INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES
ISSN : 2454-9924

Community Structure in Social Network using Multi-label Memory based Propagation Algorithm

P.Anitha#1, Mr.S.SundarRajan.,M.E.,(Ph.D).,MISTE*2,
#1Student, *2 Associate Professor and Head,
Dept.of Computer Science&Engg,Dept.of Computer Science&Engg,
Surya Groups of Institutions.Surya Groups of Institutions.

ABSTRACT
The world is deluged with various kinds of data such as scientific data, environmental data, mathematical data and financial data. Data mining refers to extracting useful information from vast amount of data. However, social media sites provide data which are vast, noisy, distributed and dynamic. The increasing online traffic in social media brings about many data mining opportunities for community detection, community notification, community comparison, anomaly detection and blocking. Detection of community structure in social network by reality mining provides useful information and also forward large amount of data in community structure. If any anomaly is detected in the community such as unwanted post, pictures, videos ect.., the anomaly is blocked in the community. The multi label propagation algorithm provides an advanced development in the social network for detecting the community structure. The anomaly diction in the community can help the people to work protectively in the social network.

Index Terms—Community Structure, Data Mining, Data forwarding, Human physical and social activity and anomaly detection.

I. INTRODUCTION

Social media, such as face book, my space, twitter, BlogSpot, YouTube, has designed to work smoothly for people to express their thoughts, voice their opinions and connect to each other anytime and anywhere. The increasing online traffic in social media brings about many data mining opportunities for user profiling, community detection, communitynotificationandcommunity Comparison [1]. Efficient data mining is part of a larger process that includes the process of discovering interesting knowledge from large amount of data stored in information repositories and deploying it into working environment. In recent years, the community structures are one of the significant characteristics in the most complex network such as social networks due to numerous trends of human being to forming large number of groups or communities. So, receiving and forwarding large amount of information via base station will make the stations overloaded and easily become the bottle necks.Label Propagation algorithms that use single labels for propagation face the problem of random tie breaking when choosing a label among a set of labels for propagation. Due to the random tie breaking nature, techniques that follow single label for propagation produces non-deterministic outputs by using Speaker-listener Label Propagation Algorithm (SLPA) for finding the community structure in the social network[2]. To overcome these issues the proposed work is done with the integration of three concepts (i.e.) finding the community structure, reality mining, and efficient data forwarding. This research work presents a multi-label propagation approach for community detection in complex networks, particularly, in the perspective of social networks and the work simulates the human pair wise communication behavior. The proposed work uses multi-label propagation technique to detect overlapping communities in social network.

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES

ISSN : 2454-9924

Designing a multi-label propagation algorithm to detect overlapping communities involves propagation of more than one label during the community detection process. The focus of this research work is based on the decision of choosing multiple labels for propagation and storing multiple labels received from the propagation process. A multi-label propagation approach which is presented in this paper is a modification of Speaker-listener Label Propagation Algorithm (SLPA)[3]. The study of spatiotemporal data mining is of great concern for the study of sensed data. Reality Mining is defined as the study ofhuman social behavior based on sensed data. SAIS (Social Attraction and Infrastructure Support) achieves a better performance thanexisting popular social community-based data forwardingalgorithms in practice, including Simbet, Bubble Rap andNguyen's Routing algorithms[4].

1.1 Community detection in social network:

A community in a network is defined as a group of nodes that have more edges among themselves than those vertices outside the group. Community detection can be classified into six categories: spectral and clustering methods, hierarchical algorithms, modularity based methods, model based methods, logical community detection methods, feature based assisted methods. In this community detection, the MLAP(multi lable propagation algrothm) algorithm proposed an adaptive spectral clustering, that will determine the number of cluster centers automatically. The neighborhood structure is analyzed by this cluster. Since many relationships are missing in the community, or not recorded, so they tried to find out hidden communities. An improved version of betweens algorithms has been presented to detect communities in networks and it applied on both weighted and unweightedNetworks. The weighted edges lead to the inter community edges, that provide greater discrimination values than intra community edges, and hence the inter community edges can be easily identified. At the end of the algorithm, most frequent adoption of label for each node is extracted to form proper communities.

1.2 Hybrid communication in social networks

The inner product similarity and design an efficient data forwarding scheme to fully demonstrate the role of community based data forwarding. A SASI (similarity attraction and infrastructure support) data forwarding scheme is proposed, which is based on the detection result of efficient data forwarding among the communities [5].

1.3 Reality mining in social network:

The process of discovering interesting, useful, non trivial patterns from large spatial datasets is called spatial data mining. Data mining is the process of finding data from different perspectives and describe into useful information that can be used to increase revenue, cut cost or both.

In this process of reality mining we can mine the details as to an individual's relationships with associates, communication community and behavior patterns [6]. The reality mining allows for the measurement of human physical and social activity. After mining, if any anomaly process is done in community such as unwanted posts, videos etc.., the process is identified the data who miss behave in community and also block that anomaly from the community.

In this paper, the above three concept is integrated and the useful information about efficient mining, community detection, anomaly detection, blocking and data forwarding is done to improve the quality of social media. The integration of these concepts provides an advanced development in the social network and people will get knowledge about the particular community they work on it. Without any tie breaking the continuous data forwarding provides effective changes. The anomaly detection in the community can help the people to work protectively in the social network. Using memory based label propagation algorithm(MLAP)for finding the community structure in social network and social attraction and infrastructure(SAIS)algorithm

ISSN : 2454-9924

include simbet, bubble rab and nguyen's routing algorithms for better performance of data forwarding and efficient mining.

II. LITERATURE SURVEY:

In this existing system we can't find the community and compare the community.Hard to detect the community in largest dataset.Label Propagation algorithms that use single labels for propagation face the problem of random tie breaking.It was shows the lower performance in reality mining.Receiving and forwarding large amount of information via base station will make the stations overloaded and easily become the bottle necks.Spatiotemporal data mining not sensing data..

A.Community detection

Extracting latent social dimensions is related to community detection. That has been an active field in social network analysis, and various methods have been pro-posed including stochastic block models, the latent space model, spec-tral clustering and modularity maximization. A comprehensive treatment is presented in. In this work, spectral clustering is employed for SocioDim, but another soft clustering method should also serve the purpose. Recently, Kumar et al. Found that real-world networks consist of a giant connected component with others being singletons and small-size connected components. Leskovec studied the statistical properties of communities on the giant connected component and found a similar pattern[7]. The optimal spectral cut always returns a community of 100 to 200 nodes, loosely connected (say, one or two edges) to the remaining network. A further comprehensive comparison of various community detection algorithms is reported in. The most community detection methods focus on discrete binary cases, i.e., extracting one community from a network based on certain criterion.WhereasSocioDim employs soft clustering to extract social dimensions, and typically many more dimensions instead of just one or two are extracted. We believe a comprehensive comparison of different soft clustering approaches for the extraction of social dimensions and their scalability is an interesting line of future work. In this paper, Community detection is used to detect the community which miss behave on their activity like uploading unwanted photos, unwanted information, Behaviour etc.

B.Communitydetection using label propagation

The main idea behind our label propagation algorithm is the following. Suppose that a node x has neighborsx1, x2, xk and that each neighbor carries a label denoting the community to which they belong to. Then x determines its community based on the labels of its neighbors [8]. We assume that each node in the network chooses to join the community to which the maximum number of its neighbors belongs to, with ties broken uniformly randomly. We initialize every node with unique labels and let the labels propagate through the network. As the labels propagate, densely connected groups of nodes quickly reach a consensus on a unique label. When many such dense (consensus) groups are created throughout the network, they continue to expand outwards until it is possible to do so. At the end of the propagation process, nodes having the same labels are grouped together as one community[9]. We perform this process iteratively, where at every step; each node updates its label based on the labels of its neighbors. The updating processes can either be synchronous or asynchronous. In synchronous updating, node x at the tth iteration updates its label based on the labels of its neighbors at iterationt1. Hence, Cx(t)=f(Cx1(t−1), ..., Cxk(t−1)), where(t) is the label of node x at time t. The problem however is that sub graphs in the network that are bi-partite or nearly bi-partite in structure lead to oscillations of labels (see figure 3)[10]. This is especially true in cases where communities take the form of a star graph. Hence we use asynchronous updating where(t)=f(Cxi1(t),...,Cxim(t), x i (+1)(t−1), ..., Cxik(t−1)) andxi1, ..., xim are neighbors of x that have already been updated in the current iteration while(m+1), ..., xik are neighbors that are not yet updated in the current iteration. The order in

ISSN : 2454-9924

which all the n nodes in the network are updated at each iteration is chosen randomly[11]. We also consider a protein-protein interaction network consisting of 2115 nodes. The communities are likely to reflect functional groupings of this network[12]. And finally we consider a subset of the World Wide Web (WWW) consisting of 325729 WebPages within the nd.edu domain and hyperlinks interconnecting them. In this paper, The existing coordinate-free orderings do not work well on social networks, we propose a new algorithm, called MLPAAlgorithms that builds on previous work on scalable clustering by MLPA the algorithm can reorder very large graphs (billions ofnodes), and unlike previous proposals, is free from parameters. The proposed work uses multi-label propagation technique to detect overlapping communities in social network.

C.Behavior and Mood Analysis

Benevenuto et al. measured the behavior of online social networks' users applying the proxy server-based measurement framework. Schneider et al. also have conducted an in-depth analysis of user behavior based on network traces across several online social networks.

Gyarmati et al. crawled the public part of users' profile pages, which contained online status information of the users. Simoes et al. proposed distance, similarity, influence and adjustments-based methods for understanding and predicting human behavior for social communities. Zhang et al.have developed a model called socioscope for predicting human-behavior in social network[13]s. Yan et al. also have presented a social network based human dynamics model to study the relationship between the social network attributes of microblog users and their behavior[14].. In this paper, An attempt could be made to develop a new model basedon Human behavior which could be used in future forpredicting some useful data e.g. Degree of influence of a personin a group, kind of friends he mayhave, type of group he mayjoin, type of relationship he may have, etc. This could be thedata useful for emotional mining.

D.Opinion Mining:

Most of works in this research area focus on classifying texts according to their sentiment polarity, which can be positive, negative or neutral. Authors in provided an in-depth survey of opinion mining and sentiment analysis. In, the problem was studied further using supervised learning by considering contextual sentiment influencers such as negation (e.g., not and never) and contrary (e.g., but and however)[15]. Wilson et al have studied several different learning algorithms such as boosting, rule learning, and support vector regression that can automatically distinguish between subjective and objective (neutral) language and also among weak, medium, and strong subjectivity. Zhang et al. presented a novel model that unifies topic-relevance and opinion generation by a quadratic combination. Zafarani et al. studied sentiment propagation in social network by making a case study of *LiveJournal*website[16]. In this method we proposed to deal with the problem of product aspects which are nouns and imply opinions using a large corpus. Authors have studied about several challenges in developing opinion mining tools for social media. Ortigosa et al. developed a hybrid approach for performing sentiment analysis in Face book with high accuracy.

III.PROPOSED COMMUNITY STRUCTURE IN SOCIAL NETWORK:

Fig.1 shows that how the user shares data in community and detecting the community which is not good. We are interested in scenarios where users can be motivated to cooperate among themselves as permitted by the network topology [17]. We express the communication expense in the form because, as described further ahead, when user decides to share information, it will be sharing the information with one neighbor at a time.

First the user will compare the community in the community list then sense the Behavior of human in social network. Reality Mining is defined as the study ofhuman social behavior based on sensed data. SAIS (Social Attraction and Infrastructure

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES

ISSN : 2454-9924

Support) achieves a better performance thanexisting popular social community-based data forwardingalgorithms in practice, including Bubble Rap andNguyen's Routing algorithm.

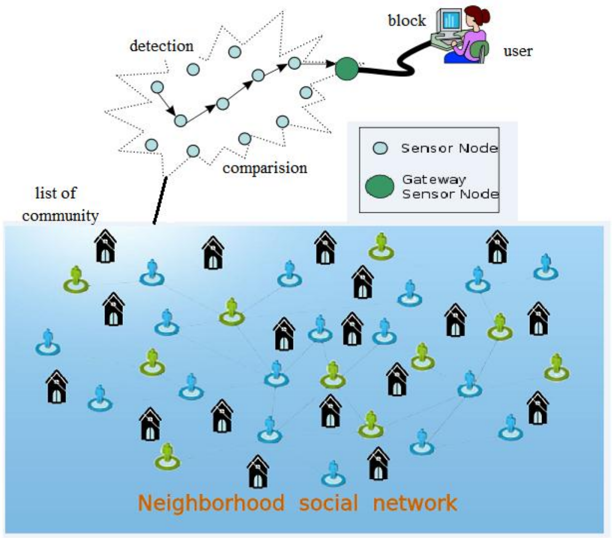


Fig.1. Architecture for finding community structure

TECHNIQUES	DESCRIPTION
Community Notification	Community notification is one of the important modules in this project. This module is used to display all active community information and also display our community information also
Community Comparison	Community comparison module is used to compare our community with other community for identify related information with other community. And also identify the behaviour of the community

Table1: Techniques involved for finding community structure.

We propose a new algorithm, called MLPAAlgorithms is used to identify (detect) the anomaly (person) who miss behave in the community. After detecting the Behavior of human based on sensed data then user will block the anomaly who miss behave in the community and also block that anomaly from the community.This involves clustering ofpeople studying the patterns of cluster formation and theircharacteristics.

In this proposed work, the above three concept is integrated and the useful information about efficient mining, community detection, anomaly detection, blocking and data forwarding is done to improve the quality of social media. The integration of these concepts provides an advanced development in the social network and people will get knowledge about the particular community they work on it. Without any tie breaking the continuous data forwarding provides effective changes. The anomaly detection in the community can help the people to work protectively in the social network. Using memory based label propagation algorithm(MLAP)for finding the community structure in social network and social attraction and infrastructure(SAIS)algorithm include simbet, bubble rab and nguyen's routing algorithms for better performance of data forwarding and efficient mining. We propose an improved label propagationalgorithm called memory-based label propagation algorithm(MLPA) for finding community structure in social networks.Thespatiotemporal data mining is of great concern for the study of sensed data. Reality Mining is defined as the study ofhuman social behavior based on sensed data.SAIS(Social Attraction and Infrastructure Support) achieves a better performance thanexisting popular social community-based data forwardingalgorithms in practice, including Simbet, Bubble Rap andNguyen's Routing algorithms.Anomaly detection isused to identify the anomaly who miss behave in the community and also block that

ISSN : 2454-9924

anomaly from the community.

IV.MULTI LABEL MEMORY BASED PROPAGATION ALGORITHM

The existing coordinate-free orderings do not work well on social networks, we propose a new algorithm, called MLPAAlgorithms that builds on previous work on scalable clustering by MLPA the algorithm can reorder very large graphs (billions ofnodes), and unlike previous proposals, is free from parameters. The proposed work uses multi-label propagation technique to detect overlapping communities in social network. Designing a multi-label propagation algorithm to detect overlapping communities involves propagation of more than one label during the community detection process. The focus of this research work is based on the decision of choosing multiple labels for propagation and storing multiple labels received from the propagation process. The experiments show that our combination of techniques provides a major increase in compression with respect to all currently known approaches. This is particularly surprising in view of the fact that we obtain the best results both on web graphs and on social networks. Our largest graph contains more than 600 millions nodes—one order of magnitude more than any published result in this area.

NGUYEN'S ROUTING ALGORITHM.

An NR algorithm is the process of selecting best paths in a network. In the past, the term routing also meant forwarding network traffic among networks[18]. However, that latter function is better described as forwarding.

a graphical map of the network is the fundamental data used for each node. To produce its map, each node floods the entire network with information about the other nodes it can connect to. Each node then independently assembles this information into a map.

BUBBLE RAP PROTOCOL

Bubble Rapfirst introduces the understanding of human mobility into the DTN design. They study the social structures of the between devices and leverage them in the design of forwarding algorithms for Pocket Switched Networks (PSNs)[19]. With experiments of real world traces, they discover that human interaction is heterogeneous both in terms of hubs and groups or communities. According to this finding, they propose Bubble Rap, a social-based forwarding algorithm, to improve the forwarding efficiency significantly compared to history-based PROPHET and social-based SimBet algorithms. This algorithm also shows how it can be implemented in a distributed way, which demonstrates that it is applicable in the decentralized environment of PSNs.

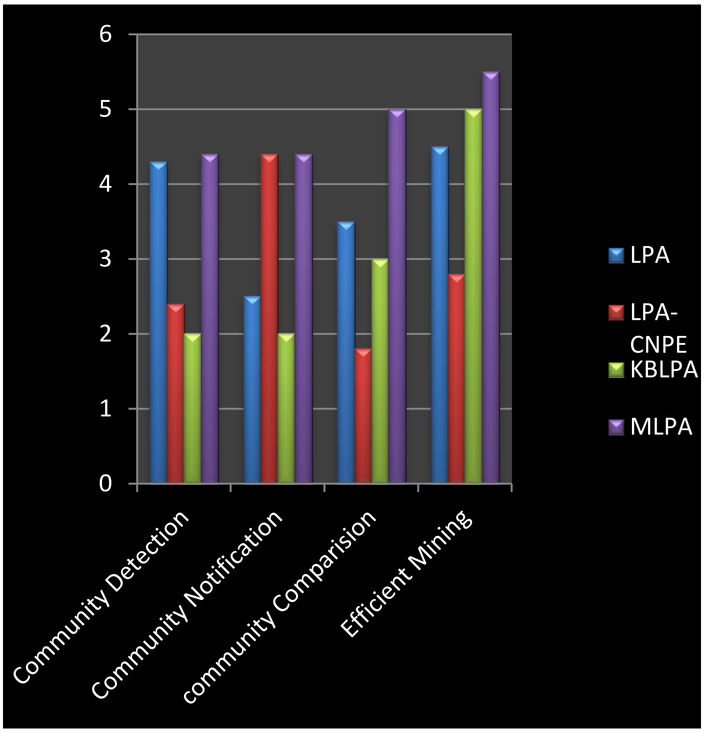
V. RESULT DISCUSSION

In this section, we experimentally examine the following questions: How is the classification performance of our proposed framework compared to that of collective community structure? Do differentiating heterogeneous connections presented in a network yield a better performance for data forwarding?

Performance on community detection

We gradually increase the number of labeled nodes from 10% to 90%. For each set-ting, we randomly sample a portion of nodes as labeled. This process is repeated 10 times, and the average performance is recorded. The performances of different methods and the standard deviation are plotted

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES
ISSN : 2454-9924



in Fig. 2. Clearly, our proposed Social network.

Data set	LPA	LPA-CNPE	KBLPA	MLPA
Community Detection	4.202	2.302	2.000	4.303
Community Notification	2.380	4.451	2.000	4.451
Community comparison	3.246	1.202	3.000	5.000
Efficient Mining	4.504	2.856	5.000	5.765

In this section, we report the results of experiments on the social network datasets described in table. The experiments are conducted on a system with windows 7 platform having configuration Intel® Core i53337U CPU 1.80 GHz and 4 GB RAM. Our results on the social network datasets are presented in Table. We set limited capacity for each memory element is 10. The results of modularity measure for proposed algorithm as MLPA in comparison with other algorithm such as LPA, LPA-CNPE , LPA-CNP1 and KLPA are listed in Table Also, the results of this experiment are demonstrated as bar chart diagram for karate, football, dolphin, political books and Blogs. As shown in Tables and bar chart, the results show a relative improvement in comparison with other community detection algorithms.

Outperforms all the other methods. wvRN, as shown in the figure, is the runner-up most of the time. MAJORITY performs even worse than RANDOM in terms of Macro-F1, as it always picks the majority class for prediction.The superiority of SocioDim over other relational learning methods with collective inference is evident. As shown in the figure, the link-based classifier (LBC) performs poorly with few labeled data. This is because LBC requires a relational classifier before the inference. This is indicated by the large deviation of LBC in the figure when labeled samples are less than 50%. We notice that LBC in this case takes many iterations to converge. wvRN is more stable, but its performance is not comparable to SocioDim. Even with 90% of nodes being labeled, a substantial difference between wvRN and SocioDim is still observed. Comparing all three methods (SocioDim, wvRN and LBC), SocioDim is the most stable and achieves the best performance.

Performance on clustering:

This indicates that the extraction of social dimensions can be crucial to our SocioDim framework. The difference of soft clustering and hard partition is evident on community structure. Both spectral clustering and modularity maximization outperform k-means partition.

When the network scales to a larger size as in LAP, modularity maximization does not show a strong superiority over hard partition. Indeed, the performance of modularity maximization and that of k-means partition are comparable on LAP. Spectral clustering, on the contrary, excels in all cases. Spectral clustering seems to capture the latent affiliations more accurately for within-network classification.

VI.CONCLUTION:

In this paper, the three conceptsis integrated and the useful information about efficient mining,

ISSN : 2454-9924

community detection, anomaly detection, blocking and data forwarding is done to improve the quality of social media. The integration of these concepts provides an advanced development in the social network and people will get knowledge about the particular community they work on it. Without any tie breaking the continuous data forwarding provides effective changes. The anomaly detection in the community can help the people to work protectively in the social network. Using memory based label propagation algorithm(MLAP)for finding the community structure in social network and social attraction and infrastructure(SAIS)algorithm include simbet, bubble rab and nguyen's routing algorithms for better performance of data forwarding and efficient mining. Finally, the algorithm extracts the most frequent common label for each node to form communities of the networks. The experimental results showed a relative improvement in comparison with other community detection algorithms. We propose a space-crossing community detection method and describe a high efficient data forwarding schemes.**In future,** an attempt could be made to develop a new model basedon Human behavior which could be used in future forpredicting some useful data e.g. Degree of influence of a personin a group, kind of friends he mayhave, type of group he mayjoin, type of relationship he may have, etc.

VII. REFERENCE

1.Lei Tang and Huan Liu, "Leveraging social media networks for classification"jan 2011.

2.WuHe,Gongjun, and Li DA Xu, "Multi label propagation for overlapping community detection"no.2.May 2014.

3.RaziehHosseini and Reza Azmi, "memory-based label propagation algorithm for community detection in social networks" alzahrauniversity,Tehran,Iran, IEEE feb 2015.

4.Jyotisunil more and chelpa lingam, "Reality mining based on social network analysis"IEEE, may2015

5.Zhong hi,chengwang,siqianyang,changjunjiang,and Ivan Stojmenovic,fellow,IEEE:"Space-crossing:community-based data forwarding in mobile social networks under the hybrid communication architecture":DOI 10.1109/TWC.2015.

6.G.Nandi and A.Das, " A survey on using data mining techniques for online social network analysis"international journal of computer science issues,vol 10,nov 2013.

7.G.T.Prabavathi and V.Thiagarasu,"Overlapping community detection algorithm in dynamic networks: An overview"(IJETCAS-13-585)2013

8.ZongqingLu,YonggangWen and GuohongCao,"Community detection in weighted networks: algorithm and applications"IEEEjan 2013.

9.P.Hui,J.Crowcroft, and E.Yoneki, "Bubble rap:social based forwarding in delay tolerant networks":a social network perspective,in ACM Mobihoc 2008.

10.B.Chen, J.Xiang,K.Hu and Y.Tang ,Enhancing betwenness algorithm for detecting community in complex networks," modern physics letters B,Vol.28,no.09,2014.

11.F.D.Malliaros and M.Vazirgiannis,Clustring and community detection in directed networks:Asurvey:"physics report,Vol.533,no.4,pp.95-142,2013.

12. T. DuBois, J. Golbeck, and A. Srinivasan, "Predicting Trust and Distrust in Social Networks", in Proc. of the 3rd IEEE Int. Conf. on Social Computing, 2011.

13. A. Ortigosa, J. M. Martín, and R. M. Carro, "Sentiment analysis in Facebook and its application to e-learning", in the Journal of Computers in Human Behavior, http://dx.doi.org/10.1016/j.cshb.2013.05.02, 2013

14. R. Zafarani, W. D. Cole, and H. Liu, Sentiment Propagation in Social Networks:Study in LiveJournal, in Advances in Social Computing, Springer, 2010, pp. 413–420.

15. T. H. Cormen, C. E. Leiserson, R. L. Rivest, and Stein. *Introduction to Algorithms*. MIT Press, 3rd edition, 2009

16. J. M. Pujol, G. Siganos, V. Erramilli, and P. Rodriguez.Scaling Online Social Networks

ISSN : 2454-9924

without Pains. In *NetDB'09: 5th International Workshop onNetworking Meets Databases*, 2009.

17. U. Zwick. Exact and Approximate Distances in Graphs – A Survey. In *ESA'01: Procceding of the 9thAnnual European Symposium on Algorithms*, LectureNotes in Computer Science 2161/2001. Springer, 2001.

18. W. Gao, G. Cao, T. La Porta, and J. Han, "On exploiting transient social contact patterns for data forwarding in delay-tolerant networks," *MobileComputing, IEEE Transactions on*, vol. 12, no. 1, pp. 151 –165, jan.2013.

19. J. Leskovec, K. Lang, and M. Mahoney, "Empirical comparison of algorithms for network community detection," in *ACM WWW 2010*.

20. A. Scherrer, P. Borgnat, E. Fleury, J. L. Guillaume, and C. Robardet, "Description and simulation of dynamic mobility networks," Elsevier Computer Networks, vol. 52, no. 15, 2008

21. T. Spyropoulos, T. Turletti, and K. Obrazcka, "Routing in delay tolerant networks comprising heterogeneous populations of nodes," IEEE TMC, 2009