



Practical Applications of Community Detection

Mini Singh Ahuja¹, Jatinder Singh², Neha³¹ Research Scholar, Department of Computer Science, Punjab Technical University, Punjab, India² Professor, Department of Computer Science, KC Group of Institutes Nawashahr, Punjab, India³ Student (M.Tech), Department of Computer Science, Regional Campus Gurdaspur, Punjab, India

Abstract: Network is a collection of entities that are interconnected with links. The widespread use of social media applications like youtube, flicker, facebook is responsible for evolution of more complex networks. Online social networks like facebook and twitter are very large and dynamic complex networks. Community is a group of nodes that are more densely connected as compared to nodes outside the community. Within the community nodes are more likely to be connected but less likely to be connected with nodes of other communities. Community detection in such networks is one of the most challenging tasks. Community structures provide solutions to many real world problems. In this paper we have discussed about various applications areas of community structures in complex network.

Keywords: complex network, community structures, applications of community structures, online social networks, community detection algorithms..... etc.

I. INTRODUCTION

Network is a collection of entities called nodes or vertices which are connected through edges or links. Computers that are connected, web pages that link to each other, group of friends are basic examples of network. Complex network is a group of interacting entities with some non trivial dynamical behavior [1]. There are many types of complex networks for example Social Networks, Technological Networks, Informational Networks, and Biological Networks etc. Social networking is an application that allows users to interact with each other. The study of complex networks is a youthful and vigorous area of scientific research stimulated largely by the pragmatic study of real-world networks such as computer networks and social networks [2]. Complex network paradigm is one of the modeling tools which have spread through several application fields such as sociology, communication, computer science, biology, and physics and so on during last decades [4].

Community structure

Community is a collection of nodes that are more densely connected as compared to the nodes outside the community. Nodes inside the community have some common properties. There are basically two categories of communities: disjoint community and overlapping community [3]. The disjoint community refers to crisp arrangement of nodes where node belongs to single community. The overlapping communities are known as fuzzy assignment of nodes where a node may overlap between two or more communities.

Community detection is the important aspect of the complex network study. Communities may relate to group of pages of World Wide Web with related topics and to functional modules in metabolic networks [4].

Community detection:

Detection of such community structures in complex network is not an easy task. There are various community detection algorithms available which detects communities in the network. Some of them are Hierarchical clustering algorithm, Divisive algorithm, Kernighan-Lin (KL) Graph Partitioning approach, Multilevel Graph Partitioning. These algorithms detect disjoint communities only. For overlapping community detection other algorithms like Clique Percolation Method (CPM), Eagle algorithm, Cluster Overlap Newman Girvan Algorithm (CONGA), Link Clustering, Label propagation algorithms etc are used.

II. APPLICATIONS OF COMMUNITY STRUCTURES

A. To detect suspicious events in Telecommunication Networks:

Charles Andre Reis pinheiro used the community structure to detect fraud events that may occur in Telecommunication networks [6]. Charles showed how community detection helps us to understand the behavior of customers on the basis of their calls and text messages over communication network. For this purpose Social Network Analysis was performed on the data from Telecommunication Company to recognize the unanticipated relations among customers. This approach disclosed the suspicious connections with in the communities and detected outliers.

I. Firstly, communities were detected using any of the basic community detection algorithms.

- II. After this step, some social measures such as Degree, Closeness, and Betweenness etc. were evaluated on these communities to detect outliers.

B. Refactoring the Software Packages:

Wei-Feng, Pan-Bo, Jianguo described a novel approach which refactors the software packages through community detection in complex software networks [7]. Refactoring refers to modifying the design of existing code or making code simple and elegant without changing the working of code.

- I. In this approach firstly an undirected weighted class dependency network of software at class level is build where nodes represents classes and edges represents interaction among classes. Weight is assigned to represent the strength of dependency among connected classes.
- II. The approach begins with the concept that at starting stage every class belongs to some specific community structure. This community structure is a package where classes are defined. Then series of class-moving operations are performed at classes that have interconnections or dependencies with other classes that are not defied in the same package. For this task author proposed Constrained Community Detection Algorithm (CCDA) to detect communities in undirected weighted class dependency network

C. Recommendation Systems:

Massimiliano Zanin presented an algorithm for recommender systems which recommends the most suitable products to the customers by predicting their interest [8]. This approach was item-based strategy where the system recommends products to customers that are related or similar to the products that the customer liked in the past. After this products were compared with other products in the network and then most related products and selected were recommended. Cosine-based similarity measure is used to compute similarity between two items. A vector of length N was generated for all items; here N represents total number of customers. The vector of nth element has value 1 only if the nth customer has brought that item in the previous time. Otherwise value remains 0. The distance between two items is formulated as:

$$dis_i(j) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{|\vec{i}| \cdot |\vec{j}|}$$

The presented approach recommends the items that have less distance between them and are similar to customer's choice. Thus community structure plays vital role for recommender systems as the members of community have similar interests and preferences [9].

D. Link Prediction:

Jorge carlos and Valverde-Rebaza proposed an approach which was based on community structure for link prediction in complex network [10]. Link prediction evaluates the possibility of existence of future links between vertices by observing vertices and links attributes in the network. Link prediction is basically used to detect missing and fake links and predicts future existence of the links with the development of network [11]. The approach was based on two steps:

- I. Firstly, network is divided into communities by using any partitioning scheme then this information which is obtained from these community structures is used for the link prediction. He used edge clustering coefficient algorithm.
- II. The second step was link prediction which was based on similarity between vertices. The similarity was evaluated through local or global information. Common Neighbor (CN) is the most common similarity measure based on local information.

E. Epidemic spreading on networks with overlapping community structure:

Jiancong Chena, Huiling Zhanga, Zhi-Hong Guana and Tao Li had explored the effect of overlapping community structure on Susceptible-Infected-Susceptible (SIS) epidemic spreading process [12]. The practical results showed that epidemics spreading process boosted with the involvement of community structure in the network as community structure involves numerous contacts between individuals. Mostly Epidemic spreading takes place between overlapping communities whose members are highly linked. Marcel Salathe and James H. Jones presented Dynamics and Control of Diseases in Networks with Community Structure [13]. They showed the impact of community structure on disease dynamics and also analyzed the networks which involves strong community structure and showed the immunization interventions targeted at individuals bridging communities are more effective than those simply targeting highly connected individuals.

F. Detection of Terrorist Groups in Online Social Networks:

Todd Waskiewicz has provided an analysis about the terrorist group activities on the social networking sites [14]. With the wide spread use of social networking sites like facebook and twitter the terrorist group's uses these sites to communicate their propaganda and to add new members to their group. The social media pages are less vulnerable to attack because these pages have feature which allows the owner of user to control unwanted access to their pages by set it to public to private. A terrorist group uses 'friend of friend' relationships to influence or add the individuals that are not directly linked with them.

Terrorists groups are identified with following ways:

- I. 'Friend of friend' activities are detected by using Social Network Analysis. For example Ego Network Analysis is used to identify the individuals in social network and the other nodes which are connected to this individual node.
- II. Then Betweenness Centrality measure is used to identify these groups. The user with highest betweenness centrality work as an intermediary between the individuals that are not directly linked.
- III. Community detection algorithms were applied for detection of such groups which evaluates the density between set of nodes and identified groups in which the density is greater inside the group than outside the group.

G. Lung Cancer Detection:

Joel J. Bechtel, MD, FCCP; William A. Kelley, MD, FCCP; Teresa A. Coons, PhD presented a community based lung cancer detection approach which basically focuses on high risk patients who undergo general care in primary care outpatient practice [15]. Lung cancer is a malignant lung tumor characterized by uncontrolled cell growth in tissues of the lung. Under this approach a simple questionnaire was answered by 1296 patients more than 50 years old to identify the patients at high risk of lung cancer. Approximately 430 patients were detected which were at high risk of lung cancer. Out of these patients 126 patients had spirometric abnormalities. Patients with air flow obstruction were offered various tests like Sputum cytology tests, chest radiographs_ and chest CT scans. Out of 126 patients 38 patients refused .only 88 patients were underwent through these tests. Out of 88 patients six cancer patients were detected and all patients were treated by radiation therapy at relatively less cost.

H. Information Diffusion:

Shuyang Lin, Qingbo Hu, Guan Wang, and Philip S. Yu conducted a research on impact of community structure on Information diffusion process [16]. Firstly Communities are detected by using agglomerative clustering algorithm. This is a bottom-up strategy started by placing each node in its own cluster and then merged these atomic clusters into larger and larger clusters, until all of the nodes are in a single cluster .They proposed community-based fast influence (CFI) algorithm to identify the effects of communities on information diffusion and thus provided accurate solution for influence maximization problem. Under this problem subset of nodes were identified from the social network that increases the spread of influence. CFI based influence maximization algorithm works by selecting seed nodes to maximize the influence [17]. Functions of links inside the community are different from links outside the community. Links among friends inside the social network are stronger but weaker links between friends of different communities provides more important information to people.

III. CONCLUSION

A community is often defined as a group of network members with stronger ties to members within the group than to members outside of the group. In this paper we are discussed about community structures and their application in complex network. Communities are used to detect fraud activities in telecommunication network, for link prediction, to detect terrorist groups in online social networks and for recommender systems where the system recommends products to customers that are related or similar to the products that the customer has liked in the past. We have also presented the impact of community structures on information diffusion process. Thus Communities provides us solutions to many real world problems. The basic purpose of this paper is to help you to understand the roles of communities in various fields.

REFERENCES

- [1] M.E.J Newman, "The structure and function of complex network" Volume 2, pp, 167-256, 2003.
- [2] L. Tang, H. Liu, "Community Detection and Mining in Social Media" Morgan & Claypool (2010).
- [3] Karsten Steinhaeuser and Nitesh v. Chawla, "Community detection in large real world networks".
- [4] Mini Singh Ahuja and Jasmine , " Community Detection: A Boom to Society" , 2015.
- [5] Mini Singh Ahuja & Jatinder Singh , "Future Prospects In Community Detection" , 2014.
- [6] Carlos André Reis Pinheiro, "Community Detection to Identify Fraud Events in Telecommunications Networks", Paper 106- 2012.
- [7] Wei-Feng, Pan Bo, Jiang Bing Li, "Refactoring Software Packages via Community Detection in Complex Software Networks", 2013.
- [8] Massimiliano Zanin, Pedro Cano, "Complex Networks in Recommendation Systems".
- [9] Jie Lu, Dianshuang Wu, Mingsong Mao, Wei Wang, Guangquan Zhang, "Recommender System Application Developments: A Survey".
- [10] Jorge Carlos Valverde-Rebaza and Alneu de Andrade Lopes, "Link Prediction in Complex Networks Based on Cluster Information".
- [11] Fei Tan, Yongxiang Xia, * and Boyao Zhu, "Link Prediction in Complex Networks: A Mutual Information Perspective", 2014.
- [12] JiancongChena, Huiling Zhanga, Zhi-Hong Guana andTao Li, "Epidemic spreading on networks with overlapping community structure", 2011.

- [13] Marcel Salath and James H. Jones, “Dynamics and Control of Diseases in Networks with Community Structure”,2010.
- [14] Todd Waskiewicz (Air Force Research Laboratory, AFRL/RIEA 525 Brooks Road, Rome, NY 13441-4505), “Friend of a Friend Influence in Terrorist Social Networks”
- [15] Joel J. Bechtel, MD, FCCP; William A. Kelley, MD, FCCP; Teresa A. Coons, PhD; Al. Gerry Klein, MD; Daniel D. Slagel, MD; and Thomas L. Petty, MD, Master FCCP,” Lung Cancer Detection in Patients With Airflow Obstruction Identified in a Primary Care Outpatient Practice”,2005.
- [16] Shuyang Lin, Qingbo Hu, Guan Wang, and Philip S. Yu, “Understanding Community Effects on Information Diffusion”.
- [17] Wei Chen, Yajun Wang,China yajun, Siyu Yang, “Efficient Influence Maximization in Social Networks”.