http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 13 (8): 1532-1539, 2014
ISSN 1812-5638 / DOI: 10.3923/itj.2014.1532.1539

Instance Analysis of Social Network Based Ucinet Tool

[1]Hua Yang, [2]Zhou-Fang Li and [3]Wei Wei
[1]Software School, Kaifeng University, Kaifeng, 475000, China
[2]School of Information Science and Engineering Henan University of Technology, ZhengZhou, 450000, China
[3]School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China

Abstract: Social network play an increasingly important role in areas such as business and social activities and is causing widespread concern in academia and industry. The complexity of its properties leads the social networking phenomenon can not be model by simple mathematical theory. It has seriously hampered the development of social network data analysis techniques. The relationship of object is analyzed by using UCINET software and instance of social network. Based on information interaction between people and all kinds of information of the members of society in the social network, a social network model is built and the social network diagram is designed. This instance and analysis method is provided that is a reliable guarantee for social networking data extraction and model.

Key words: Social network, data analysis, instance, UCINET

INTRODUCTION

As a virtual of real social network, Social network link closely relationship between people and gradually play an increasingly important role (Kossinets and Watts, 2006) in areas such as business (Varlamis *et al*., 2010) and social activities (Zheng *et al*., 2009). According to the social network topology analysis, socialization recommend, network community discovery, the dissemination of information and other aspects of the study also gradually expand. Among them, analysis of data in a social network is the foundation of social networks. Social network data is large and complex and includes a large amount of internal information at the same time, how to analysis and data mine to the social network, the type of techniques and methods is an urgent problem to be solved (Zhang *et al*., 2011).

A lot of scientists have presented their well-known research results about social network since Andersonetal (Pattison and Wasserman, 1999).

Garry Robins, Pip Pattison, Yuval Kalish and Dean Lusher (Onnela *et al*., 2007) provided an introductory summary to the formulation and application of exponential random graph models (Daraganova and Robins, 2013) for social networks in their article. They regarded all the possible ties in a social network as random variables and they assume that there are dependences in these random tie variables which determine the general form of the exponential random graph (Koskinen *et al*., 2010) model for the network. They also gave lots of examples of different dependence assumptions and their associated models, such as Bernoulli, dyad-independent and Markov random graph (Koskinen *et al*., 2010) models and so on.

Ulrik Brandes, Natalie Indlekofer and Martin Mader (Daraganova *et al*., 2012) argued that they could visualize both longitudinal networks (Snijders, 2005) and predictions of stochastic actor-oriented models, the most prominent approach for analyzing such networks by using the similar adaptations of state-of-the-art graph-drawing methods. They also illustrated on a longitudinal network (De Nooy, 2010) of acquaintanceship among university freshmen by using this proposed methods stated above. Authors of Melbourne School of Psychological Sciences in the University of Melbourne of Australia provided a tutorial review (Handcock, 2003) of some fundamental ideas and important methods for the modeling of empirical social network (Snijders *et al*., 2007) data. The authors also described the basic concepts from graph theory and central elements (De Nooy, 2010) to social network theory. In order to analyze the social networks, the authors presented models for the network degree distribution and for network roles and positions, as well as algebraic approaches, including boot-strap procedures models for testing the prevalence of network structures, basic edge and dyad-independent statistical models and more recent statistical network models that are assumed dependence, exponential random graph (Wasserman and Robins, 2005) models and dynamic stochastic actor oriented models.

Corresponding Author: Hua Yang, Software School, Kaifeng University, Kaifeng, 475000, China

Social network influence models are reviewed. The article concludes with a summary of new developments relating to models for time-ordered transactions.

Barabasi (Daraganova *et al*., 2012) said it is widespread debate that whether human social networks are central to the structure and dynamics of the contemporary social world.

In order to keep with the new enthusiasm, Brandes, Robins, McCranie and Wasserman (Koskinen and Daraganova, 2013) claimed network science as a distinctive, emerging research discipline in the first editorial of a new journal Network Science. And they pointed to familiar statements that "networks are everywhere", but argued that a science requires more than this. But network science is based on a conceptual unity across many disciplines: an ontological commitment to the primary importance of relationships between entities. Networks have had more and more notices in these fields that are mathematics, communication, information and computer science, management and organizational science, engineering, economics, psychology, political science, anthropology, medicine, public health, statistics, physics, animal behavior, sociology, biology and history. So, the encroaching disciplinary spread of network science is very wide.

At present, people are the majority in social activities, the people's initiative results in the complexity and diversity of the social network properties relationship, leading to that a simple mathematical model can not abstract the complex social network phenomenon. It has seriously hampered the development of social network data analysis techniques.

The relationship of object is analyzed by using UCINET software and instance of social network in this paper. A social network model can be built and the social network diagram can be designed on the base of information interaction between people and all kinds of information of the members of society in the social network by extracting ontology from various sources of information to determine the relationship between the ontology.

SOCIAL NETWORK AND DATA ANALYSIS

Social network and its classification: With the development of modern communication technology, based on the use of mobile phones, tablet PCs and other mobile terminal, People's communication activities and social relations, is gradually blend with the virtual network, formed a social network. According to the scope of its application, the social network can be broadly divided into the following categories.

Web-based social networks: Web page promotes information exchange and cooperation between human through the application of network and its mode is more user-centric. Input to corresponding phrase, name and information to your web pages. You can build a network of social relations based on the possibility of the existing relationship between them. Facebook, Renren, Kaixin, Friend and other network, this network contains a large number of the information of the relationship between people.

E-mail network: Through the Internet technology, E-mail solves the problems of remote mail transfer, achieves to one-to-one communication, is the most popular applications on information transferring between people and between enterprises. How to establish a social relation network by mining Email information has become a hot research topic in the current Internet.

BBS, electronic forum: BBS (Bulletin Board system) through the "mass" and "forward", in theory, achieve the freedom of expression and released to all information and discuss topics. Electronic BBS (Electronic Forum) is a many-to-many communication mode. Through its data content and reply relationships, it builds a network of social relations.

Blogs: Blog is also translated as web logs; it is a personal management site which is publishing a new article from time to time. The blog allows readers to express their views and then forms their own social network interaction. In recent years, through the mobile platform update became to realize information released by the blog.

Network real-time communication tools: It infers the social relations between chat users with the timing relationships of chat date and the introduction of similar content by using the real-time communication tools, such as QQ, ICQ, mobile phone short message, etc., real-time voice, video and text transmission can be gotten on the internet by using the tools.

It is the core task of the social network analysis that classifying the individual to society crowd and extracting the social network relationships.

Social network data analysis: Social Network Analysis (SNA) is the mapping and measuring of relationships and flows between people, groups, organizations, computers or other information/knowledge processing entities. People and groups constitute the nodes in the network.

The links between the nodes show the relationships or flows of them. SNA gives both a visual and a mathematical analysis of human relationships. Management consultants use this methodology with their business clients and call it Organizational Network Analysis (ONA). By evaluating the location of actors in the network the networks and their participants can be understand. And the centrality of a node can be found by measuring the network location. At the same time, these measures can help determine the importance or prominence of a node in the network. Network location can be different than location in the hierarchy, or organizational chart.

A social network, called the "Kite Network", which was developed by David Krackhardt, a leading researcher in social networks. Two nodes are connected if they regularly talk to each other, or interact in some way. For instance, in the network above, A regularly interacts with B, but not with C. Therefore A and B are connected, but there is no link drawn between A and C. This network effectively shows the distinction between the three most popular centrality measures: Degrees, Betweenness and Closeness.

Here is the data structure for a social network: There are multiple methods to obtain the data about social networks but a simple example will illustrate the method and the resulting data. Assume that each person in the network of N individuals is provided with a list of all other people in the network and asked to indicate with whom they interact. People with whom they interact may be assigned a value of 1 or greater to indicate the level of interaction or a zero to indicate no interaction. Each person's responses are in the form of 1-N in which each cell represents a person A's assessment of their connection with every other person.

It mainly focused on the questionnaire form in the early study of social network data, which makes the research only limited to a small range of groups in the short term, thus it greatly limits the breadth and depth of the research and the accuracy of the results. In recent years,there are research trends in communication mode for data mining from an individual in the process of social activities outside of the observation data,with E-mail, call log behavior patterns, etc. The daily behavior data of users through the phone and mobile phone on the Bluetooth function can be seen by using the social network data analysis tool UCINET, these excavate Users implied relationships in the daily behavior data and all kinds of different user groups.

SOCIAL NETWORK DATA ANALYSIS TOOLS-UCINET

UCINET software description: UCINET (University of California at Irvine Network) is powerful social network analysis software, which is embedded in the function of the social network visualization. In UCINET, data should be stored, shown and described by using the format of matrix and the maximum number that can be handled by it is 32767 points.

The advantage of UCNNET is to reveal the contact tightness and the relationship between the behaviors of the overall members. You can find the contact and decomposition mode in the whole network and find the structure equivalent actors, perform the data and its structure relationship in a more intuitive way to help the users to understand a large amount of data information and find the phenomenon implied in the data, improve the utilization and high accuracy rate.

The UCINET software contains a large number of network parameters analyses, Such as centrality, cohesion measure of the relationship between the two parties, location analysis algorithm, faction analysis, the random two parties relational model, etc. Also includes the common multivariate statistical analysis tools such as multidimensional scale, correspondence analysis, factor analysis, cluster analysis and multiple regressions for matrix data, etc. In addition, UCINET also provides tools for data management and conversion; it can convert a program from graph theory to matrix algebra language.

UCINET data environment analysis: Supported the data of the UCINET have the following three structures:

- Hierarchical structure, such as the file system, organization chart, etc
- Structure of the network type, such as network topology, website links, etc
- No connection of the data set, such as time lines, etc.

There are two UCINET ways of data input:

- Input parameters, such as factions analysis program
- Input data, such as a single data or matrix. There are a text file and data file formats of output way

Data format of UCINET which is a matrix can express many kinds of relations between the nodes within a collection by a file. The data file which name format is". # # H", in a plain text file, need into a data language documents (DL file) to realize the data processing. Data language file contains a series of Numbers (data) and a lot of key words and statements describing data, using to describe the basic information of the data. Therefore, a separate UCINET database consists of two physical files, one is to include an extension. "# # D file to reflect the actual data; another is the above ". # # H", contains the actual data.

UCINET of data visualization process: The UCINET object contains three parts: data, visual table and views. Data visualization process is divided into four steps:

- The abstract data
- Mapping the abstract data into a visible form (by filtering)
- Converting the visible table to views by the use of display (the transfer and display)
- Will be presented the views to users

In this article, by using the existing visual interface of the software, it filets out the useful information through extracting information for instance data. Then converse the data into the visual table. And finally the visual results are presented to the user through invoking UCINET development package to generate different views.

Social network data instances analysis: Here, analyzes the instance from MIT's social network data, this data set mainly comes from a random sample of 106 students who contact to SMS between each other.

First, the use of Matlab tools to extract data sets, it extract the relational data, to implement a 106×106 matrix, each node of the matrix represents the number of two SMS contacting. Then compresses this matrix by symmetrization and classification, builds a diagram of network of social relations between them, within faction classification and sub-block analysis, clearly show the relationship layout between 106 students especially the active point.

Relational data processing based on MATLAB

Data relation extraction: First it cleans the data which are provided by MIT, gets the actual number of the contact node each other through data relationships which ate extracted from MATLAB, in order to further analysis the ntimacy for between each node and mines core network nodes.

Matlab's code is as follows:

- m = zeros (106, 106)
- For n = 1:1:106
- p = size (s (n). device_macs, 2)
- For i = 1:1:p
- Hmac1 = num2hex (s (n). device_macs{i})
- For k = 1:1:size (hmac1, 1)
- for j = 1:1:106
- if strcmp (num2hex (s(j).mac), hmac1 (k,:))
- m (j, n) = m (j, n)+1
- Break
- End
- End
- End
- End
- End

There are a total of 106 nodes in data, variable s (n). Mac is on behalf of the physical address of the N node, uniquely identifying a person; s(n). Device Mac represents the node address of contacting the N node. Extracting data, matrix is shown in Fig. 1.

Data relationships cleaning: Cleaning the data to matrix, containing symmetry and proportional compression classification:

- Symmetry is mainly the use of the mean value principle, the value which the two nodes is symmetrical is equal to the average of the values which the two nodes
- To proportional compression classification. Each node in the matrix is divisible by 1000 and rounds numbers in MATLAB and converse the number of between 0 and 999 into 0, converse the number of between 1000 and 1999 into 1, by analogy, until converse the number of between 9000 and 9999 into 9, other numbers converse into 10. The main purpose is analyses more contact nodes, less ignored.

Symmetric and compression results are shown in Fig. 2.

Data relationship analysis based on UCINET

Network graph drawing: Run UCINET software, set additional weights of attribute and draw a network diagram. Link-Based Object Ranking at the same time and adjust the vertical position of the figure, the node location, size and color. Finally getting a clear social network layout which is drawn between each node and overall. As shown in Fig. 3.

By analysis of Figure 3, it can seen, the piont 1, 2, 7, 13, 14, 17, 18, 21, 24, 29, 31, 32, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 52, 54, 55, 56, 57, 59, 61, 62, 66, 67, 69, 70,

结点	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0	0	0	688	7	0	0	810	219	0	37	13	3
2	0	0	20	75	48	0	0	248	189	8	196	168	0
3	0	0	0	2504	489	78	51	430	186	8	195	34	0
4	0	6	1235	0	515	62	155	4945	3617	276	1364	1013	320
5	0	16	1218	2566	0	31	108	1750	714	138	231	167	10
6	0	0	102	249	70	0	31	62	98	1036	30	228	22
7	0	0	59	526	58	19	0	146	492	14	80	29	16
8	0	2	725	13508	554	30	11	0	2061	260	1016	58	79
9	0	0	247	5401	250	28	361	1648	0	76	997	65	0
10	0	0	220	7678	240	263	5	795	137	0	281	233	30
11	0	0	355	3752	177	8	77	1158	1037	151	0	113	1
12	0	2	67	1977	32	104	49	52	93	126	111	0	18
13	0	0	4	348	9	12	13	33	0	0	4	1	0
14	0	0	21	1384	13	1	5	102	55	0	88	13	0
15	0	0	11	150	58	394	11	107	41	2341	98	1679	20
16	0	0	3029	699	1036	779	60	1502	365	23	151	77	2
17	0	0	4	22	24	40	10	91	49	0	1	14	34
18	0	0	19	106	1	12	18	6	4	2	64	96	0
19	0	0	2486	450	925	579	40	614	392	34	123	384	2
20	0	0	39	267	22	1208	19	65	66	111	41	105	38
21	0	0	40	86	232	41	7	22	44	33	30	141	15
22	0	0	6	43	25	87	2	16	15	1046	5	88	38
23	0	0	517	7861	478	27	187	4104	2124	89	677	213	77
24	0	0	2	65	8	61	2	8	5	6	7	5	1
25	0	4	1701	6266	654	38	228	2972	1966	67	490	162	22
26	0	0	42	305	179	1427	15	105	92	28	28	69	10
27	0	0	66	393	18	60	7	69	62	246	104	1396	38
28	0	0	1	63	32	7	3	7	20	2105	4	34	0
29	0	0	21	411	14	5	3	3	65	1	44	47	1
30	0	0	1649	1562	752	81	165	1220	397	235	224	56	3
31	0	0	3	51	3	0	0	6	26	0	0	26	0
32	0	0	5	60	0	0	1	0	11	0	4	0	0

Fig. 1: Extracting matrix diagram

1	0	0	0	0	0	0
0	1	0	0	9	4	3
1	0	0	0	1	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
9	1	0	0	0	1	1
4	0	0	0	1	1	1
3	0	0	0	0	0	0
2	0	0	0	1	1	1
1	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	1
0	1	0	0	1	1	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
1	1	0	0	0	0	0
0	0	1	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	1
6	0	0	0	4	2	0

Fig. 2: Cleaning results diagram

71, 72, 76, 77, 78, 79, 80, 84, 85, 88, 89, 91, 92, 95, 96, 97, 98, 101, 104, 150, 106 are isolated points, omitted. Figure 3 reflects the relation, intimacy and partitions.

Cluster analysis: Cluster analysis of the matrix, as shown in Fig. 4, as can be seen by the chart is divided into six sub-blocks. You can see carefully that it is consistent with the network graph partition above. Just block is general, is not so detailed given the sense of layering and distance of relationship.

Faction analysis: The third step is to faction analysis of node relations Fig. 5 is shown in the form of a tree and this is consistent with block and the same of the

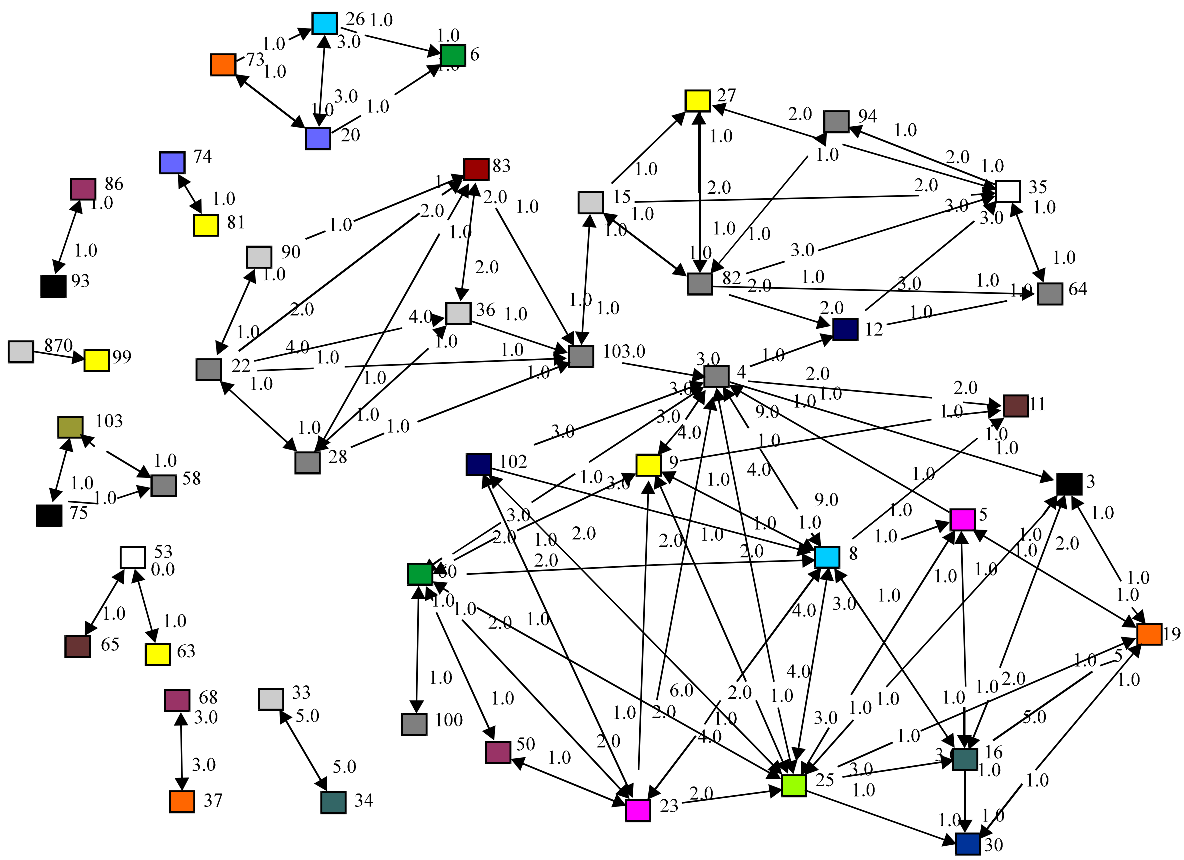

Fig. 3: Social network layout

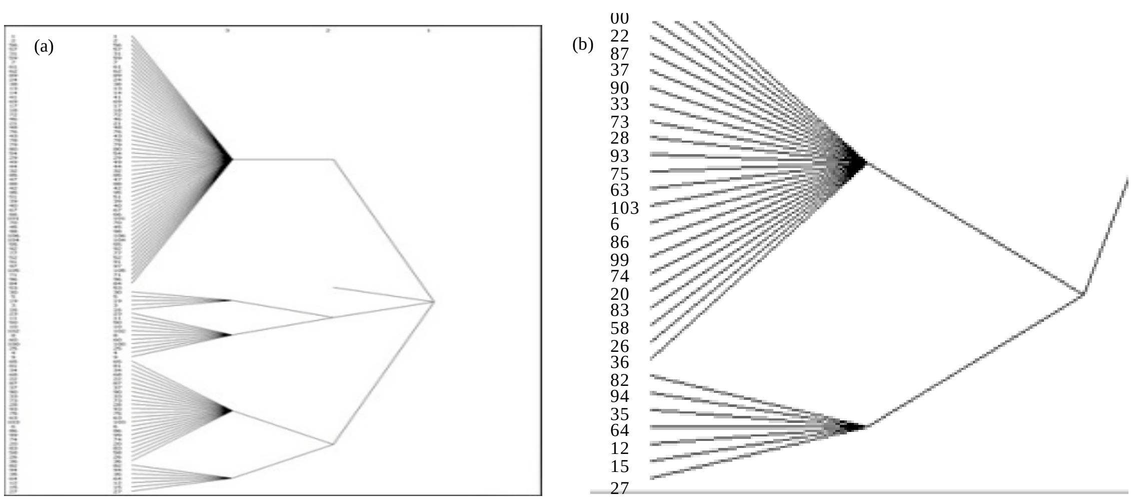

Fig. 4(a-b): Cluster analysis

subgroup distribution of node. its comparative advantage is that wing structure understands clearly relationship density between the members of the subgroup with respect to the closeness of the relationship between external members and the proximity or reach ability between the subgroup members.

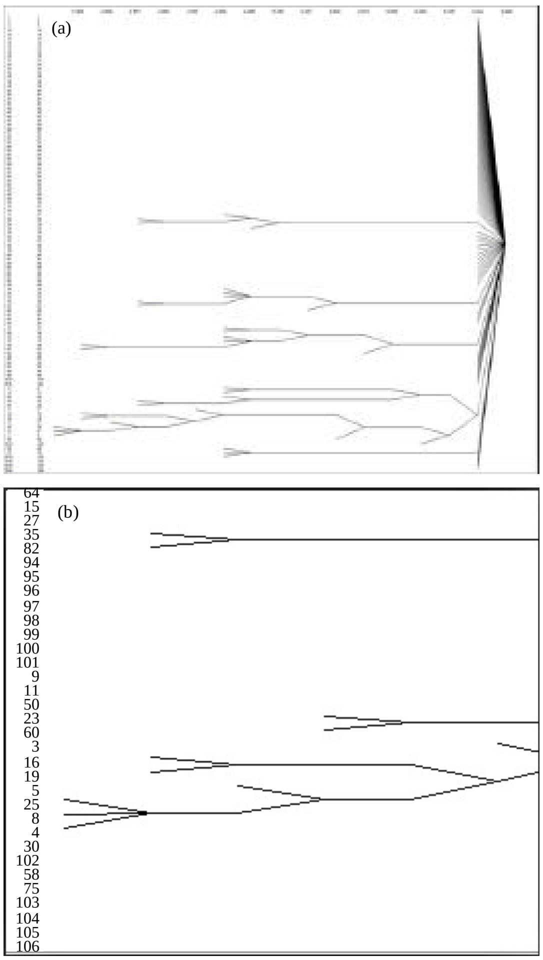

Fig. 5(a-b): Factions analysis

CONCLUSION

The data of social network and visualization process was analyzed in UCINET by using the social network data which come from MIT, example-based. The advantage of UCNNET is to reveal the contact tightness and the relationship between the behaviors of the overall members. You can find the contact and decomposition mode in the whole network and find the structure equivalent actors, perform the data and its structure relationship in a more intuitive way to help the users to understand a large amount of data information and find the phenomenon implied in the data, improve the utilization and high accuracy rate, instance and analysis method of this paper, for extracting the social network data and researching model, provides a reliable guarantee.

ACKNOWLEDGMENTS

The anonymous reviewers should be thanked for their valuable comments. This work is supported by National Natural Science Foundation of China (No. 61203094), the project of Henan province (No. 12A520009) and the science and technology development project of Henan province (No.122102210230). This work is also supported by Scientific Research Program Funded by Shaanxi Provincial Education Department (No.2013JK1139) and Supported by China Postdoctoral Science Foundation (No.2013M542370) and supported by the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20136118120010).

REFERENCES

Daraganova, G., P. Pattison, J. Koskinen, B. Mitchell, A. Bill, M. Watts and S. Baum, 2012. Networks and geography: Modelling community network structures as the outcome of both spatial and network processes. Social Networks, 34: 6-17.

Daraganova, G. and G. Robins, 2013. Autologistic Actor Attribute Models. In: Exponential Random Graph Models for Social Networks: Theory, Methods and Applications, Lusher, D., J. Koskinen and G. Robins (Eds.). CambridgeUniversity Press, New York, pp: 102-114.

De Nooy, W., 2010. Networks of action and events over time. A multilevel discrete-time event history model for longitudinal network data. Soc. Networks, 33: 31-40.

Handcock, M.S., 2003. Statistical Models for Social Networks: Degeneracy and Inference. In: Dynamic Social Network Modeling and Analysis, Breiger, R., K. Carley and P. Pattison (Eds.). National Academies Press, Washington, DC., pp: 229-240.

Koskinen, J.H., G.L. Robins and P.E. Pattison, 2010. Analysing exponential random graph (p-star) models with missing data using Bayesian data augmentation. Stat. Methodol., 7: 366-384.

Koskinen, J. and G. Daraganova, 2013. Dependence Graphs and Sufficient Statistics. In: Exponential Random Graph Models for Social Networks, Lusher, D., J. Koskinen and G. Robins (Eds.). Cambridge University Press, New York, pp: 77-90.

Kossinets, G. and D.J. Watts, 2006. Empirical analysis of an evolving social network. Science, 311: 88-90.

Onnela, J.P., J. Saramaki J. Hyvonen, G. Szabo and D. Lazer *et al*., 2007. Structure and tie strengths in mobile communication networks. Proc. Natl. Acad Sci., 104: 7332-7336.

Pattison, P. and S. Wasserman, 1999. Logit models and logistic regressions for social networks: II. Multivariate relations. Br. J. Math. Stat. Psychol., 52: 169-194.

Snijders, T.A.B., 2005. Models for Longitudinal Network Data. In: Models and Methods in Social Network Analysis, Carrington, P.J., J. Scott and S. Wasserman (Eds.). Cambridge University Press, New York, pp: 215-247.

Snijders, T.A.B., C.E.G. Steglich and M. Schweinberger, 2007. Modeling the Co-Evolution of Networks and Behavior. In: Longitudinal Models in Behavioral and Related Sciences, Van Montfort, K., H. Oud and A. Satorra (Eds.). Lawrence Erlbaum Associates, Mahwah, New Jersey, pp: 41-71.

Varlamis, I., M. Eirinaki and M. Louta, 2010. A study on social network metrics and their application in trust networks. Proceedings of the International Conference on Advances in Social Networks Analysis and Mining, August 9-11, 2010, Odense, Denmark, pp: 168-175.

Wasserman, S. and G.L. Robins, 2005. An Introduction to Random Graphs, Dependence Graphs. In: Models and Methods in Social Network Analysis, Carrington, P., J. Scott and S. Wasserman (Eds.). Cambridge University Press, New York, pp: 148-161.

Zhang, H.Q., R. Dantu and J.W. Cangussu, 2011. Socioscope: Human relationship and behavior analysis in social networks. IEEE Trans. Syst. Man Cybernetics Part A, 141: 1122-1143.

Zheng, Y., X. Xie and W.Y. Ma, 2009. GeoLife 2.0: A location-based social networking service. Proceedings of the International Conference on Mobile Data Management: Systems, Services and Middleware, May 18-29, 2009, Taipei, China, pp: 357-358.