

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

Overlapping Community Detection Model in Online Social Networks

A. Rabin¹, Mr. C.E. Mohan Kumar M.E.,²

P.G. Student, Department of ECE, GRT Institute of Engineering and Technology, Tiruvallur, Tamil Nadu, India¹

Assistant Professor, Department of ECE, GRT Institute of Engineering and Technology, Tiruvallur, Tamil Nadu, India²

ABSTRACT:-Community detection in a Social network is a trending issue in the study of network system and it helps to realize and study the overall network structure in detail. Communities are basically the partition of network nodes into some subgroups in which nodes within these subgroups are densely connected, but between the subgroups, the connections are sparser. In this paper different overlapping community detection algorithms are studied in order to detect the overlapping node in the network. Some modularity measures are also considered to measure the quality of communities detected by the algorithms

KEYWORDS: Complex network, Community Structure, Overlapping community

I. INTRODUCTION

Networks are omnipresent on the Web. The most profound Web network is the Web itself comprising billions of pages as vertices and their hyperlinks to each other as edges. Moreover, collecting and processing the input of Web users (e.g. queries, clicks) results in other forms of networks, such as the query graph. Finally, the widespread use of Social Media applications, such as Delicious, Digg, Flickr and YouTube, is responsible for the creation of even more networks, ranging from folksonomy networks to rich media social networks. Since networks originating from Social Media data are of particular interest to this study, it shall be collectively refer as Social Media networks. A social network can be defined as a set of people connected by a set of people. A social structure made of nodes (individuals and organizations) that are related to each other by various independencies like friendship, kinship etc. Social network analysis provides both visual and mathematical analysis of human relationship.

In the field of virtual community detection in online social networks, most of existing methods often detect communities from a single perspective, and ignore the influence of related characteristics of the networks on community detection. All these reduce the interpretability and accuracy of the community partition results. In order to resolve this problem, a virtual community detection model framework for online social networks is proposed. The model framework considers three key factors that affect the community detection results: the structural characteristics, the attribute information and the nodes influence levels of the network. The proposed model is not only a mapping of existing community detection models, but also a reference for designing more future models for community detection methods.

II. RELATED WORK

1) Using Spectral Clustering of Hashtag Adoptions to Find Interest-Based Communities

Aurora Schmidt ; Clay Fink ; Vladimir Barash ; Christopher Cameron ; Michael Macy

We demonstrate an efficient spectral clustering approach, that leverages power iteration on symmetric adjacency matrices, to group users based on hashtag adoptions prior to the kidnapping. Unlike follow networkbased clusters, these adoption-based clusters reveal groups of users with similar interests and prove to be more predictive of interest in future topics. We compare this unsupervised spectral clustering to spectral clustering based on symmetrized follow network relations as well as clusters induced by latent Dirichlet allocation (LDA) topics.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

2) Data Clustering and Graph Partitioning via Simulated Mixing ShahzadBhatti ; Carolyn Beck ; Angelia Nedich

In this paper, we propose a novel spectral clustering algorithm based on a mixing process on a graph. Unlike the existing spectral clustering algorithms, our algorithm does not require computing eigenvectors. Specifically, it finds the equivalent of a linear combination of eigenvectors of the normalized similarity matrix weighted with corresponding eigenvalues. This linear combination is then used to partition the dataset into meaningful clusters.

3) Community detection in social networks using a novel algorithm without parameter BinazirBalegh ;SaeedFarzi

In the present study, by determining the number of clusters and specifying initial centres of clusters, a novel algorithm has been introduced. By defining the neighbourhood of nodes, the proposed algorithm selects the initial centres of clusters in a way that the centres are at the maximum distance from each other. This algorithm has been evaluated using the three known and commonly used algorithms, i.e., K-means, k-means++ and K-Rank-D over the Dolphin, Karate club and Soccer team datasets. The results show the superiority of the proposed algorithm over other algorithms.

4) An empirical study of opinion leader in online social network - Fei Wang

Opinion leader in online social networks is important for various applications in many domains such as advertisement, administrative science and even politics. In this paper, a typical online social network is selected as empirical network to analyze its network topologies with complex network theory. Our empirical results show the topological characteristics of opinion leader nodes, which would improve accuracy of detecting opinion leader in online social network

5) Research on Mining and Detection Method of Abnormal Learning Behavior Yan Cheng ; Yong-Chun Miao ; Ping-Fei Tan ; Yan-Wen Qu

In order to mine the abnormal learning behavior of learner in virtual learning community and carry on personalized supervision and guidance, behavior filtering model based on the factor analysis of the behavior is constructed to solve the problem of the relationship between behavior factors in the learning behavior vector space model. In view of the disadvantage of neglecting local abnormal points in a global mining, a community abnormal behavior detection algorithm based on the clustering analysis method was proposed.

6) Research on online social network influence and its uncertain query technology Huijun Tang

Influence is an important part of social network. After the appearance of online social network, the data could be analyzed conveniently. Based on the influence of important nodes and communities in the network, a new influence modeling method which is based on twin support vector machines is proposed.

III. PROPOSED ALGORITHM

Page Rank Algorithm

PageRank is a link analysis algorithm and it assigns a numerical weighting to each element of a hyperlinked set of documents, such as the World Wide Web, with the purpose of "measuring" its relative importance within the set. The algorithm may be applied to any collection of entities with reciprocal quotations and references. The numerical weight that it assigns to any given element E is referred to as the PageRank of E and denoted by $\{PR(E)\}$. Other factors like Author Rank can contribute to the importance of an entity.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

A PageRank results from a mathematical algorithm based on the webgraph, created by all World Wide Web pages as nodes and hyperlinks as edges, taking into consideration authority hubs such as cnn.com or usa.gov. The rank value indicates an importance of a particular page. A hyperlink to a page counts as a vote of support.

The PageRank of a page is defined recursively and depends on the number and PageRank metric of all pages that link to it ("incoming links"). A page that is linked to by many pages with high PageRank receives a high rank itself.

Disadvantages

1. Older pages may have higher rank. It is because a new page even have some very good contents but it may not have many links in the early state.
2. Page Rank can be easily increased by the "link-farms"

Trust Rank Algorithm

TrustRank seeks to combat spam by filtering the web based upon reliability. The method calls for selecting a small set of seed pages to be evaluated by an expert. Once the reputable seed pages are manually identified, a crawl extending outward from the seed set seeks out similarly reliable and trustworthy pages. Trust Rank's reliability diminishes with increased distance between documents and the seed set. The logic works in the opposite way as well, which is called Anti-Trust Rank. The closer a site is to spam resources, the more likely it is to be spam as well. The researchers who proposed the TrustRank methodology have continued to refine their work by evaluating related topics, such as measuring spam mass.

Advantages

Automatically boost pages that have a high probability of being good, as well as demote the rankings of pages that have a high probability of being bad. Help search engines identify what pages should be good candidates for quality review

IV. RESULTS AND DISCUSSIONS

A social network is a collection of nodes representing individuals or organizations with dyadic or binary relationship between them. It is usually represented by a graph $G(V, E)$ where V is the set of vertices representing the individuals participating in the network and E is the set of edges representing the interactions between the vertices. Examples of social network can be given as scientists co-authoring a paper, employees of a company working on a common project, etc. A community represents a group of individuals such that the frequency of interactions within the group is more than that of the interactions between the groups. Community detection problem refers to the problem of finding such groups in real world social networks. A number of methods to address this problem have been proposed earlier, and new man distinguishes these into two categories: bottom-up sociological approaches and top-down computer science approaches. Modularity is a property of the network that measures when the division is good, in the sense that there are many edges within the community and only a few between them. In modularity based algorithms, each node of the graph is considered as an individual community and the communities are joined iteratively based on the increase in modularity caused by their joining. The ones producing maximum change in modularity are joined. There are few drawbacks associated with modularity based methods such as they require information regarding the entire structure of the network which is not possible to determine in case of vast real world networks. Also modularity optimization methods are not able to determine the overlapping communities. In order to detect overlapping communities clique percolation can be used. Clique percolation is based on the assumption that a community consists of fully connected sub graphs and detects overlapping communities by searching for adjacent cliques. But it is a hard method to implement well due to difficulty of producing intermediate representations of percolating structures. In this project, we implement the k-clique percolation method using a Clique matrix and binary matrix in order to store the intermediate percolating structure with an aim of simplifying the implementation.

V. CONCLUSION AND FUTURE WORK

Based on the in-depth study of the online social networks, this paper proposes a unified framework model for online social networks from three aspects: the structural characteristics, the relevant attribute information, and the nodes

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

influence levels of the network. The scalability and accuracy of the existing community detection methods can be improved according to this model. A virtual community detection model framework for online social networks is proposed. The model framework considers three key factors that affect the community detection results: the structural characteristics, the attribute information and the nodes influence levels of the network. The proposed model is not only a mapping of existing community detection models, but also a reference for designing more future models for community detection methods.

REFERENCES

- [1] Aurora Schmidt ; Clay Fink ; Vladimir Barash ; Christopher Cameron ; Michael Macy Using Spectral Clustering of Hashtag Adoptions to Find Interest-Based Communities, 2018
- [2] ShahzadBhatti ; Carolyn Beck ; Angelia Nedich,Data Clustering and Graph Partitioning via Simulated Mixing, 2018
- [3] BinazirBalegh ;SaeedFarzi, Community detection in social networks using a novel algorithm without parameter, 2017
- [4] Fei Wang, An empirical study of opinion leader in online social network, 2017
- [5] Yan Cheng ; Yong-Chun Miao ; Ping-Fei Tan ; Yan-Wen Qu, Research on Mining and Detection Method of Abnormal Learning Behavior , 2016
- [6] Huijun Tang, Research on online social network influence and its uncertain query technology, 2016
- [7] C. Gui, R. Zhang, R. Hu, G. Huang, and J. Wei, "Overlapping communities detection based on spectral analysis of line graphs," *Physica A: Statistical Mechanics and Its Applications*, vol. 498, pp. 50- 65, 2018.
- [8] S. White and P. Smyth, "A spectral clustering approach to finding communities in graphs," in *Proceedings of the 2005 SIAM International Conference on Data Mining*, 2005, pp. 274-285.
- [9] X. Zhang and M. E. J. Newman, "Multiway spectral community detection in networks," *Physical Review E*, vol. 92, p. 052808, 2015.
- [10] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Physical Review E*, vol. 69, p. 066133, 2004.
- [11] R. Guimera, M. Sales-Pardo, and L. A. N. Amaral, "Modularity from fluctuations in random graphs and complex networks," *Physical Review E*, vol. 70, p. 025101, 2004.
- [12] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, p. P10008, 2008.
- [13] T. Murata, "Modularity for heterogeneous networks," in *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*, 2010, pp.129-134.
- [14] X. Liu, W. Liu, T. Murata, and K. Wakita, "A framework for community detection in heterogeneous multi-relational networks," *Advances in Complex Systems*, vol. 17, p. 1450018, 2014.
- [15] Y. Y. Ahn, J. P. Bagrow, and S. Lehmann, "Link communities reveal multiscale complexity in networks," *Nature*, vol. 466, pp. 761-764, 2010.
- [16] C. Shi, Y. Cai, D. Fu, Y. Dong, and B. Wu, "A link clustering based overlapping community detection algorithm," *Data and Knowledge Engineering*, vol. 87, pp. 394-404, 2013.
- [17] L. Huang, G. Wang, Y. Wang, E. Blanzieri, and C. Su, "Link Clustering with Extended Link Similarity and EQ Evaluation Division," *Plos One*, vol. 8, p. e66005, 2013.
- [18] X. Zhou, Y. Liu, J. Wang, and C. Li, "A density based link clustering algorithm for overlapping community detection in networks," *PhysicaA :Statistical Mechanics and Its Applications*, vol. 486, pp. 65-78, 2017.