**COLLOQUIUM**

Department of Computer Science and Engineering

University of South Carolina

**Effective and Scalable Big Data Computing: Algorithms and Systems**

**Yang Zhou**

Date: **March 6, 2017**

Time: **10:30-11:45am**

Place: **Swearingen 1A03 (Faculty Lounge)**

# Abstract

With continued advances in science and technology, digital data have grown at an astonishing rate in various domains and forms, such as business, geography, health, multimedia, network, text, and web data. Network data are also known as graph data, such as academic collaboration, biological, communication, electrical, social, and transportation networks. Such big graph data have huge potential to reveal hidden insights and promote innovation in many business, science, and engineering domains. The reality is that people are often overwhelmed with the flood of big graph data in terms of size, type, and complexity. In order to help people quickly discover interesting knowledge and make good decisions when faced with big graph data, my research is dedicated to developing a wide spectrum of comprehensive solutions that span algorithms, systems, and applications: (1) big graph data mining and learning algorithms; (2) big graph data processing systems; and (3) domain-specific graph analytics applications.

In this talk, I will introduce problems, challenges, and solutions for collecting, processing, understanding, and learning big graph data with billions of vertices and edges. I will also discuss recent work for how to leverage algorithmic and systemic techniques to alleviate challenging bottlenecks in the development of advanced big graph data analytics tools in terms of both quality and scalability. I will conclude the talk by sketching interesting future directions for big data computing. More details can be found at: <http://www.cc.gatech.edu/~yzhou86/>

**Dr. Yang Zhou** received his Ph.D. degree in computer science at the Georgia Institute of Technology in December 2016. His primary research bridges several areas of big data algorithms and systems, including data mining, parallel and distributed computing, machine learning, database systems, and cloud computing, with a focus on the development of effective and scalable algorithms, systems, and applications that address the challenges of big data. He has also worked with researchers from diverse research fields, such as software engineering, storage systems, web services, and trust management, to build and deploy domain-driven knowledge discovery solutions that improve domain-specific system design, data management, and data analytics in real-world settings.

His research efforts have led to 30 publications with 850 citations in top venues of data mining (SIGKDD, ICDM, TKDD, DMKD), database systems (VLDB), high performance computing (HPDC, SC), networking (JSAC), and software engineering (ISSTA). Some of his research results have been included in reading lists and taught in courses at universities worldwide. He has been selected among the 20 rising stars of the KDD community by Microsoft Academic Search and Microsoft Research Asia in 2016. He has been serving as the reviewer of DMKD, JPDC, Machine Learning, TDSC, TKDD, TOIT, TSC, TWEB, and WWWJ.