On the Behavior of Rational Agents in Complex Networks

Pedro O.S. Vaz de Melo[1], **Antonio A.F. Loureiro**[1] **(advisor),**
Virgilio A.F. Almeida[1] **(co-advisor), Christos Faloutsos**[2] **(co-advisor)**

[1]Computer Science Department, Universidade Federal de Minas Gerais, Brazil
[2]Computer Science Department, Carnegie Mellon University, USA

{olmo,loureiro,virgilio}@dcc.ufmg.br, christos@cs.cmu.edu

1. Motivation

The constant advancement of information systems allows more data to be generated and stored. The Internet, for example, consists of millions of computing devices, each one of them generates, stores and transmits countless data. Clearly, the Internet is going in the direction of the ubiquitous computing paradigm, which, as envisioned in Mark Weiser's classic paper [Weiser 1999], allows any person, anywhere, at any time to interact with the environment. An important aspect to achieve this vision is the Wireless Sensor Network (WSN) that is a special type of ad hoc network, designed to collect data from the environment and provide such information to the final user.

Clearly, there will be an increase in the amount of data collected from the most diverse scenarios. Currently, there are studies on data from phone calls, online social networks, railroads, Internet websites, citation networks, movies and actors, sports leagues and many others. From these studies we now know how people link websites in their homepages and how communities of people evolve over time. It is fascinating that behind the names and numbers registered in all these data, we see the reflection of the environment itself, i.e, there is a decision made by some entity. Therefore, the knowledge of how to process this invaluable evolving dataset can lead to a better understanding of the interests and dynamics of each entity in a determined system, community or in the society.

In this work, we focus on systems that are made up of entities capable of interacting among themselves in a autonomous way, reflecting their interests and activity dynamics. We call these systems *Decision-based Complex Networks (DBCNs)* comprised of entities, nodes or agents. A DBCN is a special type of complex network [Newman 2003] that contains nodes capable of making autonomous decisions, which are guided mainly by their personal motivations. For instance, social networks formed from friendship ties or work collaborations are DBCNs, since nodes of these networks have decision power to create edges. However, semantic networks are not DBCNs, since the edges creation is guided, in these cases, by a central process.

As a complex network, DBCNs have a large number of vertices and edges that exhibit a pattern, such as communities or highly connected vertices, called hubs. While in a simple network with at most hundreds of nodes the human eye is a tool of considerable power, in a complex network, this approach is useless. Thus, to study, analyze and characterize complex networks, fast statistical methods and algorithms are necessary.

2. Objectives

This thesis aims at analyzing the behavior of agents of DBCNs. We observe real and hypothetical scenarios where environmental and social decisions play an important role on the network evolution and on the way the agents interact. When we fully understand the motivations behind the actions of the agents, we are able to devise techniques to model, predict and control the behavior of diverse and large DBCNs and of their agents. Figure 1 shows the main objectives of the thesis:

1. **Modeling.** Our goal is to design models that can accurately represent the behavior of agents in a DBCN, leading to a better understanding of the system;
2. **Predicting.** Upon understanding the reasons behind the decisions of agents, we want to understand how the system will evolve;
3. **Controlling.** Once we know how the system will behave, we design control mechanisms to make the agents to act according to a determined goal.

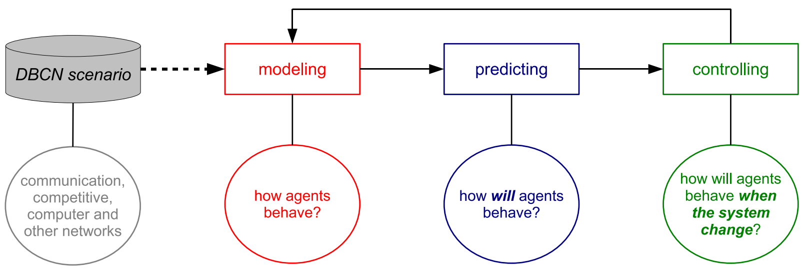

Figure 1. Thesis summary. Given a DBCN scenario, we model the behavior of agents and, from this, we predict their behavior. After that, we design control mechanisms so agents can act according to a determined goal and, thus, change the system. New modeling and predicting efforts should be made in the new system to better control its behavior.

Based on local decisions made by agents, we investigate the global evolution of the system to tackle a vast and diverse number of applications. In the thesis, we study significantly different types of network systems and propose different applications for all of them, based on the knowledge of the behavior of their agents. Thus, all the analyzed networks have in common the fact that agents are capable of: (i) making autonomous decisions, and (ii) interacting with other agents, together with the fact that their evolution significantly depends on how these decisions are made.

3. Contributions

First, we present a compact analysis of all three aspects of DBCNs we tackle in the thesis (modeling, prediction and control) using three real-world datasets of user mobility activity (Chapter 3). Next, we focus on modeling communication networks (Chapter 4). We propose models for the individual behavior of users and, from this, we propose applications for data summarization, anomaly detection and network monitoring. Next, we focus on predicting competitive networks of sports leagues such as the National Basketball Association and the Major League Baseball (Chapter 5). We propose a prediction model to identify likely teams to win/fail in a following season. Finally, we focus on controlling, investigating decision making in a scenario where WSNs are deployed at the same region and interact among themselves to request or share computational resources (Chapter 6). Using game theory, we model the cooperation problem, and present a control application that is a protocol to allow cooperation.

3.1. Mobility Networks

The main characteristic of DBCNs is that the interactions between their entities are, usually, a consequence of semi-rational decisions. We say “usually” and “semi-rational” decisions because any system is subject to random events and irrational choices. Most interactions in DBCNs arise from conscious decisions made by their entities and their evolution is significantly different from the evolution of random networks [Erdős and Rényi 1960]. Thus, while in DBCNs the edges are created from semi-rational decisions, which tend to be regular and to repeat themselves, in a *random network edges are created independently of the attributes of the nodes, i.e., the probability of connecting two nodes is constant*.

Based on this point, we propose a *Random rElationship ClASsifier sTrategy* (RECAST) to classify relationships among users to spot random events in DBCNs. RECAST examines how a real system evolves if users' decisions were random. We use the temporal graph originated from the real network dataset and its random counterpart to tell apart edges representing random events from those created by actual social relationships, such as friendship or professional interactions. By comparing both graphs in terms of metrics reflecting a pair of major social features, i.e., frequent user encounters and shared acquaintances, RECAST provides a simple yet very effective way of classifying contacts.

RECAST's classification allows us to observe many differences in the evolution of relationships when applied to three real-world datasets describing user mobility activity in city and campus networking scenarios. In fact, diversities are due to the intrinsic features of each networking scenario of interest. For instance, we show that the dataset describing the movement of cab drivers in San Francisco (USA) [Piorkowski et al. 2011, Piorkowski et al. 2009] has mostly non-social properties, which makes its graph representation similar to a random network. The same is not true for people moving in a campus. However, different campuses yield dissimilar interaction dynamics as well.

In summary, we **modeled** mobility traces into a DBCN of encounters and, from it, we proposed the RECAST. The clear classification of user relationships provided by RECAST immediately leveraged subsequent analyses. We showed that encounters between two individuals who share a social relationship are easier to **predict**. Moreover, we showed how it is possible to design a **control** application to efficiently disseminate data throughout the network using only two out of the four possible edge classes.

3.2. Communication Networks

As we show in the thesis, modeling agents of this type of network is extremely challenging. Human communication patterns are likely to change as the technological and cultural aspects of society change. For instance, the typical duration of a phone call involving two fixed phones is probably different from the one involving two mobile phones. Therefore, we use this challenging scenario in the study of modeling in DBCNs.

We analyze the rate in which two social agents communicate and the duration of their communication. First, we analyze the size of the communication flows (duration of phone calls), and we show how a good modeling effort can lead to a wide variety of applications. In summary, we tackle the following problem: given a large amount of phone records, what is the best way to summarize the calling behavior of a user? We analyze the duration of hundreds of million calls and we propose the *Truncated Lazy Contractor* (TLAC) model to describe how long are the durations of phone calls of a single user. Thus, the TLAC models the Calls Duration Distribution (CDD) of a user and is parsimonious, having only two parameters, the *efficiency* coefficient ρ and the *weakness* coefficient β. We show that the TLAC model was the best alternative to model the CDD of users of our dataset, mainly because it has a heavier tail and head than the log-normal distribution, which is the most commonly used distribution to model CDDs [Guo et al. 2007].

We also suggest the use of TLAC parameters as a better way to summarize the calls duration behavior of a user. We propose the *MetaDist* to model the population of users who have a determined calls duration behavior. The *MetaDist* is the meta-distribution of the ρ_i and β_i parameters of each user i's CDD and, when its isocontours are visualized, its shape is surprisingly similar to a bivariate Gaussian distribution. This fascinating regularity, observed in a significantly noisy data, makes the *MetaDist* a potential distribution to be explored in the direction of better understanding the call behavior of mobile users.

In the analysis of the inter-event times between communications, we considered several scenarios. Current literature has seemingly contradictory results for the marginal distribution of the inter-event times. Some studies claim good fits with power laws; others with "non-homogeneous" Poisson processes. We did an elaborate modeling that uses the knowledge of these models together with new observations in real data to propose the Self-Feeding Process (SFP) that matches very well the properties of marginals of several large, diverse and real communication datasets (phone calls, SMSs, e-mails, and online forum). More important, it unifies existing theories on human communication dynamics, exhibits power law tail behavior, burstiness, as well as locally-Poisson behavior. It is based on a novel discovery reported in the thesis that there is a strong, positive correlation between successive inter-event times. Finally, it is extremely parsimonious, using at most two parameters.

3.3. Competitive Networks

In competitive networks, agents of the system compete among themselves for a reward or limited resources. Some examples are networks of workers looking for job positions, e.g. LinkedIn, and in professional sports leagues. In these cases, the organizations, companies or teams want to sign up the best agents, workers or players, for the lowest possible cost and, by their turn, these agents want to receive the maximum possible salary. Moreover, either organizations or agents want the organizations to grow, that is, expand in the market or win titles in their leagues. This scenario presents several conflicts of interests that may unveil interesting observations over the social agents of this type of network.

In the thesis, we analyze the network formed from the teams and players of sports leagues. We start with the NBA and its 63 first years of existence. In the modeling stage, we view the NBA as a complex network in evolution. Then, in the prediction stage, we develop metrics that are correlated with the behavior of NBA teams, taking into account only the social and work relationship among players, coaches and teams. Then, based on these metrics, we propose models to predict how well a team will perform in the following season. We also evaluate the prediction models on the Major League Baseball dataset and the NetForY, one of our proposed models, performs surprisingly well on both datasets.

3.4. Wireless Sensor Networks

Finally, we analyze DBCNs formed only by computers. In this type of network, agents can be modeled as social nodes with decision capabilities [Vaz de Melo et al. 2011b]. The main aspect in this network is that in computer networks agents do not make irrational decisions due to the “trembling hand” effect [Fudenberg and Tirole 1991]. This means that decisions performed by agents are to maximize their utility and nothing else, acting in a pure selfish way, and irrational decisions made by jitter or similar causes do not exist.

We consider a scenario where distinct WSNs with different owners are deployed at the same place. In this scenario, there is the possibility that a sensor of a network cooperates with a sensor of another network. When two WSNs share their sensor nodes in the execution of some activity in a intelligent way, both networks may improve their operabilities and perform their activities in a more efficient way. Despite being obvious and simple, this idea brings with it many implications that hinder cooperation between networks. Whereas a WSN has a rational and selfish character, it will only cooperate with another WSN if it provides services that justify the cooperation.

In this specific chapter, we again tackle the three aspects of DBCNs we analyze in the thesis: modeling, prediction and control. First, we model the problem of cooperation among different WSNs using game theory that is an interesting technique to model conflicting situations among two or more rational and selfish agents [Fudenberg and Tirole 1991]. In computer networks, decisions and metrics of utility (e.g., throughput and latency) are computationally well defined, and, thus, game theory can be a powerful tool to model the behavior of social agents, which can be predicted in advance. From this knowledge, we can design incentive mechanisms to control their behaviors.

4. Discussion

In the thesis, we analyzed the behavior of rational agents in different Decision-based Complex Networks (DBCNs). We showed that the decisions made by agents of DBCNs have a crucial role in the evolution of the network. We also presented techniques to model, predict and control the evolution of such networks. Based on the analysis of the mobility networks, we published a paper in SBRC’12 [Vaz de Melo et al. 2012a]. Based on the analysis in communication networks, we published a paper in PKDD’10 [Vaz de Melo et al. 2010] and in SDM’11 [Vaz de Melo et al. 2011a]. Based on the analysis in competitive networks, we published a paper in KDD’08 [Vaz de Melo et al. 2008b] and in TKDD [Vaz de Melo et al. 2012b]. Finally, based on the analysis in WSNs, we published a book chapter [Vaz de Melo et al. 2011b], a paper in MSWiM [Vaz de Melo et al. 2008a] and two papers in SBRC [Vaz de Melo et al. 2008c, Vaz de Melo et al. 2009]. Journal versions covering the RECAST and the SFP model are being prepared.

It is important to point out the broad reach of the research performed in the thesis, covering fields apparently disconnected such as data mining, WSNs, sports analytics, human mobility dynamics, among others.

References

Erdős, P. and Rényi, A. (1960). On the Evolution of Random Graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 7:17.

Fudenberg, D. and Tirole, J. (1991). *Game Theory*. MIT Press.

Guo, J., Liu, F., and Zhu, Z. (2007). Estimate the Call Duration Distribution Parameters in GSM System Based on K-L Divergence Method. In *Int'l Conf. on Wireless Communications, Networking and Mobile Computing*, pages 2988–2991.

Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review*, 45(2):167–256.

Piorkowski, M., S.-Djukic, N., and Grossglauser, M. (2009). A Parsimonious Model of Mobile Partitioned Networks with Clustering. In *1st Int'l Conf. on Communication Systems and Networks*, pages 1–10.

Piorkowski, M., S.-Djukic, N., and Grossglauser, M. (2011). CRAWDAD data set epfl/mobility (v. 2009–02–24). `http://crawdad.cs.dartmouth.edu/epfl/mobility`.

Vaz de Melo, P. O., da Cunha, F. D., Almeida, J. M., Loureiro, A. A., and Mini, R. A. (2008a). The Problem of Cooperation Among Different Wireless Sensor Networks. In *11th ACM Int'l Symp. on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 86–91.

Vaz de Melo, P. O., Loureiro, A. A. F., do V. Machado, M., and Mini, R. A. F. (2009). Um Protocolo Baseado em Teoria dos Jogos para a Cooperação entre Diferentes Redes de Sensores Sem Fio. In *XXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 1–14.

Vaz de Melo, P. O., Viana, A. C., Fiore, M., Jaffres-Runser, K., Mouel, F., and Loureiro, A. A. (2012a). Usando Redes Aleatórias na Anlise de Mobilidade. In *XXX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuıidos*, pages 1–14.

Vaz de Melo, P. O. S., Akoglu, L., Faloutsos, C., and Loureiro, A. A. F. (2010). Surprising Patterns for the Call Duration Distribution of Mobile Phone Users. In *ECML/PKDD*, pages 354–369.

Vaz de Melo, P. O. S., Almeida, V. A., Loureiro, A. A., and Faloutsos, C. (2012b). Forecasting in the NBA and Other Team Sports: Network Effects in Action. *ACM Transactions on Knowledge Discovery and Data Mining (in print)*.

Vaz de Melo, P. O. S., Almeida, V. A. F., and Loureiro, A. A. F. (2008b). Can Complex Network Metrics Predict the Behavior of NBA Teams? In *14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, pages 695–703.

Vaz de Melo, P. O. S., Cunha, F. D., Almeida, J. M., Mini, R. A., and Loureiro, A. A. (2008c). O Problema da Cooperação entre Redes de Sensores Sem Fio. In *XXVI Simpósio Brasileiro de Redes de Computadores*, pages 1–14.

Vaz de Melo, P. O. S., Faloutsos, C., and Loureiro, A. A. (2011a). Human Dynamics in Large Communication Networks. In *11th SIAM Int'l Conf. on Data Mining*, pages 968–979.

Vaz de Melo, P. O. S., Fernandes, C., Mini, R. A. F., Loureiro, A. A. F., and Almeida, V. A. F. (2011b). *Game Theory for Wireless Communications and Networking*, chapter Game Theory in Wireless Sensor Networks. Routledge, Taylor & Francis Group.

Weiser, M. (1999). The Computer for the 21st Century. *SIGMOBILE Commun. Rev.*, 3(3):3–11.