Query Complex Graph Patterns: Tools and Applications

Hanghang Tong
CUNY City College
tong@cs.ccny.cuny.edu

ABSTRACT

In his world-widely renowned book [6], Nobel laureate Herbert Simon pointed out that it is more *the complexity of the environment*, than the complexity of the individual persons, that determines the complex behavior of humans. The emergence of online social network sites and web 2.0 applications provides a new connected environment/context, where people generate, share and search massive human knowledge; and interact and collaborate with each other to collectively perform some complex tasks.

In this talk, we focus on how to make sense of the collaboration data in the context of graphs/networks. To be specific, we will introduce a suite of tools for querying complex patterns from such graphs. Exemplar questions we aim to answer include (a) what makes a team more successful than others, and how to find the best replacement if one of its team members becomes unavailable? [2, 4]; (b) how to find a group of authors from databases, data mining and bioinformatics and they collaborate with each other in a star-shape? [8, 5]; (c) given a set of querying authors of interest, how to find somebody who initiates the research field these querying authors belong to, and how to summarize and visualize the querying authors? [3, 7, 1]; (d) how to incorporate users' preference into these complex queries [9, 10]. We will also introduce the computational challenges behind these querying tools and how to remedy them.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database applications—*Data mining*

Keywords

Complex pattern, Graph mining, scalability

1. SHORT BIO

Hanghang Tong is currently an assistant professor at Computer Science Department, City College, City University of New York. Before that, he was a research staff member at IBM T.J. Watson Research Center and a Post-doctoral fellow in Carnegie Mellon University. He received his M.Sc and Ph.D. degree from Carnegie Mellon University in 2008 and 2009, both majored in Machine Learning. His research interest is in large scale data mining for graphs and multimedia. He has received several awards, including best paper award in CIKM 2012, best paper award in SDM 2008 and best research paper award in ICDM 2006. He has published over 70 referred articles and more than 20 patents. He has served as a program committee member in top data mining, databases and artificial intelligence venues (e.g., SIGKDD, SIGMOD, AAAI, WWW, CIKM, etc).

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2579029 .

2. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. IIS1017415 Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on. Funding was provided by Defense Advanced Research Projects Agency (DARPA) under Contract Number W911NF-11-C-0200 and W911NF-12-C-0028. The content of the information in this document does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

3. REFERENCES

[1] Leman Akoglu, Jilles Vreeken, Hanghang Tong, Duen Horng Chau, Nikolaj Tatti, and Christos Faloutsos. Mining connection pathways for marked nodes in large graphs. 2013.

[2] Kate Ehrlich and Hanghang Tong. The effect of leader centrality on team performance. In *WIDS*, 2012.

[3] José Fernando Rodrigues Jr., Hanghang Tong, Jia-Yu Pan, Agma J. M. Traina, Caetano Traina Jr., and Christos Faloutsos. Large graph analysis in the gmine system. *IEEE Trans. Knowl. Data Eng.*, 25(1):106–118, 2013.

[4] U. Kang, Spiros Papadimitriou, Jimeng Sun, and Hanghang Tong. Centralities in large networks: Algorithms and observations. In *SDM*, pages 119–130, 2011.

[5] Robert Pienta, Acar Tamersoy, Hanghang Tong, and Duen Horng Chau. Mage: Expressive pattern matching in richly-attributed graphs. 2013.

[6] Herbert A. Simon. *The Sciences of the Artificial (3rd Ed.)*. MIT Press, Cambridge, MA, USA, 1996.

[7] Hanghang Tong and Christos Faloutsos. Center-piece subgraphs: problem definition and fast solutions. In *KDD*, pages 404–413, 2006.

[8] Hanghang Tong, Christos Faloutsos, Brian Gallagher, and Tina Eliassi-Rad. Fast best-effort pattern matching in large attributed graphs. In *KDD*, pages 737–746, 2007.

[9] Hanghang Tong, Huiming Qu, and Hani Jamjoom. Measuring proximity on graphs with side information. In *ICDM*, pages 598–607, 2008.

[10] Hanghang Tong, Huiming Qu, Hani Jamjoom, and Christos Faloutsos. ipog: fast interactive proximity querying on graphs. In *CIKM*, pages 1673–1676, 2009.