Collective behavior prediction in social media: A survey

M.Vasudevan
Master of Engineering
Dept of Computer Science and Engineering
K.S.R. College of Engineering
Tiruchengode

M.Tamilarasi M.E., (Ph.D)
Senior Assistant Professor
Dept of Computer Science and Engineering
K.S.R. College of Engineering
Tiruchengode

***Abstract-* Collective behavior refers to how individuals behave when they are exposed in a social network environment. This collective behavior gives the opportunity to predict online behaviors of users in a network, given the behavior information of some actors in the network. But Many social media (Facebook, Twitter, YouTube etc..) tasks can be connected to the problem of collective behavior prediction. This work studies how networks in social media can help predict some sorts of human behavior and individual preference. This can help understand the behavior patterns presented in social media, as well as other tasks like social networking advertising and recommendation.**

Keywords: Social Networks, Collective Behavior, Social Dimensions, Behavior Prediction.

I. INTRODUCTION

According to the British Encyclopedia the term collective behavior is a sociological term. Actions like riots, disaster, nationalistic movements, rumors, gossip, media hype of the public openion can all be considered to be forms of collective behavior. The collective behavior term was first used by Robert park . It is an alternative of crowd behavior [2]. He developed a theory about actions of group and this theory later named as collective behavior. This collective behavior was only one of the things designed to explain different aspect of the social change and contemporary society at that time of its birth.

The recent boom of social media (Facebook, Twitter, YouTube etc) enables human beings to interact with each other more easily than ever. User's interacting with each other by posting comments, like/dislike a product, etc . Connections in social media networks are not homogeneous. Different connections are associated with distinctive relations.

For example, one user might maintain connections simultaneously to his friends, family, college classmates, and colleagues. This relationship information, however, is not always fully available in reality. Mostly, we have access to the connectivity information between users, but we have no idea why they are connected to each other. This heterogeneity of connections limits the effectiveness of a commonly used technique — collective inference for network classification.

A recent framework based on social dimensions is shown to be effective in handling the heterogeneity problem. The framework suggests a novel way of network classification: first, capture the latent affiliations of actors by extracting social dimensions based on network connectivity, and next, apply extant data mining techniques to classification based on the extracted dimensions. This paper discuss various techniques that are used for collective behavior Prediction.

II. LITERATURE REVIEW

L.Tang and H.Liu [1] stated that collective behavior refers to how individuals behave when they are exposed in a social network environment. In this paper, they examined how they could predict online behaviors of users in a network, given the behavior information of some actors in the network. Many social media tasks can be connected to the problem of collective behavior prediction. Since connections in a social network represent various kinds of relations, a social-learning framework based on social dimensions is introduced. This framework suggests extracting social dimensions that represent the latent affiliations associated with actors, and then applying supervised learning to determine which dimensions are informative for behavior prediction. It demonstrates many advantages, especially suitable for large-scale networks, paving the way for the study of collective behavior in many real-world applications.

Collective behavior is not simply the aggregation of individuals' behavior. In a connected environment, behaviors of individuals tend to be interdependent. That is, one's behavior can be influenced by the behavior of his/her friends. This naturally leads to behavior correlation between connected users. Such collective behavior correlation can also be explained by homophily. M.McPherson, L.Smith-Lovin, and J.M.Cook [7] discussed that, the people who are interacting with each other are share certain similarities between them. The author also described that this correlated behavior information also used for prediction of online behaviors in a network.

M.E.J. Newman, A.L. Barab´asi and D.J. Watts [17] proposed a concept called collective inference. It assumes that the behavior of one actor is dependent upon that of his friends. To make prediction, collective inference is required to find an equilibrium status such that the inconsistency between connected actors is minimized. This is normally done by iteratively updating the possible behavior output of one actor while fixing the behavior output (or attributes) of his connected friends in the network. It has been shown that considering this network connectivity for behavior prediction outperforms those that do not. However, connections in social media are often not homogeneous. The heterogeneity presented in network connectivities can hinder the success of collective inference.

P.Singla and M.Richardson [4] applied data mining techniques to study this relationship for a population of over 10 million people, by turning to online sources of data. The analysis reveals that people who chat with each other (using instant messaging) are more likely to share interests (their Web searches are the same or topically similar). The more time they spend talking, the stronger this relationship is. People who chat with each other are also more likely to share other personal characteristics, such as their age and location (and, they are likely to be of opposite gender). Similar findings hold for people who do not necessarily talk to each other but do have a friend in common. Their analysis is based on a well-defined mathematical formulation of the problem, and is the largest such study they were aware of.

M.E.J.Newman [3] considered the problem of detecting communities or modules in networks, groups of vertices with a higher-than-average density of edges connecting them. Previous work indicates that a robust approach to this problem is the maximization of the benefit function known as “modularity” over possible divisions of a network. Here the author showed that this maximization process can be written in terms of the eigenspectrum of a matrix they called the modularity matrix, which plays a role in community detection similar to that played by the graph Laplacian in graph partitioning calculations. This result leads us to a number of possible algorithms for detecting community structure, as well as several other results, including a spectral measure of bipartite structure in networks and a new centrality measure that identifies those vertices that occupy central positions within the communities to which they belong. The algorithms and measures proposed are illustrated with applications to a variety of real-world complex networks.

H. W. Lauw, J. C. Shafer, R. Agrawal, and A. Ntoulas [11] study the phenomenon of homophily in the digital world. Unlike the physical world, the digital world doesn’t impose any geographic or organizational constraints on friendships. Online friends might share common interests, there’s no reason to believe that two users with common interests are more likely to be friends. A common assumption about human nature is that people have a tendency to associate with other, similar people (a phenomenon called homophily). Sociology has studied homophily in the physical world extensively. However, the studies have generally been conducted on a small scale, and the similarity factors examined have been limited mostly to easily observed or surveyed socio demographic characteristics, such as race, gender, religion, and occupation — characteristics that don't necessarily manifest themselves in online social networks. One of the strongest underlying sources of homophily in the physical world is locality due to geographic proximity, family ties, and organizational factors, such as school and work. However, in the digital world, physical locality becomes less important, and other factors such as common interests might play a greater role.

S.A.Macskassy and F.Provost [9] presented that the classified entities that are interlinked with entities for which the class is known and also they uses a modular toolkit called NetKit for classification in networked data. NetKit is based on a node-centric framework in which classifiers comprise a local classifier, a relational classifier, and a collective inference procedure. Various existing node-centric relational learning algorithms can be instantiated with appropriate choices for these components, and new combinations of components realize new algorithms. This study focuses on univariate network classification, for which the only information used is the structure of class linkage in the network (i.e., only links and some class labels).It also shows that there are two sets of techniques that are preferable in different situations, namely when few versus many labels are known initially. They also demonstrated that link selection plays an important role similar to traditional feature selection.

III. CONCLUSION

This paper has surveyed different schemes that are used for collective behavior prediction. The networks in social media are normally of big in size, involving hundreds of thousands of actors. The scale of these networks entails scalable learning of models for collective behavior prediction. To address the scalability issue, an edge-centric clustering scheme is proposed. The proposed approach can efficiently handle networks of millions of actors while demonstrating a comparable prediction performance to other non-scalable methods.

REFERENCES

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

[2] N.J. Smelser and N. Joseph, Theory of collective behavior. London (UK): Routledge & Kegan Paul, 1962

[3] M.Newman, “Finding community structure in networks using the eigenvectors of matrices,” Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), vol. 74, no. 3, 2006. [Online]. Available: http://dx.doi.org/10.1103/PhysRevE.74.036104

[4] P.Singla and M. Richardson, “Yes, there is a correlation: - from social networks to personal behavior on the web,” in WWW ’08: Proceeding of the 17th international conference on World Wide Web. New York, NY, USA: ACM, 2008, pp. 655–664.

[5]N.J. Smelser and N. Joseph, Theory of collective behavior. London (UK): Routledge & Kegan Paul, 1962.

[6]——, “Relational learning via latent social dimensions,” in KDD ’09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY, USA: ACM, 2009, pp. 817–826.

[7] M.McPherson , L.Smith-Lovin , and J. M. Cook, “Birds of a feather: Homophily in social networks,” Annual Review of Sociology, vol. 27, pp. 415–444, 2001.

[8] H. W. Lauw, J. C. Shafer, R. Agrawal, and A. Ntoulas, “Homophily in the digital world: A Live Journal case study,” IEEE Internet Computing, vol. 14, pp. 15–23, 2010.

[9] S. A. Macskassy and F. Provost, “ Classification in networked data: A toolkit and a univariate case study,” J. Mach. Learn. Res.,vol. 8, pp. 935–983, 2007.

[10] S. Gupta, R. M. Anderson, and R. M. May, Networks of sexual contacts: Implications for the pattern of spread of HIV. AIDS 3, 807–817 (1989).

[11] H. W. Lauw, J. C. Shafer, R. Agrawal, and A. Ntoulas, “Homophily in the digital world: A LiveJournal case study,” IEEE Internet Computing, vol. 14, pp. 15–23, 2010.

[12] M. E. J. Newman, Fast algorithm for detecting community structure in networks. Phys. Rev. E 69, 066133 (2004).

[13] J.Reichardt and S. Bornholdt, Statistical mechanics of community detection. Preprint cond-mat/0603718 (2006).

[14] J.Duch and A. Arenas, Community detection in complex networks using extremal optimization. Phys. Rev.E 72, 027104 (2005).

[15] M. Granovetter. Threshold models of collective behavior. American journal of sociology, 83(6):1420, 1978.

[16] T. C. Schelling. Dynamic models of segregation. Journal of Mathematical Sociology, 1:143{186, 1971.

[17] M. E. J. Newman, A.-L. Barab´asi, and D. J. Watts, The Structure and Dynamics of Networks. Princeton University Press, Princeton (2006)

[18] M.Girvan and M. E. J. Newman, Community structure in social and biological networks. Proc. Natl. Acad. Sci.USA 99,7821–7826 (2002).

[19] P.Holme, M. Huss, and H. Jeong, Subnetwork hierarchies of biochemical pathways. Bioinformatics 19,532–538 (2003).

[20] G.W.Flake, S. R. Lawrence, C. L. Giles, and F. M. Coetzee, Self-organization and identification of Web communities. IEEE Computer 35, 66–71 (2002).

[21] R. Guimer`a and L. A. N. Amaral, Functional cartography of complex metabolic networks. Nature 433, 895–900 (2005).