# New Method for Community Detection in Social Networks Extracted from the Web

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. Community detection is one of the most important and interesting research areas in social network analysis. Many works are dedicated to methods and algorithms for detecting communities in different kinds of social networks extracted on the Web. Below we give a brief description of a new method which can help to identify communities in networks.**

***Keywords*– *social network; community detection; Boolean programming problem***

## I. 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 [15].

The objective of community detection is to identify the groups with the aid of information encoded in the graph itself only. This problem has a long history and several disciplines addressed it different ways. One of the first who carried out analysis of community structure of networks was Weiss and Jacobson [16]. In their work they were detecting work groups in a government agency. Work groups were separated by removing members working with different groups thus leaving the groups by themselves. This idea was further used in a work of Girvan and Newman [17] which gave a substantial push to this research area. In their work, they identify edges which lie between communities and sequentially remove them. In this way they isolate communities from each other. Edges are removed according to their edge betweenness value which shows the frequency with which the edge is used by other edges to reach other. Currently many works propose some modifications of this algorithm. Below our method description follows.

## II. Detection of Communities in Weighted Networks

A fundamental problem in the analysis of network data is the detection of network communities, groups of densely interconnected nodes, which may be overlapping or disjoint. Network communities play important organizational and functional roles in complex networks. Consequently, the identification of communities in complex networks has become one of the most active areas of research in network theory. Complex networks are the structural skeleton of complex systems, which are ubiquitous in nature, society and technology.

A network is represented by a graph, $G=(V,E)$, where the set of nodes $E$ represents the entities of the system and the set of links $E$ represents the relationship between these entities. Given a set of data objects $O=\{o_1,o_2,...,o_n\}$ with associated positive weights $\alpha_i$ and a set of edges $(i,j)\in E$ with associated positive weights (e.g., similarities $w_{ij}$), where $i\neq j=1,2,...,n$. We address the partitional clustering problem of finding a single partition $C=\{C_1,C_2,...,C_k\}$ of $n$ data objects $O=\{o_1,o_2,...,o_n\}$ into a fixed number of $k$ clusters. That is, each data object is assigned to exactly one of those clusters and every cluster has at least one object.

This can be formulated as an optimization problem by defining binary decision variables:

$$x_{iq}=\begin{cases}1, \text{ if } o_i\in C_q\\ 0, \text{ otherwise}\end{cases},$$

$$y_q=\begin{cases}1, \text{ if an object } o_q \text{ is selected as a cluster}\\ 0, \text{ otherwise}\end{cases},$$

where $i,q=1,2,...,n$.

Then partitional clustering problem can then be formulated as a Boolean programming problem:
maximize

$$f(x,y)=\beta_1\cdot\sum_{q=1}^{n}\sum_{i=1}^{n}\alpha_i x_{iq}y_q+\beta_2\cdot\sum_{q=1}^{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}w_{ij}x_{iq}x_{jq}y_q-$$

$$-\beta_3\cdot\sum_{p=1}^{n-1}\sum_{q=p+1}^{n}w_{pq}y_p y_q \quad (1)$$

subject to

$$\sum_{q=1}^{n}x_{iq}=1,\ i=1,2,...,n, \quad (2)$$

$$\sum_{i=1}^{n}x_{iq}\geq 1,\ q=1,2,...,n, \quad (3)$$

$$\sum_{q=1}^{n}y_q=k, \quad (4)$$

$$x_{iq}\leq y_q,\ i,q=1,2,...,n, \quad (5)$$

$$x_{iq}\in\{0,1\},\ i,q=1,2,...,n, \quad (6)$$

$$y_q\in\{0,1\},\ q=1,2,...,n, \quad (7)$$

where $\beta_1,\beta_2,\beta_3$ are the weighting parameters, specifying the relative contributions of the terms to the hybrid objective function $f$, which $\beta_1,\beta_2,\beta_3\in[0,1]$ where $\beta_1+\beta_2+\beta_3=1$.

The objective in (1) is to maximize total weight of all selected clusters and minimize the similarity between the selected clusters. The first and second terms in (1) calculate the total weight of all selected clusters and the third term calculates similarity between the selected clusters. The set of constraints in (2) and integrality conditions in (6) impose that every data object is assigned to exactly one cluster. The set of constraints in (3) assures that every cluster has at least one object. The constraint in (4) ensures that only $k$ clusters are selected. The set of constraints in (5) guarantees that a cluster must be selected if there is any data object assigned to it.

In (1), the positive weight $\alpha_i$ of the object (node) $o_i$ is its information centrality which was introduced by Stephenson and Zelen [18] as a measure of node centrality of social networks. It is based on information that can be transmitted between any two points in a connected network. The motivation for this measure comes from the theory of statistical estimation. Here a path connecting two nodes is considered as a "signal", while the "noise" in the transmission of the signal is measured by the variance of this signal. It models any path from $j$ to $l$ as a signal transmission, which has a channel noise proportional to its path length.

## III. CONCLUSION

Finding communities in social networks is an important research area in social network analysis. Community detection is used in many disciplines and finds its applications in many areas. Many works are dedicated for detecting communities in social networks extracted on the Web. Each of them is applicable to some but not all kinds of networks. In this paper, we have proposed a new approach of extracting communities in networks. The details of the method were also shown.

## 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, and 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, and C.Ramakrishnan, "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, and H.Takeda, "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, and L. Yao, "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] R. Alguliev, R. Aliguliyev, and F. Ganjaliyev, "Investigation of the role of similarity measure and ranking algorithm in mining social networks", Journal of Information Science, 2011, vol. 37, no. 3, pp. 229-234.

[16] R.S. Weiss and E. Jacobson, "A Method for the Analysis of the Structure of Complex Organizations", American Sociological Review, 1995, vol. 20, no. 6, pp. 661-668.

[17] M. Girvan and M.E.J.Newman, "Community structure in social and biological networks", Proceedings of the National Academy of Sciences, USA, 2002, pp. 7821-7826.

[18] K. Stephenson and M. Zelen, "Rethinking centrality: methods and examples", Social Networks, 1989, vol. 11, no. 1, pp.1-37.