Ranking Mechanisms for Interaction Networks

Sameep Mehta, Ramasuri Narayanam, Vinayaka Pandit

IBM India Research Lab - Bangalore and New Delhi India

Abstract

Interaction networks are prevalent in real world applications and they manifest in several forms such as online social networks, collaboration networks, technological networks, and biological networks. In the analysis of interaction networks, an important aspect is to determine a set of key nodes either with respect to positional power in the network or with respect to behavioral influence. This calls for designing ranking mechanisms to rank nodes/edges in the networks and there exists several well known ranking mechanisms in the literature such as Google page rank and centrality measures in social sciences. We note that these traditional ranking mechanisms are based on the structure of the underlying network. More recently, we witness applications wherein the ranking mechanisms should take into account not only the structure of the network but also other important aspects of the networks such as the value created by the nodes in the network and the marginal contribution of the nodes in the network. Motivated by this observation, the goal of this tutorial is to provide conceptual understanding of recent advances in designing efficient and scalable ranking mechanisms for large interaction networks along with applications to social network analysis.

1 Introduction

Interaction networks are prevalent in real world applications and they manifest in several forms such as online social networks, collaboration networks, citation networks, technological networks, biological networks, and food webs [3]. In the analysis of interaction networks, it is important to understand the role played by nodes/edges either due to their structural placement in the network or due to their behavioral influence over the others in the network. Towards this end, it is important to rank nodes/edges in a given network based either on their positional power or on their behavioral influence. There exists several well known ranking mechanisms in the literature such as Google page rank and centrality measures (e.g. degree centrality, clustering coefficient and betweenness centrality) in social sciences [3]. However, we note that these traditional methods to rank nodes/edges in the network are based on the structure of the underlying network. More recently, we witness applications wherein the ranking mechanisms should take into account not only the structure of the network but also other important aspects of the networks such as the value created by the nodes in the network and the marginal contribution of the nodes in the network. Motivated by the above, we study the recent advances in the ranking mechanisms for interaction networks along with applications to social network analysis where ever possible. Due to the increasing popularity of Internet and world wide web, we tend to observe large interaction networks in recent times that consist of millions of nodes and edges. Note that the computational complexity of certain traditional ranking mechanisms is prohibitively expensive to deal with large interaction network data sets. For these reasons, the goal of this tutorial is to provide the conceptual underpinnings of designing efficient and scalable ranking mechanisms for large interaction networks.

In particular, the following provides a quick overview of the tutorial:

International Conference on Management of Data
COMAD 2011, Bangalore, India, December 19–21, 2011

- The techniques that we discuss to design the ranking mechanisms are based on
  - graph structure and graph algorithms,
  - game theoretic techniques,
  - spectral techniques, and
  - new trends such as minwise hashing.

  For each of the ranking mechanisms, we first motivate the need for that ranking mechanism, then we formally define the underlying technique for ranking the nodes/edges in the network, and finally we present certain interesting properties of that ranking mechanism.

- The applications that we consider to demonstrate the practical utility of these ranking mechanisms are such as
  - Influence maximization (or viral marketing) in social networks,
  - Graph clustering,
  - Graph sparsification, and
  - Influence attribution in citation networks.

  In this process, we also highlight how to design efficient and scalable ranking mechanisms that suit for massive networks.

2 Need for this Tutorial

Interaction network analytics has received a tremendous attention from the research community due to their ability to model several practical scenarios in a natural way. It is at most important to study the influential capabilities or positional power of the nodes/edges in the interaction networks for a better understanding of such networks and also for successful deployment of applications based on these networks. For such scenarios, ranking mechanisms play a crucial role and designing efficient and scalable ranking approaches is at high priority in the interaction network analytics. The COMAD conference has focused tracks - *social network analysis* and *novel data mining algorithms and foundations* - and the subject matter of this tutorial, we believe, is a perfect fit in these tracks.

3 Content of the Tutorial

Here we present a detailed description of the material and the results that we discuss in this tutorial.

Interaction Networks: A Quick Primer: We first present a brief introduction to interaction networks and then present a quick overview of traditional ranking mechanisms for interaction networks such as Google page rank and centrality measures in social sciences. We next present the limitations these traditional ranking mechanisms.

Ranking Mechanisms for Influence Attribution: In several practical settings of interaction networks, the ranking of nodes needs to take into account two aspects: firstly, their importance in the network structure that arises as a result of their interactions; and secondly, the value generated by the interactions involving them. Towards this end, we first discuss a ranking algorithm based on extension of eigen value methods [4, 5]. We next discuss another ranking mechanism to rank nodes in the network based on techniques from game theory. In particular, the key to this ranking mechanism is the use of Shapley value approach to determine the marginal influence of nodes in the network [7, 8]. We work with citation networks and co-authorship networks as an application this game theoretic approach. We also briefly introduce an axiomatic approach to the ranking systems on networks [2].

Ranking Mechanisms for Graph Sparsification: In recent times, massive interaction networks have become common and there is a lot of interest from research community to design efficient and scalable algorithms for the analysis of interaction networks. For such settings, one clear technique is to sparsify the given massive graph and retain only the *influential* edges for further analysis. In particular, we discuss a ranking mechanism to rank edges to sparsify the given massive networks based on Jaccard measure and Minwise hashing [9]. As an application this ranking mechanism, we study the design of scalable algorithms for graph clustering [9].

Ranking Mechanisms for Influence Maximization: In the analysis of interaction networks, we often encounter situations where we need to rank nodes/edges based their network value. In this context, we first discuss a ranking mechanism to rank edges in networks based on their network value along with an application to influence maximization in social networks [6]. The key to this ranking mechanism is to select the most influential edges in the network in a greedy fashion. We then present random walk style ranking mechanisms for designing efficient and scalable algorithms to determine authoritative nodes in massive social networks [1] and we discuss the efficacy of this ranking mechanism by applying this approach to address influence maximization problem in social networks. We also briefly discuss a ranking mechanism, namely TwitterRank [11], to identify influential users of micro-blogging services such as Twitter.

Conclusions and Future Work: Here we first summarize the contents of the tutorial and then we present a few interesting pointers to future work.

4 Targeted Audience and Expectations

This tutorial is mainly targeted at several kinds of audience such as researchers, graduate students and and industry professionals working in the areas of social networks, Internet, world wide web, and computer science and engineering. We plan the tutorial in such a way that it is self-contained and we do not assume any particular expertise from the audience. At the end of the tutorial, we hope that the attendees will be convinced to the need for designing efficient and scalable ranking mechanisms that suits for modern needs of interaction networks.

We also discuss various challenges and open problems at the end of the tutorial and we expect that this is extremely useful to the researchers. We attempt our best to maintain a striking balance between theoretical concepts and practical importance of the problems in the tutorial. Thus we hope that the practitioners also get benefited from this tutorial.

5 Biographies

Bio: Sameep Mehta is researcher at IBM Research India since 2006. Prior to joining IBM, he finished his PhD from Ohio State University. His current research interests include Data Mining, Business Analytics and Services Science.

Bio: Ramasuri Narayanam is a researcher at IBM Research - India. He received his Ph.D. in Computer Science from Indian Institute of Science (IISc), Bangalore, India in 2011. His research interests include game theory, social networks, mechanism design, and electronic commerce.

Bio: Vinayaka Pandit is a Researcher in the Analytics and Optimization department at IBM Research - India and based in Bangalore. He obtained his PhD in Computer Science from IIT-Delhi. He is primarily interested in design and analysis of algorithms. He is also interested in applying algorithmic insights to solve practical problems in domains like operations research, data mining, and databases.

References

[1] Charu C. Aggarwal, A. Khan, and X. Yan. *On flow authority discovery in social networks.* In the Proceedings of the 11th SIAM International Conference on Data Mining (SDM), pages 722-733, 2011.

[2] A. Altman and M. Tennenholtz. *Axiomatic Foundations for Ranking Systems.* Journal of Artificial Intelligence Research, 31:473-495, 2008.

[3] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World.* Cambridge University Press, Cambridge, UK, 2010.

[4] S. Kameshwaran, S. Mehta, Vinayaka Pandit, G.R. Parija, S. Singh, N. Viswanadham. *Analyses for Service Interaction Networks with Applications to Service Delivery.* In Proceedings of SIAM Data Mining Conference (SDM), pages 377-388, 2009.

[5] S. Kameshwaran, Vinayaka Pandit, S. Mehta, N. Viswanadham, K. Dixit. *Outcome aware ranking in interaction networks.* In Proceedings of 19th ACM International Conference on Information and Knowledge Management (CIKM), pages 229-238, 2010.

[6] M. Mathioudakis, F. Bonchi, C. Castillo, A. Gionis, and A. Ukkonen. *Sparsification of influence networks.* In the Proceedings of 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), pages 529-537, 2011.

[7] P. Papapetrou, A. Gionis, and H. Mannila. *A Shapley value appraoch for influence attribution.* In the Proceedings of ECML-PKDD, pages 549-564, 2011.

[8] Ramasuri Narayanam and Y. Narahari. *A Shapley Value based Approach to Discover Influential Nodes in Social Networks.* IEEE Transactions on Automation Science and Engineering (IEEE TASE), 8(1):130-147, 2011.

[9] V. Satuluri, S. Parthasarathy, and Y. Ruan. *Local graph sparsification for scalable clustering.* In Proceedings of the 2011 International Conference on Management of Data (SIGMOD), pages 721-732, 2011.

[10] S. Wasserman and K. Faust. *Social Network Analysis.* Cambridge University Press, Cambridge, 1994.

[11] J. Weng, E.-P. Lim, J. Jiang, and Q. He. *TwitterRank: Finding topic sensitive influential twitterers.* In Proceedings of the 3rd ACM International Conference on Web Search and Data Mining (WSDM), pages 261-270, 2010.