The Third International Conference "Problems of Cybernetics and Informatics"
September 6-8, 2010, Baku, Azerbaijan. Section #1 "Information and Communication Technologies"
www.pci2010.science.az/1/46.pdf

BUILDING A HETEROGENEOUS SOCIAL NETWORK OF ACADEMIC RESEARCHERS

Fadai Ganjaliyev

Institute of Information Technology of ANAS, Baku, Azerbaijan
fadaig@yahoo.com

Abstract. Social networks have attracted much attention recently. Social network analysis finds its application in many current business areas. Different studies have been conducted to automatically extract social networks among various kinds of entities from the Web. In this work we present a new method for building a social network of academic researchers. We merge single-relational networks into a single heterogeneous network which carries more exact information about the behavior of network actors.

1. Introduction

Social networks have gained high popularity recently. Social networks on the web have been extracted by retrieving relationships between entities automatically derived from multilingual news [1], from user's email inbox [2] and email communications [3]. Social networks also have been extracted from log files of online shared work-spaces [4]. A method for extracting social networks from the Web using similarity between collective contexts is proposed in [5]. FOAF profiles present a new research topic for social network researchers [6], [7], [8]. Co-occurrence based extraction also finds its applications in some works [9], [10]. Extracting social network of people on the web based on their homepages presents another approach for detecting similarity between persons [11] [12]. Some studies address the problem of extraction social network of academic researchers [10], [13]. Some researchers address not the process of extraction of a network given a set of named entities but the extraction of relations which might possibly exist between the entities [14]. The choice of the similarity coefficient may affect results in ranking entities of network as demonstrated in [23].

Community mining is one of primary research areas in social network analysis. One may define community as a set of objects sharing similar properties. In graph theory concepts community mining can be defined as identification of sub-graphs i.e. identifying sub-graphs which have higher densities compared to the whole network density. A group of studies in social network analysis has been done on community mining as well [15], [16], [17].

Extraction social network of academic researchers is one of mostly studied and interesting research areas in social network analysis on the Web. The system described in [9], for example, allows performing searches of different types: by name, by keyword, by affiliation and by research field. It also lists the researchers related to the searched person. The system uses the overlap coefficient to measure the strength between entities. The social network extraction procedure is carried out in two steps. Firstly, nodes are set, and then edges are established. Several algorithms are used to extract a social network of a researcher in Polyphonet. The presented in [18] approach in is to measure the relevance of two entities by calculating the co-occurrence measure of two names. The approach is to treat a researcher's network extraction as a procedure containing three steps: identification of relevant Web pages, profile information extraction, and integrating publications from different sources. The relevant page identification is carried out by searching pages containing the researcher's name through a search engine. Publications are not retrieved directly from Web but from online data sources.

Social network analysis is the mapping and measuring of relationships and flows between people, groups, organizations, animals, computers or other information/knowledge processing entities. The nodes in the network are the people and groups, while the links show relationships or flows between the nodes. One of the most interesting things about social structures is their substructure in terms of groupings or cliques. The number, size, and connections among the sub-groupings in a network can tell us a lot about the likely behavior of the network as a whole. How fast will things move across the actors in the network? Will conflicts most likely involve multiple groups, or two factions? To what extent do the subgroups and social structures overlap one another? All of these aspects of sub-group structure can be very relevant to predicting the behavior of the network as a whole [19].

Social networks had also found to be very powerful in identifying animal behavior [20]. Social network theory methods also applied on development of relationship between entrepreneurs [21]. Collaboration in economics community also was analyzed using social network analysis [22].

Most methods and algorithms which exist for social network analysis on the Web assume that there is only one relationship between actors i.e. consider homogeneous networks. In real, more than one relation exists in social networks (heterogeneous networks). Moreover, most methods proposed are well for specific sets of data like news articles, researchers and etc and don't demonstrate expected results for other kind of data.

In this paper we present a method which combines multiple social networks of academic researchers. The primary idea behind our proposal consists of integrating multiple homogenous networks into a single network which would contain all relations in the original networks. The resultant network edge strength can be manipulated by the way described below.

2. Building an integrated social network of academic researchers

In social network theory there are two types of networks with respect to relations: *homogenous* and *heterogeneous*. In homogenous social networks there exists only one relation between entities. Heterogeneous networks are those in which multiple relations exist between entities. Heterogeneous networks usually called multi-relational social networks.

Most social network extraction methods consider there is only one relationship between network nodes. Depending on the actors of the networks and type of the study some might be interested in particular relations. Some relations may be of more concern than others.

Academic researchers may have relations of different kinds: they may have co-authored one or more papers, may participate in the same conference, may be members of the same scientific center or might have taken part in the same project and etc. This means that a social network of academic researchers is usually multi-relational. Each of these relations may represent a different network but having a single network which would combine all these relations may give better results with respect to ranking, determining brokers and etc.

Assume that we are given a set of names of academic researchers for which we need to build a social network from available information on the Web. Denote them $a_1, a_2, ..., a_n$. Suppose there exist m relations amongst them $k_1, k_2, ..., k_m$. These can be co-authorship relation, co-citation relation, project participation relation, conference participation relation, scientific center membership relation and etc. Suppose using some predefined methods, for example, using methods based on co-occurrence based approach, for each of these relations we have build a social network of these academic researchers. To do this we for each of pair of entities we can submit a query to a general search engine. Then depending on the similarity coefficient chosen we can compute and assign weights to the edges of the graphs. As a result we shall receive m networks each of which will represent a homogenous network with respect to the k^{th} relation. Suppose for each of the networks we have received the edge weights w_{ij}^k, between the nodes i and j where $i, j = \overline{1, n}$ and $k = \overline{1, m}$. Using edge weights received we can build a heterogeneous network edge weights of which we can calculate by the following formula

$$w_{ij} = \sum_{k=1}^{m} \lambda_k w_{ij}^k$$

where $\lambda_1 + \lambda_2 + ... + \lambda_k = 1$ and w_{ij} is a weight between i^{th} and j^{th} nodes. Lambda is chosen intuitively or by experimental way.

This simple method of merging multiple homogenous social networks into a single one which will contain all relations to be extracted is very flexible in the sense that by changing values of concrete λ we can receive different heterogeneous networks. For example, if we increase the value of λ for co-author relation and respectively decrease λ's for other relations we shall receive a different heterogeneous network. In the network obtained we shall have strength of one relation weaker or stronger than others. This can also affect, for example, ranking results of the entities.

3. Conclusion

Most methods on the Web used for extracting and analyzing social networks suppose only one relationship exists between actors. In reality, on the Web resources are connected by multiple relations. We presented a model which combines multiple single-relational networks into a single multi-relation network. The resultant network would contain all relations and percentage of strength of each relation in the resultant edge strength can be changed depending on the weight coefficients chosen. This can have effect on further analysis of the network. Although this paper presents the above mentioned model for network of academic researchers, this model can be applied to other kinds of social networks as well where multiple relations may exist between network nodes. In the future, we may be able to show some practical applications of the model presented.

References

1. B. Poliqean, H. Tanev, and M. Atkinson, Extracting and learning social networks out of multilingual news, Proceedings of the Social Networks and Application Tools Workshop (SocNet-08), Skalica, Slovakia, 2008, pp. 13-16.
2. A. Culotta, R. Bekkerman, and A. McCallum, Extracting social networks and contact information from email and the web, Proceedings on the First Conference on Email and Anti-Spam, California, USA, 2004.
3. R. Rowe, G. Creamer, S. Hershkop, and S.J. Stoflo, Automated social hierarchy detection through email and network analysis, Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, San Jose, USA, 2007, pp. 109-117.
4. P. Nasirifard, V. Peristeras, C. Hayes, and S. Decker, Extracting and Utilizing Social Networks from Log Files of Shared Workspaces, Leveraging Knowledge for Innovation in Collaborative Networks, 2009, Vol. 307/2009, pp. 643-650.
5. J. Mori, Y. Matsuo, and M. Ishizuka, Extracting relations in social networks from the web using similarity between collective contexts, Proceedings of the 5th International Semantic Web Conference, Athens, GA, 2006, pp. 487-500.
6. P. Mika, Flink: Semantic Web technology for extraction and analysis of social networks, Journal of Web Semantics, 2005, vol.3, no.2, pp. 211-223.
7. T. Finin, L. Ding, L. Zou, Social networking on the semantic Web, The Learning Organization, 2005, vol. 12, no. 5, pp. 418–435.
8. B.Aleman-Meza, M.Nagarajan, C.Ramakrishnan et al., Semantic analytics on social networks: Experiences in addressing the problem of conflict of interest detection, Proceedings of the 15th International Conference on World Wide Web, Edinburgh, Scotland, 2006, pp. 407-416.
9. Y. Matsuo, J. Mori, M. Hamasaki, H. Takeda et al., POLYPHONET: an advanced social network extraction system, Proceedings of the 15th International Conference on World Wide Web, Edinburgh, Scotland, 2006, pp. 262-278.

10. M. Harada, S. Sato, and K. Kazama, Finding authoritative people from the Web, Proceedings of 4th Joint Conference on Digital Libraries, Tucson, USA, 2004, pp. 306-313.
11. Q. Li, and Y.B., Wu, People search: Searching people sharing similar interests from the Web, Journal of the American Society for Information Science and Technology, 2007, vol. 59, no.1, pp. 111-125.
12. L.A. Adamic, and E. Adar, Friends and neighbors on the Web, Social Networks, 2003, vol. 25. no. 3, pp. 211-230.
13. J. Tang, J. Zhang, L. Yao et al., Arnetminer: extraction and mining of academic social networks, Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, Las Vegas, USA, 2008, pp. 990-998.
14. J. Mori, T. Tsujishita, Y. Matsuo, and M. Ishizuka, Extracting relations in Social Networks from the Web Using Similarity Between Collective Contexts, Proceedings of the 5th International Semantic Web Conference, Athens, USA, 2006, pp. 487-500.
15. N. Du, B. Wu, X. Pei et al., Community detection in large-scale networks, Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis, San Jose, USA, 2007, pp. 50-57.
16. J. Baumes, M. Goldberg, M. Magdon-ismall, and W. Wallace, Discovering Hidden Groups in Communication Networks, Proceedings of the 2nd NSF/NIJ Symposium on Intelligence and Security Informatics, Tucson, USA, 2004, pp. 378-389.
17. D. Cai, Z. Shao, X. He et al., Mining hidden community in heterogeneous social networks, Proceedings of the 3rd International Workshop on Link Discovery, Chicago, USA, 2005, pp. 58-65.
18. J. Tang, D. Zang, and L. Yao, Social Network Extraction of Academic Researchers, Proceedings of the 2007 Seventh IEEE International Conference on Data Mining, Omaha, USA, 2007, pp. 292-301.
19. M. Jamali, and H. Abdolhassani, Different aspects of social network analysis, Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, Hong Kong, China, 2006, pp. 66-72.
20. A. Coleing, The application of social network theory to animal behavior, Bioscience Horizons, 2009, vol. 2, no. 1, pp. 32-43.
21. M. Jaafar, A. A.R. Abdul-Aziz, and M.H. Sahari, The use of social network theory on entrepreneur's linkages development, Theoretical and Empirical Researches in Urban Management, 2009, vol. 4, no. 1S, pp. 101-119.
22. K. Tomas, and B. Nisa, A social network analysis of research collaboration in the economics community, International Workshop on Webometrics, Informetrics and Scientometrics, Nancy, France, 2006, pp 333-344.
23. R. Alguliyev, R. Alguliyev, and F. Ganjaliyev, Role of similarity measures in ranking network entities, Intelligent Information Management, 2010, vol. 5, (preprint).