

Analysis of Dynamic Social and Technological Networks

Edited by

Vito Latora¹, Cecilia Mascolo², and Mirco Musolesi³

1 Università di Catania, IT, latora@ct.infn.it

2 University of Cambridge, GB, cecilia.mascolo@cl.cam.ac.uk

3 University of Birmingham, GB, m.musolesi@cs.bham.ac.uk

Abstract

With the growing popularity of online communication tools, researchers have turned their attention to the study of the networks arising between users of social networking services, between mobile phone callers and, in general, between individuals connected by technological means. Thanks to the rich set of techniques and methods developed by complex network science, and joining forces with sociologists and psychologists, the analysis of dynamic social and technological networks has sparked many important results, attracting even more interest as the importance of such systems grows over time.

This Dagstuhl seminar brought together researchers and practitioners from computer science, physics and psychology, covering the diverse areas of social and technological network analysis. The goal of the seminar was to bring together people from different areas of expertise, focusing on both mathematical aspects and practical applications of theoretical models and techniques. In particular, the evolution of this research field and of its future perspectives was a major theme of the seminar. This seminar was attended by 25 participants.

Seminar 8.–11. November, 2011 – www.dagstuhl.de/11452

1998 ACM Subject Classification H.2.8 Data Mining.

Keywords and phrases complex networks, network analysis, network data mining

Digital Object Identifier 10.4230/DagRep.1.11.39

Edited in cooperation with Salvatore Scellato

1 Executive Summary

Vito Latora

Cecilia Mascolo

Mirco Musolesi

License  Creative Commons BY-NC-ND 3.0 Unported license
© Vito Latora, Cecilia Mascolo, Mirco Musolesi

In the recent years, we have witnessed an increasing interest in the analysis of complex networks, i.e., networks composed of many interacting entities that show emergent behavior at global level. Usually, the key features of these networks can be captured by means of a statistical characterisation of their properties at local level such as their degree distribution and clustering coefficient. Models can then be built in order to describe the system in its entirety and to study the processes taking place on it, also over time. These models can be used to understand several phenomena dependent on the structure and dynamics of these networked systems, such as the spreading of computer viruses.

One of the main motivations of the explosion of interest in social networks that we are seeing is the availability of large data sets, e.g., the “snapshots” of the Internet structure or



Except where otherwise noted, content of this report is licensed under a Creative Commons BY-NC-ND 3.0 Unported license

Analysis of Dynamic Social and Technological Networks, *Dagstuhl Reports*, Vol. 1, Issue 11, pp. 39–49

Editors: Vito Latora, Cecilia Mascolo, and Mirco Musolesi



DAGSTUHL
REPORTS

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

the maps of online social networks obtained by crawling extremely popular Web sites like Facebook and Twitter. Moreover, many data collections exercises of data sets related to human interactions by means of Bluetooth radio or GPS receivers have been carried in the recent years. These large data sets provide information that is not limited to a particular instant of time, but cover a very large time interval as well as fine grained space information. These data sets can be used to study the evolution of the network and dynamic processes happening over time such as simulated epidemics. Other large data sets that have attracted considerable interest include biological, commodity and economic networks. The initial research efforts have been focussed on the analysis of the static properties of these networks, including the presence of clusters, hubs, and community structure. More recently, researchers became interested in studying dynamic processes taking place on these networks such as information diffusion in the Internet, disease epidemics and malware propagation.

The goal of this seminar was broad, including both mathematical aspects and practical applications of theoretical models and techniques. The seminar focused on two key classes of networks that are of fundamental importance not only in computer science but also in the everyday life of millions of people, namely technological networks and online social networks.

The main contributions of this seminar were:

- Presenting a wide range of recent research results on the dynamics of processes and structure of technological and social networks.
- Exchanging solutions and practices in the different areas of computer science and other disciplines in order to find novel solutions and start fruitful long-term collaborations among the seminar attendants.
- Exploring the new challenges and opportunities arising from the analysis of data from mobile devices and social network tools, which offer the change to collect very rich data sets of information about the everyday life of people including their movements, their contacts and their social network.
- Discussing how the social networks extracted from mobile device interactions are time and location dependent, requiring new models and techniques to study them.
- Examining the application of machine learning and data mining techniques to the analysis of technological and social networks, bringing together researchers and practitioners working on the massive available networking data sets and machine learning experts interested in real-world problems.
- Studying and considering the computational challenges presented by the scale of these data sets, which impose the design of novel algorithms and the rethinking of existing techniques.
- Discussing the ethical problems arising from the treatment of privacy-sensitive user data and potential technical and legal solutions to overcome them.

2 Table of Contents

Executive Summary

Vito Latora, Cecilia Mascolo, Mirco Musolesi 39

Overview of Talks

Entropy of dynamical social networks

Ginestra Bianconi 42

Bistability Through Triad Closure

Des Higham 42

The network position of MRSA risk wards in a hospital system

Petter Holme 42

Suppression and explosive phase transitions

Byungnam Kahng 43

Components in time-varying graphs

Vincenzo Nicosia 43

A tale of many cities: universal patterns in human urban mobility

Anastasios Noulas 44

Personality and Language in Social Media

Daniele Quercia 44

Geography of Personality

Jason Rentfrow 45

Distance Matters: Socio-spatial Properties of Online Social Networks

Salvatore Scellato 45

Understanding mobility in a social petri dish

Roberta Sinatra 46

Temporal metrics and applications in real networks

John Tang 46

Working Groups 47


Panel Discussion 47

Participants 49

3 Overview of Talks

3.1 Entropy of dynamical social networks

Ginestra Bianconi (Northeastern Univ. – Boston, US)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Ginestra Bianconi

Main reference K. Zhao, M. Karsai and G. Bianconi, Entropy of dynamical social networks, PlosOne (in press)


Human dynamical social networks encode information and are highly adaptive.

To characterize the information encoded in the fast dynamics of social interactions, we introduce the entropy of dynamical social networks. By analysing a large dataset of phone-call interactions we show evidence that the dynamical social network has an entropy that depends on the time of the day in a typical week-day. Moreover we show evidence for adaptability of human social behavior showing data on duration of phone-call interactions that significantly deviates from the statistics of duration of face-to-face interactions.

This adaptability of behavior corresponds to a different information content of the dynamics of social human interactions. We quantify this information by the use of the entropy of dynamical networks on realistic models of social interactions.

3.2 Bistability Through Triad Closure

Des Higham (The University of Strathclyde – Glasgow, GB)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Des Higham

Joint work of Higham, Des; Grindrod, Peter; Parsons, Mark;


Main reference P. Grindrod P, D.J. Higham, M.C. Parsons, “Bistable evolving networks,” University of Strathclyde, Maths and Stats Report #6, 2011.

URL <http://www.mathstat.strath.ac.uk/downloads/publications/6techbis-11t.pdf>

I will describe a new dynamic network model that incorporates the concept of triad closure. This feature, where new friendships are more likely to emerge between those who have more current friends in common, has been observed empirically in the social sciences. Using a discrete time Markov chain setting, mean field analysis reveals that two different types of long term behavior are possible—the system may evolve either into a sparse or a highly triangulated regime, depending on the initial data and the microscale details. Computer simulations confirm this bistability phenomenon.

3.3 The network position of MRSA risk wards in a hospital system

Petter Holme (University of Umeå, SE)





License  Creative Commons BY-NC-ND 3.0 Unported license
© Petter Holme

We study a dataset of patient flow in a regional Swedish hospital system. We can follow how about 300,000 patients move between 8507 wards over about ten years. The dataset also reveals when a patient test positive with Methicillin Resistant Staphylococcus aureus (MRSA).

To simplify the complex flow of patients, we represent it as a network of wards where two wards are connected if they a patient moves from one ward to the other without visiting a third ward in between. From this network we characterize the typical network position of wards with a high prevalence of MRSA, and the how the patient's location in the network changes upon testing positive with MRSA. We note that the many types of network centrality measures are positively, albeit weakly, correlated with the average prevalence of MRSA at a ward. On average, wards with medium values of these centrality measures have the highest average prevalence. We see a weak effect of the hospital system's response to the patient testing positive—after testing positive the patient move to wards with lower degree (number of links to other wards) and longer average durations of stay. One can use ward networks to aid the discovery of potential hot spots for MRSA epidemics. However, this cannot be done very effectively because even though the hospital system is administratively organized in a hierarchical way, the patient flow is too random to make the network structure a strong predictor of MRSA epidemics.

3.4 Suppression and explosive phase transitions

Byungnam Kahng (Seoul National University, KR)





License     Creative Commons BY-NC-ND 3.0 Unported license
© Byungnam Kahng

When dynamics of networks proceeds under the suppression of formation of a given target pattern, the pattern can be formed eventually at a certain transition point. This formation can occur drastically, and exhibits a discontinuous (explosive) phase transition. I show some examples of such explosive transitions in percolation transition, data packet transition and epidemic spreading models.

References

3.5 Components in time-varying graphs

Vincenzo Nicosia (Università di Catania, IT)

License     Creative Commons BY-NC-ND 3.0 Unported license
© Vincenzo Nicosia

Joint work of Nicosia, Vincenzo; Tang, John; Musolesi, Mirco; Russo, Giovanni; Mascolo, Cecilia; Latora, Vito
Main reference V. Nicosia, J. Tang, M. Musolesi, G. Russo, C. Mascolo, V. Latora, "Components in time-varying graphs," under review on Physical Review E, arXiv:1106.2134v2 [physics.soc-ph]

URL <http://arxiv.org/abs/1106.2134v2>


Real complex systems are inherently time-varying. Thanks to new communication systems and novel technologies, it is today possible to produce and analyze social and biological networks with detailed information on the time of occurrence and duration of each link.

However, standard graph metrics introduced so far in complex network theory are mainly suited for static graphs, i.e., graphs in which the links do not change over time, or graphs built from time-varying systems by aggregating all the links as if they were concurrent in time. In this paper, we extend the notion of connectedness, and the definitions of node and graph components, to the case of *time-varying graphs*, which are represented as time-ordered sequences of graphs defined over a fixed set of nodes. We show that the problem of finding strongly connected components in a time-varying graph can be mapped into the problem of discovering the maximal-cliques in an opportunely constructed static graph, which we name the *affine graph*. It is therefore a NP-complete problem. As a practical example, we

have performed a temporal component analysis of time-varying graphs constructed from three data sets of human interactions. The results show that taking time into account in the definition of graph components allows to capture important features of real systems. In particular, we observe a large variability in the size of node temporal in- and out-components. This is due to intrinsic fluctuations in the activity patterns of individuals, which cannot be detected by static graph analysis.

3.6 A tale of many cities: universal patterns in human urban mobility

Anastasios Noulas (University of Cambridge, GB)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Anastasios Noulas

Main reference A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, C. Mascolo, “A tale of many cities: universal patterns in human urban mobility,” arXiv:1108.5355v4 [physics.soc-ph]

URL <http://arxiv.org/abs/1108.5355v4>


The advent of geographic online social networks such as Foursquare, where users voluntarily signal their current location, opens the door to powerful studies on human movement. In particular the fine granularity of the location data, with GPS accuracy down to 10 meters, and the worldwide scale of Foursquare adoption are unprecedented. In this paper we study urban mobility patterns of people in several metropolitan cities around the globe by analyzing a large set of Foursquare users. Surprisingly, while there are variations in human movement in different cities, our analysis shows that those are predominantly due to different distributions of places across different urban environments.

Moreover, a universal law for human mobility is identified, which isolates as a key component the rank-distance, factoring in the number of places between origin and destination, rather than pure physical distance, as considered in some previous works. Building on our findings, we also show how a rank-based movement model accurately captures real human movements in different cities.

Our results shed new light on the driving factors of urban human mobility, with potential applications for urban planning, location-based advertisement and even social studies.

3.7 Personality and Language in Social Media

Daniele Quercia (University of Cambridge, GB)

License  Creative Commons BY-NC-ND 3.0 Unported license
© Daniele Quercia

In Twitter, we tested whether users can be reduced to look-alike nodes (as most of the spreading models would assume) or, instead, whether they show individual differences that impact their popularity and influence. Again, one aspect that may differentiate users is their character and personality. For 335 users, we gather personality data, analyze it, and find that both popular users and influentials are extroverts and emotionally stable (low in the trait of Neuroticism). Also, since it has been shown that personality is linked to the use of language (which is unobtrusively observable in tweets), we carry out a study of tweets and show that popular and influential users linguistically structure their tweets in specific ways.

3.8 Geography of Personality

Jason Rentfrow (University of Cambridge, GB)

License © © ⊖ Creative Commons BY-NC-ND 3.0 Unported license
© Jason Rentfrow

Everybody knows that New Yorkers are outspoken, neurotic, and always in a hurry, and that Texans are slow-talking, friendly, and proud members of the National Rifle Association. Obviously such characterizations are nothing more than stereotypes, but they raise the question of whether there are psychological differences across the US. Recent research suggests there are statewide differences in personality and that those differences are linked to a host of important social indicators. However, that work was based on one sample so the robustness of the findings is unclear. Using data from over 1.5 million respondents from five independent samples, I examined the reliability and validity of state-level personality. Analyses of the convergent validity of the state-level personality scores revealed a considerable level of convergence for each of the Big Five personality domains, with Conscientiousness displaying the least and Openness displaying the most convergence across samples. Consistent patterns of relationships across samples were observed between the state-level personality domains and conceptually relevant social indicators. For example, state-level Agreeableness was negatively related to rates of violent crime and positively related to community involvement; state-level Neuroticism was negatively related to psychological well-being and positively related to rates of cancer and mental illness; and state-level Openness was negatively related to votes cast for conservative politicians and positively related to markers of cultural diversity. Overall, these findings indicate that state-level personality is a robust and stable construct.

3.9 Distance Matters: Socio-spatial Properties of Online Social Networks

Salvatore Scellato (University of Cambridge, GB)

License © © ⊖ Creative Commons BY-NC-ND 3.0 Unported license
© Salvatore Scellato


Joint work of Scellato, Salvatore; Noulas, Anastasios; Lambiotte, Renaud; Mascolo, Cecilia;
Main reference S. Scellato, A. Noulas, R. Lambiotte, C. Mascolo, "Socio-spatial Properties of Online Location-based Social Networks," in Proc. of Fifth Int'l AAAI Conference on Weblogs and Social Media (ICWSM'11), 2011.
URL <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2751>

The spatial structure of large-scale online social networks has been largely inaccessible due to the lack of available and accurate data about people's location. However, with the recent surging popularity of location-based social services, data about the geographic position of users have been available for the first time, together with information on their online social connections.

In this work we present a comprehensive study of the spatial properties of the social networks arising among users of three main popular online location-based services. We observe robust universal features across them: while all networks exhibit about 40% of links below 100 km, we further discover strong heterogeneity across users, with different characteristic spatial lengths of interaction across both their social ties and social triads. Our results constitute the first large-scale study to unravel the socio-spatial properties of online location-based social networks.

3.10 Understanding mobility in a social petri dish

Roberta Sinatra (Università di Catania, IT)


License  Creative Commons BY-NC-ND 3.0 Unported license
© Roberta Sinatra

Joint work of Szell, Michael; Sinatra, Roberta; Petri, Giovanni; Thurner, Stefan; Latora, Vito

Despite the recent availability of large data sets on human movements, a full understanding of the rules governing motion within social systems is still missing, due to incomplete information on the socio-economic factors and to often limited spatio-temporal resolutions. Here we study an entire society of individuals, the players of an online-game, with complete information on their movements in a network-shaped universe and on their social and economic interactions. Such a “socio-economic laboratory” allows to unveil the intricate interplay of spatial constraints, social and economic factors, and patterns of mobility. We find that the motion of individuals is not only constrained by physical distances, but also strongly shaped by the presence of socio-economic areas. These regions can be recovered perfectly by community detection methods solely based on the measured human dynamics. Moreover, we uncover that long-term memory in the time-order of visited locations is the essential ingredient for modeling the trajectories.

3.11 Temporal metrics and applications in real networks

John Tang (University of Cambridge, GB)

License  Creative Commons BY-NC-ND 3.0 Unported license
© John Tang

Joint work of Tang, John; Mascolo, Cecilia; Musolesi, Mirco; Latora, Vito; Nicosia, Vincenzo; Scellato, Salvatore

The study of important nodes in social and technological networks is an important research question, however the current state-of-the-art is based on static or aggregated model of the network topology. We argue that dynamically evolving network topologies are inherent in many systems, including real online social and technological networks: fortunately the nature of these systems is such that they allow the gathering of large quantities of fine-grained temporal data on interactions amongst the network members.

In this talk we shall present a temporal graph model and reformalise the concepts of shortest paths taking into account time information such as duration, frequency and time order. From this we propose novel temporal centrality metrics, namely closeness and betweenness, which take into account such dynamic interactions over time. These metrics can be applied to a large variety of dynamic networks, including mobile networks, online social networks, and in general, for the study of human interactions. In particular, using a real corporate email and human contact dataset we evaluate the important individuals selected by means of static and temporal analysis taking two perspectives: firstly, from a semantic level, we investigate their corporate role in the organisation; and secondly, from a dynamic process point of view, we devise two short-range mobile malware containments schemes based on blocking (using betweenness) and opportunistic patch spreading (using closeness). We find that temporal analysis provides a better understanding of dynamic processes and a more accurate identification of important nodes compared to traditional static methods.

4 Working Groups

During the seminar the participants split in three working groups, centered around three broad discussion themes. In the following paragraphs there is a summary of the discussion held by each working group.

- **Access to data** There is an important need to acquire large datasets depicting social interactions between individuals. However, such information is almost always controlled by large online companies or mobile phone operators: thus, it is often hard for academic researchers to get access to this kind of data, hampering research efforts. An important point is that anonymization techniques for such datasets are not mature enough to ensure no sensitive personal data can be extracted from dataset: thus, companies are not willing to share information. A potential solution might be to find new ways of collaborating with companies, so that company data can be used by academic researchers without publicly sharing any user information.
- **New models** Even though the temporal evolution and the dynamics of social networks have been largely studied, a new aspect of user behavior is increasingly becoming connected to social interactions: mobility over geographic space. Hence, new models are needed to merge together information about how users interact, when they do so and where they go. While social network structure and user mobility have been separately investigated, with important outcomes, new efforts must be spent to define new models which include both aspects.
- **Practical applications** The practical outcomes of research efforts can be roughly divided in two broad classes: models of user behavior and design of new systems and applications. Companies are more suitable to pursue the latter class, as they benefit from larger amount of resources and data than academic entities. However, academic research has an invaluable advantage in understanding and modeling the universal factors behind user behavior, providing insights that can fuel further research, both in academic and company settings. A mutual and lively exchange of ideas and results between the two fields is hence mandatory.

5 Panel Discussion

The seminar hosted a panel discussion, which centered around four main themes.

- **Theoretical challenges** The analytical challenges arising from recent results regard how to deal with time and space. In particular, the effect of space on user behavior must be better understood, exploiting the huge literature of geography studies. At the same time, dealing with how these systems change over time requires extending models and techniques to include time scales. In general, these two aspects must be intimately combined to provide new spatio-temporal models which also include social behavior.
- **Application challenges** The growing popularity of online services requires users to maintain several versions of their identity, depending on the services they are using. This generates the problem of managing the connection between the real, offline identity and the multiple online identities, protecting privacy and personal information at the same time. In particular, as personal data become scattered across different companies, new mechanisms to manage trust and reputation must be designed, tested and implemented.

- **Future directions** Even though it might be difficult to have an important impact as an academic researcher, since large-scale companies have more resources to solve real-world problems, there is still a lot of value in academic research output. In particular, academic efforts should try to focus on tasks that are usually outside the scope of company laboratories, such as theoretical models and interdisciplinary studies.
- **A psychology perspective** Even though a vast number of works and studies consider social networks as large-scale systems where the nature of individual nodes can be neglected, all nodes are not equal, since they represent human entities. Thus, there are two important consequences: the introduction of psychological and sociological theories can further help research efforts and, on the other hand, large-scale data analysis can help to study psychological problems. At the end, an important result could be to connect user behavioral patterns with personality aspects.

Participants

- Ginestra Bianconi
Northeastern Univ. – Boston, US
- Vincent D. Blondel
UC Louvain-la-Neuve, BE
- Sonja Buchegger
KTH Stockholm, SE
- Krishna P. Gummadi
MPI für Software Systeme –
Saarbrücken, DE
- Des Higham
The University of Strathclyde –
Glasgow, GB
- Petter Holme
University of Umeå, SE
- Byungnam Kahng
Seoul National University, KR
- Renaud Lambiotte
Imperial College London, GB
- Vito Latora
Università di Catania, IT
- Cecilia Mascolo
University of Cambridge, GB
- Alan Mislove
Northeastern Univ. – Boston, US
- Mirco Musolesi
University of Birmingham, GB
- Vincenzo Nicosia
Università di Catania, IT
- Anastasios Noulas
University of Cambridge, GB
- Eamonn O’Neill
University of Bath, GB
- Alessandro Panconesi
University of Rome “La
Sapienza”, IT
- Daniele Quercia
University of Cambridge, GB
- Jason Rentfrow
University of Cambridge, GB
- Stefan Saroiu
Microsoft Res. – Redmond, US
- Salvatore Scellato
University of Cambridge, GB
- Roberta Sinatra
Università di Catania, IT
- Michael Sirivianos
Telefonica Res. – Barcelona, ES
- Thorsten Strufe
TU Darmstadt, DE
- John Tang
University of Cambridge, GB
- Walter Willinger
AT&T Res. – Florham Park, US

