng, Chung-Ching Chio, and Chih-Chiang Chien. **The new Comorbidity Index for predicting survival in elderly dialysis patients: a long-term population-based study**. *PloS one*, **8**(8):e68748, 2013. 86

[103] Nicola Luigi Bragazzi. **From P0 to P6 medicine, a model of highly participatory, narrative, interactive, and augmented medicine: some considerations on Salvatore Iaconesis clinical story**. *Patient preference and adherence*, **7**:353, 2013. 89, 90, 92

[104] Gabriella Pravettoni and Alessandra Gorini. **A P5 cancer medicine approach: why personalized medicine cannot ignore psychology**. *Journal of evaluation in clinical practice*, **17**(4):594–596, 2011. 89, 90, 92

[105] Albert-László Barabási, Natali Gulbahce, and Joseph Loscalzo. **Network medicine: a network-based approach to human disease**. *Nature Reviews Genetics*, **12**(1):56–68, 2011. 91

[106] Nitesh V Chawla and Darcy A Davis. **Bringing big data to personalized healthcare: a patient-centered framework**. *Journal of general internal medicine*, **28**(3):660–665, 2013. 93

[107] Kwang-Il Goh, Michael E Cusick, David Valle, Barton Childs, Marc Vidal, and Albert-Laszlo Barabasi. **The human disease network**. *Proceedings of the National Academy of Sciences*, **104**(21):8685–8690, 2007. 95

[108] D-S Lee, J Park, KA Kay, NA Christakis, ZN Oltvai, and A-L Barabási. **The implications of human metabolic network topology for disease comorbidity**. *Proceedings of the National Academy of Sciences*, **105**(29):9880–9885, 2008. 95

[109] César A Hidalgo, Nicholas Blumm, Albert-László Barabási, and Nicholas A Christakis. **A dynamic network approach for the study of human phenotypes**. *PLoS Comput Biol*, **5**(4):e1000353, 2009. 95

[110] Alessandro Di Stefano, Aurelio La Corte, and Marialisa Scatà. **A Novel Multi-agent Social Multilayer Framework for Improving Health Information Exchange and Management**. In *International School and Conference on Network Science (NetSci 2015)*, 2015. 95

[111] Marialisa Scatà, Alessandro Di Stefano, Evelina Giacchi, Aurelio La Corte, and Pietro Liò. **The Bio-Inspired and Social Evolution of Node and Data in a Multilayer Network**. In *International Conference on Data Communication Networking (DCNET 2014)*, 2014. 98, 99

[112] Ludwig Johann Neumann and Oskar Morgenstern. *Theory of games and economic behavior*. Princeton University Press Princeton, NJ, 1947. 105

[113] Eric Van Damme and Jörgen W Weibull. **Evolution in games with endogenous mistake probabilities**. *Journal of Economic Theory*, **106**(2):296–315, 2002. 112

[114] Peter D Taylor and Leo B Jonker. **Evolutionary stable strategies and game dynamics**. *Mathematical biosciences*, **40**(1):145–156, 1978. 114

[115] Carlos P Roca, José A Cuesta, and Angel Sánchez. **Evolutionary game theory: Temporal and spatial effects beyond replicator dynamics**. *Physics of life reviews*, **6**(4):208–249, 2009. 115, 121

[116] David B Fogel, Gary B Fogel, and Peter C Andrews. **On the instability of evolutionary stable strategies**. *Biosystems*, **44**(2):135–152, 1997. 116

[117] Karen M Page and Martin A Nowak. **Unifying evolutionary dynamics**. *Journal of theoretical biology*, **219**(1):93–98, 2002. 116

[118] Matjaž Perc and Paolo Grigolini. **Collective behavior and evolutionary games–An introduction**. *Chaos, Solitons & Fractals*, **56**:1–5, 2013. 118, 133, 134

[119] William H Press and Freeman J Dyson. **Iterated Prisoners Dilemma contains strategies that dominate any evolutionary opponent**. *Proceedings of the National Academy of Sciences*, **109**(26):10409–10413, 2012. 118, 133

[120] Elizabeth Pennisi. **On the Origin of Cooperation**. *Science*, **325**(5945):1196–1199, 2009. 120

[121] David G Rand and Martin A Nowak. **Human cooperation**. *Trends in cognitive sciences*, **17**(8):413–425, 2013. 110, 120, 121

[122] Milena Tsvetkova, Michael W Macy, and Attila Szolnoki. **The social contagion of generosity**. *PloS one*, **9**(2):e87275, 2014. 120

[123] Keke Huang, Xiaoping Zheng, Yeqing Yang, and Tao Wang. **Behavioral evolution in evacuation crowd based on heterogeneous rationality of small groups**. *Applied Mathematics and Computation*, **266**:501–506, 2015. 120

[124] Keke Huang, Yuan Cheng, Xiaoping Zheng, and Yeqing Yang. **Cooperative behavior evolution of small groups on interconnected networks**. *Chaos, Solitons & Fractals*, **80**:90–95, 2015. 120

[125] Cheng-Yi Xia, Sandro Meloni, Matjaž Perc, and Yamir Moreno. **Dynamic instability of cooperation due to diverse activity patterns in evolutionary social dilemmas**. *EPL (Europhysics Letters)*, **109**(5):58002, 2015. 120

[126] James H Fowler and Nicholas A Christakis. **Cooperative behavior cascades in human social networks**. *Proceedings of the National Academy of Sciences*, **107**(12):5334–5338, 2010. 121

[127] Zhen Wang, Michael A Andrews, Zhi-Xi Wu, Lin Wang, and Chris T Bauch. **Coupled disease–behavior dynamics on complex networks: A review**. *Physics of life reviews*, 2015. 121

[128] Joaquín Sanz, Cheng-Yi Xia, Sandro Meloni, and Yamir Moreno. **Dynamics of interacting diseases**. *Physical Review X*, **4**(4):041005, 2014. 121

[129] Alessandro Vespignani et al. **Predicting the behavior of techno-social systems**. *Science*, **325**(5939):425, 2009. 121

[130] Damon M Centola. **Homophily, networks, and critical mass: Solving the start-up problem in large group collective action**. *Rationality and Society*, **25**(1):3–40, 2013. 121, 122, 123, 125

[131] Mark Granovetter. **Threshold models of collective behavior**. *American journal of sociology*, pages 1420–1443, 1978. 121

[132] Sandra González-Bailón, Javier Borge-Holthoefer, and Yamir Moreno. **Broadcasters and hidden influentials in online protest diffusion**. *American Behavioral Scientist*, page 0002764213479371, 2013. 122

[133] Miller McPherson, Lynn Smith-Lovin, and James M Cook. **Birds of a feather: Homophily in social networks**. *Annual review of sociology*, pages 415–444, 2001. 122, 125

[134] Aris Anagnostopoulos, Alessandro Bessi, Guido Caldarelli, Michela Del Vicario, Fabio Petroni, Antonio Scala, Fabiana Zollo, and Walter Quattrociocchi. **Viral misinformation: the role of homophily and polarization**. *arXiv preprint arXiv:1411.2893*, 2014. 122

[135] Amos Tversky. **Features of similarity**. *Psychological Review*, **84**:327–352, 1977. 122

[136] Damon Centola, Juan Carlos Gonzalez-Avella, Victor M Eguiluz, and Maxi San Miguel. **Homophily, cultural drift, and the co-evolution of cultural groups**. *Journal of Conflict Resolution*, **51**(6):905–929, 2007. 122

[137] Michael T Hannan and John Freeman. **The population ecology of organizations**. *American journal of sociology*, pages 929–964, 1977. 122

[138] Michael E Porter. *Competitive strategy: Techniques for analyzing industries and competitors*. Simon and Schuster, 2008. 122

[139] Feng Fu, Martin A Nowak, Nicholas A Christakis, and James H Fowler. **The evolution of homophily**. *Scientific reports*, **2**, 2012. 123

[140] Ginestra Bianconi, Sergey N Dorogovtsev, and José FF Mendes. **Mutually connected component of networks of networks with replica nodes**. *Physical Review E*, **91**(1):012804, 2015. 123

[141] Cheng-Yi Xia, Xiao-Kun Meng, and Zhen Wang. **Heterogeneous Coupling between Interdependent Lattices Promotes the Cooperation in the Prisoners Dilemma Game**. *PloS one*, **10**(6):e0129542, 2015. 123

[142] Ernesto Estrada and Jesús Gómez-Gardeñes. **Communicability reveals a transition to coordinated behavior in multiplex networks**. *Physical Review E*, **89**(4):042819, 2014. 125, 131, 132, 141

[143] Phillip Bonacich and Paulette Lloyd. **Eigenvector-like measures of centrality for asymmetric relations**. *Social Networks*, **23**(3):191–201, 2001. 126

[144] György Szabó and Csaba Tőke. **Evolutionary prisoners dilemma game on a square lattice**. *Physical Review E*, **58**(1):69, 1998. 134