 http://doi.org/10.36893/JES.2024.V15I10.010

ROBUST DETECTION OF LINK COMMUNITIES WITH SUMMARY DESCRIPTION IN SOCIAL NETWORKS

[1]MRS B DURGA BHAVANI, [2]NANDIGAMA NIHARIKA, [3]LOYAPALLY PRANITHA, [4]KONDU ARCHITHA

[1]Assistant Professor, Department of Information Technology, **MALLA REDDY ENGINEERING COLLEGE FOR WOMEN**, Maisammaguda, Dhulapally Kompally, Medchal Rd, M, Secunderabad, Telangana.

[2, 3,4] Student, Department of Information Technology, **MALLA REDDY ENGINEERING COLLEGE FOR WOMEN**, Maisammaguda, Dhulapally Kompally, Medchal Rd, M, Secunderabad, Telangana.

ABSTRACT

Community detection has been extensively studied for various applications. Recent research has started to explore node contents to identify semantically meaningful communities. However, links in real networks typically have semantic descriptions and communities of links can better characterize community behaviors than communities of nodes. The second issue in community finding is that the most existing methods assume network topologies and descriptive contents carry the same or compatible information of node group membership, restricting them to one topic per community, which is generally violated in real networks. The third issue is that the existing methods use top ranked words or phrases to label topics when interpreting communities, which is often inadequate for comprehension. To address these issues altogether, we propose a new Bayesian probabilistic approach for modeling real networks and developing an efficient variational algorithm for model inference. Our new method explores the intrinsic correlation between communities and topics to discover link communities and extract semantically meaningful community summaries at the same time. If desired, it is able to derive more than one topical summary per community to provide rich explanations. We present experimental results to show the effectiveness of our new approach and evaluate the method by a case study.

1 INTRODUCTION

Social networking has become increasingly important for connecting people of diverse background. It is prevalent over the Internet for geographically dispersed users. As a result, large quantities of network data, particularly from social sciences, have been accumulated. Analysis of such large quantities of data is in demand to

http://doi.org/10.36893/JES.2024.V15I10.010

help reveal underlying social structures, discern organizational behavior and predict future trends.

Graph is the simplest form of a network. It represents basic units as nodes and relationships between them as links. A growing interest in social networks has revived graph mining algorithms[1].An important problem in analyzing social networks is the problem of community detection [2]. The primary objectives of community finding is to identify groups of nodes with common functions and meaningful functional structures of the groups of nodes. A group of nodes form a dense region of closely related entities in a graph, and thus constitute a community. Finding such communities is an effective means to social network analysis, e.g., personalized recommendations and recognition of abnormal activities.

The majority of the existing community detection methods use exclusively information of network topologies, as reviewed in [3]. These include hierarchical clustering[1], [4], modularity-based methods [5], [6], [7], spec-tral optimization algorithms [8], [9], Markov dynamic algorithms [10], [11], and statistical inference methods[12], [13],[14].

However, content information, e.g., descriptions of interactions among the entities in a network, may also pro-vide useful information on network communities. Indeed, it has been shown recently that the use of node contents can significantly improve the quality of the resulting communities. Along this line are topic model-based methods [15], [16], generative models [17]and heuristic methods[18]. Information of network topologies and node contents are also complementary to each other; if one is missing or inaccurate, the other can be used to make up for the missing or noisy data. More importantly, node contents can also be used to discover interpretable community descriptions to help reveal the latent functions of individual communities. Communities with functional descriptions are desirable in practice and have attracted attention lately, e.g., in the latest works[19], [20], [21].

While some progress has been made, the results of the existing methods are far from satisfactory in several aspects .In addition to network topology, content information on links is also prevalent and appears more often than information on nodes in social and other real-world networks. Content information on links may often include descriptions of interactions among the

ISSN:0377-9254

 http://doi.org/10.36893/JES.2024.V15I10.010

entities in a network, providing information of network communities that is complementary and orthogonal to information of network topology. Furthermore, a group of users in a social network may interact over more than one topic of common interest and consequently form a community. They often exchange text messages, which are naturally represented by links connecting them in the network. These links may cover more than one topic. That is, a community may have more than one topic, as illustrated in Figure 1(a). This observation helps reveal several limitations of the existing methods.

First, supported by the observation above, contents on links carry more information of community structures than contents on nodes. However, the existing methods typically exploit node contents. In order to apply an existing algorithm to networks with information on links, we may combine all messages sent by a user as the content of the node (Figure 1(b)). This conversion of link content to node content may lose information, e.g., addressees in Figure 1(b), and reduce the effectiveness of the method because the user may exchange messages with others in different communities, some of which may not even be associatedor consistent at all.

Second, it is usually assumed in the existing methods that network structures and node contents have the same information of node group memberships, i.e., communities and topical clusters provide the same information. This assumption is often violated. For example, social relations in Twitter often directly reflect users' groups, whereas user-generated contents may be diverse[22]. Therefore, when node contents do not match well with community structures, the performance of these algorithms may deteriorate.

Third, the existing methods aim at finding one topic for one community, despite that communities in real social networks may have multiple topics, providing limited interpretability of resulting communities. For example, there are red and blue user communities in Figure 1(b). The red community naturally has two topics, which are difficult to distinguish by the existing methods. In-stead, these methods will interpret this community by a mixture of these two topics, which is difficult to comprehend.

Finally, the existing methods use individual words or short phrases to summarize communities, even though the text messages exchanged among users are typically complete sentences that have more information than individual words. It may not be straightforward to under-stand communities using a few words. Take Figure 1(b) as an example. It is difficult to appreciate the listed topics without knowing how the words used are related. This issue is exacerbated when the top-ranked terms for the topics are also overlapped, e.g., the mixed topic of red community.

Although link contents are more informative than node contents and the former have unique characteristics that are missing in the latter[23], there are few methods that have been developed to use both network topologies and link contents for finding link communities.

We developed a new Bayesian probabilistic model for finding link communities with informative explanations. We transformed model inference to the problem of *maxi-mum a posteriori*(MAP) learned by an efficient variational algorithm to exploit network topologies, link contents and their intrinsic correlations. We like to highlight that our method addresses altogether the four problems discussed above. It does not assume that topologies and con-tents share the same community memberships, is able to interpret a community by more than one topic, and uses whole sentences to summarize communities, as illustrated in Figure 1(a).

An initial version of our method and some preliminary results were presented earlier in a conference [38].These results are significantly extended in this paper by adopting a Bayesian approach to treat the parameters in [38] as random variables, explicitly modeling the intrinsic uncertainty between model parameters, an improved model generative process, new mathematical derivation as well as more experimental validations.

 http://doi.org/10.36893/JES.2024.V15I10.010

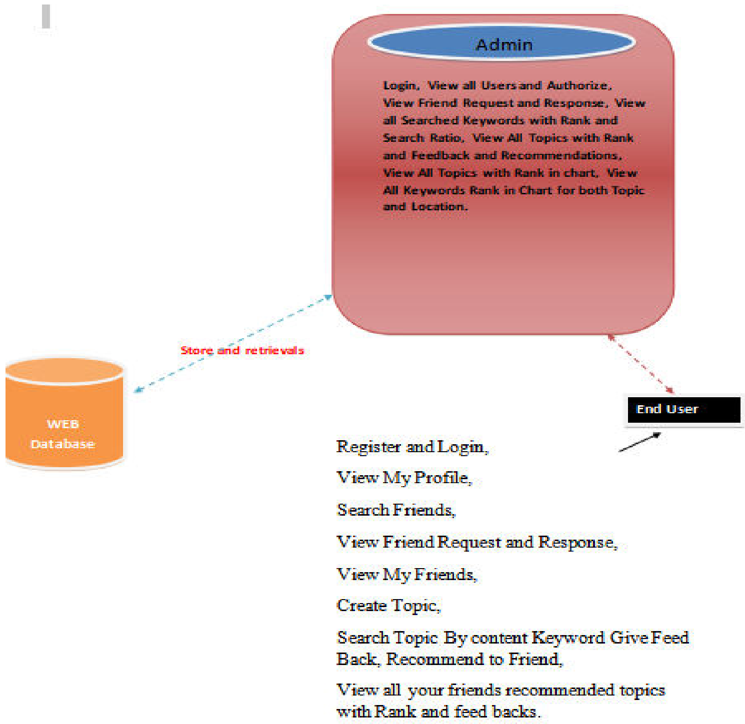

Fig1: System Architecture

II.EXISTING SYSTEM

However, these methods typically use the information of network topology alone, and ignore the content information which is also valuable for inferring communities, e.g., individuals with similar contents are more likely to belong to the same community. Thus, some methods have been recently proposed to use both of these two sources of information to better find communities. For instance, several topic model-based methods have been applied to combine content analysis and link analysis in a unified framework. Examples include generative models that combine a generative link model and a generative content model through some shared hidden variables and then estimate the conditional distributions to find the community assignments.

Block-LDA [16] and Link-PLSA-LDA [25] are both representatives of this type of methods. In addition, Yang et al. [26] introduced an alternative discriminative probabilistic method to detect communities, which is different from the generative models. This model uses node popularities to calculate the probability whether two nodes are connected, and then incorporates content information in the link model to estimate community memberships. Neville et al.

[27] proposed a weighted adjacency matrix to take into account the content similarity measure used. The weight of each edge is defined as the number of

attribute values shared by the two end vertices. They then applied three existing graph clustering algorithms to the weighted adjacency matrix to perform a network clustering. Ruan et al. [18] proposed a heuristic method that hinges upon the intuition that many realworld networks contain erroneous links and that content information can help strengthen signals of community identities. However, these methods combine node contents with network topologies to find communities without explanations.

With further development of community detection using network contents, it has attracted much interest to introduce descriptions to interpret communities identified and make them understandable in practice. Several algorithms using both network and node contents have been proposed recently for this task. For example, Liu et al. [28] proposed a method that treated a network as a dynamic system and considered its community structure as a result of interactions among nodes. They further described the interactions among nodes in two different ways, including a linear model to approximate influence propagation, and a model of random walks characterizing the interactions among nodes.

By modeling interactions, the nature of communities is described by analyzing the stable states of the dynamic system. Wang et al. [19] proposed a nonnegative matrix factorization model with two matrices, one for community memberships and the other for community attributes, and developed efficient updating rules to derive the matrix parameters with a convergence guarantee. This method uses the node attributes to improve the effectiveness of community detec-tion and provides semantic interpretations to the result-ing network communities as well. CESNA [17] by Yang et al. uses a generative model for networks with node attributes. It avoids the assumption of soft membership that nodes sharing multiple common communities are less likely to be connected, but assumes that communities “generate” both the network and attributes.

Disadvantages

1) It is usually assumed in the existing methods that network structures and node contents have the same information of node group memberships.
2) Finding one topic for one community, despite that communities in real social networks

 http://doi.org/10.36893/JES.2024.V15I10.010

may have multiple topics, providing limited interpretability of resulting communities.

III.PROPOSED SYSTEM

The proposed system developed a new Bayesian probabilistic model for finding link communities with informative explanations. We transformed model inference to the problem of *maxi-mum a posteriori* (MAP) learned by an efficient variational algorithm to exploit network topologies, link contents and their intrinsic correlations. We like to highlight that our method addresses altogether the four problems dis-cussed above. It does not assume that topologies and con-tents share the same community memberships, is able to interpret a community by more than one topic, and uses whole sentences to summarize communities, as illustrat-ed in Figure 1(a).

An initial version of our method and some preliminary results were presented earlier in a conference [38]. These results are significantly extended in this paper by adopt- ing a Bayesian approach to treat the parameters in [38] as random variables, explicitly modeling the intrinsic uncertainty between model parameters, an improved model generative process, new mathematical derivation as well as more experimental validations.

Advantages

1) The system is more effective due to presence of Generating Observed Quantities and Latent Quantities.

2) The system is fast due to Iterative Optimization Algorithm.

IV.IMPLEMENTATION

Admin

In this module, the Admin has to login by using valid user name and password. After login successful he can perform some operations such as Login, View all Users and Authorize, View Friend Request and Response, View all Searched Keywords with Rank and Search Ratio, View All Topics with Rank and Feedback and Recommendations, View All Topics with Rank in chart, View All Keywords Rank in Chart for both Topic and Location.

Friend Request & Response

In this module, the admin can view all the friend requests and responses. Here all the requests and responses will be displayed with their tags such as Id, requested user photo, requested user name, user name request to, status and time & date. If the user accepts the request then the status will be changed to accepted or else the status will remains as waiting.

.

Social Network Friends

In this module, the admin can see all the friends who are all belongs to the same site. The details such as, Request From, Requested user's site, Request To Name, Request To user's site.

All Recommended Posts

In this module, the admin can see all the posts which are shared among the friends in same and other network sites. The details such as post image, title, description, recommend by name and recommend to name.

User

In this module, there are n numbers of users are present. User should register before performing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user can perform some operations like Register and Login, View My Profile, Search Friends, View Friend Request and Response, View My Friends, Create Topic, Search Topic By content Keyword Give Feed Back, Recommend to Friend, View all your friends recommended topics with Rank and feed backs.

Searching Users

In this module, the user searches for users in Same Site and in Different Sites and sends friend requests to them. The user can search for users in other sites to make friends only if they have permission.

Adding Posts

In this module, the user adds posts details such as title, description and the image of the post. The post details such as title and description will be encrypted and stores into the database.

V.CONCLUSION

We have proposed a new Bayesian probabilistic model for link community detection that explores network to apologies and link contents, and developed an efficient variational algorithm for learning the model. The new algorithm was developed particularly to address the four critical issues that have not been adequately considered in the current methods for community detection, i.e., 1) se mantic or text contents in social networks are often associated with links rather than with nodes, and hence may form multiple communities of links; 2) network to apologies and link contents may not share the same information of community memberships, so that a community may have more than one topic; 3) when contents do not match well with network communities, the results of detected communities deteriorate even with additional node contents; and 4) it is desirable to have more contextual summaries for better describing the topics of a communi-ty. For these issues, our method was designed to discover link communities, extract summaries for topic labeling, and explore the intrinsic correlation of communities and topics, all at once. By exploiting this correlation, our mod-el can not only combine network topologies and contents to accurately identify link communities, but also interpret each community using more than one topical summary if necessary, providing richer explanations. We evaluated the new method on two types of real networks, on which the new method out performed nine state-of-the-art methods.

In this paper we focused mainly on designing an accurate model, and getting the number of communities $c$via model selection when c is unknown. We considered cases where the numbers of communities c and topics k are the same since the datasets we used have those two numbers being equal. Our method is also suitable for situations when $c \neq k$.

VI. REFERENCES

1] A C.C. Aggarwal and H. Wang, *Managing and mining graph data. Advances in Database Systems*, Springer, 2010.

[2] M. Girvan and M.E.J. Newman, “Community structure in social and biological networks,” *Proceedings of the National Academy of Sciences of the United States of America,* vol. 99, no. 12, pp. 7821-7826, Jun. 2002.

[3] S. Fortunato and D. Hric, “Community detection in networks: A user guide,” *Phys. Rep,* vol. 695, pp. 1-44, Nov. 2016.

[4] S. Jia, L. Gao, Y. Gao, J. Nastos, Y. Wang, X. Zhang and H. Wang, “Defining and identifying cograph communities in complex networks,” *New Journal of Physics,* vol. 17, no. 1, pp. 013044, 2015.

[5] M.E.J. Newman and M. Girvan, “Finding and evaluating community structure in networks,” *Phys. Rev. E,* vol. 69, no. 2, pp. 026113, 2004.

[6] P. Zhang and C. Moore, “Scalable detection of statistically significant

communities and hierarchies, using message pass-ing for modularity," *Proceedings of the National Academy of Sci-ences of the United States of America,* vol. 111, no. 51, pp. 18144-18149, 2014.

[7] L. Yang, X. Cao, ; D. He, and W. Zhang, "Modularity based community detection with deep learning," *Proc. 25th Interna-tional Joint Conference on Artificial Intelligence*, 2016, pp. 2083-2089.

[8] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zde-borová, and P. Zhang, "Spectral redemption in clustering sparse networks," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 110, no. 52, pp. 20935-20940, 2013.

[9] Y. Li, K. He, D. Bindel, and J.E. Hopcroft, "Uncovering the small community structure in large networks: A local spectral approach," *Proc. 24th Int. Conf. World Wide Web,* 2015, pp. 658-668.

[10] J.C. Delvenne, S.N. Yaliraki, and M. Barahona, "Stability of graph communities across time scales," *Proceedings of the Na-tional Academy of Sciences of the United States of America,* vol. 107, no. 29, pp. 12755-12760, 2010.

[11] M. Rosvall, A.V. Esquivel, A. Lancichinetti, J.D. West, and R. Lambiotte, "Memory in network flows and its effects on spreading dynamics and community detection," *Nature Com-munications*, vol. 5,no. 1, pp. 4630, Aug. 2014.

[12] B. Karrer and M. Newman, "Stochastic block models and community structure in networks," *Phys. Rev. E,* vol. 83, no. 2, pp. 211-222, 2011.

[13] A. Anandkumar, R. Ge, D. Hsu, and S.M. Kakade, "A tensor approach to learning mixed membership community models," *Journal of Machine Learning Research,* vol. 15, no. 6, pp. 2239-2312, 2014.

[14] D. He, D. Liu, D. Jin, and Z. Wang, "A stochastic model for detecting heterogeneous link communities in complex net-works," *Proc. Twenty-Ninth AAAI Conf. Artificial Intelligence*, 2015, pp. 130-136.

[15] Z. Zhao, S. Feng, Q. Wang, J.Z. Huang, G.J. Williams, and J. Fan, "Topic oriented community detection through social ob-jects and link analysis in social networks," *Knowledge-Based Systems,* vol. 26, pp. 164-173, 2012.

[16] R.Balasubramanyan and W.W. Cohen, "Block-LDA: Jointly modeling entity-annotated text and entity-entity

http://doi.org/10.36893/JES.2024.V15I10.010

links," *Proc. 11th SIAM International Conference on Data Mining*, 2011, pp. 450-461.

[17] J. Yang, J. McAuley, and J. Leskovec, "Community detection in networks with node attributes," *Proc. 13th IEEE Int. Conf. Data Mining*, 2013, pp.1151-1156.

[18] Y. Ruan, D. Fuhry, and S. Parthasarathy,"Efficient community detection in large networks using content and links," *Proc. 22nd Int. Conf. World Wide Web,* 2013, pp. 1089-1098.

[19] X. Wang, D. Jin, X. Cao, L. Yang, and W. Zhang, "Semantic community identification in large attribute networks," *Proc. Thirtieth AAAI Conf. Artificial Intelligence*, 2016, pp. 256-271.

[20] D. He, Z. Feng, D. Jin, X. Wang, and W. Zhang, "Joint Identifi-cation of Network Communities and Semantics via Integrative Modeling of Network Topologies and Node Contents," *Proc. Thirty-First AAAI Conf. Artificial Intelligence*, 2017, pp. 116-124.

She received B.Tech degree in the stream of IT from Sarada Institute Of Technology and Science in 2012 and M.Tech degree in the stream of CSE from Keshav Memorial Institute Of Technology in 2014. She is good academic Experience of 10 years. Her Area Of Research includes Network Security, Machine Learning. She attended various workshops and Faculty Development Programs.