

Sangkeun Lee

• Computer Scientist
Computational Data Analytics Group
Computational Sciences & Engineering Division
Oak Ridge National Laboratory
Phone: 718-877-0486 • E-Mail: lees4@ornl.gov, leesangkeun@gmail.com

Research Interests

- **Large-scale Graph/Data Analysis** – High performance computing for graph analysis, homogeneous & heterogeneous graph analysis (e.g, node ranking, link analysis, clustering, path mining, etc.)
- **Big Data & NoSQL Technology for Graph/Data Analysis**– Scalable ETL (Extract, Transform, Load) and graph construction, heterogeneous data integration
- **Information Retrieval & Recommender Systems** – Collaborative filtering, content-based filtering, multidimensional recommender systems, flexible recommender systems, context-aware recommender systems

Education

- **Ph.D.**, Computer Science and Engineering, Seoul National University, Seoul, Korea, 2005.3-2012.8
Dissertation: “A Heterogeneous Graph-based Framework for Flexible Hybrid Recommender Systems” (Advisor: Sang-goo Lee)
- **B.S.**, Computer Science, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, 2001-2004

Work Experiences

- **Computer Scientist** at Computational Data Analytics Group, Computational Sciences & Engineering Division, Oak Ridge National Laboratory managed by UT-Battelle, LLC (2015.9 – present)
- **Postdoctoral Research Associate** at Computational Data Analytics Group, Computational Sciences & Engineering Division, Oak Ridge National Laboratory managed by UT-Battelle, LLC (2014.3 – 2015.8)
- **Postdoctoral Research Associate** at Biomedical Science and Engineering Center, Oak Ridge National Laboratory (Post-doctoral program managed by ORISE/ORAU) (2013.3-2014.2)
- **Postdoctoral Research Associate** at Institute of Computer Technology, Seoul National University (2012.8- 2013.2)

Academic Experiences

- **External Reviewer** – Expert Systems with Applications – Journal (ISSN: 0957-4174)
- **Program Committee** – The Asia Pacific Web Conference, APWEB 2015, Guangzhou, China
- **Program Committee** – The ACM Conference Series on Recommender Systems, RecSys 2014, Silicon Valley
- **Instructor** – Introduction to IT Venture Creation (2012, fall), Seoul National University

- **Teaching Assistant** – Topics in Database: Context-Aware Personalization (2008, fall), Seoul National University
- **Teaching Assistant** - Seminar on Computer Applications: Semantic Technology (2008, fall), Seoul National University
- **Teaching Assistant** – Database (2008, spring), Seoul National University
- **Teaching Assistant** - Advanced Database (2008, spring), Seoul National University

Recent Experiences in Grant-funded Projects

Pattern Discovery and Predictive Modeling in Massive Heterogeneous Graphs Using Cray's Urika

- (2014.7- Present) Post-doctoral Researcher, funded by ORNL's Lab-directed Research & Development (LDRD) fund: A project on enabling fast and scalable graph analysis on the graph processing specialized supercomputer, *Urika*. I have been implementing a set of well-known graph mining algorithms using *SPARQL* that can be executed on the *Urika* system.

Healthcare (Medicare & Medicaid) Data Mining

- (2013.6- 2014.6) **Post-doctoral Researcher**, funded by The Centers for Medicare and Medicaid Services (CMS) of U.S: A project on health care data processing, management, and mining. I contributed the project mainly focusing on retrieving or detecting redundant or similar providers from the *LinkedCMS* that is a large-scale graph constructed from various heterogeneous Medicare and Medicaid provider databases.

Web Mining for Deriving Geospatial Cancer Mortality Trends in the US

- (2013.3-2013.5) **Post-doctoral Researcher**, funded by National Institutes of Health (NIH) of U.S: A project on utilizing web mining for cancer mortality trend analysis. I developed a reporting system, which automatically generates cancer mortality trend reports by analyzing the text of publically available local newspaper obituaries. I received an *honorable mention poster award* from the 1st ORNL Postdoc Research Symposium.

Large Scale Graph Data Processing for Smart Reality

- (2011- 2013.2) **Researcher**, funded by National Research Foundation of Korea: A project on graph data processing and management. I have developed a heterogeneous graph-based recommender system, which aims to enable flexibility of processing various recommendations with different semantics. Several related publications have been published.

Open-source Software Project

Graph Mining Using SPARQL

- The Resource Description Framework (RDF) and SPARQL Protocol and RDF Query Language (SPARQL) were introduced about a decade ago to enable flexible schema-free data interchange on

the Semantic Web. Today, data scientists use the framework as a scalable graph representation for integrating, querying, exploring and analyzing data sets hosted at different sources. With increasing adoption, the need for graph mining capabilities for the Semantic Web has emerged. We address that need through implementation of popular iterative Graph Mining algorithms (degree distribution, diameter, radius, node eccentricity, triangle count, connected component analysis, and PageRank). We implement these algorithms as SPARQL queries, wrapped within Python scripts. These graph mining algorithms (that have a linear-algebra formulation) can indeed be unleashed on data represented as RDF graphs using the SPARQL query interface.

- Please see our github project page for more details: <https://github.com/ssrangan/gm-sparql>

Other Projects

Context-Aware Recommendation for In-Vehicle Infotainment

- (2010- 2013.2) **Researcher**, funded by Bell Labs in Seoul. Recent efforts to provide context-aware services in smart cars have mainly focused on easing drivers' inconvenience. We did researches on a context-aware recommendation system to provide various kinds of information on head-up displays.

KT Semantic Web Technology Roadmap

- (2007) **Researcher**, funded by a Korean telecommunication company KT. It was a consulting project for a telecommunication company. I mainly contributed the survey and writing consulting reports.
(http://ids.snu.ac.kr/w/images/1/1c/An_Enterprise_Strategy_for_Semantic_Technology_Adoption.pdf)

Prompt Supplier Relationship Management System

- (2006) **Research Assistant**, A project for developing a supplier relationship management system. (Publication: <http://ids.snu.ac.kr/w/images/c/cf/DJ-2007-01.pdf>), I worked for this project as a Java programmer, implementing a production level system.

Intelligent Information System for Government Procurement

- (2005-2007) **Research Assistant**, A project on client/server system development of product search engines. I have developed a module for managing massive catalog index for efficiently processing e-catalogs. (Publication: http://ids.snu.ac.kr/wiki/Massive_Catalog_Index_based_Search_for_e-Catalog_Matching)

Publications

Topics related to recommender systems and graph mining

- Y Kim, S Atchley, GR Valle'e, **S Lee**, GM Shipman, Optimizing End-to-End Big Data Transfers over Terabits Network Infrastructure, IEEE Transactions on Parallel and Distributed Systems (IEEE TPDS 2016) (Accepted)

- S Hong, **S Lee**, SH Lim, SR Sukumar, Comprehensive Evaluation of Graph Pattern Matching in Graph Analysis Systems, High-Performance Parallel and Distributed Computing (HPDC) 2016 (Accepted)
- **Sangkeun Lee**, Sreenivas R. Sukumar, Seokyong Hong, Seung-Hwan Lim, Enabling Graph Mining in RDF Triplestores using SPARQL for Holistic In-situ Graph Analysis, Expert Systems with Applications (SCIE, 5-Year IF: 2.254 , ISSN: 0957-4174)
- Rina Singh, Jeffrey A. Graves , **Sangkeun Lee**, Sreenivas R. Sukumar, Mallikarjun Shankar, Enabling Graph Appliance for Genome Assembly, IEEE Big Data Workshop on Mining Big Data to Improve Clinical Effectiveness in conjunction with IEEE Big Data 2015, Santa Clara, CA, USA
- **Sangkeun Lee**, Sanghyeb Lee, Byung-Hoon Park, PathMining: A Path-based User Profiling Algorithm for Heterogeneous Graph-based Recommender Systems, FLAIRS-RS 2015 : Florida Artificial Intelligence Research Society - RecSys Special Track (FLAIRS-RS 2015) (Accepted)
- **Sangkeun Lee**, Seung-Hwan Lim, and Byung-hoon Park, Table2Graph: A Scalable Graph Construction From Relational Tables using Map-Reduce, The First IEEE International Conference on Big Data Computing Service and Applications (IEEE BigDataService 2015) (Accepted)
- **Sangkeun Lee**, Seung-Hwan Lim, Tyler C. Brown, Sreenivas R. Sukumar, Graph Mining Meets the Semantic Web, 6th International Workshop on Data Engineering meets the Semantic Web In Conjunction with ICDE 2015 (ICDE 2015-DESWEB Workshop)
- Seung-hwan Lim, Sreenivas R. Sukumar, **Sangkeun Lee**, Gautam Ganesh, Tyler C. Brown, Graph Processing Platforms at Scale: Practices and Experiences, IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS 2015) (Accepted)
- **Sangkeun Lee**, Minsuk Kahng, Sang-goo Lee, Constructing Compact and Effective Graphs for Recommender Systems via Node and Edge Aggregations, Expert Systems with Applications (SCIE, 5-Year IF: 2.254 , ISSN: 0957-4174) (Accepted)
- **Sangkeun Lee**, Mallikarjun Shankar, Byung-Hoon Park, Clustering Providers Across Disparate Healthcare Datasets using a Path-based Pseudo Similarity Measure, KDD Workshop on Data Science for Social Good (KDD-DSSG 2014)
- Yeonchan Ahn, Sungchan Park, **Sangkeun Lee**, and Sang-goo Lee, A Heterogeneous Graph-based Recommendation Simulator, 2013, Proceedings of the 5th ACM Conference on Recommender Systems (RecSys 2013)
- **Sangkeun Lee**, Georgia Tourassi, Hong-Jun Yoon, Songhua Xu, David Svitz, Assessing the Utility of Web Mining for Deriving Geospatial Cancer Mortality Trends in the US, 2013, 1st ORNL Postdoc Research Symposium (**Honorable Mention Poster Award**)
- Chunggrim Kim, **Sangkeun Lee**, Sungchan Park, Sang-goo Lee, Influence Maximization Algorithm Using Markov Clustering, 2013, The 4th International Workshop on Social Networks and Social Web Mining (DASFAA 2013)
- Youngki Park, ByoungJu Yang, Sang-il Song, **Sangkeun Lee**, Seungseok Kang, Sang-goo Lee, Context-Aware Recommendation in Smart Cars, 2013, Proceedings of the 6th Biennial Workshop on Digital Signal Processing for In-Vehicle Systems (DSP in Vehicles 2013), Page 64-67

- **Sangkeun Lee**, Sungchan Park, Minsuk Kahng, Sang-goo Lee, PathRank: Ranking Nodes on a Heterogeneous Graph for Flexible Hybrid Recommender Systems, 2012, Expert Systems with Applications (SCIE, 5-Year IF: 2.254, ISSN: 0957-4174)
- **Sangkeun Lee**, Sungchan Park, Minsuk Kahng, Sang-goo Lee, PathRank: A Novel Node Ranking Measure on a Heterogeneous Graph for Recommender Systems, 2012, Proceedings of the Conference on Information and Knowledge Management 2012 (CIKM 2012)
- ByoungJu Yang, **Sangkeun Lee**, Sungchan Park, Sang-goo Lee, Exploiting Various Implicit Feedback for Collaborative Filtering, 2012, Proceedings of the 21th International Conference on World Wide Web 2012 (WWW 2012), Page 639-640
- **Sangkeun Lee**, Sang-goo Lee, A Generic Graph-based Multidimensional Recommendation Framework and Its Implementations, 2012, Proceedings of the 21th International Conference on World Wide Web 2012 (WWW 2012)
- Sang-il Song, **Sangkeun Lee**, Sungchan Park, Sang-goo Lee, Determining User Expertise for Improving Recommendation Performance, 2012, Proceedings of the 6th International Conference on Ubiquitous Information and Communication 2012 (ICUIMC 2012)
- **Sangkeun Lee**, Sang-il Song, Minsuk Kahng, Dongjoo Lee, Sang-goo Lee, Random Walk based Entity Ranking on Graph for Multidimensional Recommendation, 2011, Proceedings of the 5th ACM Conference on Recommender Systems (RecSys 2011), Page 93-100
- Minsuk Kahng, **Sangkeun Lee**, Sang-goo Lee, Ranking Objects by Following Paths in Entity-Relationship Graphs, 2011, Proceedings of the 4th Workshop for Ph.D. Students in Information and Knowledge Management (PIKM 2011 in conjunction with CIKM 2011), Page 11-18 (**Best Paper Award**)
- Sung Eun Park, **Sangkeun Lee**, Sang-goo Lee, Session-based Collaborative Filtering for Predicting the Next Song, 2011, Proceedings of the First ACIS/JNU International Conference on Computers, Networks, Systems, and Industrial Engineering (CNSI 2011), Page 353-358
- Minsuk Kahng, **Sangkeun Lee**, Sang-goo Lee, Ranking in Context-Aware Recommender Systems, 2011, Proceedings of the 20th International Conference Companion on World Wide Web (WWW 2011), Page 65-66
- **Sangkeun Lee**, Minsuk Kahng, Sang-goo Lee, Flexible Recommendation Using Random Walks on Implicit Feedback Graph, 2011, Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication (ICUIMC 2011)
- Dongjoo Lee, Sung Eun Park, Minsuk Kahng, **Sangkeun Lee**, Sang-goo Lee, Exploiting Contextual Information from Event Logs for Personalized Recommendation, 2010, Computer and Information Science 2010, Studies in Computational Intelligence, Volume 317

Topics related to lifelog, context-aware systems

- **Sangkeun Lee**, Juno Chang, Sang-goo Lee, Survey and Trend Analysis of Context-Aware Systems, 2011, Information-An International Interdisciplinary Journal, Volume 14(2), Page 527-548, SCIE
- **Sangkeun Lee**, Gihyun Gong, Inbeom Hwang, Sang-goo Lee, LifeLogOn: A Practical Lifelog System for Building and Exploiting Lifelog Ontology, 2010, Proceedings of 2010 IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC

2010) and 2010 IEEE International Workshop on Ubiquitous and Mobile Computing (UMC 2010), Page 367-373

- **Sangkeun Lee**, Gihyun Gong, Sang-goo Lee, Entity-Event Lifelog Ontology Model (EELOM) for Lifelog Ontology Schema Definition, 2010, Proceedings of the 12th Asia-Pacific Web Conference (APWeb 2010), Page 344-346
- **Sangkeun Lee**, Gihyun Gong, Sang-goo Lee, LifeLogOn: Log on to Your Lifelog Ontology!, 2009, Proceedings of the 8th International Semantic Web Conference (ISWC 2009)
- **Sangkeun Lee**, Sungchan Park, Sang-goo Lee, A Study on Issues in Context-Aware Systems Based on a Survey and Service Scenarios, 10th ACIS International Conference on Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing, 2009
- **Sangkeun Lee**, Dongjoo Lee, Seungseok Kang, Sang-goo Lee, A Pragmatic Approach to Realizing Context-aware Personal Services, 2008, ECBS Workshop in conjunction with IAT'08

Other topics

- Seungseok Kang, Jung-Yeon Yang, **Sangkeun Lee**, Gihyun Gong, Jaeseok Myung, Sungchan Park, Sang-goo Lee, An Enterprise Strategy for Semantic Technology Adoption, 2008, The 5th International Conference on Information Technology and Applications
- Jae-won Lee, Taehee Lee, **Sangkeun Lee**, Ok-ran Jeong, Sang-goo Lee, Massive Catalog Index based Search for e-Catalog Matching, 2007, E-Commerce Technology and the 4th IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services, 2007. CEC/EEE 2007. The 9th IEEE International Conference, Page 341-348
- Dongjoo Lee, Seungseok Kang, **Sangkeun Lee**, Young-gon Kim, Sang-goo Lee, BestChoice SRM: A Simple and Practical Supplier Relationship Management System for e-Procurement, 2007, Proceedings of the 3rd international workshop on Data engineering issues in E-commerce and services: In conjunction with ACM Conference on Electronic Commerce (EC '07), Volume 236, Page 5-11

International Conference Participations

- CIKM 2012 in Maui, Hawaii
- WWW 2012 in Lyon, France
- RecSys 2011 in Chicago, USA
- WWW 2011 in Hyderabad, India
- APWEB 2010 in Busan, Korea
- ISWC 2009 in Chantilly, USA
- WI-IAT 2008 in Sydney, Australia

Fellowships, Honors and Awards

- **Honorable Mention Poster Award** at The 1st ORNL Annual Postdoctoral Research Symposium 2013
- **Best Paper Award** at PIKM 2011 in conjunction with CIKM 2011

- 2005-2011 **Brain Korea 21 Fellowship**, Korea Ministry of Education and Human Resources Development
- 2008-2012 **Samsung Electronics Fellowship**, Samsung Electronics
- 2008 **Seoul National University Foundation Scholarship**

Languages

- English (Full professional proficiency)
- Korean (Native or bilingual proficiency)

Skills

- Languages & Frameworks: Python, Java, C++, C, SQL, JSP
- Database & Tools: RDBMS (Oracle, MySQL, etc.), NoSQL DB (MongoDB, Neo4J), Big Data Analysis (Hive, Map-Reduce)

References

- Dr. Sang-goo Lee, Professor, School of Computer Science and Engineering, Seoul National University, Korea, +82-2-880-5517, sglee@snu.ac.kr
- Dr. Han-joon Kim, Associate Professor, School of Electrical and Computer Engineering, University of Seoul, Korea, +82-2-2210-5632, khj@uos.ac.kr
- Dr. Byung H. Park, Research Staff Member, Computer Science and Mathematics Division, Oak Ridge National Laboratory, parkbh@ornl.gov