IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A FRAMEWORK FOR COMMUNITY IDENTIFICATION IN DYNAMIC SOCIAL NETWORKS

D.Ragini*, P. Latha
* ME-CSE Kingston Engineering College Vellore, India.
Assistant Professor –CSE Kingston Engineering College Vellore, India.

ABSTRACT
A unified framework is provided for tracking the evolution of hierarchical and overlapping communities in dynamic online social networks. HOCTracker is used for tracking such communities. HOCTracker adapts a preliminary community structure towards dynamic changes in social networks. It uses a novel density-based approach for detecting overlapping community structures. It automatically tracks evolutionary events like birth, growth, contraction, merge, split, and death of communities. It does not require an ageing function to remove old interactions for identifying community evolutionary events. HOCTracker is applicable to directed/undirected and weighted/un-weighted networks.

KEYWORDS: HOCTracker, Dynamic Social Communities, OSLOM, Evolving Networks, mesoscopic structure.

INTRODUCTION
Communities in social networks often map to important functional groups making their detection a highly desirable task. The problem of community detection in social networks depends on various factors. Data mining research from social networks has recently gained significant popularity with an aim of analyzing mesoscopic structure of networks and their evolution. The problem of community detection in social networks depends on various factors, including whether the definition of community relies on global or local network properties, whether communities overlap, whether communities possess a hierarchical structure, whether link weights are utilized, whether outliers are considered, and whether dynamic nature of networks/communities is considered. The objective of the paper is to avoid overlapping and to detect communities in dynamic social networks such as facebook which creates a problem in secret information leaking and reduce the storage in database server.

PROBLEM DOMAIN
Data mining is the entire process of applying computer-based methodology, including new techniques for knowledge discovery. Knowledge Discovery is a methodology for extracting useful knowledge from data and Knowledge Prediction uses known data to forecast future trends, events etc. Inferring new information from already collected data is termed to be Data Mining. It is the practice of examining large pre-existing databases in order to generate new information. Data Mining is a process used by companies to turn raw data into useful information. The overall goal of data mining process is to extract information from a dataset and transform into understandable format.

PROBLEM CHARACTERIZATION
The Proposed HOC which uses topological overlap criteria to define the similarity between two arbitrary nodes in a network to identify clique-based communities. Lancichinetti presented OSLOM which is able to detect a hierarchical community structure by repeatedly applying the community detection algorithm on intermediate super-networks of detected communities. Both HOC and OSLOM are multi-resolution community detection methods as they have a freely tunable resolution parameter which allows them to identify communities at varying levels of resolutions, thus forming a community structure hierarchy. HOCTracker features include:

1) It automatically finds the overlapping and detects the hierarchical structures of communities.
2) It uses novel density based approach to identify the changes occurring in network and processes only active nodes for new time-step, instead of processing all nodes for every time-step.

3) It presents a design of an intermediate evolution log (IEL) and maintains the evolutionary mappings between community structures of two consecutive time steps.

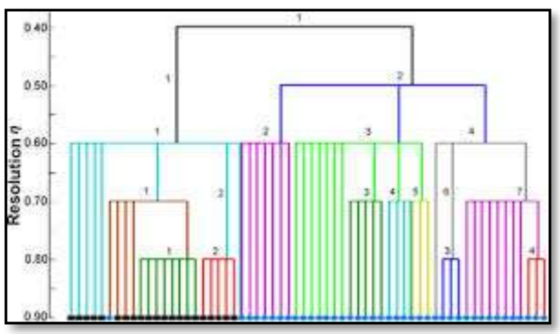
The IEL has four fields

1) **ID**: It represents the label of resulting community.
2) **Parents**: It includes the list of ID's of the actual communities.
3) **Transition**: It indicates immediate event including ID's of immediate communities that resulted in this community. This field can be used to track the transitions of particular community on processing the candidate nodes.
4) **Live**: This is the flag which represents the status of community. A value of 1 represents community is currently alive and 0 entry represents community has transitioned into some other state. For new entry this field is set to 1(for death set to 0)

FEATURES OF HOCTRACKER

The HOCTracker has a rich set of features. It includes;

1) The lack of a commonly accepted definition of what constitutes a community, allowing for communities to overlap.
2) A quantitative (statistical) analysis of large-scale social network demonstrating significant numbers of communities which have non-trivial overlap.
3) This is generalized to identify community evolution by detecting birth, death, merge, split, growth, and shrink events of communities.
4) The contents of the original community don't change.



LITERATURE SURVEY

A body of literature has been conducted by several authors and a list of them is given below;

1. **Overlapping community detection in networks: the state of the art and comparative study**
   A thorough comparison of different algorithms (a total of fourteen) is provided. In addition to community-level evaluation, we propose a framework for evaluating algorithms' ability to detect overlapping nodes, which helps to assess over-detection and under-detection. For low overlapping density networks, SLPA, OSLOM, Game, and COPRA offer better performance than the other tested algorithms. For networks with high overlapping density and high overlapping diversity, both SLPA and Game provide relatively stable performance.
2. **Analysis and mining of online social networks: emerging trends and challenges**
   Social network analysis (SNA) is a multi-disciplinary field dedicated to the analysis and modeling of relations and diffusion processes between various objects in nature and society, and other information/knowledge processing entities with an aim of understanding how the behavior of individuals and their interactions translate into large-scale social phenomenon. Due to exploding popularity of online social networks and availability of huge amount of user-generated content there is a great opportunity to analyze social networks and their dynamics at resolutions and levels not seen before. This has resulted in a significant increase in research literature at the intersection of the computing and social sciences leading to several techniques for social network modeling and analysis in the area of machine learning and data mining.
3. **Towards linear time overlapping community detection in social networks**
   Membership diversity is a characteristic aspect of social networks in which a person may belong to more than one social group. For this reason, discovering overlapping structures is necessary for realistic social analysis. In this paper, we present a fast algorithm, called SLPA, for overlapping community detection in large-scale networks. SLPA spreads labels according to dynamic interaction rules. It can be applied to both unipartite and

bipartite networks. It is also able to uncover overlapping nested hierarchy. The time complexity of SLPA scales linearly with the number of edges in the network. Experiments in both synthetic and real world networks show that SLPA has an excellent performance in identifying both node and community level overlapping structures.

4. **Finding overlapping communities in networks by label propagation**
Finding overlapping communities within complex networks that leverages swarm intelligence for decentralized multi-threading processing with label propagation for its fast identification of communities. The combination of the two technologies offers a high performance approach to overlapped community detection that allow for the processing of very large networks in tractable time.

5. **Finding statistically significant communities in social networks**
Community structure is one of the main structural features of networks, revealing both their internal organization and the similarity of their elementary units. Despite the large variety of methods proposed to detect communities in graphs, there is a big need for multi-purpose techniques, able to handle different types of datasets and the subtleties of community structure. In this paper we present OSLOM (Order Statistics Local Optimization Method), the first method capable to detect clusters in networks accounting for edge directions, edge weights, overlapping communities, hierarchies and community dynamics. It is based on the local optimization of a fitness function expressing the statistical significance of clusters with respect to random fluctuations, which is estimated with tools of Extreme and Order Statistics. OSLOM can be used alone or as a refinement procedure of partitions/covers delivered by other techniques. We have also implemented sequential algorithms combining OSLOM with other fast techniques, so that the community structure of very large networks can be uncovered. Our method has a comparable performance as the best existing algorithms on artificial benchmark graphs. Several applications on real networks are shown as well. OSLOM is implemented in a freely available software, and we believe it will be a valuable tool in the analysis of networks.

6. **static community detection algorithms for evolving networks**
A common feature of many real networks is the existence of groups of nodes, or communities, which are crucial in understanding the underlying structure of large networks. Communities can be, for example, groups of friends in a social network, websites dealing with the same subject or scientists working on similar topics in co-authorship networks. Formalizing the notion of community and detecting communities is a harsh task which has attracted a lot of attention in the recent years. In particular, many algorithms based on the network topology have been proposed. However, most studies usually deal with a single static network which is a snapshot of the data. Very few studies have given attention to temporal features. As real data are always evolving (in phone networks, people do not call each other permanently; on the web, pages appear, disappear or are updated continuously for instance), a lot of information is lost. It tracks communities between successive snapshots of the evolution of a network. We will first use classical community detection algorithms whose main issue in this context is stability.

7. **Community detection in Social Media ,Performance and application considerations**
Community detection constitutes a significant tool for the analysis of complex networks by enabling the study of mesoscopic structures that are often associated with organizational and functional characteristics of the underlying networks. To this end, this survey first frames the concept of community and the problem of community detection in the context of Social Media, and provides a compact classification of existing algorithms based on their methodological principles. The survey places special emphasis on the performance of existing methods in terms of computational complexity and memory requirements. It presents both a theoretical and an experimental comparative discussion of several popular methods. In addition, it discusses the possibility for incremental application of the methods and proposes five strategies for scaling community detection to real-world networks of huge scales. Finally, the survey deals with the interpretation and exploitation of community detection results in the context of intelligent web applications and services.

8. **Community-Based Features for Identifying Spammers in Online Social Network**
The popularity of Online Social Networks (OSNs) is often faced with challenges of dealing with undesirable users and their malicious activities in the social networks. The most common form of malicious activity over OSNs is spamming wherein a bot (fake user) disseminates content, malware/viruses, etc. to the legitimate users of the social networks. The common motives behind such activity include phishing, scams, viral marketing and so on which the recipients do not indent to receive. It is thus a highly desirable task to devise techniques and methods for identifying spammers (spamming accounts) in OSNs. With an aim of exploiting social network characteristics of community formation by legitimate users, this paper presents a community-based framework to identify spammers in OSNs. The framework uses community-based features of OSN users to learn classification models for identification of spamming accounts. The preliminary experiments on a real-world dataset with simulated spammers reveal that proposed approach is promising and that using community-based node features of OSN users can improve the performance of classifying spammers and legitimate users.

9. **Overlapping community structures and their detection on social networks**
Network communities have long been believed to be groups of tight-knit nodes having more internal than external connections. In social networks, a social community usually consists of people sharing common interests who tend to interact more frequently with other members than to the outside world. The discovery of network communities provides us a much better understanding about the structural topology of each community as well as its organization principles. For example, on online social network sites, e.g., Facebook, Twitter or MySpace, a user with favors in movies, music and art can join in and becomes an active member of those communities of interests. A disjoint community detection method, when applied to a network with overlapping communities, shall misleadingly classify overlapped nodes into different communities, thus fails to reveal the original network's structure. This makes the detection of overlapping communities an interesting, yet challenging problem. This problem also drives the need for a different concept of overlapping community structure.
10. **Dense sub graph extraction with application to community detection**
A challenging problem in the analysis of graph structures is the dense sub graph problem, where given a sparse graph, the objective is to identify a set of meaningful dense sub graphs. This problem has attracted much attention in recent years due to the increased interest in studying various complex networks, such as the World Wide Web (information network), social networks, and biological networks, etc. The dense sub graphs are often interpreted as "communities" based on the basic assumption that a network system consists of a number of communities, among which the connections are much fewer than those inside the same community. The recent data mining literature has seen various techniques for approaching this network analysis problem from different aspects. It focus on the methodology of graph partitioning/clustering, with the goal of obtaining dense partitions/clusters. This may not accurately represent human communities, since it is common for social connections not to be evenly divided.

CONCLUSION AND FUTURE ENHANCEMENT
A novel density-based framework, HOCTracker, is proposed to track community evolution Heat-map representing year-wise distribution of the number of communities and evolution events identified by HOCTracker on Wikipedia Election network. It uses an efficient log-based approach to map evolutionary relations between communities identified at two consecutive time-steps of a dynamic network. It does not require an ageing function to remove old interactions for identifying these events.

ACKNOWLEDGEMENT
I would like to take this opportunity to express my profound gratitude and deep regard to my guide, *Prof.P.Latha CSE, Kingston Engineering College*, for her exemplary guidance, valuable feedback and constant encouragement in completing this paper. Her valuable suggestions were of immense help in getting this work done. Working under her, was an extremely knowledgeable experience. Also, I would like to extend my sincere gratitude to my parents and my brothers for their constant support and encouragement in completing this paper.

REFERENCE
[1] J. Xie, S. Kelley, and B. K. Szymanski, "Overlapping community detection in networks: The state of the art and comparative study," ACM Comput. Surv., vol. 45, p. 43, 2011.
[2] S. Y. Bhat and M. Abulaish, "Analysis and mining of online social networks: Emerging trends and challenges," WIREs: Data Min. Knowl. Discovery, vol. 3, no. 6, pp. 408–444, 2013.
[3] G. Palla, I. Der_enyi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," Nature, vol. 435, no. 7043, pp. 814–818, 2005.
[4] A. McDaid and N. Hurley, "Detecting highly overlapping communities with model-based overlapping seed expansion," in Proc. Int. Conf. Adv. Soc. Netw. Anal. Min., 2010, pp. 112–119.
[5] J. Huang, H. Sun, J. Han, and B. Feng, "Density-based shrinkage for revealing hierarchical and overlapping community structure in networks," Physica A: Statist. Mech. Appl., vol. 390, no. 11, pp. 2160–2171, 2011.
[6] J. Xie and B. K. Szymanski, "Towards linear time overlapping community detection in social networks," in Proc. 16th Pacific-Asia Conf. Adv. Knowl. Discovery Data Min., 2012, vol. 2, pp. 25–36.
[7] S. Gregory, "Finding overlapping communities in networks by label propagation," New J. Phys., vol. 12, no. 10, p. 103018, 2010.
[8] P. Kumar, L. Wang, J. Chauhan, and K. Zhang, "Discovery and visualization of hierarchical overlapping communities from bibliography information," in Proc. IEEE 8th Int Conf. Depended, Autonom. Secure Comput., 2009, pp. 664–669.

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

[9] A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato, “Finding statistically significant communities in networks,” PLoS ONE, vol. 6, no. 4, pp. 1–18, 2011.
[10] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, “Group formation in large social networks: Membership, growth, and evolution,” in Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Min., 2006, pp. 44–54.
[11] C. Tantipathananandh, T. Berger-Wolf, and D. Kempe, “A framework for community identification in dynamic social networks,” in Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discovery Data Min., 2007, pp. 717–726.
[12] D. Greene, D. Doyle, and P. Cunningham, “Tracking the evolution of communities in dynamic social networks,” in Proc. Int. Conf. Adv. Soc. Netw. Anal. Min. 2010, pp. 176–183.
[13] Y. Wang, B. Wu, and N. Du, “Community evolution of social network: Feature, algorithm and model,” arXiv:0804.4356 physics. soc-ph, pp. 1–16, 2008.
[14] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, and B. L. Tseng, “Analyzing communities and their evolutions in dynamic social networks,” ACMTrans. Knowl. Discovery Data, vol. 3, pp. 8:1–8:31, Apr. 2009.
[15] R. Cazabet, F. Amblard, and C. Hanachi, “Detection of overlapping communities in dynamical social networks,” in Soc. Comput. (SocialCom), 2010 IEEE Second Int. Conf., Aug. 2010, pp. 309–314.
[16] D. Greene, D. Doyle, and P. Cunningham, “Tracking the evolution of communities in dynamic social networks,” in Proc. Int. Conf. Adv. Soc. Netw. Anal. Min. Aug. 2010, pp. 176–183.
[17] T. Falkowski, Community Analysis in Dynamic Social Networks. Gottingen, Germany: Sierke, 2009.
[18] S. Y. Bhat and M. Abulaish, “OCTracker: A density-based framework for tracking the evolution of overlapping communities in OSNs,” in Proc. IEEE/ACM Int. Conf. Adv. Soc. Netw. Anal. Min., Aug. 2012, pp. 501–505.