Review On Finding Relevant Content and Influential Users based on Information Diffusion

Ms.Ashwini Sopan Shidore,Ms.Sindhu M.R

ME(II),Computer,
G.H.Raisoni College OfEngg,
Ahmednagar, Maharashtra,India
ashidore419@gmail.com

Abstract— Understanding information diffusion processes that take place on the Web, specially in social media, is a fundamental step towards the design of effective information diffusion mechanisms, recommendation systems, and viral marketing. Two key concepts in information diffusion are influence and relevance. Ability to popularize content in an online community is the influence . Influentials introduce relevant content, in the sense that such content satisfies the information needs of a significant portion of this community.We describe the problem of identifying influential users and relevant content in information diffusion data and also study how individual behavior data may provide knowledge regarding influence relationships in a social network in this paper.

Keywords— Influence, Relevance, Information diffusion , Social Networks, Twitter, Page-Rank, Profile-Rank.

INTRODUCTION

Powered by the remarkable success of Twitter, Facebook, Youtube, and the blogosphere, social media is taking over traditional media as the major platform for content distribution. The combination of user-generated content and online social networks is the engine behind this revolution in the way people share news, videos, memes, opinions, and ideas in general. As a consequence, understanding how users consume and propagate content in information diffusion processes is a very important step towards the design of effective information diffusion mechanisms, viral marketing and recommendation systems on the Web. Two key concepts in information diffusion are influence and relevance. In social networks, influence can be defined as the capacity to affect the behavior of others [1]. However, in information diffusion scenarios, influence is usually a measure of the ability of popularizing information. Relevance is a relationship between a user and a piece of information, in the sense that relevant information satisfies a user's information needs/interests, being a fundamental concept also in information retrieval and recommender systems [4, 11]. This work focuses on the link between user influence and information relevance in information diffusion data, which describe how users create and propagate information across time. As we are interested in the diffusion of content (e.g., news, videos) on the Web, we use the terms 'content' and 'information' interchangeably.

An important challenge in the study of influence and information diffusion in social networks is the lack of data at a large enough scale. Most of the social network datasets available for research contain just static topological information (i.e., persons and rela- tionships) and do not contain key information for analyzing influence and information diffusion. Further, social influence analysis requires temporal information that indicates, for instance, eventual association of persons to information items. As a consequence of such scarcity, a significant part of the existing models and analy- sis of social influence are based on synthetic data, which is frequently based on epidemiological models [17]. Blogs, news media websites, viral marketing campaigns, photo and video sharing services, and online social networks in general have provided rich datasets that supported several interesting findings regarding social influence and information diffusion in real scenarios.

LITERATURE SURVEY

Meeyoung Cha∗ Hamed Haddadi† Fabr´ıcio Benevenuto‡ Krishna P. Gummadi∗[24] described that using a large amount of data collected from Twitter, we present an in-depth comparison of three measures of influence: indegree, retweets, and mentions. Based on these measures, we investigate the dynamics of user influence across topics and time. We make several interesting observations. First, popular users who have high indegree are not necessarily influential in terms of spawning retweets or mentions. Second, most influential users can hold significant influence over a variety of topics. Third, influence is not gained spon- taneously or accidentally, but through concerted effort such as limiting tweets to a single topic. We believe that these findings provide new insights for viral marketing and suggest that topological measures such as indegree alone reveals very little about the influence of a user.

Jie Tang ,Jimeng Sun,Chi Wang and Zi Yang [2] shows that in large social networks, nodes (entities,users) are influenced by others for different reasons. For ex:, the colleagues have strong influence on one's work, while the friends have strong influence on one's daily life. How to differentiate the social influences from various angles(topics)? How to quantify the strength of those social influences? How to estimate the model on real large networks? To address these questions, we propose Topical Affinity Propagation (TAP) to model the topic-level social influence on large networks. TAP can take results of any topic modeling and the existing network structure to perform topic-level influence propagation. With the help of the influence analysis, we present several important applications on real data sets such as 1) what are the representative nodes on a given topic? 2) how to identify the social influence so neighboring nodes on a particular node? To scale to real large networks, TAP is designed with efficient distributed learning algorithms that is implemented and tested un- der the Map-Reduce framework. We also present the common characteristics of distributed learning algorithms for Map-Reduce. Finally, we describe the effectiveness and efficiency of TAP on real large data sets.

Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon[21] start with the network analysis and study the distributions of followers and followings, the relation between followers and tweets, degrees of separation. Next we rank users by the number of followers, the number of retweets and PageRank and present quantitative comparison among them. The ranking by retweets pushes those with fewer than a million followers on top of those with more than a million followers. Through our topic analysis we show what different categories trending topics are classified into, how long they last, and how many users participate. Finally, we study the information diffusion by retweet. We construct retweet trees and examine their spatial and temporal characteristics. This work is the first quantitative study on the entire Twittersphere and information diffusion on it..

Arlei Silva, Hérico Valiati, Sara Guimarães, Wagner Meira Jr.[28] described how individual behavior data may provide knowledge regarding influence relationships in a social network and also define what we call the influence network discovery problem, which consists of identifying influence relationships based on user behavior across time. Several strategies for influence network discovery are proposed and discussed. a case study on the application of such strategies using a follower-followee network and user activity data from Twitter, which is a popular microblogging and social networking service. We consider that a follower-followee interaction defines a potential influence relationship between users and the act of posting a tweet, a URL or a hashtag show an individual behavior on Twitter. The results described that, while tweets may be used effectively in the discovery of influence relationships, hashtags and URLs do not lead to good performance in such task. Moreover, strategies that consider the time when an individual behavior is observed outperform those that do not and by combining such information with the popularity of the behaviors, even good results may be achieved.

Arlei Silva ∗, Sara Guimarães, Wagner Meira , Mohammed Zaki [20] study the problem of identifying influential users and relevant content in information diffusion data. We propose ProfileRank, a new information diffusion model based on random walks over a user-content graph. ProfileRank is a PageRank inspired model that describes the principle that relevant content is created and propagated by influential users and influential users create relevant content. One good property of ProfileRank is that it can be adapted to provide personalized recommendations. Experimental results show that ProfileRank makes accurate recommendations, outperforming baseline techniques. We also illustrate relevant content and influential users discovered using ProfileRank. Our analysis shows that ProfileRank scores are more correlated with content diffusion than with the network structure. We also demonstrate that our new modeling is more efficient than PageRank to perform these calculations.

ACKNOWLEDGMENT

We would like to thank all the authors of different research papers referred during writing this paper. It was very knowledge gaining and helpful for the further research to be done in future.

CONCLUSION

In this paper, we have studied how individual behavior data may be applied in the identification of influence relationships in social networks. This paper surveys different research papers that proposed various methods which are basis for future research in the field of information diffusion data. Identifying influential user efficiently from large datasets are the challenging tasks in the field of information diffusion data. ProfileRank, is a random walk based information diffusion model that computes user influence and content relevance using information diffusion data. ProfileRank scores are more correlated with content diffusion than with the network structure. We also described that our new modeling is more efficient than PageRank .

REFERENCES:

[1] N. Friedkin. A structural theory of social influence, volume 13. Cambridge University Press, 2006.

[2] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In KDD, 2009.

[3] R. Baeza-Yates, B. Ribeiro-Neto, et al. Modern information retrieval, volume 82. Addison-Wesley New York, 1999.

[4] F. Alkemade and C. Castaldi, "Strategies for the diffusion of innovations on social networks," Comput. Economics, vol. 25, no. 1-2, pp. 3–23, 2005.

[5] M. Gomez Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In SIGKDD, 2010.

[6] H. Tong, S. Papadimitriou, P. S. Yu, and C. Faloutsos. Proximity tracking on time-evolving bipartite graphs. In SDM, 2008.

[7] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In WSDM, 2010.

[8] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In WWW, 2004.

[9] J. Hannon, M. Bennett, and B. Smyth. Recommending twitter users to follow using content and collaborative filtering approaches. In RecSys, 2010.

[10] F. Ricci, L. Rokach, and B. Shapira. Introduction to recommender systems handbook. Recommender Systems Handbook, 2011.

[11] H. Tong, C. Faloutsos, and J.-Y. Pan. Fast random walk with restart and its applications. In ICDM, 2006.

[12] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: finding topic-sensitive influential twitterers. In WSDM, 2010.

[13] A. Anagnostopoulos, R. Kumar, and M. Mahdian. Influence and correlation in social networks. In KDD, 2008.

[14] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In KDD, 2003.

[15] J. Yang and J. Leskovec. Patterns of temporal variation in online media. In WSDM, 2011.

[16] M. R. Subramani and B. Rajagopalan. Knowledge-sharing and influence in online social networks via viral marketing. Commun. ACM, 2003.

[17] B. Taskar, M. fai Wong, P. Abbeel, and D. Koller. Link prediction in relational data. In NIPS, 2004.

[18] D. Watts and P. Dodds. Influentials, networks, and public opinion formation. Journal of Consumer Research, 2007.

[19] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman. Influence and passivity in social media. In PKDD, 2011.

[20] Arlei Silva, Sara Guimarães, Wagner Meira, Jr., Mohammed Zaki "ProfileRank:finding relevant content and influential users based on information diffusion,2013

[21] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In WWW '10: Proceedings of the 19th international conference on World Wide Web, pages 591–600, 2010.

[22] M. J. Pazzani and D. Billsus. Content-based recommendation systems. In The Adaptive Web, pages 325–341. Springer Verlag, 2007.

[23] N. E. Friedkin, A Structural Theory of Social Influence. Cambridge University Press, 1998.

[24] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. In ICWSM, 2010.

[25] E. Bakshy, B. Karrer, and L. A. Adamic, "Social influence and the diffusion of user-created content," in Proceedings of the 10th ACM conference on Electronic commerce. ACM, 2009, pp. 325–334.

[26] M. Gomez Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2010, pp. 1019–1028.

[27] F. Bonchi, "Influence propagation in social networks: A data mining perspective," IEEE Intelligent Informatics Bulletin, vol. 12, no. 1, pp. 8–16, 2011.

[28] A. Silva, H. Valiati, S. Guimarães, and W. Meira Jr. From individual behavior to influence networks: A case study on twitter. In Webmedia, 2011.

[29] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, "The role of social networks in information diffusion," in Proc. of the 21st international conference on World Wide Web. ACM, 2012, pp. 519–528.

[30] S. Aral, L. Muchnik, and A. Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. PNAS, 2009.

[31] F. Alkemade and C. Castaldi, "Strategies for the diffusion of innovations on social networks," Comput. Economics, vol. 25, no. 1-2, pp. 3–23, 2005.

[32] M. De Choudhury, S. Counts, and M. Czerwinski. Identifying relevant social media content: leveraging information diversity and user cognition. In HT, 2011.

[33] J.Wortman, "Viral marketing and the diffusion of trends on social networks," University of Pennsylvania, Tech. Rep. Technical Report MS- CIS-08-19, May 2008