Application for Doctoral Study

RESEARCH PROPOSAL

Information Influence Modeling and Risk Forecasting

based on Complex Social Network

Applicant Name: Zhang Xueting

Date: 2016/12/07

Abstract

There are close association between online social network and offline real world with interaction and mapping relationship, which is feasible to build influence model for evaluation. Social network study, especially based on the mobile Internet, focuses on dynamical evolving network with different information transmission mechanism and influences model from the previous research. Big data approaches forecast the change trend and realize the risk event surveillance and prediction.

Key word: big data, mobile Social Network, information transmission, risk forecast, complex system, optimization algorithm

Motivation

Online Social Network plays a significant role to improve the society development in commercial and political level. Network Science, Web Science and Internet Science are three interdisciplinary areas which enhance our understanding of complex systems from biological to technological and social [1,2].

Data collection in current world is ubiquitous and universal, whereas the interaction of human-data is mundane both as a foreground technology (e.g., smartphones, iPad, wearable devices) and a background technology (e.g., financial system, traffic management) [3]. Complex network analysis leverages big data to expand the study field about the influence of information transmission and important event (or risky event) forecasting, which is beneficial to both social stabilization and commercial development.

Big data makes us better understand the real world, including satellite data, genetic data, GPS data, and social media data, thus it is feasible to transfer "sampling" to "complete set". Mobile Internet and social network are the main trend of modern society, and it is of practical significance to study on the information spreading mechanism and influence modeling, finding the human behavior change and forecasting the potential risks [4].

Literature Review

Network Science can be traced to 18th century as mathematical study on graph theory by Leonard Euler [5]. The concept of "Sociometry" was proposed by Jascob Moreno[6], and soon after that, "Network Science" drew on the anticipation of multi-disciplines such as physics, mathematics, computer science, economies, neuroscience or even astronomy [7].

Albert-László Barabási asserts that they try to find the hidden pattern behind everything we do, and the power law dominates most human behaviors that the characteristic of heterogeneity are manifested both in the dimension of time and space [8]. With new tools provided by powerful but affordable computers and computing software, researchers have made impressive gains in modeling and theorizing the nature of networks as complex, large-scale, interactive systems.

Alex Pentland is the director of the MIT Human Dynamics Lab and the Media Lab Entrepreneurship Program, and major in the study of complexity science and network science from evolving dynamics aspect [9]. In fact, the development of social physics has experienced more than two decades, which indicates the expectation to make society study definable and predictable by the method of mathematization and theorization [10]. The study about human behavior, group behavior, thinking flow transmission, and public opinion proliferation, all of which are similar to that energy changes the grain movement by flowing among them [11,12].

A computational composite network model is developed to better predict human behavior (such as app installation), which captures individual variance and exogenous factors [13].The traditional influence model is expanded with different factors in different cases, and after processing, integrating and analyzing the information, the risk prediction is more precise and general.

Xin Lu is my teacher and colleague in the same laboratory. He focuses on mobile data analysis and human behavior after disasters. Hidden Markov Chain and Movement Entropy are the main approach to make population displacement predictability [14]. RDS (Respondent-driven sampling) is used for the study of HIV/AIDS-related high risk populations and the estimation precision is improved by ego network data about their personal networks [15]. To protect the privacy of mobile data and better supervise the public health in the hyper-connected world, researchers construct Flowminder Foundation and pursue further studies [16].

Objectives and Methodology

The model of influence is unobservable network where the explanation and prediction can be extracted by the observable (or physical) networks. In my study, the research can be divided into three main parts: the dynamical network modeling, the information transmission modeling (the influence model), and the prediction algorithm.

Albeit many studies show that the human social network is always scale-free network with heterogeneity, there are still many difficulties to build the network model reflecting the real world [17,18]:
- Single-level network is hard to reflect the real relationship among individuals;
- Behavior change and influence transmission is not only affected by the network

structure but also interfered by the extra factors;

♦ It is not denial about the individual difference that "rumors stop by the sage", so some characteristics in specific cases of small groups are not general and extendible.

Different contagion hypothesis are made under different conditions, that different cases have different optimal parameters. Some information (or thoughts) are easily transferred with only one contact, while others are spread along with the relationship of network and always have cascade transformation.

Prediction algorithms are numerous and we choose the one that can be corrected constantly by collected data with strong robustness. Probability theory based on composite network model can improve the Spread Trend Forecast precision by intelligent optimization algorithm, changing the estimation problem to optimization problem during the model-training step [19, 20]. Distributed management of data also plays an important part. EMBER, an automated 24*7 continuous system for forecasting civil unrest across 10 countries of Latin America using open source indicators such as tweets, news source, blogs, economic indicators and other data sources. Through numerous evaluations, the superiority of EMBERS demonstrates the capability to forecast significant societal happenings [21,22].

Moreover, to adjust the model parameters, we need an index system to evaluate the prediction precision and efficiency.

Preliminary Work and Prospective Outcomes

Due to the interdisciplinary characteristics of this research field, I have already prepared some essential knowledge and skills and published some papers as preliminary study.

I propose the approximate maximum network lifetime by establishing the WSN (wireless sensor network)routing strategy planning model, and the network energy consumption also gets reduction by controlling the weight coefficient [23]."*Epidemic Spreading Characteristics and Immunity Measures Based on Complex Network with Contact Strength and Community Structure*" demonstrates the modeling of epidemiology with community structure based on contact strength and SI model, and a weighted network based on BBV network model. After the verification by the real data of MERS, it is found that the spreading rate is closely related to the average weight of network but not the number of communities [24].

During the research, I plan to publish 3-4 top ranked Journal and/or conference papers such as (but not limited to):

♦ Nature Communications

- PNAS (*Proceedings of the National Academy of Sciences*)
- TOIS (*ACM Transactions on Information Systems*)
- TKDD (*ACM Transactions on Knowledge Discovery from Data*)
- SIGMOD (*ACM Conference on Management of Data*)
- *IEEE transactions on Systems, Man and Cybernetics*

It is expected that the findings and outcomes in this research will have much more significance and deliver several advantages in both theory and practice:
- Find out the transmission mechanism of mobile social network;
- Enhance the big data technology for social physics analysis;
- Integrate the data mining approach with complex network science;
- Provide support for political and commercial services.

Reference

[1] T. Tipopanis, W. Hall, J. Crowcroft, et al. "Network Science, Web Science, and Internet Science." *Communications of the Acm*. 2015.

[2] A.Barabasi, and R. Albert. "Emergence of scaling in random networks, Science 286." *Science*, pages: 509-512, 1999.

[3] R. Mortier, H. Hadddadi, T.Henderson, et al. "Human-Data Interaction: The Human Face of the Data-Driven Society." *Social Science Electronic Publishing*, 2014.

[4] H. Pan, J. Crowcroft, and E. Yoneki. "BUBBLE Rap: Social-Based Forwarding in Delay-Tolerant Networks." *IEEE Transactions on Mobile Computing,* pages:1576-1589. 2010.

[5] Euler, L. Konigsberg Bridge problem Commentariiacademiae scientiarum Petropolitanae, pages:128–140, 1741.

[6] J.Moreno, et al. "Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Socio-Drama (2nd ed.)."*Beacon House, Oxford, England,* 1953.

[7] Newman, M.E.J. "Networks: An Introduction."*OxfordUniversity Press*, 2010.

[8] A. Barabási. "Scale-Free Networks: A Decade and Beyond." *Science*pages:412-413, 2010.

[9] Y. Altshuler,N. Aharony,Y. Elovici, et al. "Stealing Reality: When Criminals Become Data Scientists (or Vice Versa). " *Security and Privacy in Social Networks. Springer New York*, pages:133-151, 2013.

[10] N. Eagle, and A. Pentland. "Reality mining: sensing complex social systems." *Personal & Ubiquitous Computing*, pages:255-268, 2006.

[11] W. Pan, W. Dong, M. Cebrain, et al. "Modeling Dynamical Influence in Human Interaction."*Eprint Arxiv*, pages:77 - 86, 2011.

[12] N.Eagle, A. Pentland,and D. Lazer. "Inferring friendship network structure by using mobile phone data." *Proceedings of the National Academy of Sciences,* 2009.

[13] P. Wei, N. Aharony, and A. Pentland. "Composite Social Network for Predicting Mobile Apps Installation." *Eprint Arxiv*, 2011.

[14] X. Lu, L. Bengtsson, P. Holme. "Predictability of population is placement after the 2010 Haiti earthquake."*Proceedings of the National Academy of Sciences*, pages:11576-11581, 2012.

[15] X. Lu. "Linked Ego Networks: Improving Estimate Reliability and Validity with Respondent-driven Sampling."*Social Networks*, pages:669-685, 2013.

[16] Buckee, O. Caroline, T. Andrew, et al. "Society: Protect Privacy of Mobile Data."*Nature*, pages:35-35, 2014.

[17] A. Ritter, Mausam, O. Etzioni, and S. Clark. "Open domain event extraction from twitter."*inKDD '12*, 2012.

[18] M. Skjegstad, B. Elligsater, T. Maseng,J. Crowcroft, et al. "Large-Scale Distributed Internet-based Discovery Mechanism for Dynamic Spectrum Allocation." *2014 IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN)*, pages:404-415, 2012.

[19] J.She, J. Crowcroft, H. Fu, F. li. "Convergence of interactive displays with smart mobile devices for effective advertising: A survey." *ACM Transactions on Multimedia Computing Communications & Applications,* pages:294-304, 2012.

[20] Y. Keneshloo, J. Cadena, G. Korkmaz, et al. "Detecting and forecasting domestic political crises: a graph-based approach." *Proceedings of the 2014 ACM conference on Web science*, 2014.

[21] N.Ramakrishnan, P. Butler, S. Muthiah, et al. 'Beating the news' with EMBERS: Forecasting Civil Unrest using Open Source Indicators." *Eprint Arxiv* pages:1799-1808, 2014.

[22] J. Xu, et al. "Civil unrestprediction: A tumblr-based exploration."*In Social Computing,Behavioral-Cultural Modeling and Prediction,* 2014.

[23] X. Zhang, Y. Chen, J. Jiang, et al. "A Routing Algorithm based on Dynamic Weight of the Wireless Sensor Network."*in proceedings of IEEE 6th International Conference on Software Engineering and Service Sciences*, Beijing, 2015.

[24] X. Zhang, B. Ge, Q. Wang, et al."Epidemic Spreading Characteristics and Immunity Measures based on Complex Network with Contact Strength and Community Structure."*Mathematical Problems in Engineering,* 2015.