SHORT SUMMARY

For many IR applications, one needs to deal with large-scale graphs such as Web graph, user-page bipartite graph, and social network graph. All these graphs are of very large scale and contain rich information. As a result, it is non-trivial to perform efficient and effective mining and learning on them. On one aspect, we need to design scalable algorithms. On another aspect, we also need to develop powerful computational infrastructure to support these algorithms. This tutorial aims at giving a timely introduction to the related works, and provides the audiences with a comprehensive view on the literature.

BIO:

Dr. Bin Gao is a researcher at Microsoft Research Asia. His research interests include online advertising, information retrieval, and large-scale graph learning and mining. He has authored two book chapters, 25 papers in top conferences and journals, and over 20 granted or pending patents. He co-authored the best student paper at SIGIR (2008). He serves as PC for SIGIR (2009~2012), WWW (2011~2012), and senior PC for CIKM (2011). He is a reviewer for TKDE, TIST, PRL, IRJ, etc. He is a tutorial speaker at WWW (2011).

Mr. Taifeng Wang is a researcher at Microsoft Research Asia. He is the author of several quality papers on large-scale graph mining and learning, and he has filed several patents on designing distributed systems for large-scale graph computations. He has been the program committee members for international conferences like SIGIR, and he designed and developed a distributed system supporting large-scale graph mining and learning algorithms in a commercial search engine.

Dr. Tie-Yan Liu is a lead researcher at Microsoft Research Asia. His research interests include learning to rank, large-scale graph learning, and most recently Internet advertising. He has authored two books, about 70 papers, and over 20 granted patents. He co-authored the best student paper for SIGIR (2008) and the most cited paper for JVCIR (2004~2006). He is a PC co-chair of RIAO (2010), a track chair of WWW (2011), an area chair of SIGIR (2008~2011) and AIRS (2009, 2010), and a co-chair of several workshops at SIGIR, ICML, and NIPS. He is an associate editor of ACM TOIS, and an editorial board member of IRJ. He is a keynote speaker at PCM (2010).
and a tutorial speaker at SIGIR (2008, 2010) and WWW (2008, 2009). He is a senior member of the IEEE and a member of the ACM.

ABSTRACT:

For many IR applications, one needs to deal with large-scale graphs. For example, to compute page importance for search, one may need to analyze the link structure of the Web graph; to understand Web user behaviors, one may need to cluster user-page bipartite graphs extracted from search engine logs; and to make recommendations in online community, one may need to conduct data mining on the social network graph. All these graphs are of very large scale and contain rich information. As a result, it is non-trivial to perform efficient and effective mining and learning on them. On one aspect, we need to design scalable algorithms. On another aspect, we also need to develop powerful computational infrastructure to support these algorithms. We observe that in recent years, there are some promising advances in the aforementioned aspects, which can potentially enhance many important IR applications, and greatly advance the state of the art of IR related research. This tutorial aims at giving a timely introduction to these works, and provides the audiences with a comprehensive view on the related literature. We believe many IR researchers would have interest in listening to this tutorial, and we hope that they can be motivated to participate in the research of large-scale graph learning.