MINING CHALLENGES ON SOCIAL NETWORKS DATA

Gabriel TUTOKY, Ján PARALIČ

Technical University of Kosice
Faculty of Electrical Engineering and Informatics
Letná 9/B, 042 00 Košice, Slovakia
{gabriel.tutoky,jan.paralic}@tuke.sk

Abstract. Social Network Analysis is still a popular research area where researchers are faced to large scale and data heterogeneity challenges. This article deals with two methods which we experimentally used on real life networks for solving some of the mentioned challenges. First method "ageing of the ties among actors" was used for transforming heterogeneous two-mode community network data into one-mode social network where friendship strength was determined on previous actor's collaborations on events and also on the time spent between events. Second method "local structure analysis" was used for analysing large scale Slovak company networks where we identified common local structures (patterns) in this network.

1. Introduction

Social Networks are usually large-scale networks with thousands of nodes and edges. Analysis of these networks, usually based on global properties, can lead to interesting and helpful results. There are many ways how to analyse social networks and how to use obtained information e.g. movie recommendation [6] or analysis of Twitter microblogs [7]. Nevertheless, there exist many different situations in network analysis where data used for analysis did not carry sufficient information. In this paper we describe two approaches how to process and analyse networks with additional time related information [1], [2] and how to analyse a network with respect not to its global but to its local properties [3]. First approach is based on the principle of projection of two-mode (affiliation or actor-to-event) networks onto one-mode (actor-to-actor) networks. In the projection phase we used additional information from heterogeneous resources such as time and length of the event, actor's role on the event, event's type etc. Second approach is focused on mining and analysing of local clusters in networks which are not very dense, but density of clusters hidden inside them is much larger. This analysis consists of two processes: first is decomposition of large network with good-enough local structure preservation and second is local structure pattern detection.

2. Ageing of the Ties among Network Actors

Collaboration network indirectly describes social relations among actors in form of two-mode (affiliation) network where the first mode is a set of actors (first type of nodes in the network), and the second mode is a set of events which affiliate the actors (second type of nodes in the network). Analysis of such network focused on predicting participation of actors on events is presented in work [5]. In other way, two-mode network can be projected onto one-mode network where actors are linked to one another (i.e. a network without nodes of the second type). Weights in projected network have usually binary values, but we can use several measures to estimate weights of the relations among actors by e.g. number of actors which affiliate on the same event. Additionally we can have more information about actors and events e.g. actor's role on particular event or event properties like event type, time of the event, event duration etc. In our work [1], [2] we proposed several approaches how to use such data for better estimation of the relations' weights among actors. We also proposed time dependent weights in our representation of one-mode projected affiliation network – a kind of aging of the ties. Our proposed weights' aging method is based on assumption that past collaborations among network actors are less important than lately created collaborations. These past collaborations after passing sufficiently long time have no more influence in the present and they are next removed from the network – old ties (without refreshing) among collaborators are than "forgotten" in such a way. From the social network analysis point of view our proposal of aging of the edges can lead to new opportunities in network analysis like *Tracking collaboration over the time* or *Creation of network snapshots in given time*.

3. Local Structure Analysis and Pattern Mining

Local structure analysis is closely connected to common analysis called positional analysis in (social) networks and definition of equivalency is prerequisite for positional analysis. There are several types of equivalencies defined, like local role equivalency (LRE), automorphic (AE), regular (RE) and structural equivalency. If we want to define equivalency, we need to understand local structure of the network and vice versa. One way to achieve understanding of this structure and determining structural equivalencies is to use blockmodeling technique, but blockmodeling of large networks can be difficult, very time consuming and problematic. Proposed Local Structure Analysis is more time effective with similarly good results. This analysis consists of two phases: first one is decomposition of a large network (i.e. disconnecting the network into sub-networks with good-enough local structures' preservation) and second phase is Local Structure Pattern Mining (LSPM). During the first phase we needed to choose suitable actors for removal and experimentally best criterion for decomposition showed to be actors' betweeness centrality. After removing of only 5% of actors in Slovak Company Network [4] (used in our experiments) with highest betweeness centrality, the largest component was disconnected into a group of small enough components suitable for our LSPM. In this second phase we analysed small components in order to extract some regular patterns in their local structures. After analysis we were able to observe several most frequent patterns over smaller components.

4. Conclusions

In this work we summarized our research devoted with the aim to solve some of the mining challenges in analysis of Social Network Data. Both approaches *Ageing of the ties among network* actors and *Local Structure Analysis and Pattern Mining* were presented with experimental results in our works listed below.

Acknowledgement: This work was partially supported by Slovak Research and Development Agency under the Contract No. APVV-0208-10 and partially by the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic under Grant No. 1/1147/12.

References

to other papers publishing the results that are summarized here

[1] Tutoky, G., Paralič, J.: Modeling of Collaboration Social Networks. In: *Computational Collective Intelligence: Technologies and Applications: Third International Conference*, ICCCI 2011: Gdynia, Poland, September 21-23, 2011: Proceedings: Part 1. - Berlin: Springer-Verlag, LNAI, 2011, pp. 407-418. ISBN 978-3-642-23934-2, ISSN 0302-9743

[2] Tutoky, G., Paralič, J.: Weights Aging in Social Networks. *Aspects of Computational Intelligence: Theory and Applications*. Topics in Intelligent Engineering and Informatics, Volume 2, Springer Berlin Heidelberg, 2013, p. 207-218

[3] Repka, M., Paralič, J. : Advanced Analysis of Local Structures in Company Network. In: *Journal of Convergence Information Technology*. Vol. 8, no. 11 (2013), p. 645-654. - ISSN 1975-9320

[4] Repka, M., Paralič, J.: Company Networks Analysis. *LAP Lambert Academic Publishing, Saarbrucken*, Germany, 2013, 139 p.

Other references

[5] Michalco, J., Navrat, P.: Arrangement of Face-to-Face Meetings using Social Media. Studies in Informatics and Control. (2012), Vol. 21, No. 4, pp. 383-392.

[6] Demovic, Lubos; Fritscher, Eduard; Kriz, Jakub; Kuzmik, Ondrej; Proksa, Ondrej; Vandlikova, Diana; Zelenik, Dusan; Bielikova, Maria, "Movie Recommendation Based on Graph Traversal Algorithms," Database and Expert Systems Applications (DEXA), 2013 24th International Workshop on , vol., no., pp.152,156.

[7] F. Ricci, L. Rokach, and B. Shapira, “Introduction to recommender systems handbook,” in Recommender Systems Handbook. Springer US, 2011, pp. 1–35.