U Kang

Computer Science Department
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

Office : 412.268.3070
ukang@cs.cmu.edu
http://www.cs.cmu.edu/~ukang

RESEARCH INTERESTS

Data mining in big graphs, with an emphasis on bridging data mining and systems research for extremely scalable graph analysis with applications on knowledge discovery and anomaly detection. Specifically: distributed mining and managing billion-scale graphs, graph indexing, graph compression, spectral graph analysis, tensor analysis, and anomaly detection in graphs.

EDUCATION

Ph.D. Computer Science, Carnegie Mellon University 2012
Dissertation: Mining Tera-Scale Graphs: Theory, Engineering and Discoveries
Advisor: Professor Christos Faloutsos

M.Sc. Information Technology, Carnegie Mellon University 2009

B.Sc. Computer Science and Engineering, Seoul National University 2003

AWARDS AND HONORS

1. **Best Application Paper Award** at Pacific-Asia Conference on Knowledge Discovery and Data Mining 2011
2. **Open Source Software Award** (silver) at Open Source Software World Challenge 2010
3. **Best Application Paper Award** (runner-up) at IEEE International Conference on Data Mining 2009
4. **JeongSong Foundation Fellowship** 2007-2009

RESEARCH EXPERIENCE

Research Assistant, Carnegie Mellon University 2008-present
Advisor: Professor Christos Faloutsos
Ongoing research on mining massive graphs, and finding patterns and anomalies.

Research Intern, IBM T.J. Watson Research, Hawthorne, NY May-August 2011
Mentors: Dr. Hanghang Tong and Jimeng Sun
Large scale graph mining

Research Intern, IBM T.J. Watson Research, Hawthorne, NY May-August 2010
Mentors: Dr. Spiros Papadimitriou, Jimeng Sun, and Hanghang Tong
Large scale centrality computation and graph management

Research Intern, Microsoft Research, Redmond, WA May-August 2009
Mentors: Dr. Mikhail Bilenko
Axiomatic analysis of similarity functions

Research Intern, Microsoft Research, Redmond, WA May-August 2008
Mentors: Dr. Manuel Reyes Gomez
Attribute clustering of service data

Researcher, Korea Telecom Research Lab, Seoul, Korea January 2004-July 2007
Researched data mining and computer security, especially in finding anomalous data in time series intrusion logs. Designed and developed large scale integrated security management software using C# which is used by more than 100 companies in Korea.

TEACHING EXPERIENCE

GUEST LECTURES

1. "Mining Tera Scale Graphs: Theory, Engineering, and Discoveries" Spring 2012
   Course: Machine Learning with Large Datasets (10-605), Carnegie Mellon University.
   Instructor: Prof. William Cohen.

TEACHING ASSISTANT

1. Database Applications (15-415), Carnegie Mellon University Spring 2012
   Instructor: Prof. Christos Faloutsos. 41 students.
2. Science of the Web (15-396), Carnegie Mellon University Fall 2011
   Instructor: Prof. Luis Von Ahn and Brendan Meeder. 74 students.

PUBLICATIONS

BOOK CHAPTER

1. **U Kang** and Christos Faloutsos, *Mining Tera-scale graphs with 'Pegasus': algorithms and discoveries*, in Large Scale Data Analytics, Springer May 2012 (to appear). Editors: Aris Gkoulalas-Divanis and Abdel Labbi.

REFEREED JOURNALS

J1. **U Kang**, Charalampos E. Tsourakakis, and Christos Faloutsos, *PEGASUS: Mining Peta-Scale Graphs*, Knowledge and Information Systems (KAIS), Springer, 2011.
J2. **U Kang**, Charalampos E. Tsourakakis, Ana Paula Appel, Christos Faloutsos, and Jure Leskovec, *HADI: Mining Radii of Large Graphs*, ACM Transactions on Knowledge Discovery from Data (TKDD), 2011.
J3. **U Kang**, Hanghang Tong, Jimeng Sun, Ching-Yung Lin, and Christos Faloutsos. *GBASE: An Efficient Analysis Platform for Large Graphs*, VLDB Journal, 2012

REFEREED CONFERENCES

C1. Charalampos E. Tsourakakis, **U Kang**, Gary Miller, and Christos Faloutsos, *DOULION: Counting Triangles in Massive Graphs with a Coin*, ACM SIGKDD International Conference On Knowledge Discovery and Data Mining (KDD) 2009, Paris, France.
C2. **U Kang**, Charalampos E. Tsourakakis, and Christos Faloutsos, *PEGASUS: A Peta-Scale Graph Mining System - Implementation and Observations*, IEEE International Conference on Data Mining (ICDM) 2009, Florida, Miami, USA.
Best Application Paper Award (runner-up).
*NSF Student Travel Award***.**

C3. **U Kang**, Charalampos E. Tsourakakis, Ana Paula Appel, Christos Faloutsos, and Jure Leskovec, *Radius Plots for Mining Tera-byte Scale Graphs: Algorithms, Patterns, and Observations*, SIAM International Conference on Data Mining (SDM) 2010, Columbus, Ohio, USA.
NSF Student Travel Award.

C4. **U Kang**, Mary McGlohon, Leman Akoglu, and Christos Faloutsos, *Patterns on the Connected Components of Terabyte-Scale Graphs*, IEEE International Conference on Data Mining (ICDM) 2010, Sydney, Australia.
NSF Student Travel Award.

C5. **U Kang**, Duen Horng Chau, and Christos Faloutsos, *Mining Large Graphs: Algorithms, Inference, and Discoveries*, IEEE International Conference on Data Engineering (ICDE) 2011, Hannover, Germany.

C6. **U Kang**, Spiros Papadimitriou, Jimeng Sun, and Hanghang Tong, *Centralities in Large Networks: Algorithms and Observations*, SIAM International Conference on Data Mining (SDM) 2011, Mesa, Arizona, USA.
SDM 2011 Travel Scholarship.

C7. **U Kang**, Brendan Meeder, and Christos Faloutsos, *Spectral Analysis for Billion-Scale Graphs: Discoveries and Implementation*, Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) 2011, Shenzhen, China.
Best Application Paper Award.

C8. Robson L. F. Cordeiro, Caetano Traina Jr., Agma J. M. Traina, Julio Lopez, **U Kang**, and Christos Faloutsos. *Clustering Very Large Multi-dimensional Datasets with MapReduce*, ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2011, San Diego, CA, USA.

C9. **U Kang**, Hanghang Tong, Jimeng Sun, Ching-Yung Lin, and Christos Faloutsos. *GBASE: A Scalable and General Graph Management System*, ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2011, San Diego, CA, USA.

C10. Danai Koutra, Tai-You Ke, **U Kang**, Duen Horng (Polo) Chau, Hsing-Kuo Kenneth Pao, and Christos Faloutsos. *Unifying Guilt-by-Association Approaches: Theorems and Fast Algorithms*, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD) 2011, Athens, Greece.

C11. **U Kang** and Christos Faloutsos. *Beyond `Caveman Communities': Hubs and Spokes for Graph Compression and Mining*, IEEE International Conference on Data Mining (ICDM) 2011, Vancouver, Canada.

C12. **U Kang**, Duen Horng Chau, and Christos Faloutsos. *PEGASUS: Mining Billion-Scale Graphs in the Cloud*, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2012, Kyoto, Japan

C13. **U Kang**, Hanghang Tong, and Jimeng Sun. *Fast Random Walk Graph Kernel*, SIAM International Conference on Data Mining (SDM) 2012, Anaheim, California, USA.

OTHER PUBLICATIONS

O1. **U Kang**, Charalampos Tsourakakis, Ana Paula Appel, Christos Faloutsos, and Jure Leskovec. *HADI: Fast Diameter Estimation and Mining in Massive Graphs with Hadoop*, CMU Machine Learning Tech Report CMU-ML-08-117, December 2008.

O2. **U Kang**, Duen Horng Chau, and Christos Faloutsos. *Inference of Beliefs on Billion-Scale Graphs*, The 2nd Workshop on Large-scale Data Mining: Theory and Applications (LDMTA) 2010, in conjunction with KDD 2010, Washington D.C., USA

O3. **U Kang**, Mikhail Bilenko, Dengyong Zhou, and Christos Faloutsos.
Axiomatic Analysis of Co-occurrence Similarity Functions,

CMU Computer Science Tech Report CMU-CS-12-102, February 2012.

DEMO

D1. Leman Akoglu*, Duen Horng Chau*, **U Kang***, Danai Koutra*, and Christos Faloutsos. *OPAvion: Mining and visualization in large graphs*, ACM SIGMOD Conference 2012, Scottsdale, AZ, USA.

D2. Leman Akoglu*, Duen Horng Chau*, **U Kang***, Danai Koutra*, and Christos Faloutsos. *OPAvion: Large Graph Mining System for Patterns, Anomalies & Visualization*, Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) 2012, Kuala Lumpur, Malaysia.

PATENTS

UNITED STATES

1. **U Kang**, Spiros Papadimitriou, Jimeng Sun, Ching-Yung Lin, "Determining the Importance of Data Items and Their Characteristics Using Centrality Measures" (filed on April 2011).
2. **U Kang**, Hanghang Tong, Jimeng Sun, Ching-Yung Lin, "Method and System for Managing and Querying Large Graphs" (filed on August 2011).

KOREA

1. **U Kang**, "Virtual Web-Server Based Intrusion Enticement System for Early Detection of Internet Web Attack and Method Thereof", Korean patent number: 10-2004-0074724, issued Sep 17, 2004.
2. Sang-Kon Kim, Yoon-Ho Choi, Jong-Ho Park, Ho-Kun Moon, Myung-Soo Rhee, **U Kang**, Jin Gi Choe and Seung-Woo Seo "Network Security System and Method for Tracing Attacker on the Encrypted Connection Chain", Korean patent number: 10-2005-0021125, issued Mar 14, 2005.
3. Yoon-Ho Choi, Jong-Ho Park, Sang-Kon Kim, Ho-Kun Moon, Myung-Soo Rhee, **U Kang**, Jin Gi Choe and Seung-Woo Seo, "System and Method of Forensics Evidence Collection at the Time of Infringement Occurrence", Korean patent number: 10-2005-0129965, issued Dec 26, 2005.

TALKS

INVITED CONFERENCE SPEAKER

1. *Peta Scale Graph Mining with Pegasus*, Invited workshop speaker, Workshop on High Performance Analytics - Algorithms, Implementations, and Applications, Columbus, OH, Apr. 2010.

INVITED LECTURES

1. *Pegasus: mining peta-scale graph*, Dept. of Computer Science and Engineering, Hanyang University, Seoul, Korea, Nov. 2010.
2. *Mining Billion-Scale Graphs: Algorithms and Discoveries*, Dept. of Electrical Engineering, Seoul National University, Korea, May 2011.
3. *Mining Billion-Scale Graphs: Algorithms and Discoveries*, SAP R&D Center, Korea, May 2011.
4. *Mining Tera-Scale Graphs with MapReduce: Theory, Engineering and Discoveries*, Dept. of Computer Science and Engineering, Hanyang University, Seoul, Korea, Dec. 2011.
5. *Mining Tera-Scale Graphs with MapReduce: Theory, Engineering and Discoveries*, Dept. of Computer Science, KAIST, Daejeon, Korea, Dec. 2011.

TUTORIALS

1. *Managing and Mining Large Graphs: Patterns and Algorithms* (with Christos Faloutsos), SIGMOD 2012, Scottsdale, AZ, May 2012

PROFESSIONAL SERVICE

PROGRAM COMMITTEE MEMBER

1. International Workshop on Finding Patterns of User Behaviors in NEtwork and MObility Data, 2011, in conjunction with ECML/PKDD 2011
2. ICDM 2011 Workshop on Data Mining Technologies for Computational Creative Intelligence, 2011
3. ICML 2012 Workshop on Mining and Learning With Graphs (MLG), 2012

JOURNAL REVIEWS

1. ACM Journal on Experimental Algorithmics
2. ACM Transactions on Knowledge Discovery from Data

CONFERENCE REVIEWS

1. ACM SIGMOD International Conference on Management of Data, 2010
2. ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2010
3. European Conference on Machine Learning and Principles of Knowledge Discovery in Databases(ECML/PKDD), 2010
4. 21st International Conference on World Wide Web (WWW), 2012

REFERENCES

Prof. Christos Faloutsos (advisor)
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213
+1 (412) 268-3505
christos@cs.cmu.edu

Prof. Tom Mitchell
Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA 15213
+1 (412) 268-5196
tom.mitchell@cs.cmu.edu

Prof. Garth Gibson
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213
+1 (412) 268-6645
garth@cs.cmu.edu

Prof. Robert Grossman
Computation Institute
University of Chicago
Chicago, IL 60637
+1 (773) 702-0473
robert.grossman@uchicago.edu