ARS '15

**FIFTH INTERNATIONAL WORKSHOP
ON SOCIAL NETWORK ANALYSIS
APRIL 29 - 30, 2015
ANACAPRI - CAPRI ISLAND - ITALY**

BOOK OF ABSTRACTS

Analyzing Social Interaction in Online and Offline Networks

Martin Atzmueller, University of Kassel, Germany

Abstract

In social network analysis, there are a variety of options for investigating social interactions. This talk discusses the analysis and grounding of social interactions in online and offline networks. We consider distributional semantics, structural network correlation and network inter-dependencies. Specifically, we focus on the analysis of user relatedness, community structure, and relations on online and offline networks. We discuss findings and results that justify the use of even implicitly accruing social interaction networks for the analysis of user-relatedness, community structure, etc. Furthermore, we provide insights into relating online and offline interactions and focus on characteristics of their interrelations. Our results provide for valuable insights, e.g., for the development of link analysis methods, community detection, and the connection of online and offline information.

Aglomerative clustering of large networks

Vladimir Batagelj, University of Ljubljana, Slovenia

Abstract

In the papers Ferligoj and Batagelj (1982,1983) adapted algorithms for solving the clustering problem with relational constraints were proposed. We will first overview the main results from these papers.

The main problem in extending the proposed approach to large networks is that it is based on the dissimilarity matrix and therefore too much space (and time) consuming. To obtain an efficient agglomerative clustering algorithm for large networks we: (1) compute the dissimilarities between units (vertices of network) only for endpoints of existing links of constraining relation; (2) define the dissimilarities between clusters based only on the dissimilarities of the corresponding links and derive the update relations.

We also show that for selected dissimilarities between clusters, the Bruynooghe reducibility property holds. This allows us to speed-up the hierarchical clustering procedure by using the RNN (reciprocal nearest neighbors) approach.

On the same ideas another agglomerative algorithm for clustering the network links was developed. The developed algorithms are illustrated by clustering some real life large networks.

Purely speculative behaviour and heterogeneous agents: a look at the Bitcoin transactions network

Pietro Battiston, Scuola Superiore Sant'Anna, Pisa, Italy

Abstract

Bitcoin and other cryptocurrencies have characteristics which make them extremely interesting from an economic point of view. First, the network of transaction is public. Second, such currencies represent a purely speculative investment, since virtually all goods which can be bought with bitcoins have prices which are indexed in dollars or euros. By analyzing the history of transactions of Bitcoin, this paper characterizes the different types of agents which take part in them, showing that they differ enormously in the volume and in pattern of trade, and that the two aspects are related. Moreover, by comparing such network with the history of prices of Bitcoin with respect to national currencies, the speculative strategies of these different types are compared and shown to also differ.

Comparing fields of sciences: multilevel networks of research collaborations in Italian academia

Elisa Bellotti, University of Manchester, UK
Luigi Guadalupi, Irat-Cnr, Italy
Guido Conaldi, University of Greenwich, UK

Abstract

Much of the work in the sociology of science observes scientific communities from a micro perspective, focusing on interactions in laboratories. In doing so, they try to uncover the impact of social and cultural norms in the everyday production of scientific results. Other studies approach the topic from a macro perspective, analysing scientific organizations and the reciprocal influence they have with wider society. Less attention has been paid to the meso level of interactions within and between scientists and the environments they work in. Methodologically, the gap in the literature can be filled by the recent advancements in multilevel analytical approaches, especially by the combination of statistical multilevel analysis with social network analysis. This combination allows to model structural effects on individual behaviours, where these effects are at work at different levels of social interactions being it between individuals, disciplines, organizations. The paper adopts the structural approach of Lazega et al. (2008) and analyses the local system of public funding to academic disciplines in Italy using bipartite networks. Such analysis has been already done for the two academic areas of physics (Bellotti 2012) and philosophy (Bellotti 2014). Here we extend the analysis to all the areas of research in Italian Academia, in order to compare the results across different scientific fields. By doing this, we observe the variability of structural effects across disciplinary areas, that we expect to be organized in different but comparable ways. In particular, previous analysis of physicists and philosophers showed in both cases the overarching importance of academic ranks and of brokerage roles in obtaining research funding, together with some other interesting effects, like the less impacting but still significant importance of working with a long term established group of colleagues, and the advantages of working on specific sub-disciplines (Bellotti 2012 and 2014). Here we want to see if these results replicate across other disciplinary areas, and/or some interesting differences can be found. For this purposes, we analyse 12 years (1999 – 2011) of the Italian Ministry of University and Research funding of Projects of National Interest (Prin) in all the disciplinary areas of academia. The micro (collaborations between scientists), macro (collaborations between institutions and between disciplines) and meso level (the combination of network measures at a micro and macro level) of interactions are firstly independently analysed, and results are used to model the total amount of money researchers have received over the 12 years against the variables that meaningfully describe the network structures of collaborations to research projects.

Interdisciplinarity and specialisation: Roads to scientific success

Moreno Bonaventura, Queen Mary University of London,UK
Vincenzo Nicosia, Queen Mary University of London,UK
Vito Latora, Queen Mary University of London,UK
Pietro Panzarasa, Queen Mary University of London,UK

Abstract

As the increasing complexity of large-scale research requires the combined efforts of scientists with expertise in different fields, the ad- vantages and costs of interdisciplinary scholarship have taken center stage in current debates on scientific production. Here we conduct a comparative assessment of the scientific success of specialized and interdisciplinary researchers in modern science. Drawing on a comprehensive data set on scientific production, we propose a two-pronged approach to interdisciplinarity. For each scientist, we distinguish be- tween background interdisciplinarity, rooted in knowledge accumulated over time, and social interdisciplinarity, stemming from expo- sure to collaborators' knowledge. We find that, while abandoning specialization in favor of moderate degrees of background interdisciplinarity deteriorates performance, very interdisciplinary scientists always outperform specialized ones, at all career stages. Moreover, successful scientists tend to intensify the heterogeneity of collaborators and to match the diversity of their network with the diversity of their background. Collaboration sustains performance by facilitating knowledge diffusion, acquisition and creation. Successful scientists tend to absorb a larger fraction of their collaborators' knowledge, yet at a slower pace, than less successful ones. Collaboration also pro- vides successful scientists with opportunities for the cross-fertilization of ideas and the synergistic creation of new knowledge. These results can inspire scientists to shape successful careers, research institutions to develop effective recruitment policies, and funding agencies to award grants of enhanced impact.

The network of Italian Erasmus Agreements

Kristijan Breznik, International School for Social and Business Studies, Slovenia
Giancarlo Ragozini, University of Naples Federico II, Italy

Abstract

Student exchange programme called Erasmus determines a large network of European higher education institutions. As Italy is among the most important stakeholders in Erasmus program, paper analyses student mobility to/from Italian higher education institutions, which have been involved in the Erasmus mobility program between the academic years 2007/08 and 2012/13. Aim of the paper is to present preliminary results about the underlying structure of the network. By analytical methods we first identify countries and institutions that are the most attractive for Italian students, as well as countries and institutions that are the best exporters of Erasmus students into the Italian higher education system. Furthermore, in order to get further insights of the Erasmus system, we consider the 2- mode networks defined by the Italian Erasmus Agreements' network of Italian higher education institutions and foreign institutions that exchanged at least one student in the observed period.

In order to analyze this large network we combine advanced network analytical and graphical tools tailored for the analysis of 2-mode networks. In particular we consider the multiplication of 2-mode networks and the multiple correspondence analysis (and its extensions). Both methods allow the researcher to include into the analysis attributes of the nodes and other information, such as the disciplines related to the agreements, and to consider both the country level and the single institution level.

A taxonomy of industrial networks

Tom Broekel, Leibniz University of Hannover, Germany

Abstract

Few studies exist that systematically compare and evaluate the structures of industrial knowledge networks. Two exceptions are the studies by Roediger-Schluga and Dachs (2006) and Broekel and Graf (2012). The first study compares the telecommunication research and the agro-industrial research network. The networks are constructed on the basis of EU-Framework Programmes participation. The study's key finding is that systematic differences seem to exist in the structure of R&D collaboration (knowledge) networks in the two industries.

In a similar manner, Broekel and Graf (2012) compare R&D collaboration networks in ten technological research areas. They make use of a database on joint R&D projects subsidized by the German ministry of education and research. They confirm the finding by Roediger-Schulga and Dachs (2006) and show that variations in network structures relate to the participation of public research organizations.

The present paper aims at extending this line of research by means of establishing a taxonomy of industrial networks. For this, we exploit the same database as Broekel and Graf (2012), namely the database on joint R&D projects subsidized by the German ministry of education and science. The data can be used to construct a great number of industrial networks.

In the initial step of the empirical analysis, we estimate a wide range of network characteristics. The correlations between the measures are subsequently explored with model-based cluster analysis. The derived clusters of industries with similar network structures are evaluated against Pavitt's famous taxonomy (Pavitt, 1984) as well as against industry-specific insights from industrial and innovation economics. Consequently, a taxonomy of industrial knowledge networks is derived.

Broekel, T. and Graf, H. (2012), Public research intensity and the structure of German R&D networks: A comparison of ten technologies. *Economics of Innovation and New Technology*, 21(4): 345-372

Pavitt, K., 1984. Sectoral patterns of technical change: towards a taxonomy and a theory. Research Policy, 13(6): 343–373.

Roediger-Schluga T. and Dachs B. (2006). Does technology affect network structure? *NEMO Working Paper* #2

Reducing Information Overload: News Feed Ranking in a On-line Social Network: Case from OK.ru

Dmitry Bugaychenko, Saint Petersburg State University, Russia

Abstract

Nowadays social networks are the main source of information for fast growing majority of people. In a social network we receive an endless stream of social updates regarding our friends, communities of interests, politics and world events. The stream of social updates is virtually endless and most users are not capable to manage it at whole, thus SN providers are working hard to select the most important, interesting and exciting updates into a personal news feed. The selection algorithm is crucial for users satisfaction and thus to the SN success. However, it is a very difficult task and suffers from nearly all machine-learning problems: cold start, imbalanced and noisy dataset, lack of ground truth and etc. During the talk we will discuss some of algorithms we tried for news feed personalization at OK.ru, approaches we applied to workaround aforementioned problems and results of on-line and offline experiments.

Tweets as source of information for public health. A critical analysis starting from the use of e-cigarettes to quit smoking

Stefano Campostrini, Ca' Foscari University of Venice, Italy
Federica Fedrigo, Ca' Foscari University of Venice, Italy

Abstract

The quantity of social-economics data created every day is rocketing because of the increasing diffusion of Internet connections and use of electronic devices that leaves different kinds of digital tracks: pictures, emails, call records, information about online purchases, financial transitions, interactions on social networks and many others. The term Big data is referred to data characterized by the so-called 3Vs (volume, velocity and variety) and that seem to be easily and quickly accessible relatively cheaply. There are some fields, like economics and finance, where Big Data are already widely and intensively analyzed and exploited for many purposes. But there are others, like medical science or public health, where data obtained from social networks have rarely been exploited to better understand public opinion or social behavior and to analyze ways to influence them. The sociologist Zeynep Tufekci stated that the Big Data analysis applied to social platforms could be as important to social sciences as it was the introduction of the microscope to the biology, but what information can be achieved? With what quality? What the most appropriate tools for analysis? We chose Twitter as social network to study the potentialities and limits of the analysis of online behavior of users. We focused on tweets containing words referring to the e-cigarettes phenomenon, to understand if in Italy the digital smoke is used as method to decrease or stop the tobacco consumption, a quite discussed issue in public health. We have extract tweets from January 2014 to April 2014 using a Twitter REST API, conducting both quantitative and qualitative analyses, and then we have tried to relate observed trends and changes n the tweets with events occurred in the same time period.

Beside the substantial – interesting and relevant – results from the quantitative analysis (the e-smoke seems very little correlated with the attempts to quit), a linguistic analysis has shown the limits of the standard quantitative approaches: polarity, intensity and the intrinsic idea of the message posted have shown peculiarities that, if ignored, could be really misleading. Certainly social network produce big data that potentially are precious sources of information also for fields, like public health, in which the knowledge of people's attitude and behaviors is essential. Caution should be taken in interpreting results from mere quantitative analysis: a combination with qualitative approaches related to linguistic – narrative analysis seem promising.

Who knows about it? Unpacking the Structure of Knowledge Diffusion in 52 Wikipedia Versions

Niccolò Casnici, University of Brescia, Italy
Pierpaolo Dondio, DIT-Dublin Institute of Technology, Ireland
Flaminio Squazzoni, University of Brescia, Italy

Abstract

Wikipedia is the online free encyclopaedia collaboratively written by web users, ranking among the top 10 most visited websites, with about 240 languages, with 122 languages hosting at least 100.000 articles (as of June 2014). According to Wikipedia guidelines, the reliability of the entries is guaranteed by the rule of the neutral point of view (NPOV), according to which every article should "represent fairly, proportionately, and, as far as possible, without bias, all of the significant views that have been published by reliable sources on a topic" (en.wikipedia.org/wiki/ Wikipedia: Neutral_point_of_view). Although the NPOV policy, a recent and growing literature from both academia and popular media showed that the Wikipedia content is not exempt from bias. Wikipedia's topical coverage is mainly driven by the specific interest of the contributors and the reliability of the articles strongly depends on the macro-area of the article. Furthermore, (i) the similarities between the editions of Wikipedia decrease depending on the increasing geographic distance between the countries, (ii) increase when the size of the Wikipedias (i.e. the number of articles) is more similar and finally (iii) increase when the language is more similar. The presence of such bias raises the importance to understand how the knowledge is shared, accessible and distributed among the various versions of Wikipedia. Starting from a sample of 99.841 articles about writers, poets and novelists from 52 different languages, we constructed a Global Wikipedia Network (GWN), in which Wikipedias of different languages represented the nodes of the graph. We considered two Wikipedias linked if one of them hosts articles belonging to the other one. Hosting is the crucial mechanism at the base of knowledge spreading processes in the GWN. For instance, if an article about a Korean poet is hosted by the French Wikipedia, it means that the French speaking population can potentially access that page. Otherwise the content of the article would have been restricted to the Korean speaking. In other words, we claim that the structure of hosting relationships between the Wikipedias of different languages strongly determines the possibility of a population to access to the common knowledge embodied in the whole encyclopaedia. For that reason, it is essential to understand the mechanisms explaining the presence (absence) of the ties between the Wikipedias and, even more importantly, to investigate the dynamics that drive the generation of such ties. In our research we explained the presence (absence) of ties in the GWN by referring to endogenous structural mechanism, like network centralization around the main Wikipedias, reciprocity and network closure. We also looked at exogenous actor-relation effect, like homophily toward linguistic similarities and geospatial effects. An exponential random graph model (ERGM) approach allowed us to investigate such mechanisms at the same time.

Comorbidity Networks from General Practitioner Prescriptions

Pierpaolo Cavallo, University of Salerno, Italy
Mario De Santis, Osservatorio Sanità e Salute, Italy
Sergio Pagano, University of Salerno, Italy

Abstract
Comorbidity is a major health and societal issue, but it has so far captured much less attention than other disease-related studies. A major reason is the lack of agreed upon methods to understand the complex interdependent relationships between morbidities at various dimensions: the patient, the reference population, the observational time, the geographic location, etc. The emerging field of Network Medicine may offer a natural representation of such complexity. In this study we report a network analysis of the comorbidity networks that can be extracted from General Practitioner prescriptions data referred to a sample population of about 15000 patients and spanning several years. The relevant network aspects are analyzed and discussed.

A proposal for the evaluation of 'social' reputation: a financial example

Paola Cerchiello, University of Pavia, Italy
Paolo Giudici, University of Pavia, Italy

Abstract

The role of online opinions and of social forum is continuously gaining great interest and influence. Thus it is important to track such information and to employ efficient statistical methods to analyse it.

We propose a combined approach based on both sentiment analysis and graphical models.

In particular we focus our attention on Twitter data that are typically available in huge amount whose maximum number of character is 140, and this avoids the explosion of the data matrix. However this peculiarity does not imply an easier task since Twitter lexicon tends to be very topic dependent, continuously evolving and being particularly prone to the massive usage of slang expression and neologism.

Thus the challenge is twofold: on one hand we need to train an appropriate lexicon based on the chosen topic, on the other hand we propose a graphical model based on the employed words and on the associated sentiment. Handcock & Raftery (2007) is our inspiring paper, where the authors take advantage of network models to represent relations between interacting units or actors. In particular they propose the so-called latent position cluster model in which a probability of a link between two actors depends on their distance on an unobserved social space and their positions come from a mixture of distributions, each representing a cluster.

We move from such approach and we apply graphical models not on actors (i.e. people twitting) but on the employed bag of words. Thus we work on an adjacency matrix that represents the co-occurrence of words along the analysed tweets but at the same time we take into account the sentiment associated to such tweets. The sentiment is measured on a scale positive-negative-neutral and is trained according to several competing classifiers ranging from the popular naïve Bayes to the more efficient Support Vector Machines and Bagging. The relative performance are evaluated and compared in terms of misclassification error, recall and precision.

The approach has been tested in two different contexts, on United Kingdom listed banks and on Italian popular banks and the classifiers have been trained on a corpus of 14,000 tweets manually labelled as positive negative neutral. The final aim is to evaluate the reputation of listed banks according to the twitter community perception.

Understanding empathy in a Twitter community

Valerio Cestrone, University of Naples Federico II, Italy
Simona Balbi, University of Naples Federico II, Italy
Agnieszka Stawinoga, University of Naples Federico II, Italy

Abstract

Social interactions are the foundation of mankind and social networking sites are the evolution of personal relations allowing anyone the opportunity to influence others in the stream beyond geographical distances.

The aim of this work is giving a statistical strategy for understanding the reactions of a community to a stimulus given by an influent individual. We focus our attention on the mechanism of "retweeting" on Twitter, in a Sentiment Analysis perspective.

We investigate how the text content affects the relations in a network of followers, during a defined period. A graphical display allows representing the positive/negative reactions.

Countries' Market Power In The International Multi-Product Trade Network

Isabella Cingolani, Politecnico Milan, Italy
Pietro Panzarasa, Queen Mary University of London, UK
Lucia Tajoli, Politecnico Milan, Italy

Abstract

We endorse a multi-layer network perspective to provide a new model for measuring the competitiveness of the countries involved in the international multiproduct network of trade relations. Each layer of the network corresponds to a specific product that a country can export to, or import from, other countries. In each layer, a country's market power towards other countries can be expressed as a function of the position the country occupies in the trade network. First, for each product we model a country's market power in terms of its ability to secure the highest local market shares in the export of that product to other countries. Second, a country's market power becomes stronger as the dependence of the importing countries increases. Third, a country's dependence is modelled as a function of the number and market shares of the countries from which products are imported. Concentration of imports on few and powerful sources of supply thus contributes towards the market power and dependence of countries by affecting the degree of substitutability among potential alternative competitors. We apply our measures of market power and dependence to the trade relations extracted from the BACI data set. To this end, we constructed a multi-layer network with 5, 212 layers, each corresponding to a specific product category. The nodes of the network are the 221 countries. For each layer, a directed link is established from one country to another when the former exports the corresponding product to the latter. The weight of each link is based on the nominal export value of the corresponding trade, converted to US dollars in the year 2010. Results suggest that adjusting a country's market power through the dependence of the importing countries can unmask relevant information that would otherwise remain undetected, both in the aggregate network and at the level of a specific layer. Our measure of market power does not simply reflect an exporter's relative market share held in the destination countries, but also tries to capture the actual competitive advantage of the exporter with respect to alternative sources of supply competing in the same markets. Findings indeed indicate that the ranking of countries obtained with our measure of market power does not trivially replicate the ranking obtained with more traditional measures of exporters' market shares in the destination markets.

Birds of a feather? Revisiting homophily and tie creation in social networks

Valerio Ciotti, Queen Mary University of London, UK
Pietro Panzarasa , Queen Mary University of London, UK
Vito Latora, Queen Mary University of London, UK

Abstract

A number of network growth mechanisms have been suggested to explain how social connections are forged and severed over time. Among them, a key role is played by homophily, namely the principle that similarity breeds connection. Homophily has been empirically documented in a variety of domains, including marriage, friendship, work advice, support, and information transfer. Among the various forms of homophily, cognitive homophily suggests that the convergence of beliefs, interests, and interpretations are the cognitive antecedent that sparks the creation and dynamics of social ties. We propose and test the hypothesis that the probability of a social tie increases with cognitive similarity, as suggested by the theory of homophily, but only up to a critical threshold. Above this level, the effects of cognitive similarity reverse. Individuals with highly similar beliefs, meanings, and interpretations are not likely to provide each other with the information they seek, and will direct their attention to other less cognitively similar partners.

We analyse how the cognitive similarity among users in an online forum affects the probability that they will communicate with one another. Our study draws on a dataset from an online social network of users from the University of California, Irvine, who posted messages to discussion groups. We measure the cognitive similarity between all pairs of users in terms of the semantic distance between the words they posted, regardless of which groups they belong to. Users can be cognitively but not socially, connected: this occurs when they post semantically related words, but do not share any group in the forum and thus do not have any opportunity to communicate. If homophily were the only force governing tie creation, we should expect some increasing function linking the probability of tie creation to cognitive similarity: people who are similar would be more likely to communicate with one another than dissimilar people. By contrast, our findings suggest that homophily seems to affect tie creation, but only up to a certain threshold, beyond which the effects of similarity seem to reverse. In order to cast light on these mixed effects of homophily on tie creation, we propose a model that integrates homophily, social influence, and social dependence into a unified growth mechanism underpinning the evolution of a social network over time.

Networks without Graphs
In search of novel visualization approaches to tame complexity

Paolo Ciuccarelli, Politecnico Milan, Italy

Abstract

When it comes to visualize network-data the solutions is often - if not always – to represent it as a graph, especially for "other" disciplines that are nowadays facing – with a certain excitement - the new era of digital born data (e.g. Computational Sociology, Digital Humanities etc.).

With data becoming "big" graphs show all their limits as cognitive tools: when the number of edges and links increase the legibility and the understandability of the graph decrease. Also network analysis related principles are not obvious for non-experts.

Thus, the research activity in the field of network visualization is growing, addressing especially the still open question of how to overcome the so called "hairball" effect (Schultz et al. 2013), where too complex and tightly connected networks become difficult to explore and observe.

A number of different techniques are deployed in this sense, mainly focusing on the design and use of different spatialization algorithms (Jacomy 2009, Martin 2011, Jacomy at al 2011), in order to define the most suitable and efficient layout visualization according to the features of the network (i.e. number of nodes in the network or type of connections). Still, the problem of many layout algorithms remains: typically they cannot be adjusted to address user's specific needs, they can produce different layouts of the same network and layouts of the same network created by different algorithms cannot be easily compared (Krzywinski, 2011).

Few research efforts have been dedicated so far to the study of new visualization patterns. Although alternative visualization patterns for network data are historically known (Bertin, 1969) still the use of graphs is rarely questioned and often taken for granted.

The most relevant cases of alternative visualization patterns will be presented and discussed, especially Hiveplot and BioFabric: Hiveplot aims at taming the 'hairball' effect and the visual complexity arising from large number of edges and exposing both trends and outlier patterns in network structure (Krzywinski, 2011); BioFabric proposes a novel network visualization technique that depicts nodes as one-dimensional horizontal lines arranged in unique rows; each edge in the network is then depicted using a vertical line assigned to its own unique column (Longabaugh, 2012).

A third research line focuses on the attempt to ease the exploration of too complex networks is the one reflecting on "how to serve" the network to the end user. These experiment often result in the design of interfaces with filter and search functionalities, enabling more detailed views of the network (Van Ham & Perer 2009), or in the design of different views of the same network that are shown to the user one after the other, serving the purpose of telling a story through the different visualization of the network. The now popular idea of telling stories with data – or "storify" the data – has been applied also network visualization, as in the tool "Manylines" developed by the MediaLab at Sciences-Po.

Leveraging Network Science to address Grand Societal Challenges

Noshir Contractor, Northwestern University, USA

Abstract

The increased access to big data about social phenomena in general, and network data in particular, has been a windfall for social scientists. But these exciting opportunities must be accompanied with careful reflection on how big data can motivate new theories and methods. Using examples of his research in the area of networks, Contractor will argue that Network Science serves as the foundation to unleash the intellectual insights locked in big data. More importantly, he will illustrate how these insights offer social scientists in general, and social network scholars in particular, an unprecedented opportunity to engage more actively in monitoring, anticipating and designing interventions to address grand societal challenges.

Modelling time evolving interactions in networks through a non stationary extension of stochastic blockmodels

Marco Corneli, Université Paris 1 Panthéon Sorbonne, France
Pierre Latouche, Université Paris 1 Panthéon Sorbonne, France
Fabrice Rossi, Université Paris 1 Panthéon Sorbonne, France

Abstract

In this paper, we focus on the stochastic block model (SBM), a probabilistic tool describing interactions between nodes of a network using latent clusters. The SBM assumes that the network has a stationary structure, in which connections of time varying intensity are not taken into account.

In other words, interactions between two groups are forced to have the same features during the whole observation time. To overcome this limitation, we propose a partition of the whole time horizon, in which interactions are observed, and develop a non stationary extension of the SBM, allowing to simultaneously cluster the nodes in a network along with fixed time intervals in

which the interactions take place. The number of clusters (K for nodes, D for time intervals) as well as the class memberships are finally obtained through maximizing the complete-data integrated likelihood by means of a greedy search approach. After showing that the model works properly with simulated data, we focus on a real data set. We thus consider the three days emph{ACM Hypertext} conference held in Turin, June 29th - July 1st 2009. Proximity

interactions between attendees during the first day are modelled and an interesting clustering of the daily hours is finally obtained, with times of social gathering (e.g. coffee breaks) recovered by the approach.

We compare our results with those obtained through similar methods and discuss improvements obtained with respect to the classical SBM. Applications to large networks are limited due to the computational complexity of the greedy search which is dominated by the number of clusters used in the initialization. Therefore, advanced clustering tools are considered to reduce the number of clusters expected in the data, making the greedy search applicable to large networks.

The evaluation of policies towards innovative networks and social network analysis

Ivan Cucco, University of Naples Federico II, Italy
Alfredo Del Monte, University of Naples Federico II, Italy

Abstract

In this paper we show how Social Network Analysis (SNA) can be used to evaluate the opportunity that governments subsidize innovative networks through Network-Based Policies (NBP). We argue that, although SNA is helpful to assess how efficiently information is transferred within a network, it does not by itself provide a clear estimation of the generated welfare effects. We propose for this reason a methodology that integrates SNA indicators with an evaluation of the welfare generated by both direct and indirect effects.

In the first part of the paper we discuss the conditions under which a policy towards innovative networks is additional or, conversely, a waste of resources. We then apply our methodology to an empirical case and evaluate ex post the effects of the Italian policy in support of technological districts. Our results confirm the necessity to integrate social network indicators with a more appropriate measure of the welfare effects of NBPs. We also suggest that the lack of such integration may explain why previous empirical analyses have given conflicting answers as to the questions of whether public R&D is additional to private spending or rather a substitute for it, and whether even additional public policy might be inefficient.

Are regional innovation policies promoting related variety? An ERGM analysis of collaborative R&D projects in Italian convergence regions

Ivan Cucco, University of Naples Federico II, Italy
Alessandro De Iudicibus, University of Naples Federico II, Italy
Francesco Savoia, University of Naples Federico II, Italy

Abstract

As highlighted by systemic approaches to innovation, regions are playing an increasingly important role in the design and implementation of place-based innovation policies. A wide debate has recently emerged on the limits and validity of different policy models, distinguishing for example between 'platform' and 'cluster' regional policies, or between a 'corporatist' and an 'evolutionary' interpretation of the Triple Helix.

Within the framework of EU cohesion policy, a number of Technological Districts (TDs) and Public-Private Laboratories (PPLs) have been established since 2005 in the four Southern Italian convergence objective regions (Calabria, Campania, Puglia and Sicilia) with the aim of fostering competitiveness, systemic innovation and research-industry linkages. Technological Districts and Public-Private Labs have become the key actors in collaborative R&D projects subsidized with both EU and national funds.

In this work we use Exponential Random Graph Models to analyze the collaboration networks elicited from data on the co-participation in publicly funded R&D projects involving TDs and PPLs. We seek to understand to which degree the overall policy design has been able to promote organizational and sectoral heterogeneity among project participants. We find that different network architectures coexist under the same policy umbrella, and relate their variety to the ideal models identified in the literature and to the characteristics of the regional innovation systems.

Stability of co-authorship blockmodels

Marjan Cugmas, University of Ljubljana, Slovenia
Anuška Ferligoj, University of Ljubljana, Slovenia
Luka Kronegger, University of Ljubljana, Slovenia

Abstract

Since early works of Price (1963, 1965) and of Garfield (1979), sociologists introduced a series of theories dealing with scientific collaboration. This is often operationalized through the co-authorship of one or more published scientific productions. The structure of the collaboration can be studied usefully by blockmodeling co-authorship networks. Kronegger et al. (2011) studied the co-authorship networks of four scientific disciplines in Slovenia and identified the most typical collaboration structure as having three basic positions: multi-core, semi-periphery and a periphery.

Here, we propose a new measure of stability of the obtained cores across two time period by using pre-specified blockmodeling. A new index ARIM (Adjusted Rand Index Modified), based on ARI (Adjusted Rand Index) (Hubert and Arabie 1985), is proposed to measure the stability of cores. This proposed index is applied to empirical data for co-authorship networks featuring most of the scientific disciplines in Slovenia across two time periods (1991-2000 and 2001-2010).

Propensity score matching with social network data

Luca De Benedictis, EIEF and University of Macerata, Italy
Bruno Arpino, University Pompeu Fabra, Spain
Alessandra Mattei, University of Florence, Italy

Abstract

Propensity score matching (PSM) is a widely used technique in causal inference studies to adjust for bias arising from an unbalanced distribution of observed confounders between a treatment and a control group. Both theoretical and applied works has recently considered the PSM for longitudinal or clustered data, but the analysis of interlinked data is still missing. In this paper we consider the implementation of PSM in the context of social networks. In some applications, together with individual unit characteristics, also features of the social network in which units are embedded may be considered as confounders (i.e., variables that impact on both the probability of receiving the treatment and the outcome). Failing to adjust for these social networks factors may lead to biased estimates. Therefore, we study how to specify the propensity score model when characteristics of the social network cannot be ignored for unconfoundedness to hold. As a motivating case study we consider estimating the effect of liberal trading order on bilateral trade in the XXth century.

The evolution of policy-driven innovation networks at regional level: evidence from a stochastic actor oriented approach

Domenico De Stefano, University of Trieste, Italy
Annalisa Caloffi, University of Padova, Italy
Federica Rossi, Birkbeck, University of London, UK
Margherita Russo, University of Modena and Reggio Emilia, Italy
Susanna Zaccarin, University of Trieste, Italy

Abstract

The last twenty years have witnessed the diffusion of regional innovation policies that support networking among heterogeneous organisations (e.g.: firms and universities; small and large firms).

The implicit assumption is that policies are needed to stimulate interactions that would not occur spontaneously, but whose presence would be desirable.

Since it would not be possible, nor appropriate, to identify the specific partners with whom an organisation should collaborate, what policies can do is provide incentives in order to encourage collaboration among heterogeneous organisations, establish a general framework of rules and then leave the participants free to organise their innovative activity. However, the presence of general and flexible incentives does not ensure that undesirable behaviours (such as the formation of an oligarchic core of agents driving the funded projects, or the reiteration of established linkages) are not adopted anyway.

Our study seeks to deepen the reflection on these points, introducing two main original elements with respect to the existing literature. First, we observe small scale regional policies targeting SMEs, instead that the large-scale EU programmes. Second, we adopt a dynamic approach, rather than a comparative static approach as usually adopted.

Our empirical analysis explores how collaboration has evolved within a set of policies targeting innovation networks, which have been implemented by the regional government of Tuscany (Italy) in the programming period 2000-2006.

By using a stochastic actor oriented approach, we investigate how their relationships have evolved according to several aspects: (i) reiteration of pre-existing linkages as opposed to experimentation of new relationships (*stability or experience effect*); (ii) collaboration with agents possessing complementary rather than similar knowledge and competencies or abilities (*homophily effects*); (iii) formation of a core of agents "controlling" the policy programmes (*Matthew effect*).

Predicting leadership networks from little words and big motives

Leslie DeChurch, Georgia Tech, USA
Dorothy Carter, Georgia Tech, USA
Zachary Gibson, Georgia Tech, USA

Abstract

Leadership is pivotal to the success of organizational teams. Recent attention has shifted from a "leader-centric" to a "network-centric" view. Leadership networks are influence relations wherein individuals, who may or may not have formal authority, "claims" influence over others who in turn, "grant" these claims. We explore how members' motives and semantic communication predict the formation of leadership network ties in multiteam systems (MTSs). As MTS members collaborate, they express their dominant motivations, thus activating others' perceptions of their leadership capacity and willingness to acquiesce to influence attempts. MTS members who prioritize MTS goals above the goals of their team or themselves will be more likely to express other oriented sentiment and thus, be more likely to be accepted as leaders. Additionally, function words, pronouns that signal group membership, also ought to shape leadership relations. Emphasizing a broader group identity by communicating about shared superordinate goals signals that a would-be leader is "other focused." Use of second-person singular pronouns (e.g., you, yours), on the other hand, signals other-focused intentions. In fact, referring to other teams may suggest a focus on aligning efforts with another group. We tested these ideas in a sample of 252 students performing a laboratory task as members of one of 21 MTSs. Individuals' dominant motives were manipulated with a reward structure (i.e., individual, team, MTS). A manipulation check (i.e., "What is your dominant goal?") confirmed that participants internalized their assigned dominant motive as assigned. Communication occurred across an instant messaging platform. The semantics of messages (N = 97,499 messages) were coded for function words. Leadership networks were assessed by asking: "Who did you rely on for leadership?" ERGM analyses indicate that, controlling for network density, reciprocity, and common team membership, members' motives significantly predict leader emergence early on; superordinate goal communication predicts leader emergence throughout the task; and other group references predict leader ties in the final phases. We discuss the implications of our findings for leader emergence and multiteam collaboration in organizations.

Macro blockmodel stability with micro network instability

Patrick Doreian, University of Pittsburgh, USA, University of Ljubljana, Slovenia

Abstract

Blockmodeling is used to discern the structure of a social network by identifying positions and the ties between positions that are called blocks. Most often the presumption is that the overall network is relatively stable as are the ties between the units comprising the network. This need not be the case. This presentation considers two networks. One unsigned and one is signed. Both feature nations for very different relations. The unsigned network features investment by states in other states in electricity infrastructure over five years before 2000. The signed network deals with ties between states after WWII from 1946 through 1999. Both networks are highly unstable at the micro level yet have a remarkable stability at the macro level. Reasons for this are discussed both substantively and technically. Implications regarding the coupling substance and technique when blockmodeling are presented.

Large Network Analysis: Representing the Community Structure by Means of Interval Data

Carlo Drago, University Nicolò Cusano, Italy

Abstract

In recent years the advances in computer technology and the lower costs of digital storage has determined a growth in the data collections on an exceptional scale. In this process large network becomes ubiquitous in various different fields. Currently large networks can also be characterized by continuous operations (which lead, for example, to textual contents) and many interactions (which become streams of edges). By considering the streaming framework, the different networks can be analyzed in two ways: by taking in the account the structure and also by mining the different operations (for example the textual contents of the posts). These complex data typologies lead to new challenges for the data analysts. A relevant one for example is to understand the structure of the interaction between the different nodes. In this sense one of the most relevant problems on large networks is to understand and explore the community structure and the interaction which can occur on the communities. In particular we can assume there could be frequent and significant interactions on the communities between the nodes. Furthermore the relevant problem is to understand the entire network organization and the functions associated to the different network partitions as a way to understand the way the different nodes interact. One approach proposed on the symbolic data analysis framework considers the networks as symbolic data. In this framework we move considering a method to represent effectively the network community structure. In this context the idea presented is to represent the different relevant network components as the communities as interval data. This data representation allows obtaining a stylized description of the network community structure. The different representation permits to characterize adequately the different communities and also to analyze the different relationships between actor's attributes of the communities. In this way the objective is to understand the communities functioning and their role on the network. The method is useful to represent the network hierarchical structure. Applications of this approach can be considered on real social networks and communication networks.

Bayesian Inference on Group Differences in Brain Networks

Daniele Durante, University of Padova, Italy
David Dunson, Duke University, USA

Abstract

Network data are increasingly available along with other variables of interest. Our motivation is drawn from neurophysiology studies measuring a brain activity network for each subject along with a categorical variable, such as presence or absence of a neuropsychiatric disease, creativity groups or type of ability. We develop a Bayesian approach for inferences on group differences in the network structure, allowing global and local hypothesis testing adjusting for multiplicity. Our approach allows the probability mass function for network-valued data to shift nonparametrically between groups, via a dependent mixture of low-rank factorizations. An efficient Gibbs sampler is defined for posterior computation. We provide theoretical results on the flexibility of the model and assess testing performance in simulations. The approach is applied to provide novel results showing relationships between human brain networks and creativity.

Mixed-methods for the analysis of scientific collaboration networks

Cristiano Felaco, University of Naples Federico II, Italy
Rita Cimmino, University of Naples Federico II, Italy

Abstract

In social sciences, last years have witnessed the emergence of methodological debates about the opportunity for mixing methods in Social Network Analysis (SNA).

Recent mathematical developments and computing technologies have enabled statistical computations of large network structures. Despite much of development of SNA has occurred using a quantitative perspective, however, more researchers have studied the social networks using qualitative methods. In this perspective, the social network is seen as a "social world", a world of shared meanings, feelings, conventions, norms and identities. The qualitative methods allow to understand the process of creating of social network. For this reason, networks are both structure and process at same time and mixing methods can be more useful when considering the various aspects of networks.

This paper presents a research of sociologists' collaboration network in Campania Region adopting a mixed approach. On the hand, we have chosen to analyze the co-authorship network to detect the formal collaboration between the authors and, on the another hand, have used in-depth interviews to understand informal collaboration. We will show how a mixed approach made it possible to detect in depth the dynamics related to the scientific collaboration network.

Community structure in complex networks

Santo Fortunato, Aalto University in Espoo, Finland

Abstract

Complex systems typically display a modular structure, as modules are easier to assemble than the individual units of the system, and more resilient to failures. In the network representation of complex systems,modules, or communities, appear as subgraphs whose nodes have an appreciably larger probability to get connected to each other than to other nodes of the network. In this talk I will address three fundamental questions: How is community structure generated? How to detect it?
How to test the performance of community detection algorithms? I will show that communities emerge naturally in growing network models favoring triadic closure, a mechanism necessary to implement for the generation of large classes of systems, like e.g. social networks. I will discuss the limits of the most popular class of clustering algorithms, those based on the optimization of a global quality function, like modularity maximization.
Testing algorithms is probably the single most important issue of network community detection, as it implicitly involves the concept of community, which is still controversial. I will discuss the importance of using realistic benchmark graphs with built-in community structure.

Reputation and impact in academic careers

Santo Fortunato, Aalto University in Espoo, Finland

Abstract

Reputation is an important social construct in science, which enables informed quality assessments of both publications and careers of scientists in the absence of complete systemic information. However, the relation between reputation and career growth of an individual remains poorly understood. We show that the impact of a paper is boosted by the reputation of its coauthors up to a given threshold, above which the impact of the work is sustained by its own reputation. This suggests that investing on quality over quantity is still a promising strategy, but that the choice of collaborators is an important added value.

A new algorithm for community detection in oriented and weighted networks

Félicité Gamgne Domgue, Université of Yaoundé I, Camerun
Norbert Tsopze, Université of Yaoundé I, Camerun
Rene Ndoundam, Université of Yaoundé I, Camerun

Abstract

Community detection in social networks, has gained a lot of attention in last few years. Most of existing approaches focus on defining communities on non-oriented and non-weighted networks. Some assumptions for oriented-weighted networks have been proposed. These ignore the orientation of edges, and apply non-oriented graph algorithms for community detection. However, non-oriented graphs are not adapted to real studied systems, like the citation network, stars and fans (on Twitter and Facebook), prey and predators(in trophic network)…, where the omission of edges' orientation leads to the loss of important information. In this work, we propose a new efficient algorithm for communities' detection based on the idea that vertices belonging to the same community have common node-neighborhood. The new algorithm is structured in two steps: identifying kernels (i.e. sets of leader nodes) in the network that are sets of nodes with common neighbors, then extracting edges connecting kernels by using flow betweenness, and evaluating the obtained partition. Both steps are repeated until the maximization of an evaluating density function. To show the efficiency of the new algorithm, we test it on some real networks and the results are compared to those obtained by some community detection algorithms on oriented networks described in the state of the art.

A new and efficient architecture for the collection of big data: Unity.

Giulio Gatti, University of Pavia, Italy
Paola Cerchiello, University of Pavia, Italy

Abstract

The issue of collecting efficiently and rapidly huge amount of data is crucial in big data applications. 'Velocity', 'Variety' and 'Volume' constitute the 3 main dimensions to be considered when dealing with big data & analytics. 'Velocity' means rapid obsolescence of the data, a clear example is represented by social networks and blogs that very rapidly 'eat' topic trends that can rise and fall according to a one day horizon. 'Variety' deals with the heterogeneity of data structures and sources; typically you have to harmonize different granularity of data registration (second, hour, day, week and so on). But at the same time you can have access at very different sources: media, social networks, economic indicators, balance sheets, internal questionnaires, opinions, ratings. The dimension 'Volume' of course suggests the idea of huge amount of data from terabytes to petabytes and zetabytes.

After having considered the most important and used software for Big data, we will depict our new proposal, i.e. the architecture of the system 'Unity' a software able to collect data from any kind of on-line source by means of proprietary sondes. Moreover, we will explain how to cope with the 'volume' dimension by using a MapReducer, a method able to reduce both the dimensions of the stored data matrices avoiding storage problems and CPU overflow. In particular we will show which elements of the architecture design should be fine tuned to get the maximum benefit from the relative putting in place.

Finally we will show a practical example, considering a complete workflow of analysis, from the collection by means of sondes, the reduction through the MapReducer, to the connection to the most important open source statistical software like R and Knime for the statistical analysis of such data.

Efficient Influence Optimization in Social Networks

Luisa Gargano, University of Salerno, Italy
Gennaro Cordasco, Second University of Naples, Italy
Marco Mecchia, University of Salerno, Italy
Adele Rescigno, University of Salerno, Italy
Ugo Vaccaro, University of Salerno, Italy

Abstract

Given a network represented by a graph $G = (V,E)$, we consider a process of influence diffusion in G that evolves as follows: Initially only the elements of a given subset of the nodes in the network are influenced; subsequently, at each time instant the set of influenced elements is augmented by all the nodes in G that have a sufficiently large number of already influenced neighbors.

In this paper, we consider the problem of finding a small subset of nodes that can influence the whole network. This is a relevant and widely studied problem that abstracts many phenomena in the social, economic, biological, and physical sciences. In particular, it is strictly relevant to Viral Marketing in Social Networks.

It is known that the above optimization problem is a very hard computational problem. In fact, it was proven that it is highly unlikely the existence of a polynomial time algorithm with approximation guarantee better than a polylogarithmic factor.

We present a fast and surprisingly simple algorithm that exhibits the following features:

1) It provably produces an optimal solution (i.e, a minimum size subset of nodes that can influence the whole network) in case G is either a tree, or a cycle, or a complete graph;

2) in real-life networks it produces solutions that outperform the ones produced by previously published algorithms. In particular, we conducted experiments on several datasets of various sizes from the Stanford Large Network Dataset Collection (SNAP) and from the Social Computing Data Repository at Arizona State University. We compared the performance of our algorithm towards that of the best, to our knowledge, computationally feasible   algorithms in the literature. All test results consistently show that our algorithm presents the best performances on all the considered data.

Big Data Analytics and Social Mining

Fosca Giannotti, University of Pisa, Italy

Abstract

The Big Data originating from the digital breadcrumbs of human activities, sensed as a by-product of the technologies that we use, let us observe the individual and collective behavior of people at an unprecedented detail.

Big Data at a societal scale provide a powerful microscope, which together with Social Mining – the ability of discovering knowledge from these data – can help us understand and forecast many complex and hidden socio-economic phenomena, from the diffusion of information, innovation and crises to the unequal distribution of resources and opportunities. Scientific research is being revolutionized by this new wave, and policy making is next in line, because big data and social mining are providing novel means for measuring and monitoring well-being in our societies – more realistically, beyond the GDP, more precisely, continuously, everywhere. Social mining is about making sense of Big Data by extracting meaningful information from large, messy and noisy data (originally produced for other purposes than analysis) and it is a complex process that requires high-level analytics, modeling and reasoning techniques.

My seminar discusses the novel questions that big data and social mining allow to raise and answer, how a new paradigm for scientific exploration, statistics and policy making is emerging, and the major scientific, technological and societal barriers to be overcome to realize this vision. I will focus on concrete projects with telecom providers and official statistics bureau in Italy and France aimed at measuring, quantifying and possibly predicting key demographic and socio-economic indicators based on nation-wide mobile phone data: the population of different categories of city users (residents, commuters, visitors) in urban spaces, the inter-city mobility, the level of well-being and economic development of geographical units at various scales.

Visual network storytelling

Paul Girard, MediaLab, Sciences-Po, Paris, France

Abstract

We love networks! Networks are powerful conceptual tools, encapsulating in a single item multiple affordances for computation (networks as graphs), visualization (networks as maps) and manipulation of data (networks as interfaces). In the field of mathematics, graph theory has been around since Euler's walk on Königsberg's bridges (Euler 1736). But it is not until the end of the last century that networks acquired a multidisciplinary popularity. Graph computation is certainly powerful, but it is also very demanding and for many years its advantages remained the privilege of scholars with solid mathematical fundamentals.

In the last few decades, however, networks acquired a new set of affordances and reached a larger audience, thanks to the growing availability of tools to design them. Drawn on paper or screen, networks became easier to handle and obtained properties that calculation could not express. Far from being merely aesthetic, the graphical representation of networks has an intrinsic hermeneutic value. Networks can become maps and be read as such.

Combining the computation power of graphs with the visual expressivity of maps and the interactivity of computer interface, networks can be used in Exploratory Data Analysis (Tukey, 1977). Navigating through data becomes so fluid that zooming in on a single data-point and out to a landscape of a million traces is just a click away.

Increasingly specialized software has been designed to support the exploration of network data. Tools like Pajek (vlado.fmf.uni-lj.si/pub/networks/pajek), NetDraw (sites.google.com/site/netdrawsoftware), Ucinet (www.analytictech.com/ucinet), Guess (graphexploration.cond.org) and more recently Gephi (gephi.org) have progressively smoothed out the difficulties of graph mathematics, turning a complex mathematical formalism into a more user-friendly point-and-click interface[1].

If visual exploration of networks can output to confirmatory statistics, what about sharing one network exploration with others?

We developed Manylines (https://github.com/medialab/manylines), a tool allowing you to share the visual analysis of a network with a wide audience by publishing it on the web. With Manylines, you can not only easily publish a network on the web but also share its exploration by describing the network's visual key findings. Through a set of examples, we will illustrate how the narrative opportunities of Manylines can contribute to the enunciation of a visual grammar of networks.

---

1   A simple look at the URLs of the subsequent tools reveals the efforts deployed to make network-manipulation tools user-friendly and thereby available to a larger public.

Unveiling Mobility Complexity through Network Analysis

Riccardo Guidotti, University of Pisa, Italy
Salvatore Rinzivillo, KDD LAB – Knowledge Discovery and Data Mining Lab, CNR, Italy
Anna Monreale, KDD LAB – Knowledge Discovery and Data Mining Lab, CNR, Italy
Dino Pedreschi, University of Pisa, Italy
Fosca Giannotti, KDD LAB – Knowledge Discovery and Data Mining Lab, CNR, Italy

Abstract

One of the most fascinating challenges of our time is to understand the complexity of the global interconnected society, especially, to understand human mobility. Movements can be collected with various techniques (e.g. GPS and GSM) generating huge quantity of data. As consequence, the analysis of mobility has been promoted as a worth application. It is important to point out that "places" are not only geographical objects, but they are also part of people lives. Indeed, there is a relationship between how people's movements are affected by visited places, and how the places themselves are characterized and connected by human mobility. Our task is to study this relationship.

We name mobility complexity the knowledge governing the interplay between drivers and places and we model the relationship among them with a bipartite graph called Driver-Place network. There is a link between a driver and a place if the driver visited that place. We used an ad-hoc ranking method called Method of Reflection (MOR) to evaluate the nodes relationship.

By mapping MOR on the Driver-Place network we obtain a mutual reinforcing definition of complexity. The produced scores are the indicators of drivers and places complexity.

A driver with a high mobility complexity visits places with an average high mobility complexity; a place with a high mobility complexity is visited by drivers with an average high mobility complexity. Finally, we used community discovery on the network projections to study how groups of places and drivers are characterized in terms of mobility complexity. The Driver projection connects two drivers if they visited at least a common place, while, the Place projection links two places if they were visited at least by a common driver.

We analyzed Pisa and Florence mobility through GPS and GSM dataset. The mobility complexity of both drivers and places characterizes the dataset movements. In addition, the mobility complexity distributions are long tailed different ways in the two provinces. Using mobility complexity, we can distinguish a large set of common drivers, a consistent set of very systematic drivers and several explorers. Symmetrically, also places can be differentiated according to their complexity and to the complexity of their visitors. Moreover, we found that the more complex a community is, the weaker are the ties among their nodes: heterogeneous community; while the less complex a community is, the stronger are the links: homogeneous community.

An Estimation of the Shortest and Largest Average Path Length in Graphs of Given Density

László Gulyás, Loránd Eötvös University, Hungary
Gabor Horvath, Loránd Eötvös University, Hungary
Tamas Cseri, Loránd Eötvös University, Hungary
George Kampis, Loránd Eötvös University, Hungary

Abstract

Many real world networks (graphs) are observed to be 'small worlds', i.e., the average path length among nodes is small. On the other hand, it is somewhatunclear what other average path length values networks can produce. In particular, it is not known what the maximum and the minimum average path length values are. In this paper we provide a lower estimation for the shortest average path length (l) values in connected networks, and the largest possible average path length values in networks with given size and density. To the latter end, we construct a special family of graphs and calculate their average path lengths. We also demonstrate the correctness of our estimation by simulations.

Analysis of a Network of Hungarian Political Actors

László Gulyás, Loránd Eötvös University, Hungary
Zsolt Jurányi, Loránd Eötvös University, Hungary
George Kampis, Loránd Eötvös University, Hungary

Abstract

"aHalo" (theNetwork) is a web portal maintained and run by the K-Monitor anti-corruption NGO of Budapest, Hungary. It collects information about the network of actors of Hungarian politics (or more generally, public life), both persons and organizations. The project capitalizes on a variety of information sources, including, among others, news articles, company ownership, public tender reports and court documents: public tenders are downloaded from official databases and processed automatically, news reports are collected and analysed manually, etc. The emerging result is a continuously growing, rich but diverse set of information about Hungarian political actors and their internal relationships. The proposed presentation will provide a first aliscor of the database, currently holding 169.459 nodes and 1.748.040 (directed) links, with the toolset of complex network analysis. In particular, it will detail the continuing efforts to present the network to the public in a fo!rm that is accessible and comprehensible to the visitors of the web portal who are untrained in social network analysis.

Social media mining for business applications

Alexandra Ioanid, University Politehnica of Bucharest, Romania
Andrei Niculescu, University Politehnica of Bucharest, Romania

Abstract

Social media platforms represent an important source of user-generated data that can be used by companies to raise their market share. Social media data is chaotic, as well as variable in size and form, so it requires special data mining techniques to gather and analyze it. Analyzing big data generated by targeted customers helps companies to understand the needs of their customers and to create brand awareness. The current research paper focuses on the recommendation power of influential users, after analyzing the predictive rankings of interest level in certain products. Interest level in certain products are measured after a big amount of public user-generated data is being analyzed, taking into consideration positive and negative reactions to certain activities that take place in a social network. Using as methodology the exploration of recognized literature sources, the paper analyzes the practical implications of social network data mining for managers.

Ensemble selection for enhancing graph coarsening quality

Rushed Kanawati, Universitè Paris 13, France

Abstract

Graph coarsening refers to the process reducing the scale of a graph by first partitioning graph nodes sets then merging each obtained partition in one node. This process has been widely applied in order to reduce the complexity of large-scale graph analysis. The idea is to reduce the graph scale allowing to apply suitable algorithms on the coarse graph then expand the results to the initial graph settings. For exemple, in order to partition a graph using a robust community detection algorithm, one can first coarse the target graph, apply the selected community detection algorithm on the reduced graph then expand found communities in order to find communities in the initial target graph.

One coarsening approach consists on applying a very fast but highly unrobust community detection algorithm on the target graph n times. One known example of fast unrobust community detection algorithm is the label propagation algorithm. n different partitions of the graph is then obtained. An ensemble clustering approach is then applied in order compute a consensus clustering out of the n basic clusterings. However, the quality gain that can be obtained from applying ensemble approaches is known to be tightly linked to both quality and diversity of the clusterings provided at the input. Most of existing work simply ignore this important issue of ensemble selection.

In this paper we propose a graph-based ensemble selection approach that allows to take into account both criteria of quality and diversity. The basic idea is to construct a similarity graph on the set of available clusterings. Different clustering similarity measures can be used (ex. NMI,VI, ARI, etc) and different similarly graphs constructing approaches can also be applied (ex. K-nearest neighbors graphs, e--similarity graphs, Relative neighborhood graphs). A community detection algorithm is applied on the obtained graph. The output is a clustering of the set of clusterings of the initial graph. The final ensemble to be used is then constructed by taking from each cluster of clusterings the one that maximize a given quality function (ex. The modularity). This ensures having a set of clusterings that are of high quality and with great diversity. Experiments on both real and artificial network datasets show the validity of our approach.

Applications of Social Network analysis in Research & Innovation Policy: Experiences across Europe

Martina Kauffeld-Monz, iit - Institute of Innovation and technology, Germany

Abstract

There is found to be a lots of questions with respect to R&D and innovation that Could be Studied with Instruments of Social Network analysis. This results from Innovation Systems approaches (regional Systems of Innovation, national Systems of Innovation, Sectoral Systems of Innovation, system innovations) that determinate Innovation policy in most European countries since the last to decades. This Paper 1) Reviews how Network analysis is Applicated in German Research&Innovation Policy and Its evaluation compared to other European countries. 2) Refering to this essential shortcomings and further options of development are outlined. 3) Finally the question is aliscor, how the discussion about big data will draw the topic of large networks and network analysis in general to rising attention of innovation Policy.

Publication productivity and publication excellence of researchers in small scientific community

Luka Kronegger, University of Ljubljana, Slovenia
Franc Mali, University of Ljubljana, Slovenia
Rok Platinovšek, University of Ljubljana, Slovenia
Anuška Ferligoj, University of Ljubljana, Slovenia

Abstract

The questions how to be among the most productive, who are the best evaluated researchers and what makes them so successful may intrigue members of scientific community. It concerns not only researchers but also policy makers, founders and other stakeholders who basically run modern science.

To study factors that could stimulate of inhibit publication productivity and scientific excellence as evaluated by national research agency we used data of complete scientific bibliography of all Slovenian researchers having research ID in Slovenian Research Agency. These factors are: 1) processes of research collaboration of each researcher, 2) allocation of financial means for R&D, and 3) differences on external R&D evaluation procedures among scientific disciplines. Two response variables (publication productivity and publication excellence) were analysed separately using multilevel regression models with information on affiliation to research field, several network and demographic characteristics measured on the level of each researcher and fragmentation of R&D funds measured on the level of scientific disciplines. Semi-continuous distribution of response variables slightly complicated the use of methods but enabled us to respond to four different research questions: which factors influence publication of the al least one scientific contribution, which influence the productivity, which factors contribute to publish the first excellent publication and what makes a scientists the best scientists.

Large Network Collections: The Power of Many Datasets

Jerome Kunegis, University of Koblenz-Landau, Germany

Abstract

In this talk, I will present multiple analysis methods that can be performed on network dataset collections which contain a large number of different datasets. Although network datasets (both social and otherwise) are used in a large fraction of studies in data mining and other areas, many papers base their work on only a single, or a small number of datasets. While this is adequate when answering research questions about a specific community or a specific dataset, the use of individual datasets cannot give insight into network analysis problems as a whole. In particular, statements such as 'social networks are scale-free', 'hyperlink networks have larger diameters than social networks' and 'the matrix exponential is a good algorithm for predicting the formation of new links' cannot be answered in a generalizable way using only a few datasets. This talk will thus review multiple recent results from research performed at the Institute for Web Science and Technologies (WeST) at the University of Koblenz-Landau based on the Koblenz Network Collection (KONECT), a collection of 230+ network datasets of varying sizes, covering many different graph types and application areas. These recent results cover the verification of well-known graph models, structural differences between graphs from different categories (as in hyperlink vs social network), and the performance of link prediction-like algorithms.

Elite Universities form a core

Vito Latora, Queen Mary University of London, UK

Abstract

The Engineering and Physical Sciences Research Council (EPSRC) is the main agency for funding research in engineering and the physical sciences in the UK. In this talk we investigate how funding has shaped research partnerships by examining projects funded by EPSRC over the past three decades.

A net of Moving People: Network Analysis of International Migration Flows

Valerio Leone Sciabolazza, University of Rome La Sapienza, Italy

Abstract

This paper delves into the determinants of global migration flows by applying a social network methodology. In order to explicitly address the multidimensional aspects of this phenomenon, we employ data on bilateral flows between 196 countries from 1990 through 2010 in combination with information about language and colonial history of each population, and bilateral distance between countries. Our aim is twofold. First, we integrate geographic contextual effects into a broader framework with respect to gravity models, so to consider the spatial dimension of the network as both produced by and productive of social relations. In order to control for different effects, we also analyze (dis)assortative phenomena at different geographical distances. Second, we control for the indirect effects on sending and receiving countries operating through the network evolution, in order to take into account the alteration of the structural conditions under which migration takes place, i.e. immigrants are not indifferent about destinations given the historical migration of their country. After a preliminary evaluation of the data using a Bayesian approach, so to find the correct specification of our models, our analysis are performed using exponential random graph techniques (ERGM, STERGM). The expected result of this paper is to contribute to the body of knowledge with a deeper explanation of migration evolution by grasping empirically its spatial and temporal processes.

A computational approach to the study of EU case law: analysis and implementation

Nicola Lettieri, University of Sannio, Italy
Delfina Malandrino, University of Salerno, Italy
Armando Faggiano, University of Salerno, Italy
Carmine Spagnuolo, University of Salerno, Italy

Abstract

In the last few years, network analysis (NA) has acquired a central role in many different research areas ranging from computer science to the study of social phenomena. In the legal field, in particular, new insights are going to be produced by the intersection between graph theory, NA techniques and legal science research questions. This process increases the need of both theoretical approaches and tools. The paper presents an ongoing research project aiming to explore the potentialities of NA in supporting the study of EU case law. The goal is to define a computational approach offering new ways to measure the relevance of legal precedents, also experiencing new methods of visual information retrieval. According to this aim, the project has moved in a twofold direction. The first one consisted in the design of an instrument allowing to apply NA techniques to the entire corpus of EU case law (judgments, orders, opinions and conclusions of the General Advocate of the Court !
 of Justice and of the General Court, from the date of their creation) in XML format. Data, made available by the EUR-Lex database, are represented by the full texts of the documents of interest and by a rich and varied set of metadata (e.g. related legal doctrine). The second direction, still to be explored, aims to devise new metrics and algorithms to be validated according to the state of the art of legal literature. After a brief introduction to the state of the art of the application of NA techniques in the legal field, the paper presents and discusses both the domain-related and technical issues of the project sketching an overview of ongoing works and future directions.

The role of social support networks in everyday coping strategies of single mothers: a multilevel analysis

Rosaria Lumino, University of Naples Federico II, Italy
Giancarlo Ragozini, University of Naples Federico II, Italy
Maria Prosperina Vitale, University of Salerno, Italy

Abstract:
There is a strong research evidence showing the crucial role of social support networks in defining successful coping strategies by female-headed families, especially in South European Countries. In such a context, single mothers are more vulnerable to the risk of social isolation due to the lack of childcare, the limited rights to even a low subsistence income and the labor market rigidity -which discourages women's participation in the work force.

The present contribution aims at analyzing social support networks of single mothers by considering the role of kinship and friendship ties as essential sources of support and generators of social integration. An ego-centered network approach is adopted to reconstruct social support networks of a sample of single mothers in Naples. The pattern of relationships on which social support networks are based, within and across family boundaries, are analyzed by means of multilevel logistic regression models. More specifically, we aim at understanding: i) on which kind of social support networks single mothers rely for daily living; ii) which characteristics of ego and alteri explain the presence of non-kin among alteri in the personal networks; iii) which factors affect the strength of ego-alteri ties in terms of closeness and of importance.

Understanding the development trends of renewable energy technologies: A patent network analysis.

Marianna Marra, University of Salerno, Italy

Abstract

The governance of renewable energy technologies is a hot topic in developed countries and emerging economies. This paper aims to shed light on innovation and technological dynamics of the renewable energy industry using EPO patent data published between 2000 and 2013 in Europe, Japan and US. The dataset consists of directed citations network of 18,000 patents. As results, (i) by using the critical path algorithm (CPM) the main path of the technological development is traced, (ii) the most successful inventions and (iii) their best developments are identified by the hubs and authorities algorithm, (iv) the most active countries and inventors are identified, (v) finally policy implications for the governance of renewable energy industry are presented.

Visualizing innovation networks: An overlay mapping analysis approach to Internet of Things (IoT)

Marianna Marra, University of Salerno
Stefano Perna, University of Salerno

Abstract

The importance of investing in emerging and innovative technologies is widely recognised by governments. Understanding dynamics of these technologies is also important for both decision and policy makers. This paper investigates the development of the scientific domain of the so called Internet of Things (IoT) by means of a scientometric overlay mapping analysis (Leydesdorff, Porter, Rafols) and by tracing its intellectual and technological trajectory. We examine a sample of academic publications retrieved by Isl Web of Science and patents retrieved by USTPO (United States Patent and Trademark Office). Results show - through a set of maps,
data visualizations and network graphs - the geographical, social and cognitive space of this technological domain allowing the decision-maker to do inferences about dynamics of emergence and future trends of development. Finally policy implications are presented.

Scalable algorithms for analyzing large networks in NetworKit

Henning Meyerhenke, Karlsruhe Institute of Technology (KIT), Germany
Christian L. Staudt, Karlsruhe Institute of Technology (KIT), Germany
Elisabetta Bergamini, Karlsruhe Institute of Technology (KIT), Germany
Moritz von Looz, Karlsruhe Institute of Technology (KIT), Germany
Aleksejs Sazonovs, Karlsruhe Institute of Technology (KIT), Germany
Roman Prutkin, Karlsruhe Institute of Technology (KIT), Germany

Abstract

We present NetworKit, an open-source software package for high-performance analysis of large complex networks. Complex networks are equally attractive and challenging targets for data mining, and novel algorithmic solutions, including parallelization, are required to handle datasets containing billions of connections. Our goal for NetworKit is to package results of our algorithm engineering efforts and put them into the hands of domain experts. NetworKit is a hybrid combining the performance of kernels written in C++ with a convenient Python frontend. The package targets shared-memory platforms with OpenMP support. The current feature set includes various fundamental analytics kernels as well as a collection of graph generators. Selected algorithms from NetworKit for community detection, semi-dynamic betweenness centrality, and the generation of hyperbolic unit disk graphs will be presented in some detail. In comparison with related software, we propose NetworKit as a package geared towards large networks and satisfying three important criteria: High performance, interactive workflows and integration into an ecosystem of tested tools for data analysis and scientific computation.

Predicting economic networks: Gravity or Radiation?

Lorenzo Napolitano, Scuola Superiore Sant'Anna, Pisa, Italy
Pietro Battiston, Scuola Superiore Sant'Anna, Pisa, Italy
Giorgio Fagiolo, Scuola Superiore Sant'Anna, Pisa, Italy

Abstract

This paper investigates the relative performance of gravity vs radiation models in explaining observed macroeconomic network structures and predicting their out-of-sample evolution. We employ longitudinal data about a class of macroeconomic networks, including international trade, migration and mobility, and finance. Traditionally, prediction of this class of networks has been performed using the gravity model (GM), which is considered the workhorse when it comes to replicating empirical data in the above-mentioned contexts. The GM owes part of its success to the very parsimonious use of covariates: typically, a set of geographically located economic entities, a measure of individual node size or importance and an array of pairwise distances are enough to obtain a basic estimation. More recently, however, the radiation model (RM) has been proposed as a possible alternative. The RM shares with the former a minimal specification setting but it often outperforms the GM when it comes to replicating properties of observed data. This paper employs various specifications of the two models in an attempt to compare them both in terms of their ability to reproduce the properties of empirical bilateral flows and to yield accurate out-of-sample predictions as far as the topology of economic networks are concerned. Furthermore, we put our results in perspective, discussing how the GM and the RM compare to null random-network and p* models.

Depression Ideation in VK.com: Statistical and Network Approaches to Detection

Alexey Natyokin, Data Mining Labs, Russia
Aleksandr Semenov, National Research University Higher School of Economics, Russia
Philipp Upravitelev, Consultant Plus, Russia
Mikhail Trofimov, Russia
Maxim Kharchenko, Russia

Abstract

Online social networks provide a rich amount of information about users: their profiles, social connections and their wall posts. This information can be further used to predict high-level features of user behavior, including their traits of character. Recent publications used data from online social networks to detect people with depression ideation and diagnosis. However these papers used data from different online social networks (Facebook, Twitter and Mixi) and biased towards Japan and English speaking users.

In this article we implement methods and metrics, used in these studies to Russian online social network VK.com in order to test the models and compare the results.

For this purpose we gathered data from the profiles of users from most popular communities about suicide and depression on the vk.com and compared their features with random users.

In our analysis we used not only standard attributes of a user as age, sex, number of friends but also structural properties of their egocentric networks.

We analyze the resulting classifiers to identify phenomenological hypotheses on features that distinct users with depression ideation in order to provide tools for early depression ideations detection.

Algorithms for Backbone Extraction in Large Social Networks

Mark Ortmann, University of Konstanz, Germany
Arlind Nocaj, University of Konstanz, Germany
Ulrik Brandes, University of Konstanz, Germany

Abstract

A generic approach to the identification of cohesive groups in social networks consists of the extraction of sparse subnetworks that have a more pronounced variation in local density and are thus easier to cluster. Simmelian backbones, for instance, are obtained by locally adaptive filtering of structurally embedded ties and thus implement the assumption that redundant ties are more likely to link actors within the same group.

We present efficient computational methods for determining the degree of structural embeddedness and the extraction of sparse subgraphs based on various criteria. Since the efficiency of these methods depends on graph invariants relating to local density, we provide empirical evidence on their range in networks obtained from social media.

Healthcare and I.C.T. in the modern Southern Italy's landscape

Francesco Pascuzzo, University of Salerno, Italy

Abstract

An essential aspect of clinical governance, about security and e-health, is undoubtedly that of the mapping of diseases in the territory. A certain policy of prevention by health authorities of the area depends on epidemiological and demographic variables are important. In particular it has been investigated the case of the geographic information system (GIS) used by Cotugno Hospital in Naples for traceability of some infectious diseases arising from immigration. In a geographic area as well as the densely populated Naples and its surroundings, spreading of diseases caused by bacteria and viruses from direct contact is very rapid and should therefore be kept under close observation. Today it's true, when we see that Mare Nostrum is bringing thousands of migrants from North Africa and the Middle East. From Salerno, a city chosen as one of the main ports for the reception of migrants, has set off the alarm infections following the arrival of illegal immigrants fleeing wars and poverty. The situation in Naples, has long been the destination of immigrants mainly from Senegal, Bangladesh, Pakistan and Nigeria, is monitored by Prevention Department of A.O. Cotugno. The high rate of infection of certain areas of the capital city of Campania is the reason why we decided to use the Gis system, which uses mostly images digitally, through the provision of prevention methods referred to by the abbreviation CREG, similar to the Drg used into the hospitals. Hence the careful clinical risk analysis of certain treatments, in view of the delicate situation that today, especially after the outbreak of Ebola in Africa, is to be kept strictly under control. The screening of a disease of African origin as the Hiv virus also is at the heart of the work of Cotugno served then that mapping of migrants present in the city. Vaccinations, careful assessments of demography, mortality rate, dependencies on other phenomena such as alcoholism (for cirrhosis of the liver, for example) and other elements of analysis are essential tools to better meet certain health needs of the population. Facilitating decision-making processes the Gis system proves to be a useful ally in the fight against infections and epidemics. It is in particular due to its nature as a management tool, read and understand complex phenomena such as infections by bacteria or viruses.

Mapping Longitudinal Network in Twitter Conversation

Stefano Perna, University of Salerno, Italy
Giuseppe Giordano, University of Salerno, Italy
Pierluigi Vitale, University of Salerno, Italy

Abstract

This work aims at analyzing and visualizing online social network dynamics. We consider Twitter conversations and we extract network data related to a specific main topic of interest.

We are interested in collecting new data as new tweets are posted so that we collect a relational data flow as a continuous process observed in discrete points in time.

According to specific data analysis, we extract the main features of such process and transform the whole network as a single point projected onto the subspace spanned by the network features. The network evolution can be analyzed through the different trajectories that successive projections leave on the mapping.

Empirical analysis and dynamic tools of data visualization will be used to put emphasis on classical conversation-prototype that are found in liquid environments as online social networks are.

The Evolution of Inter-organization Networks in Competitiveness Clusters in France

Fabienne Picard, UTBM Economics, Trévenans France
Sana Mrizak, UTBM Economics, Trévenans France

Abstract

There is an increase in the number of studies applying network theory and network analysis to economic geography. These applications consider three relevant topics: Inter-organization networks in clusters, Regional innovation systems, Agglomeration Economies.

Our objective in this work is to use network analysis to better understand inter-organizations networks in clusters in France.

In few words, the cluster literature asserted that regions are drivers of innovation and economic growth: firms in clusters benefit almost automatically from knowledge externalities that are 'in the air', as Marshall once put it. This is because tacit knowledge travels more easily across short distance, and shared institutions at the cluster level further facilitate the effective transfer of knowledge. In this work, we analyze how the configuration of network in a cluster evolves over time and we try to find what motives do explain that?

To reach our goal, we consider the collaborative projects provided by the Competitive in France. Our study shows that the inter-organization network becomes more denser across time. Moreover, we analyze the concentration of the network through cliques and cohesive groups. The goal of this analysis is to check whether the clustering around several projects coincide with the innovative niches in each region. Finally, our study puts into emphasis the role of the competitiveness cluster in the smart specialization strategy.

Unfolding Networks. Communication Design strategies for public engagement in controversy mapping

Azzurra Pini, Politecnico Milan, Italy

Abstract

Network data have been by far one of the most common ways to represent and analyze different aspects of reality. Nevertheless, their understanding and interpretation for non-expert users is still a challenging task, due to the complexity of their numerous linked entities and the resulting dynamics.

The increasing developments on analysis methods as well as the availability of large databases has also provoked the potential escalation of graph complexity, calling for effective solutions for the communication of complex networks. Most of the current perspectives on network visualization have been approaching this issue by developing algorithms and visual solutions for reducing the visual clutter created by the massive number of entities and links, and increase the readability of large graphs. A number of contributions also explored possible alternatives to graphs by developing different visual patterns. Only a few contributions though took on the communication aspect from a design point of view, specifically addressing a public of non-expert users.

The present contribution introduce a different approach which, working on the graph configuration, focuses on interaction techniques in order to unfold the construction steps of the network and to provide the user the tools to autonomously explore it.

The two projects presented have been developed within the framework of controversy mapping[2], a research method that provides a series of tools and techniques[3] to allow the investigation of socio-technical debates. This method is centered on the elaboration of networks of actors, emerged on the web around controversial topics, often resulting in particularly complex visualizations. For the public communication of such analyses the mere visualization not suffice: most users claim a lack of readability and understanding, due to a lack of background knowledge.

A major problem that was highlighted during these projects concerns the need to engage users through two main different aspects. The first one consists in opening access points to the users, by providing the context of the searched issue as well as the analysis tools to interpret the graph and its components (e.g. nodes, links, degree, etc.). On the other hand the necessity of building confidence, by making the method understandable and extensively describing the research protocols and theoretical assumptions.

The research presents in detail the specific problems in the comprehension of the network visualizations emerged during a series of workshop sessions set up with the users. The criticalities have been tackled in a second phase, where the development of an interactive platform allowed to set up the context of the analyzed issue and to systematically present the methodology applied as well as the type of data. More importantly, the interactive platform supported both the exploration of the network visualizations in its different components (clusters, categories, etc.) and provided some interpretations to bolster the identification of patterns and the creation of insights.

---

2   http://www.mappingcontroversies.net/

3   https://wiki.digitalmethods.net/Dmi/ToolDatabase

Health migration and performances of Italian health system: An exploratory Social Network Analysis

Claudio Pinto, University of Salerno, Italy

Abstract

This work concern the application of Social Network Analysis to explore the structure of health migration networks Considering the type of care and the regime of care we define five direct health migration networks with 21 actors. For each health migration network a sociogram is displayed offering the visualization of health migration networks among three geographical area, also. The performance's indicator cover several dimension as efficiency, access, quality of care, and organizational appropriateness. One of the assumption is that the health migration networks are also a communication network inside of which different type of information circulate, assuming that each migration's patient travel with own information set on clinical conditions, pathology, cost of treatment among the other possible. General information on the networks are density (with and without loops) and average degree, and structural analysis conduct triadic census and lines values. Density and average degree suggest that the more cohese networks are the acute, in ordinary and day hospital regime, and the rehabilitative in ordinary regime. Triadic census show that the 300 is the most frequent in acute in ordinary and day hospital health migration networks and, together 210, these two triad characterize the rehabilitative in ordinary regime health network. The remaining health migration show also different type of triad. Betweenness centrality suggest that the central positon role of regional health care systems change inside each networks. Correlation analysis reveal ,in general, a positive effect among access, efficiency and quality performance's indicators and betweenness centrality measure. However, very interesting is the correlation among the betweenness centrality measure and each single indicators for each specific health migration network. For example a positive correlation exist between betweenness centrality measure and the performance's indicator "Rate of hospitalization for not controllable diabete" in the acute (in ordinary and day hospital regime) and rehabilitative (in ordinary regime) health migration network, but a negative correlation exist in the case of rehabilitative (in day hospital regime) and long-term care health migration network with the same performance's indicator. Concluding the work, trough structural analysis, betweenness centrality measure, and correlation analysis as here developed, suggest interesting consideration for public policies indications, as well as the work show as the SNA is an useful instrument to deal with public policy and public health issues.

The Effects of Civility on Advice, Leadership, and Performance

Christine Porath, Georgetown University, USA
Alexandra Gerbasi, Grenoble Ecole de Management, France
Sebastian Schorch, Grenoble Ecole de Management, France

Abstract

Workplace incivility is rampant and on the rise—with costs to individuals and organizations. Despite the increased need for civility, little is known about potential benefits of civility, defined as behavior involving politeness and regard for others in the workplace, within workplace norms for respect. Research suggests that being civil may be hazardous to influence, power, and income. Yet, throughout history, civil behavior was extolled because it paid back dividends. Our focus is whether that holds true in organizations. Using social exchange theory, we investigate this in two studies. First, in a two-wave social network study of a research and development department (n = 31) of a biotechnology firm, we apply stochastic actor oriented models (SIENA), and find that people who perceive a member as civil, will be more likely to seek that member out for work advice and see him as a leader. Then, to test the effects of cumulative network ties on performance we estimated a series of nodal regression models. We find that the more the employee is perceived as civil by others in his or her network, the better his or her performance. Being sought for work advice and being viewed as a leader mediate this effect. In a second experimental study (n = 162), we extend our understanding of what drives these benefits. We find that people who are civil are perceived as warm and competent-- even after controlling for positive emotions. These positive perceptions explain why those perceived as civil are more likely to be sought for work advice and seen as leaders. Civility is fairly unique in drawing perceptions of both warmth and competence. By evoking perceptions of both warmth and competence, civility may enhance influence and effectiveness in the workplace. Through respectful engagement employees increase the likelihood that others seek—and presumably exchange—information and advice, which in turn, increase performance. Rather than hurting themselves by appearing weak or deferential, behaving respectfully seems to garner influence—paving the way for future interactions and opportunities (e.g., recommendations for jobs). Contrary to some theorizing and research, our findings suggest that through civility, people may be perceived as leaders. Our research suggests that civility pays. Given the potential consequences, civility should not be ignored—but honed to manifest connections, relationships, and influence. We discuss theoretical and practical implications.

On the sampling distribution of the ML estimators in Network Autocorrelation Models

Giovanni C. Porzio, University of Cassino and Southern Lazio, Italy
Patrick Doreian, University of Pittsburgh, USA, University of Ljubljana, Slovenia
Michele La Rocca, University of Salerno, Italy
Maria Prosperina Vitale, University of Salerno, Italy

Abstract

This work investigates the finite sample properties of the maximum likelihood (ML) estimators for network autocorrelation models (NAMs), a class of auto-regressive models used to study the effect of networks on dependent variables of interest when the data points are interdependent.

It is known that the estimated autocorrelation parameter has a finite sample negative bias, the amount of which is positively related with the network density (Mizruchi, Neuman, 2008, 2010). A recent study on the statistical power of a test on the same parameter is also available (Wang et al. 2014).

We examine the whole finite sample distribution of both the ML estimator of the autocorrelation parameter and the regressor parameters. More specifically, through an extensive simulation study, this work investigates whether – and the conditions under which – the ML estimators are normally distributed. The finite sample distributions are evaluated with respect to the network density and topology, the distribution of error terms, and the strength of the autocorrelation parameter. It turns out that the ML estimators of the autocorrelation parameter and of the intercept are not normally distributed in case of small sample size, even in presence of normally distributed errors.

Furthermore, the network density has some effect on the variability of the estimators. On the other hand, it seems that other features of the network topologies, in the main, have little effects on the estimator distributions.

Also, proper methods to deal with the bias of autocorrelation parameter are introduced and studied. Particularly, a residual-based bootstrap is proposed, in line with a related literature (Lin et al., 2011; Yang, 2013). A further simulation study shows that the bootstrap based distributions are more accurate and should be preferred in case of low density and moderate network effects.

Substantive uses and interpretations of blockmodeling inconsistencies

Laura Prota, University of Salerno, Italy

Abstract

A blockmodel is a hypothesis on the roles and positions of a social system with blockmodeling a set of methods for fitting a blockmodel. When pre-specified blockmodeling is used, this hypothesis is made explicit by defining an ideal role system a priori and fitting it with empirical data. The number of inconsistencies between the ideal and the observed blockmodel is not solely a measure of fit. In addition, the location and the structural implications of inconsistencies can be evaluated from a substantive perspective. In this study, pre-specified blockmodeling is used to fit an ideal multiple-cores configuration with data on R&D collaborations linking firms and research centers in a technological district of Italy during 2005-2013. Explicit hypotheses were made in advance on the expected location of inconsistencies and on their structural implications with respect to structural change. Results show that the evolutionary trajectory of R&D collaborations changed over time by approaching initially a core-periphery model before forming a small-world system.

Quali-quantitative methods for the analysis of a regional innovation experience: the case of Imast Technological District

Laura Prota, University of Salerno, Italy
Maria Prosperina Vitale, University of Salerno, Italy
Eva Milella, IMAST S.carl, Italy
Maria Rosaria D'Esposito, University of Salerno, Italy

Abstract

This work aims at exploiting the potentials of both qualitative and quantitative methods to understand whether social and cognitive proximity overlap in regional innovation networks. The showcase of IMAST Technological District in Campania region is examined. First we undertake an ego-centered network analysis of the actor-by-project collaboration network to derive social proximity measures and trace their changes over the period 2005-2013. Second, we adopt a lexical correspondence analysis on qualitative in-depth interviews gathered from managers of all public and private research institutions and firms in the district. Finally, we triangulate these social and cognitive proximity measures to understand to what extent these two dimensions overlap over time. Results highlight the importance of interpreting network collaboration data in the light of their social context.

Prototyping Networks Through Archetypes

Giancarlo Ragozini, University of Naples Federico II, Italy
Maria Rosaria D'Esposito, University of Salerno, Italy

Abstract

In this paper we propose a method to analyze and synthetize a set of N networks, that refer to a common scenario and that are comparable among each other. Examples of this type of data are set of collaboration networks, each defined for a different scientific field, or set of ego networks, where egos belong to a same category, set of governance networks, etc. For this set of networks can be of interest to find a small number of representative networks that can serve as a distillation or condensed view of the data set. In a statistical perspective this goal amount to find a set of prototypes (where a prototype is thought an ideal exemplar that summarize and represent a group of data, or a category, in terms of their most relevant features and their specificity in contrast to other groups or categories). The prototypical networks can serve as benchmarks for the other networks and are useful also to compare networks among each other.

Given the set of N networks, we propose to find a set of prototypical networks by using the tools proper of archetypal analysis. In details, the procedure we propose is as follows:

i)   describe each network through a vector of p parameters characterizing the whole network structure and topology;

ii) map each network in a multivariate Euclidean space Ep by using the vectors of parameters in point i);

iii) find in E^p a set of archetypes of the N data points corresponding to the original N networks;

iv) find the prototypical networks by using the tools proper of the archetypal analysis.

We exemplify the proposed procedure by analyzing a set of 48 governance networks of public structures devoted to provide youth services and referring to 48 different territorial districts in Campania region in Italy. Our results highlight the presence of different network structures that can be interpreted in terms of the governance forms established in literature.

The Forward Search for Very Large Datasets

Marco Riani, University of Parma, Italy
Andrea Cerioli, University of Parma, Italy
Domenico Perrotta, ISPRA, Italy

Abstract

The identification of atypical observations and the immunization of data analysis against both outliers and failures of modelling are important aspects of modern statistics. The forward search is a graphics rich approach that leads to the formal detection of outliers and to the detection of model inadequacy combined with suggestions for model enhancement. The key idea is to monitor quantities of interest, such as parameter estimates and test statistics, as the model is fitted to data subsets of increasing size. In this paper we propose some computational improvements of the forward search algorithm and we provide a recursive implementation of the procedure which exploits the information of the previous step. The output is a set of efficient routines for fast updating of the model parameter estimates, which do not require any data sorting, and fast computation of likelihood contributions, which do not require any inverse matrix or qr decomposition. It is shown that the new algorithms enable a reduction of the computation time by more than 80%. Furthermore, the running time now increases almost linearly with the sample size. All the routines described in this paper are included in the FSDA toolbox for MATLAB (http://www.riani.it or http:// http://fsda.jrc.ec.europa.eu) which is freely downloadable from Internet.

Functional data Analysis for longitudinal network data

Elvira Romano, Second Unievrsity of Naples, Italy
Giuseppe Giordano, University of Salerno, Italy

Abstract

This proposal describes challenges of using functional data analysis (FDA) for the exploration and analysis of longitudinal social network data, as well of metadata about network dynamics. We discuss the special data structures that arise in longitudinal social network data and why FDA is a suitable approach for analyzing such data. Due to the dynamics and complexity of data structure that arise in longitudinal networks, we propose a functional model for describing and monitoring data patterns. We describe an application originating from social network and discuss some promising future directions.

A fast subgraph isomorphism algorithm for social networks graphs

Alessia Saggese, University of Salerno, Italy
Vincenzo Carletti, University of Salerno, Italy
Pasquale Foggia, University of Salerno, Italy
Mario Vento, University of Salerno, Italy

Abstract

Social network data is naturally represented using graphs because of its intrinsic structure composed of a set of entities (people or organizations) and the relationships among them, being so represented as the nodes and the edges of a graph.

The search for substructureswithin a structure is a recurring problem in several applications. If the structures are represented as graph, subgraph isomorphism is a technique that is used very often for this task. A subgraph isomorphism algorithm attempts to find a mapping between the nodes of a so-called pattern graph and a subset of the nodes of a larger target graph such that the structural interconnection information is preserved (i.e. the mapped subset of the target has the same structure as the target). In the general case, subgraph isomorphism has been demonstrated to be an NP-complete problem, and thus the in worst case it requires an exponential time; however, there are several algorithms that are fast enough on the graphs obtained from real applications to be practically usable. However, the effectiveness of these algorithms depends on the structural properties of the considered graphs: there is no single algorithm that definitely outperforms the others on every kind of graphs. Thus it is important to choose an algorithm that is tailored for the kind of graphs the application is expected to work with. In graphs obtained from social networks, the probability of the presence of an edge between any two nodes has not a uniform distribution. Several algorithms presentin the literature have been evaluated on uniform random graphs; however their performance on such graphs is not indicative of their effective usability on social network data. Two probabilistic models that have been experimentally found to reflect the edge distribution of several kinds of social networks are the scale-free graphs, introduced by A. Barabási, and the small-world graphs, introduced by Watts and Strogatz.

In this work, we propose a novel subgraph isomorphism algorithm that has been designed so as to be particularly effective on these kinds of graphs, and thus is well suited to applications of social network analysis. The algorithm is based on a backtracking approach, exploring the search space in a depth-first order. The algorithm is designed to have a low memory footprint with respect to the number of graph nodes, so as to be able to process very large graphs, that are often encountered in social network analysis. Several heuristics have been devised to prune the search space, using both structural information and the labels attached to the nodes and edges of the graphs. In order to keep low the time required to process each search state, the implementation of these heuristics relies on the use of a pre-computed reordering of the pattern nodes, so as to optimize the search space visit, and on a node classification procedure that partitions the nodes into disjoint sets so as to reduce the dimension of the sets considered by the algorithm for each intermediate state of the search. The effectiveness of the algorithm has been assessed by an experimental evaluation over two datasets of scale-free and small-world graphs, in comparison with other state-of-the-art subgraph isomorphism algorithms; the results confirm the advantages of the proposed approach.

Which analytic methods for Big Data ?

Gilbert Saporta, Conservatoire National des Arts et Métiers, CEDRIC, France

Abstract

Classical inference is not fitted for massive data: statistical tests reject any reasonable model, confidence intervals are reduced to nothing . Model validation should be done through cross validation or split sampling. Explicit, parcimonious generative models are replaced by predictive algorithms. Model choice is driven by statistical learning theory and not by penalized likelihood. The analyst's toolbox includes revisited classical data analysis techniques (PCA, MCA as particular cases of SVD, clustering) mainly for exploratory purposes as well as machine learning methods (SVM, boosting, ensemble learning) for prediction. In the case of high dimensional data where the number of variables exceeds the number of units, sparse methods based on L1 regularization provide elegant and simple solutions; we will present a sparse generalization of multiple correspondence analysis. Is the data deluge making the scientific method obsolete, as C.Anderson claimed some years ago? We will conclude by some comments on correlation and causality.

A dynamical perspective on community detection in networks

Michael Schaub, University of Namur, Belgium, Université catholique de Louvain, Belgium

Abstract

Detecting communities in complex networks originating from various disciplines has received significant interest recently.

One of the main hopes underlying this interest is that communities can provide a simplified functional and dynamical description of the system under investigation.

However, despite this interest in the functional behavior, the dynamical information about the processes occurring on the network is often ignored when clustering networks.

While flow-based algorithms have aimed to incorporate dynamics into the clustering process, many behaviors may not be modeled by a simple di usion.

For instance, dynamics often do not conserve the total amount of flow in the system but may instead have a multiplicative character such as encountered during the spread of a disease, or a rumor.

In many cases nodes may moreover have negative (inhibitory) as well as positive (excitatory) effects on each other, as is the case in many social, neural or genetic networks. In all those scenarios standard flow-based methods are thus no longer directly applicable.

Here we will present a dynamical graph clustering framework that accounts for this diversity of dynamical dependencies when partitioning a network and can be applied to weighted, directed networks with both positive and negative weights (signed networks).

Importantly our measure can incorporate any type of dynamics that can be described by an at least marginally stable transfer function.

Using this framework we can thus account for a plenitude of physical phenomena which are not of diffusive type, and go beyond standard Markovian processes.

The quest for opinion manipulation in massive online communities: leveraging Big Data and process mining to discover coalition formation in social networks

Giovanni Paolo Sellitto, Independent Scholar, Sapienza University of Rome, Italy
Riccardo Onori, Sapienza University of Rome, Italy

Abstract

Social network platforms are social spaces where the agents (human agents and possibly programmatic BOTs) generate, exchange, diffuse and debate contents, information and opinions. Usually long-term interactions in the context of massive online social networks emerge from a sequence of repeated short-term interactions. The dynamics of coalitions' formation are part of the evolution of the network and in turn, the network structure and the contents exchanged bear an effect on the inner dynamics of the social groups, influencing the opinions, the interaction patterns and the relative positions of the agents. We apply process-mining techniques on the data gathered from On-line Social Networks like twitter or other micro-blogs to discover the co-evolution of the semantic network made up by the contents, opinions and their relations, and of the underlying social network among the agents that embrace, produce and reshape the contents. We adopt a multi-paradigm approach to integrate, process, search and visualize structured and unstructured data from various sources. The final goal of our analysis is gathering insights on the opinion manipulation phenomenon. The unstructured big data sets are handled through Apache Hadoop open-source software framework with MapReduce architecture. Based on the Dynamic Network Analysis paradigm, we explore what kind of micro-processes (e.g. reciprocity, attention or repeated interactions) may have the highest impact on the emergence of larger relations of trust between the agents interacting trough a social networking platform.

Tracking the emergence of lobbying, progressive opinion biasing or sentiment hijacking in the dynamic evolution of a social network is a tough task, and it requires the processing of Big Data and the application of machine learning, either because the data are produced with a fast rate or the number of interactions is huge. This kind of analysis is of growing importance, e.g. to avoid bully behavior or to detect polarization among online communities that can reverberate also in the context of real life.

The adequate response to these latter threats is perhaps at a higher level, through culture, education and knowledge, which require time, and possibly an ad-hoc training. Nevertheless, some cognitive tools can help us in gaining awareness of the dynamics taking shape on a social platform, and this can at least help people in the detection of some threats for the democratic outcomes like populism, lobbying, community segregation, shifting of concepts and manipulation of language.

Social Network Analysis of Russian Protest in Twitter: Topoligy and/or Sociology?

Aleksandr Semenov, National Research University Higher School of Economics, Russia
Igor Zakhlebin, Russia
Alexander Tolmach, Russia

Abstract

While there is already a decent amount of research on "Twitter revolutions" in general, there is almost no scientific publications on Russian protest movements in this social media service. We present the descriptive social network analysis of tweets, gathered during protest meetings on the 24th Dec 2011 in Moscow, Russia. The data was obtained from both Twitter Streaming API and Firehouse, which makes this dataset the most possible complete on that topic.

From both sources of data we construct 3 types of networks: "reply," "retweet" and "mention," and study their topological properties, positions and characteristics of key users, information propagation and discussion. Then we compare the results and estimate the biases, caused by different sources of data.

Among the key findings of the analysis were very low percent of discussions and information diffusion in the overall communication, almost no overlap in key persons of "reply" and "retweet" networks and noticeable "echo-chamber" effect in the overall "mention" network.

Besides that we compare our results with similar cases of social network analysis of Twitter data from other protest movements in Egypt and Spain. One of the most interesting findings is difference in degree correlation with Spanish case of 'indignados', which is reported to be disassortative, while our network have positive degree correlation, which stays the same for both Streaming API and Firehose datasets.

We discuss this and other results from both sociological and technological perspectives.

Collaboration networks in the theatre industry: Creativity and spatial patterns in stage co-productions among Italian theatres

Marco Serino, University of Naples Federico II, Italy
Valerio Ciotti, Queen Mary University of London, UK
Pietro Panzarasa, Queen Mary University of London, UK

Abstract

The association between creativity and collaboration has been widely documented by a number of studies in the social, organizational, and network sciences. Collaboration may also be facilitated by the spatial location of the partners involved in creative projects. In the theatre industry, one interesting form of collaboration, albeit so far underrepresented in the literature, is the stage co-production. This refers to the creative work designed and performed jointly by two or more producing theatres, theatre companies, festivals, or even non-theatrical organizations and institutions, such as local authorities, universities, and cultural associations. This complex form of inter-firm collaboration generates indirect linkages among organizations through chains of various co-productions. The resulting network can be thought of as a two-mode affiliation network including two classes of nodes: 1) actors (i.e., the organizations); and 2) events or affiliations (i.e., the co-product!

 Ions). Despite the vast body of literature on social capital and the effects of networks on various measures of performance, the degree to which the creativity of theatre productions is nurtured by the underlying collaboration network still remains largely unexplored. Moreover, the relationship between creativity and network structure may be mediated by the geographic location of the organizations collaborating on the joint production. In this paper we present an ongoing study of the network of co-productions involving Italian government-supported theatres. Data were collected through an Internet-based survey of both the productions and co-productions these theatres released during the 2013/2014 season. For each organisation, a number of relevant attributes were recorded, including the geographical location and the seating capacity of the venue. In addition, we defined an index of creativity based on the characteristics of each play. Preliminary findings suggest that alisc!

 Vity is indeed associated with the structural properties of the collaboration network. For instance, intense flows of communication and recurrent partnerships may favour creative achievements. Moreover, cohesive structures, rich in third-party relationships, could result in fewer opportunities for newness in artistic outcomes than open structures, rich in structural discontinuities and brokerage opportunities. Further, we tested the effect of geographical clustering, such as regional or urban-based niches of collaboration, on the creative efforts of the organisations.

On the structure of collaboration among Italian theatres

Marco Serino, University of Naples Federico II, Italy
Giancarlo Ragozini, University of Naples Federico II, Italy

Abstract

In the study of collaborations among theatres, a well-suited way of defining and operationalizing such collaboration is to consider the co-production of theatre plays as an event involving a number of co-producing organizations. This mechanism gives rise to the so-called affiliation network, a type of two-mode network characterized by a set of actors (the co-producing organizations) and a set of events (the co-productions), in which the events affiliate the actors (i.e., the actors are connected to each other through the events). In this paper, we propose an analysis of the affiliation network of co-productions involving Italian government supported resident theatres during the 2013/2014 theatre season. To perform our analysis, instead of the conversion approach – in which the two-mode network is converted in two distinct one-mode networks, each of which is analysed separately – we adopt the direct approach, in order to analyse the two modes at the same time, providing an investigation of the bipartite network that maintains its relevant properties. We make use of Multiple Correspondence Analysis (MCA) and consider the affiliation matrix as a two-way case-by-variable matrix, thanks to the similarity between the structures of the two matrices. We thus propose a joint representation of actors and events in a common space, allowing visualization and analysis of the relationships between the two modes. Furthermore, by the inclusion of both node and event attributes, we study the association between relational patterns and the characteristics of the actors and the events. As actors can be either similar or different types of organizations, we are interested in studying how they jointly participate in co-productions with regard to their characteristics, comprising their location in a regional, national or even international (for some co-producers) context. The events, i.e., the co-produced plays, may show patterns of association with one another that can be of interest according to several key attributes like theatre genre, play script, or performing style. Therefore, using these techniques we attempt to understand both relational and attribute similarities among actors and events, in order to provide an insightful exploration of the bipartite network.

Discovering SIFIs, a temporal complex approach

Alessandro Spelta, Università Cattolica, Milan, Italy
Pablo Rovira Kaltwassery, University of Leuven, Belgium

Abstract

This paper proposes an intrinsically temporal methodology to identify Systemically Important Financial Institutions (SIFIs) and to track their common activity over time. The introduction of the temporal dimension gives the opportunity to develop an early warning indicator for the risk associated to the whole network where the strength of contemporaneous spillover e ects is summarized in a single variable: the time score.

Moreover, the same solution method can be used to identify the community structure of the network and the systemic importance of banks within each community, tracking their activity through time.

The technique is based on tensor decomposition, where a temporal network is naturally represented as a time-ordered sequence of adjacency matrices, each one describing the state of the financial network at a given point in time. Our method is able to take into account both the spatial and the temporal distribution of the links that represent flows of funds between institutions. We apply the method to the e-Mid dataset, the most central banks turn out to be Italian banks. The cross-correlation between the traded volume and the time score coefficient founded by the decomposition shows how this series anticipates the movement in the overall traded volume.

Analysis of brand image with the use of network tools

Agnieszka Stawinoga, University of Naples Federico II, Italy
Simona Balbi, University of Naples Federico II, Italy
Germana Scepi, University of Naples Federico II, Italy

Abstract

In Text Mining, large corpora are explored, in order to discover and synthesize their content, in an automatic, timesaving way. If the aim is to understand the similarity of the documents in the corpus, it is interesting to represent them as a network, measuring their proximity in relation to the words they use. Therefore, statistical tools, developed for the analysis of Social Networks, can be applied. In Social Network Analysis, it is often important to introduce information related to the nodes, such as characteristics identifying their belonging to a specific group. In literature different indices have been proposed in order to measure the tendency of actors to have relations with actors similar to themselves. Analogously, in the analysis of textual data, better results can be achieved by introducing information related to each document (e.g. characteristics of the author, or when the document has been written). Here we focus our attention on advertisements, and their te!

Xtual component. In competitive markets there is a complex relation between the image of a brand, and the messages of the advertising campaigns over time. In the paper we propose a statistical procedure for analyzing the evolution of the brand image through the different campaigns of a famous brand, namely Coca Cola, on the basis of the homophily measures.

Large network community detection in practical scenarios

Lovro Šubelj, University of Ljubljana, Slovenia

Abstract

Network community structure is a thoroughly investigated concept with various practical applications. However, due to the lack of data, many of the past studies were focused on networks of rather small or moderate size. Thus, only recent research has shown that community structure revealed in large networks does not actually coincide with some ground truth clusters [Phys. Rev. E 90(6), 062805 (2014)]. Despite this disturbing fact, we here show that community information can still be beneficial in practical scenarios. As our first example, we consider a citation network of over 500 thousand papers published by American Physical Society until 2013. We predict the journal information of all papers in 2013 based on the journal information of the papers published until 2012 and complete citation information. Constructing a simple majority classifier for each paper based merely on its cited papers, and thus merely on the neighbors of the corresponding node, gives 67% prediction accuracy. However, extending nodes' neighborhoods to the entire communities in which they reside, boosts the accuracy for additional 5%. As our second example, we consider a collaboration network between over 300 thousand authors compiled from DBLP computer science repository. We predict the entire list of publication venues of an author (e.g., journal or conference) based on the publication venues of other authors and complete collaboration information. The accuracy of a classifier based merely on nodes' neighborhoods, and thus merely on immediate collaborations, is 31%, whereas the classifier based on community information has 35% prediction accuracy. As our final example, we consider a reference network between over 100 thousand US diplomatic cables until 2010 released by WikiLeaks. We predict the classifications of all cables in 2010 (e.g., secret or confidential) based on the classifications of the cables until 2009 and complete reference information. The classifier based merely on nodes' neighborhoods, and thus merely on referenced cables, has 28% prediction accuracy, whereas the community information improves the accuracy by almost 20%. We stress that above superior performance is obtained by the algorithms based on label propagation [Physica A 397, 144-156 (2014)], in contrast to more standard community detection approaches like spectral methods, modularity optimization and map equation algorithm.

Global and local centrality of emerging countries in the World Trade Network

Lucia Tajoli, Politecnico Milan, Italy
Luca De Benedictis, EIEF and University of Macerata, Italy

Abstract

Among the many changes that characterized the world trading system in the past decade, a very relevant and much discussed one concerns the role of the so-called "emerging countries". There is no consensus on the exact definition of this term (see Kearney, 2012), as they are referred to as middle-low income countries under rapid transformation, or rapidly growing countries in terms of GDP, or countries with increasing market shares in world export, etc. Certainly they are playing an important role in changing the trading system (see Hanson, 2012; Iapadre and Tajoli, 2014).

In this paper, we want to assess the position of a number of countries in the World Trade Network (WTN) in order to assess whether they are truly "emerging" within the system and they occupy central positions so to affect the overall connectivity of the WTN. In fact, in spite of high and increasing market shares in terms of exports (equivalent to strength centrality in network analysis, a local centrality measure), emerging countries might not always be central in global terms. Therefore, to make this assessment, we compare the market shares to the position of countries in the network in terms of eigenvector centrality: within the network, this is a comparison between a local centrality measure and a global centrality measure.

Overall, unsurprisingly we find a positive relationship between the local centrality measure and the global centrality measure (eigenvector centrality). But quite a few important countries are not in similar positions from a local and global point of view. Furthermore, most smaller and less developed countries display poor connections in spite of increasing market shares.

The analysis also shows that the core-periphery structure of the WTN apparent in the mid-1990s is no longer so evident. But even if the number and relevance of trade connections of most emerging and developing countries has increased significantly, China is still an exception in terms of extremely high connectivity and there is strong heterogeneity among the emerging economies. Finally, we analyze whether the trade evolution of the emerging economies is related to their trade partnerships and to their original position in the network.

Multi-Scale Centrality Measures for the Global Network of Corporate Control

Frank Takes, Leiden University, LIACS, The Netherlands
Eelke Heemskerk, University of Amsterdam, The Netherlands

Abstract

Although often depicted as atomistic and individualistic market actors, corporations are tightly embedded in networks of power and control. Foundational elements of these networks are interlocking directorates, where officers of one firm serve on the board of another firm. Increasingly, these hitherto national business communities' networks now form a new complex global system of corporate control.

Centrality gives an indication of the importance of a firm within the global economy and may indicate those firms that play a larger role in economic cascades such as one that led to the financial crisis. But in a large, global network, standard centrality measures are difficult to interpret due to the multi-level network structure. In previous work we found that if a hierarchical community detection algorithm is applied to the global network of interlocking directorates (largest 1 million firms worldwide), the resulting communities have a clear regional or even national character. This indicates that the footprint of the national networks is still visible in the global network. For corporate network research, it becomes important to consider how a firms centrality is derived from local, national, or global connections. The key challenge is therefore how centrality measures should be compared on the various scales. For example, when numeric centrality values are compared, ad! Justments based on the overall structure of the network may be necessary. Even if only the relative rank of the firm matters, a comparison between two different rankings is far from trivial, as many different techniques with different outcomes can be employed to compare rankings.

This paper presents both theoretical insights in how centrality measures should be interpreted at multiple scales, as well as domain-specific results with respect to how centrality measures rank both firms and countries within the global corporate network. First, we alisco the structure of global corporate network on both a national and a global scale by using various network analysis techniques and measures. We study a network consisting of more than 1 million companies (nodes) and more than 3 million social ties (edges) between companies. Second, we look at the aliscor of different centrality measures when applied to various regional subgraphs and the full global network. By comparing the centrality values of firms on a global scale to centrality values based on a certain region, we generate alternative rankings of companies based on their importance on a national, regional or global scale. Furthermore, aggregating to the country level, we can generate a ranking of how c! ountries contribute to the connectedness on a regional or global scale based on the centrality of the firms of the considered country in the respective regional or global network. Third, we suggest new measures for comparing centrality on a regional and a global scale as part of a broader multi-scale centrality computation framework. This allows us to gain insight in which companies within (groups of) countries contribute most to the connectedness of the global economic order.

Network visualization: Fine-tuning layout techniques for different types of networks

Nees Jan van Eck, Leiden University, The Netherlands
Ludo Waltman, Leiden University, The Netherlands

Abstract

An important issue in network visualization is the problem of obtaining a good layout for a network. For a given network, which may be either weighted or unweighted, the problem is to position the nodes in the network in a two-dimensional space in such a way that an attractive layout is obtained. Many layout techniques have been proposed. In the visualization of bibliometric networks, multidimensional scaling and the layout technique of Kamada and Kawai have for instance been used a lot. More recently, the VOS (visualization of similarities) layout technique, implemented in our VOSviewer software (www.vosviewer. com), is often used for bibliometric network visualization.

There is no layout technique that is generally considered to give optimal results. One reason for this is that comparisons between layouts produced by different techniques involve a lot of subjectiveness. Someone may consider one layout to be more attractive than another, but someone else may have an opposite opinion on this. In addition, the attractiveness of a layout may depend on the type of visualization that is needed. For instance, some layouts may be more attractive for interactive visualizations, while other layouts may be more attractive for static visualizations. Furthermore, different types of networks may benefit from different layout techniques.

In recent studies, the idea of parameterized layout techniques has been introduced. Parameterized layout techniques produce different types of layouts depending on the values chosen for their parameters. In this research, we present a comprehensive study of a parameterized version of our VOS layout technique. Two parameters are included. These are referred to as attraction and repulsion parameters. We compare the layouts obtained for different parameter values. Comparisons are made both subjectively using the VOSviewer software (i.e., which layout do we find most appealing?) and more objectively using so-called meta-criteria. Sensitivity to local optima is taken into account as well. Comparisons are made for all important types of bibliometric networks, in particular co-authorship, citation, co-citation, bibliographic coupling, and co-occurrence networks. Both smaller and larger networks are considered.

The findings of our analysis can be summarized as follows:

1. For most networks, the default parameter values (i.e., attraction = 2 and repulsion = 1) yield an acceptable layout.

2. For many but not all networks, the most attractive layout is obtained by setting attraction to 1 and repulsion to 0.

3. Compared with other networks, co-authorship networks of researchers require very different parameter values (e.g., attraction = 2 and repulsion = -2). Layouts obtained using these parameter values are sensitive to local optima.

4. To some extent, the optimal parameter values for a given network can be predicted based on characteristics of the network such as average degree, diameter, etc.

5. Meta-criteria tend to favor layouts that subjectively we do not regard as the most attractive ones.

Although our study focuses on the visualization of bibliometric networks, we also analyze how our findings generalize to other types of networks.

Mapping historical networks of the creative industry: Connecting Creative Agency in Early Modern Amsterdam, Venice, Florence, and Rome

Ingeborg van Vugt, Amsterdam University, The Netherlands

Abstract

Here we will report on the first (preliminary) outcomes of the pre-PhD project Mapping historical networks of the creative industry: Connecting Creative Agency in Early Modern Amsterdam, Venice, Florence, and Rome', funded by the University of Amsterdam as part of the research programme 'Creative Amsterdam: An E-Humanities Perspective' (CREATE). By using the software NODEGOAT, which platform is aimed at integrating heterogeneous datasets, we try to create multi-layered networks of actors and documents that are potential relevant for the history of the creative industry in Amsterdam, Florence, Venice and Rome in the Early Modern Period. Those multi-layered networks are created and contextualized by means of data integration from various datasets and complementary archival research, focusing in particular on the analysis of the rich epistolary correspondences between agents of the book industry (scholars – individual and in academies –, publishers, printers in early modern Italy and the Dutch Republic). Since it will be impossible to integrate all data necessary for the study of the history of creative industries we believe that this research can provide a valuable source to identify pairing of networks of agents involved and determining whether any combinations of these pairings would result in important clusters which enables strategic choices for further research.

In this research, the examination of co-citation networks is particularly valuable because it presumes that the occurrence of clustering pairs of publications and persons mentioned in the same correspondence can identify additional networks that have not otherwise been characterized. For instance, when Cosimo III sent engineer Pietro Guerrini to spy on the latest developments in fortifications and the linen industry in the Low Countries, he stipulated that the contact person to introduce him would be the book trader Pieter Blaeu. This raises questions about how and why specific people in different networks were able to connect and contribute to each other, and to what extent those networks overlapped. Thus, by providing a richer version of the properties of the network (for instance: who is introducing whom, which names are co-cited in letters around specific products) one could identify several interesting angles for in-depth research and comparison of processes occurring in societal developments and networks.

Moreover, we are particularly interested in various clustering patterns of societies in hubs and in intermediary persons with few connections but linking various communities (weak ties) who were relevant for the transfer of knowledge from one network to another. This aspect will be studied combining epistolary networks with data on intellectual and artistic organizations, the Accademie. Data from correspondences includes references to objects, such as books and engravings, which circulated throughout networks and reveal new networks and new insights on how people are placed between industrious and scientific networks of a more general nature and those of the creative industry.

Assessing Reputational Risk with Functional Data Analysis

Simone Vantini, Politecnico Milan, Italy
Marika Arena, Politecnico Milan, Italy
Alice C. L. Parodi, Politecnico Milan, Italy

Abstract

Reputational risk assessment has always attracted the attention of both academics and practitioners, since reputation is generally considered a critical asset for a company. It could influence indeed the behavior of the company's stakeholders either positively or negatively. Generally speaking, reputation can be conceived as the system of stakeholders' perceptions and expectations towards the corporation, and reputational risk can be defined as the risk of having this system of perceptions / expectations altered or damaged. However, perceptions and expectations are hardly measurable, hence requiring the definition of novel approaches to measure reputational risk.

In this work we analyze two proxies of corporate reputation associated to two different groups of stakeholders: share market value and social media activity (i.e., number of tweets sent to the official company account and the number of Google queries for the company name); and focus on negative reputational events pertaining to a multinational company of the energy sector. In detail:

1. We identify a list of negative events potentially relevant from a reputational perspective occurred between October 2008 and September 2013;

2. We explore – by means of functional data analysis – the reactions of the share market and social media to those events;

3. We identify possible underlying factors determining the shapes and magnitudes of share market and social media reactions (e.g. type of event, location where the event happens, business unit);

4. We identify the existence of possible relationships between the reaction of the share market and the reaction on social media.

Improving co-authorship network structure by combining different data sources: issues and practical considerations

Maria Prosperina Vitale, University of Salerno, Italy
Domenico De Stefano, University of Trieste, Italy
Vittorio Fuccella, University of Salerno, Italy
Susanna Zaccarin, University of Trieste, Italy

Abstract

The bibliographic archives used to study scientific collaboration can affect the derived bibliometric indicators as well as the co-authorship network structure. Indeed, the most used international databases might not be able to cover all kinds of works, especially for those disciplines having a more national orientation in their scientific production. In this case, the integration of high-impact journals databases with specialized and local bibliographic archives may be the best compromise to obtain a good coverage of whole research products of scientists involved in a specific field. To carry out the above task, two main challenges have to be addressed: 1) how to combine information by identifying and linking duplicate records, i.e. record linkage, and 2) how to deal with authors name disambiguation, i.e. synonyms and polysems.

In this study, we aimed at discussing main issues and practical considerations when these two features are dealt to reach a better quality of co-authorship data for network analysis. Specifically, the bibliographic archives used in De Stefano et al. [2013] are joined to obtain a unified co-authorship network, based on both top-international as well as nationally oriented scientific production of Italian academic Statisticians. To this aim, in the first step a semi-automatic method was adopted to merge three bibliographic archives. Due to the lack of training data, in the second step a modified version of the techniques described in Strotmann et al. [2009] provided promising results for author name disambiguation. Once we assessed how well the two procedures fared in achieving high quality results in the constructed co-authorship network, further statistical analyses will be devoted to identify the co-authorship characteristics of the emerging groups of statisticians under analysis.

Mapping cultural networks to detect a localized 'entrepreneurial discovery project'. Some empirical evidences

Maria Patrizia Vittoria, IRISS CNR, Italy
Pasquale Napolitano, IRISS CNR, Italy

Abstract

With the objective of rendering the policy paradigm and policy prescriptions more useful, recent smart specialization policy is aimed at the localized 'entrepreneurial discovery process' (Trajtenberg 2015; Foray 2006, 2015). Entrepreneurial discovery precedes the innovation stage and comprises exploration and establishment of a new domain of technological and market opportunities and the potential for innovation (Foray 2015). This process has been examined at the macro analytical level, i.e. from existing productive structures to new domains of potential competitive advantage, as structural change or related variety (Frenken et al. 2007). In these works, policy usefulness depends on how the localized entrepreneurial project (i.e. the capacity of the periphery to re-invent itself) is discovered and analyzed. As a result, diagnostic analytic is a crucial component of the smart specialization policy framework.

This paper considers the case of a cultural industry (Tagliacarne/Unioncamere 2009) to explore the usefulness of a network approach to reveal a localized entrepreneurial project. We are interested in how to recognize smart specialization in some emergent and consolidated cultural networks. We provide a literature review and discuss some case studies. The literature in the field of art and communication science highlights the hidden (but persistent) entrepreneurial character of the urban peripheral master artisan (Napolitano 2015) while organizational theories have been applied mainly to an institutional framework that includes industrial district (plus the additional concept's developments, until those of the 'advanced cultural district') (see Sacco, Ferilli 2006) and macro analytical dimension. In contrast, our cases explore the network dynamics in terms of changes to the structure and strength of ties throughout a cycle of entrepreneurial discovery (learning by communication and collaboration, creativity and experiential learning) (Teece 2000) and changing network focus (Nooteboom 2000).

Analyzing large bibliometric networks: The structure and evolution of network science

Ludo Waltman, Leiden University, The Netherlands

Abstract

Network science is a highly interdisciplinary field of research that has been attracting more and more attention. In this talk, I will present a large-scale bibliometric network analysis focusing on network science research. My starting point is the full body of scientific literature available in the Web of Science database, including many millions of publications from all fields of science. Using a combination of different techniques for the analysis and visualization of large networks, I will study the structure and evolution of the scientific literature, in particular the network science literature. My aim is to provide insights from a bibliometric point of view into the following three questions: (1) What is the scope of network science? In particular, which research areas may or may not be considered to belong to network science? (2) How is network science organized? What are the most important research areas within network science? How do these areas relate to each oth!
 Er? (3) How has network science developed over time? For instance, which research areas have grown most over time? Are certain areas converging toward each other? In addressing these questions, I hope to show the powerful nature of network analysis techniques for revealing the structure and evolution of scientific research.

Decentralized cooperation programs as practices of territorial networks evolution: inclusive or centralized dynamics?

Armin Wiedenhofer, Training Centre for International Cooperation, Italy
Carla Inguaggiato, Training Centre for International Cooperation, Italy

Abstract

This paper aims at showing that international cooperation projects can produce benefits and effects not only on direct beneficiaries in developing countries but also in the communities that promote such interventions. These effects on civil society can be particularly relevant in development programs that try to aggregate different territorial actors and promote collaborative environments. Those effects can be material and immaterial, social and economic, relational and organizational.

In recent literature there are no studies that describe and analyze this kind of effects. The study proposes to fill in this gap by evaluating the effects on social networks of decentralized cooperation programs using a combination of the territorial approach, network theories and the use of social network methodology.

This paper presents a longitudinal analysis of the relational and structural assets of Trentino's International cooperation activities and the specific case of "Tavolo Trentino con il Kosovo" (TCK). The study aim at assessing the benefits that this experience of decentralized cooperation can produce on the province of Trento. In particular the analysis focuses on the effects in terms of socio-organizational innovation and empowerment capacities on the actors.

Specifically, through a secondary collection of data obtained by the reading of Local Government and Association official documents, the analysis produces 10 bimodal matrixes that describe networks of the experience in different years (1999 – 2009). The matrixes are composed by 92 actors and 7 types of events (projects, formation days, tasks in Kosovo, awareness campaigns, meetings with territorial and international actors, consultative assemblies and cultural exchanges), and permit to describe the evolution of the networks and the TCK capacity to create a common collaborative structure between the different territorial actors.

The results of this research demonstrate that this program has enhanced its capacity to include new actors into the network. However problems in terms of empowerment and ownership of the single actors remain. The network has been distinguished by dynamics as centralization process, creation of a intra-organizational elite and mimetic isomorphism esteem other Trentino's decentralized cooperation programs. This kind of evolution has affected program's potentiality to promote effective forms of collaboration and consequently its capacity to produce socio-organizational innovation.

Analysis and Visualization of Communication Profiles for Online Community Members

Rostislav Yavorskiy, National Research University Higher School of Economics, Russia

Abstract

Our general goal is to develop a tool for visual representation of communication patterns of users in online communities. To make the tool independent from the language the analytics is built on the communication graph structure and ignores the textual content of the messages. In this talk several our projects related to profiling online user activities will be presented.

Impact of actor non-response treatments in valued networks on clustering

Anja Žnidaršič, University of Maribor, Slovenia
Patrick Doreian, University of Pittsburgh, USA, University of Ljubljana, Slovenia
Anuska Ferligoj, University of Ljubljana, Slovenia

Abstract

Valued networks have higher potential to represent a real world dynamics of relations among units than simplified binary networks. While in binary networks only presence or absence of relations is recorded, values on ties can represent the strength of relations or the frequency of contacts. All collected network data are prone to measurement errors regardless of the of data collection source. If all outgoing ties are missing, while incoming ties are available the error is known as actor non-response. Actor non-response treatments have been investigated for binary networks. Here, this approach is extended to valued networks with eight actor non-response treatments via simulations. The first treatment is the complete-case approach where beside the row of absent ties also the corresponding column is deleted and the result is a smaller network. A null tie imputation procedure replace all absent ties by zeroes. If the modal value of incoming ties for a non-respondent is used instead of absent tie the procedure is called imputations based on modal values, while also the mean value of incoming ties could be employed (imputations based on mean values). In the reconstruction procedure, an absent outgoing tie from actor I to actor j is replaced by the incoming tie from actor j to actor i.

Reconstruction of ties between two non-respondents is not possible, therefore in the simplest case the null tie imputations are used, while the second option uses imputations based on modal values for ties between non-respondents. The seventh procedure is imputation of a total mean where the valued density of the network is imputed instead of absent ties. The k-nearest neighbors approach searches for k closest actors according to their incoming ties and then calculates the outgoing ties of the non-respondent as a median of outgoing ties of selected nearest neighbors.

The impact of all these non-response treatments for hierarchical clustering of valued networks is presented. The networks in the study will be simulated based on three well known blockmodel structures: a core-periphery model, a cohesive subgroups model, and a hierarchy model.

List of authors