ISSN:0975-9646

Kavita G. Gare et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 7257-7261

A Review on Recommendations and Overlapping Communities for Location Based Social Networks

Kavita G. Gare[#1], Nilesh V. Alone[*2]

[#1]*Student, Computer Engg. Department, GES's R. H. Sapat COE Nashik*
Pune University, India
[*2]*Professor, Computer Engg. Department, GES's R. H. Sapat COE Nashik*
Pune University, India

***Abstract*— Today, a location based social networks is a drastically growing area which attracts users attention towards itself. Location based social networks(LBSN) assist between real world and online social networks by allowing users to check-in at a physical place and share the locations/location related contents with their friends. This Location sharing can be through GPS, mobile email or text. Location related contents can be geo-tagged photos and notes. LBSN sites includes foursquare, brightkite, GyPSii, Citysense etc. Many other online social networks provides activities (such as following, grouping, voting, tagging, etc.) that helps to interact with the virtual world but, "check-in" reflects a user's geographical action in the real world, residing where the online world and real world intersect. Location data helps to understand the users preferences and behaviour.**

People in the social structure naturally forms a community among themselves. For example, a person usually belongs to several social groups such as family, friends, and colleges. Usually, these communities in a social network can overlap each other. Detecting overlapping communities is very important to understand and analyze the structure of social network.

Recommendations help to suggest the opinions to the friends and family members. Friends have a good relationship among themselves. Hence, they try to recommend the things that can be useful to the persons closest or nearer to them. This paper reviews the overlapping communities structure, algorithms for overlapping community detection and recommendation based on location and friend.

***Keywords*— Location Based Social Networks, Overlapping Communities, Friend Recommendation, Location Recommendation.**

I. INTRODUCTION

With the extensive use of mobile devices and location-based services in the world, there is new way for online social interaction, namely location-based social networks (LBSNs). Location-based social networking sites uses GPS, Web 2.0 technology and mobile devices to allow people to share their locations (usually referred to as "check-in"), find out local Points of Interest and discounts, leave comments on specific places, connect with their friends, and find other friends who are nearby.

The distinct location-based social networking sites (e.g., Foursquare and Facebook Places) have attracted billions of users around the world and generated massive location-based social network data, providing us with both opportunities and challenges for investigating a user's mobile behavior, with the purpose of designing more advanced location-based services such as location-based marketing [1]. and disaster relief[2].

People in LBSN are structured in the form of community. Community is a group of nodes which has dense and sparse relations with other parts of a network[3]. Identifying these communities helps to better understand the structure of social network. There can be two types of groups one, where each user can belong to one or more communities. Another, where a user can belong to more than one group, also called as overlapping communities. First group is unable to represent the as it is structure of social network. But overlapping communities provides a clear understanding about the structural aspects of social networks. So, over the recent years, detection of overlapping communities is a key attention.

Recommendations are designed to recommend items to users in various situations such as online shopping, dating, and social events. Recommendation helps for decision making by filtering the uninterested things. By recommendation, one can save time in selecting the item which he/she wants. Recommendations also assist to establish communications in between two users, by allowing friend recommendation. Furthermore, recommendations could also benefit virtual marketing, since the appropriate recommendations could attract users with specific interests. Recommender systems on location-based social networks are comparatively new and mainly locations and friends are recommended.

Section II presents related work. Section III discusses overlapping communities detection algorithms. Section IV reviews details on recommendation systems over LBSNs. Finally, section V concludes the paper.

.

II. RELATED WORK

In this section , research works on location based social networks is discussed. Scellato et al. [4], presented a graph analysis based approach to study social networks with geographic information and new metrics to charaterize how geographics distance affects social structure. Noulas et al. [5] gives a details about a users behavior in foursquare. This users behavior helps to know the users check-in nature. Also the author reveals spatio-temporal patterns and urban spaces demonstration. He also conveys ideas about recommender systems.

Noulas et al. [6] provides a way of modeling human activity and geographical areas. For this, place categorization needs to be performed. Foursquares dataset and spectural clustering algorithm is also applied. It helps to find user communities visiting to similar categories of places. Also demonstrates way of using semantic information for applications such as recommender systems.

Huiji Gao et.al [7] introduced the first extensive study of temporal effects on LBSNs. Here, a general structure to utilize and deal with temporal cyclic patterns is provided. Two real world datasets are used to generate the results. Results demonstrate the frameworks ability to choose the effective location prediction algorithm among various other prediction models

Eunjoon Cho et.al [8] presents a model towards human mobility. Model combines users regular short range movements in a travel due to the social network structure. Also describes how the model gives better performance by reliably estimating the dynamics and location of the future human movement.

Zhu Wang et.al [9] provides a framework to trace the overlapping as well as hierarchical communities in LBSNs. Work is done based on the user check-in traces at venues. Framework groups same interests or like minded users from social perspectives. For this intermode and intramode features are extracted from social network. Foursquare dataset is used to evaluate the performance of the framework.

III. Overlapping Communities

Community overlapping is an important characteristic of many real-world social networks. A user may be a part of more than one community. Communities are of family members, friends and can be of co-workers. An individual/user can belong to a number of communities. There is no limit on a number of communities to a user as it is a users choice to associate with a group to which he/she wishes.

A multimode multi-attribute edge-centric coclustering framework : This is the recent work to detect overlapping communities in LBSNs. In this framework, 1) LBSNs dataset is collected and based on the characteristics of this dataset features are extracted to perform fusion as well as feature normalization. 2) the edge clustering algorithm is proposed to detect the overlapping community structure. Finally, detected communities are combined together by considering data about user/venue. This obtained community profiles helps to understand the social and semantic structure/meaning of communities in LBSNs. Following are the early overlapping communities detection algorithms, based on different categories.

Overlapping Community Detection Algorithms

Overlapping community detection algorithms are reviewed in this section. The work on community overlapping was started by Palla in 2005[10]. Focus is on finding overlapping communities where each node can belong to one or more communities. After this work, many algorithms were found for the overlapping community detection. There are five classes namely Clique Percolation algorithms, Agent and Dynamic based algorithms, Fuzzy based algorithms, Local expansion and Optimization algorithms and Line graph and Link partitioning algorithms .

A. Clique Percolation Method

Clique Percolation Method (CPM) is a deterministic community detection method, which allows for overlapping communities. CPM exploits local topological properties of a network [10]. It is a first attempt over an overlapping community. CPM identifies all cliques of size k in a network at the initial stage. Once CPM done with identification, a new graph is formed where each vertex represents one of these *k-cliques*. If the *k-cliques* representing the vertex shares *k-1* members, then only two nodes can connect to each other. The connected components from the resultant graph seeks which cliques compose the communities. There can be overlapping between communities, as a vertex can be in multiple *k-cliques* simultaneously. There is an assumption in CPM that the graph has huge number of cliques and it is suitable only for networks which considers densely connected segments. If a graph involves a few cliques, then it is not possible for CPM to detect meaningtful social structure.

CPM is conceptually simple, but CPM-like algorithms are seems to be for pattern matching rather than finding overlapping communities as they aim to find specific and restricted/limited structure in a network

B. Fuzzy Detection Algorithm

Fuzzy community detection algorithms evaluate the strength of association between all pairs of nodes and communities. These types of algorithms calculates, a soft membership vector, or belonging factor [Gregory 2010], for each node. There is a need to find out the dimensionality k of the membership vector, this is the drawback of such algorithms. The value k can be determined from the data and provided as a parameter to the algorithm. These algorithms include proposing a method for combining spectral mapping, fuzzy clustering and optimization of a quality function[12], allowing each vertex of the graph to belong to multiple communities at the same time[13], disjoint community detection[14].

C. Agent and Dynamic based algorithms

Label Propagation Algorithm (LPA) is an agent and dynamic based algorithm proposed by Raghavan et al in 2007. LPA finds communities from a large networks and runs linearly in the number of edges. At first, a unique label is assigned to each node in a network. The vertex replaced the label which is used by same maximum number of neighbors and updates its own label. This process is repeated after every iteration. The neighbor is chosen randomly. After the several iterations performed, all the members of a community is assigned with a label and all the vertices having similar label are added to one separate community. LPA uses only the network structure to guide itself, it does not require optimization details and prior information about the communities in a network. The drawback of LPA is, it can detect only disjoint communities.

Gregory S provides a Cluster-Overlap Newman Girvan Algorithm (CONGA) which is an “overlapping” version of existing disjoint community detection algorithm[15]. CONGA is an extension to the Girvan and Newman’s algorithm, which divides a vertex into two vertices repeatedly during the process of divisive clustering. This algorithm considers both split betweenness and the conventional edge betweenness.

D. Local Expansion and Optimization

Algorithms in this category trust on a local benefit function that characterizes the quality of a densely connected group of nodes. Baumes et al. [2005] uses a two phase method to iteratively improve the candidate cluster of CONGA. The method first smashed the network into a number of disjoint seed communities and keeps adding and removing vertices to and from candidate set respectively. The process continues till the density of a candidate set is not maximixed[16]. It depends on finding a local maximum of density.

Lancichinettia A., Fortunato S, proposed LFM method to find both overlapping communities and the hierarchical structure[17]. In this method, after identifying the highest fitness value the node is distributed to different communities. There can be many visits to one node, this places the node in more than one community. Proper tuning of resolution parameter determines the size of each community. This gives a meaningful hierarchical communities. After comparing this algorithm with that of Baumes[16], the only difference found is that a seed community is only a vertex that is not yet allotted to any community. This algorithm provides a way to yield a large class of algorithms by choosing a different expression for the fitness function or a different optimization procedure of the fitness as a single cluster.

E. Line Graph and Link Partitioning[11]

Not only the nodes but also partitioning of links helps to discover the community structure in LBSNs. A node in the original graph is called overlapping if links connected to it are put in more than one cluster.

In Ahn et al. [2010][2], link partitioning is done via hierarchical clustering of edge similarity. Given a pair of links e_{ik} and e_{jk} incident on a node k, a similarity can be computed via the Jaccard index defined as

$S(e_{ik}, e_{jk}) = |Ni \cap Nj| / |Ni \cup Nj|$, where Ni : is in proximity of node i including i. A link dendogram is form using a single-linkage hierarchical clustering. At some threshold when this dendogram is cut, it produces link communities.

Evans and Lambiotte [2009, 2010] forms a network having a weighted line graph, where nodes are the links of the original graph. Further, the disjoint community detection algorithms can be applied. The node partition of a line graph results into an edge partition of the original graph. CDAEO [Wu et al. 2010] gives a further processing procedure to determine the extent of overlapping. Once the prior partitioning on the line graph is done, for a node i with $|Eicmin|/|Eicmax|$ below some predefined threshold, where $Eicmin(cmax)$ is the set of edges in the community with which i has the minimum (maximum) number of connections, links in $Eicmin$ of the line graph are removed. This essentially reduces node i to a single membership.

Kim and Jeong [2011] provides the line graph by extending the map equation method (also known as Infomap [Rosvall 2008]), which applies Minimum Description Length (MDL) principle to the path of random walk on the line network. Clique graph [Evans 2010] is an extension work of line graph, wherein given order cliques are represented as nodes in a weighted graph. Fraction of cliques gives the membership strength of a node i to community c. Fraction of cliques contains i and assigned to c.

As these algorithms rely on cryptic definition of community, there is no surity that it can perform better than node based overlapping detection [Fortunato 2010].

IV. RECOMMENDATION SYSTEMS OVER LBSNS

A. Location Recommendation

Location recommendations purpose is to recommend a set of locations to a user based on the user's interests. In the context of location recommendation, location prediction is one another concept. Location prediction usually predicts the next location to an existing location that the user has been before and location recommendation aims to recommend a new location to which the user has never visited. From a research point of view a location prediction on LBSNs is about utilizing the social information, while the research in location recommendation on LBSNs mainly focuses on the geo-spatial and temporal influence, and the social network information is usually utilized through traditional collaborative filtering[18],[19], which considers the location as an item such as that on Epinions[20],[21]. For evaluation, performance@N[22] is usually adopted to assess the location recommendation performance. The performance@N is a metric which consists of precision@N and recall@N, where "N" is the top highest ranked point of interests(POIs) as recommendation to a target user. It consider all the locations that should be recommended as uncovered locations, and the set of correctly recommended locations as recovered locations. The precision@N evaluates the ratio of recovered locations to the N recommended locations, and the recall@N calculates the ratio of recovered locations to uncovered locations

Location recommendation in location based social networks is primarily introduced by Ye et al.[23]. In this, the major focus is on efficiency of location recommendation. There are two essential contents : 1) only friendship information was used for collaborative filtering; and 2) instead of calculating the user similarity based on historical behavior (e.g., check-in history), the correlations between geographical distance and user similarity were captured, and leveraged them for user similarity calculation. This work is later extended in[22], which considers both spatial influence and social friendships for location recommendation. Three factors are investigated and combined together to recommend locations. The first factor represents influence from similar users, the second factor indicates influence from friends, and the third factor captures geographical influence, under the hypothesis that people tend to visit close places more often than distant places. A spatial constraint is generated to capture the geographical influence by exploiting the relationship between a user visiting two places and the geographical distance of these two places. These three factors are then represented by three probabilities, and linearly combined together with corresponding weights. The results demonstrated that the most influential factor actually comes from the similar users, while friendship and geographical distance together have around 30% influence.

B. Friend Recommendation

Friend recommendation is a way to suggest one user to another user having similar properties among themselves. Friend recommendation aims to inspect the similar patterns between a target user and other users, and then recommends users with the most similar patterns to the target user. Similarities between two users are in terms of common interests, travelling trajectories, shopping habits etc. For link prediction between two users in LBSN, supervised learning is used mostly. By analyzing historical data for each pair of users, a features set is first extracted and based on the extracted features a classifier is trained to predict the link between two users. To evaluate proposed approaches the social network information is used as base and ROC curves[24],[25] are usually used as evaluation metrics.

Ongoing work on friend recommendation vary in how to choose the feature space and classifier. To predict the link among two users having co-locations, logistic regression by Jonathan et al.[19] is used. Feature extraction was based on the tuples. Touples consist of place x, actor1, actor2. Touples indicates that actor1 and actor2 have checked-in into place x at least once. Based on the touple, three features are extracted : 1. the total number of check-ins at place x, 2. Numbers of check-ins of actor1 and 3. Numbers of check-ins of actor2. For each co-location inspection among two users Justin et al.[26] extracted 67 features from the data on Locaccino[27]. With respect to user attributes and co-location properties, extracted features include structure properties, location diversity, intensity and duration, mobility regularity, etc. Once, they have completed with features extraction, three classifiers are selected for predicting the link between two users. But Final results shows that AdaBoost has the best classification performance. Their opinion is that only considering the number of co-locations is not enough for friend recommendation and also reported that there is a positive correlation between the location diversity and the number of social ties a user has in the social network. Sadilet et al.[25] takes the same scenario while in addition considered the content features from tweets. Scellato et al.[24] utilizes the place features such as common check-ins, social features like common friends, and global features such as distance between homes, then selected various classifiers in WEKA for link prediction on Gowalla. Their results demonstrated that the purely social based features contribute least to the prediction performance, while space features and global features lead to better performance, indicating the importance of location-based activities on location-based social networking analysis.

V. CONCLUSIONS

In this paper, to understand what the LBSN is, we have discussed the research work on Location based social networks. We have also reviewed an overlapping community detection algorithms based on five different categories. Overlapping communities provides the structure of real world social networks, so to understand the relationship structure among nodes/users it is essential to identify an overlapping communities in a LBSNs. Recommendations plays an important role by giving suggestions to the users. This reduces time to seek new things at a location nearer to user. Recommendations also assist users to make a new friends. So, in this paper recommendations over LBSNs and their algorithms are also discussed. Furthermore, recommendations based on overlapping communities profiling can also be possible.

ACKNOWLEDGMENT

We are glad to express our sentiments of gratitude to all who rendered their valuable guidance to us. We would like to express our appreciation and thanks to Prof. Dr. P. C. Kulkarni, Principal, G. E. S. R. H. Sapat College of Engg., Nashik. We are also thankful to Prof. N. V. Alone, Head of Department, Computer Engg., G. E. S. R. H. Sapat College of Engg., Nashik. We thank the anonymous reviewers for their comments.

REFERENCES

[1] A. Research. Mobile advertisers forecast to spend $1.8 billion on location-based campaigns in 2015. 2010.

[2] H. Gao, G. Barbier, and R. Goolsby. Harnessing the crowdsourcing power of social media for disaster relief. Intelligent Systems, IEEE, 26(3):10–14, 2011.

[3] Girvan M., and Newman M.E.J, Community structure in social and biological networks, PNAS, Vol. 99, 2002, pp.7821–7826.

[4] S. Scellato, C. Mascolo, M. Musolesi, and V. Latora, “Distance matters: Geo-social metrics for online social networks,” in Proc. WOSN, 2010, p. 8.

[5] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, “An empirical studyof geographic user activity patterns in Foursquare,” in Proc. ICWSM, 2011, pp. 570–573.

[6] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, “Exploiting semantic annotations for clustering geographic areas and users in location-based social networks,” in Proc. ICWSM, 2011, pp. 32–35.

[7] Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu, “Modeling Temporal Effects of Human Mobile Behavior on Location-Based Social Networks ”, CIKM’13, Oct. 27–Nov. 1, 2013, San Francisco, CA, USA.

[8] Eunjoon Cho, Seth A. Myers, Jure Leskovec ,”Friendship and Mobility: User Movement In Location-Based Social Networks”, KDD’11, ACM 978-1-4503-0813-7/11/08, August 21–24, 2011, San Diego, California, USA.

[9] Zhu Wang, Daqing Zhang, Xingshe Zhou, Dingqi Yang, Zhiyong Yu, and Zhiwen Yu, “Discovering and Profiling Overlapping Communities in Location-Based Social Networks”, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, VOL. 44, NO. 4, APRIL 2014.

[10] Palla G, Derenyi I, Farkas I, Vicsek T, Uncovering the overlapping community structure of complex networks in nature and society, Nature, Vol. 435(7043), 2005, pp. 814–818.

[11] Xie J., Kelly S., and Szymanski, B. K.,.Overlapping Community Detection in Networks: the State of the Art and Comparative Study, arXiv: 1110.5813v4[cs.SI], 2012.

[12] Zhang S., Wang R.S., Zhang X.S, Identification of overlapping community structure in complex networks using fuzzy c-means clustering, Physics A, Vol. 374, 2007, pp. 483-490.

[13] Nepusz Y., Petrocz A., Negyessy L., and Bazso F, Fuzzy communities and the concept of bridgeness in complex networks, Physics Review E 77, 2008, pp. 016107.

[14] Wang X., Jiao L., and Wu J,, Adjusting from disjoint to overlapping community detection of complex networks, Physica A , Vol. 388, 2009, pp. 5045–5056.

[15] Gregory S, An algorithm to find overlapping community structure in networks, In Proc. the 11th PKDD, Sept. 2007,pp.91-102.

[16] Baumes J., Goldberg M., Krishnamoorthy M., Magdon-Ismail M., and Preston N, Finding communities by clustering a graph into overlapping subgraphs, In Proceedings of IADIS Applied Computing, 2005, pp. 97–104.

[17] Lancichinettia A., Fortunato S, Community detection algorithms: A comparative analysis, Physics Review E, Vol. 80, 2009, 056117.

[18] Berjani, B., Strufe, T.: A recommendation system for spots in location-based online social networks. Proc. of SNS11 (2011).

[19] Zhou, D., Wang, B., Rahimi, S., Wang, X.: A study of recommending locations on location-based social network by collaborative filtering. Advances in Artificial Intelligence pp. 255-266 (2012).

[20] Tang, J., Gao, H., Liu, H.: mTrust: Discerning multi-faceted trust in a connected world. In:Proceedings of the fifth ACM international conference on Web search and data mining, pp. 93–102. ACM (2012).

[21] Tang, J., Gao, H., Liu, H., Sarma, A.: eTrust: Understanding trust evolution in an online world. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (2012).

[22] Ye, M., Yin, P., Lee, W., Lee, D.: Exploiting geographical influence for collaborative point-of interest recommendation. In: Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 325–334 (2011).

[23] Ye, M., Yin, P., Lee, W.: Location recommendation for location-based social networks. In:Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 458–461. ACM (2010).

[24] Scellato, S., Noulas, A., Mascolo, C.: Exploiting place features in link prediction on location based social networks. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1046–1054. ACM (2011).

[25] Sadilek, A., Kautz, H., Bigham, J.: Finding your friends and following them to where you are. In: Proceedings of the fifth ACM international conference on Web search and data mining, pp. 723–732. ACM (2012).

[26] Cranshaw, J., Toch, E., Hong, J., Kittur, A., Sadeh, N.: Bridging the gap between physical location and online social networks. In: Proceedings of the 12th ACM international conference on Ubiquitous computing, pp. 119–128. ACM (2010).

[27] Sadeh, N., Hong, J., Cranor, L., Fette, I., Kelley, P., Prabaker, M., Rao, J.: Understanding and capturing people's privacy policies in a mobile social networking application. Persona and Ubiquitous Computing 13(6), 401–412 (2009).