International Journal of scientific research and management (IJSRM)
||Volume||2||Issue||7||Pages|| 1139-1142||2014||
Website: www.ijsrm.in ISSN (e): 2321-3418

Scalable Learning Of Collective Behaviour

Ahmed ullah khan[1], MD Anwar Ali[2], MD khaliq[3]
M.Tech student , Associate Proffesor , Associate Proffesor
ahmedkhush2006@gmail.com, md.anwar.ali123@gmail.com, ahmedkhandba1@gmail.com

Abstract--- In a multimode consists of heterogeneous types of actors among which various types of interactions between them In a multimode, both actor membership and interactions evolve, which causes a problem for identifying the evoving communites. Identifying communites can help understanding the structural properties of the network,address the data shortage, unbalanced problems and tasks lik targeting marketing and finding within or between groups. In a proposed system to address the issue by employing temporal information to anayze a multimode network an scalability issue are carefully studied.

***Key words*-- Data mining, community detection, community evolution, multimode networks.**

I. INTRODUCTION

A network data can be produced from social networks, technology networks,genetic networks, network analysis and also increasing in many fields. Examples include targeted marketing, recommendation systems, relational learning and behavior prediction. Applications like web mining and online targeted marketing contains more than one type of entities. There are different type of interactions. This kind of network is called multimode network.

In existing system friendship networks and mobile networks are one mode networks. The networks of one mode, that is only one type of actors (nodes) is present in a networks and the connections (interactions) between actors are of same type. A multimode network are also called as heterogeneous network. Within multimode network different type of entities tend to form groups or communities. The heterogeneous entities with interactions, evolving groups can be understanding of interactions between modes as well as long term evolution patterns. In social media understanding group structures and properties.

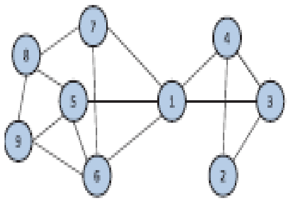
In social computing and communication technologies enables people to get together and share information in innovative ways. Social networking sites empower people of different ages and backgrounds with new forms of collaboration, communication, and collective intelligence. Prodigious numbers of online volunteers collaboratively write encyclopedia articles of unprecedented scope and scale; online marketplaces recommend products by investigating user shopping behavior and interactions; and political movements also exploit new forms of engagement and collective action. In the same process, social media provides opportunities to study human interactions and collective behavior on an unprecedented scale. In this work, To study how networks in social media can help predict some human behaviors and individual preferences. In particular, given the behavior of some individuals in a network, how can we infer the behavior of other individuals in the same social network This study can help better understand behavioral patterns of users in social media for applications like social advertising and recommendation

II. RELATED WORK

Social media contains huge amount of data generated by face book, You Tube, flicker etc. It facilitates people of all walks of life to connect to each other. Social media involves hundreds of thousands of actors. It can help predict some sorts of human behavior. It is capable of handling networks of millions of actors.

In existing system is based on social dimensions. The extraction of social dimensions is based on the network topology to capture the potential affiliation of actors. These extracted social dimensions represent the each actor is involved in diverse affiliation. It can be treated as features of actors for discriminative learning, determines the social dimensions and assign some proper weights. The same affiliation of actors connect to each other. To address the heterogeneity of connections, a social dimensions for collective behavior. For latent affiliation of actors, need to find out the group of people interact with each other more frequently. A network is converted into features; a typical classifier such as support vector machine can be used

Edge-Centric Clustering



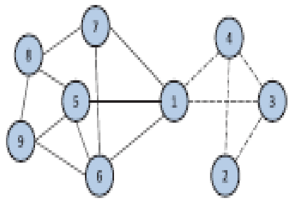


Fig1:Toy Example   Fig2:Egde cluster

For instance, a toy network with two communities shown in fig1 and partition the edges into two clusters shown in fig2. A Edge-Centric clustering schema can be used to extract social dimensions. Each edge can be treating as data instances and nodes that define edges as features. Based on the features of each edge, can be clustering the edges into disjoint sets. In social dimensions extraction cluster the edges rather than nodes in a network into disjoint sets. Because the affiliation of one actor is no more than the connections the social dimensions based on edge-centric clustering are guaranteed to be sparse.The sparse social dimensions shows comparable prediction as earlier approaches to extract social dimensions

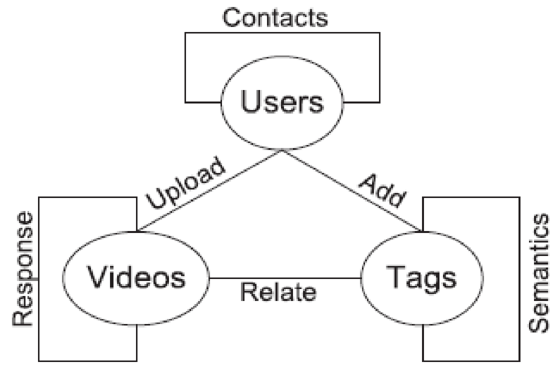
III. PROPOSED SYSTEM

In social media multiple modes of actors can be involved in a network, resulting as a multimode network. In a multimode network, actors of different modes can evolve differently.

1.Multimode Networks

For instance, a network in You Tube, users, videos, and tags can be constructed in a 3-mode network (shown in fig3). In this network both Videos and tags are considered as "actors", user can be the major mode. In You Tube network different types of interactions exist between them. Users can upload

videos and tags are correlated to each other. In a multiple type of entities exist in the same network and entities relate to similar interests are likely to form a group, videos are clustered, and they relate to similar contents and tags are clustered they are associated with similar users and videos. A user group interact with multiple groups of another mode. Users might have multiple interests, thus relating to various videos or tags.A friendship, users can uploaded the video clips and respond to another video, tags can also connect to each other based on their semantic meaning.



:

Fig 3: An Example of a 3-mode network

Another example of a multimode network is the field of academia as shown in fig .A entities are researchers, conference/journals, papers and words are correlated to each other. A Scientific literature, paper connects by citations; papers are published at different places like conference, journals, workshops, thesis etc.

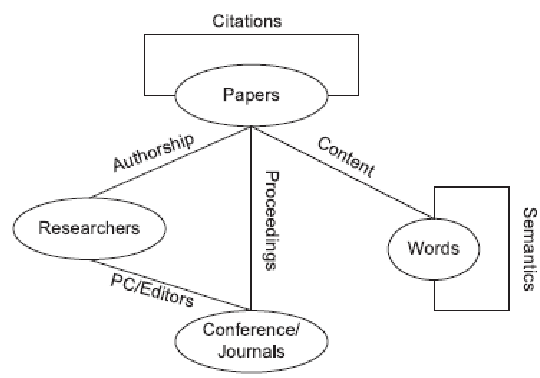
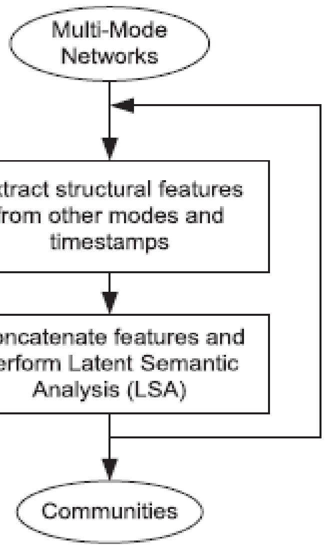


Fig 4: A multimode network of academic publications

In a multimode networks, partitioning the communities apply the K-means clustering schema can be used. The attributes for certain modes of actors, adding these attributes as features in P_i^t that is

$$\widetilde{P}_i^t = \left[\left\{\sqrt{w_a^{(i,j)}} R_{i,j}^t C^{(j,t)}\right\}_{j\neq i}, \sqrt{w_b^{(i)}} C^{(i,t\pm1)}, \sqrt{w_c^{(i)}} F_i^t\right],$$

Where F_i^t denotes the attributes for actors in mode i, and $w_c^{(i)}$ the weight for actor attributes.



The features in a multimode networks are updated based on features from other modes and timestamps are obtained from Latent Semantic Analysis (LSA). LSA is iterative process, until an equilibrium state is reached.

IV. EXPERIMENT RESULTS

A Synthetic network consists of three modes with 2,3, and 4 clusters, and 50, 100, and 200 actors, respectively. Interactions occur between the modes.

- It determines the latent community for each actor.
- It generates interactions between groups identity for entities, interactions follows a Bernoulli distribution.

To simulate community evolution, generate interactions at different timestamps as follows.

- For community evolution, at each timestamp a certain percentage (5-20 percent) of actors switch to another group from that of the previous timestamp.
- The interactions between groups probability with a Gaussian distribution on the probability of the previous timestamp

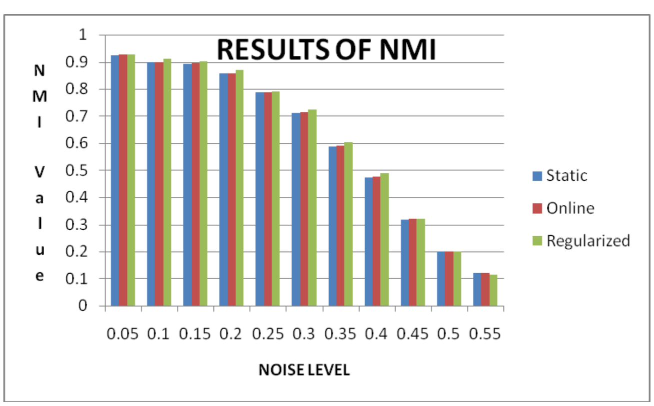


Fig. 5: Results of NMI

- The noise level is gradually increased from the top to bottom. Noise is inserted at each timestamp. For example noise level 0.2 approximately 20 percent of entries in the matrix to be set to 0 or 1.
- Normalized Mutual Information (NMI) is to evaluate clustering. Let π^a, π^b denote two different partitions of communities. NMI is defined as

$$NMI = \frac{\sum_{h=1}^{k^{(a)}}\sum_{\ell=1}^{k^{(b)}} n_{h,\ell}\log\left(\frac{n\cdot n_{h,\ell}}{n_h^{(a)}\cdot n_\ell^{(b)}}\right)}{\sqrt{\left(\sum_{h=1}^{k^{(a)}} n_h^{(a)}\log\frac{n_h^a}{n}\right)\left(\sum_{\ell=1}^{k^{(b)}} n_\ell^{(b)}\log\frac{n_\ell^b}{n}\right)}},$$

Where n is the total number of data instances, $k^{(a)}$ and $k^{(b)}$ represent the number of communities in partitions π^a and π^b, respectively, n_h^a, n_ℓ^b and $n_{h,\ell}$ are the number of actors. NMI equals 1 when partitions are equivalent.

It can be notice that when the noise is high (e.g.noise level is 0.55), the smoothness of community structure between consecutive timestamps is destroyed

V. CONCLUSION AND FUTURE WORK

- In network different modes of actors from their own communities evolve gradually.
- A real world applications various types of actors and interactions with each other in a multimode networks.
- In a existing system, a clustering algorithm assign entity to a single group. Actors can be involved in multiple communities. To handle the community in a edge-centric clustering for community detection.
- Community can be defines as set of edges and nodes. In a existing system one node can be associated with one community. In a edge-centric clustering schema partitioning the edges into disjoint sets, one node can be associated with multiple communities. On extending the edge centric schema for detecting single mode networks to multimode networks
- It also investigating the user grouping and behavioral patterns in multimode networks

A further researchcan be required for capturing the micro level evolution of actors and macro level evolution of groups

(1) References

[1] Lei Tang ,Xufei Wang and Huan Liu, "Scalable Learning of Collective Behavior", IEEE Transactions on Knowledge and Data Engineering, 2011

[2] L. Tang and H. Liu, “Toward predicting collective behavior via social dimension extraction,” IEEE Intelligent Systems, vol. 25, pp. 19–25, 2010

[3] J. Baumes, M. Goldberg, M. Magdon-Ismail, and W. Wallace, “Discovering Hidden Groups in Communication Networks,”Proc. Second NSF/NIJ Symp. Intelligence and Security Informatics,2004.

[4] L. Tang and H. Liu, “Scalable learning of collective behavior based on sparse social dimensions,” in CIKM ’09: Proceeding of the 18th ACM conference on Information and knowledge management. New York, NY, USA: ACM, 2009, pp. 1107–1116.

[5] L .Tang, H. Liu, J. Zhang, and Z. Nazeri, “Community Evolution in Dynamic Multi-mode Networks,” Proc. 14th

[6] ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining (KDD ‘08),pp.677-685, 2008.

[7] M. Newman, “The Structure and Function of Complex Networks,” SIAM Rev., Vol 45, pp. 167-256, 2003.

[8] L. Tang and H. Liu, “Relational Learning via Latent Social Dimensions”, proc 15th ACM SIGKDD Int’l Conf Knowledge Discovery and Data Mining (KDD ’09),pp 817-826, 2009.

[9] H.Zha , X. He, C.H.Q. Ding, M. Gu, and H.D. Simon, “Spectral Relaxation for K-Means Clustering “, Proc. Neural Information Processing Systems (NIPS), pp. 1057-1064, 2001.

[10] L. Backstrom, D. Huttenlocher , J. Kleinberg, and X. Lan, “Group Formation in Large Social Networks: Membership, Growth, and Evolution,” Proc. 12th ACM SIGKDD Intl’l Conf. Knowledge Discovery and Data Mining (KDD ’06) pp 44=54, 2006.

[11] B. Long, Z.M. Zhang, X. Wu, and P.S. Yu, “Spectral Clustering for Multi-Type Relational Data,” Proc. 23rd Int’l Conf. Machine Learning (ICML ’06), pp. 585-592, 2006.

[12] J. Wang, H. Zeng, Z. Chen , H. Lu, L.Tao, and W. –Y, Ma, “Recommendation Reinforcement Clustering for Multi-Type. Interrelated Data Objects,” Proc, 26th Ann. Int’l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR ’03), pp. 274-281, 2003.

[13] T. Evans and R. Lambiotte, “Line Graphs, link partitions, and overlapping communities,” Physical Review E, Vol. 80, no. 1,p. 16105, 2009

[14] S. White and P. Smyth, “A Spectral Clustering approach to finding communities in graphs,” in SDM, 2005.

[15] J. Neville and D. Jensen, ”Leveraging relational autocorrelation with latent group models ,” in MRDM ’05; Proceeding of the 4th international workshop on Multi relational mining. New York, NY, USA; ACM, 2005, pp 49-55.

[16] R-E. Fan and C,=J Lin, “A study on threshold selection for multi-label classification,” 2007