

# JialuLiu

70 Christopher Columbus Drive  
Apt 4005  
Jersey City, NJ, 07302

Cell: (+1) 217-979 7852  
E-mail: [jialu@google.com](mailto:jialu@google.com)  
WWW: [jialu.info](http://jialu.info)

## Education

- 2011 - 2016      **Ph.D.**      University of Illinois at Urbana-Champaign  
Majoring in Computer Science, Specialization in Data Mining
- 2007 - 2011      **B.Eng.**      Zhejiang University  
Majoring in Computer Science

## Research Interests

Data Mining and Machine Learning on networked and textual data, especially with limited supervision.

## Professional Experience

- 2016.5 -      **Structured Data@Google Research, NYC**      Research Assistant  
Enrich Google search and news with web-scale structured data
- 2011.9 - 2016.5      **Data Mining Research Group at UIUC**      Research Assistant  
Thesis: Constructing and Modeling Text-Rich Information Networks: A Phrase Mining-based Approach      *Advisor: Jiawei Han*
- Proposed an embedding method for text-rich information networks. (ICDM16)
  - Modelled document representations via latent keyphrase inference. (WWW16)
  - Designed a phrase mining framework from massive text corpora. (SIGMOD15)
  - Proposed a generative clustering algorithm for information networks. (SDM15)
  - Built an entity disambiguation system for information networks. (KDDCup13)
  - Proposed a spectral clustering algorithm for information networks. (IJCAI13)
  - Proposed a multi-view clustering algorithm for information networks. (SDM13)
- 2015.6 - 2015.8      **Structured Data@Google Research, NYC**      Research Intern  
Enrich Google search results by showing site-specific exploratory linkable entities relevant to the (query, page/entity) pair in a named carousel, so that users can further explore relevant pages on the site more conveniently within the search results.      *Host: Cong Yu*
- 2014.5 - 2014.8      **Rephil@Google Los Angeles**      Eng Intern  
Designed and implemented a robust framework for adding trending topics into the incremental Rephil model, showing extraordinary quality improvements. Rephil is a topic modeling system used widely within Google, probably the largest in the world in terms of model scale.      *Host: Don Metzler*

- 2013.5 - 2013.8 **DSA@Google Mountain View** Eng Intern  
 Designed and implemented a module for detecting main product prices in web pages; a module for detecting near duplicate web pages selling same main products.  
*Host: Xiaoxin Yin*
- 2012.5 - 2012.8 **IBM T.J. Watson Research Center** Research Intern  
 For community detection problem on subnetworks of large social or information networks, we suggest to incorporate knowledge from the whole network via iteratively querying the neighborhood.  
*Host: Charu Aggarwal*
- 2009.8 - 2011.5 **State Key Lab of CAD & CG at Zhejiang University** Research Assistant  
 Proposed to consider intrinsic geometry of conditional probability distribution along the geodesics of data manifold by incorporating a smoothing regularizer into the objective of the traditional Gaussian Mixture Model. *Advisor: Deng Cai*

## Publications

### book and chapter

**Jialu Liu**, Jingbo Shang and Jiawei Han, "Phrase Mining from Massive Text and Its Applications", (Book) in Synthesis Lectures on Data Mining and Knowledge Discovery, Morgan & Claypool Publishers, March 2017.

**Jialu Liu** and Jiawei Han, "Spectral Clustering", (Book Chapter) in Charu C. Aggarwal and Chandan K. Reddy (eds.) Data Clustering: Algorithms and Applications., CRC Press, 2013, pp.177-200.

### journal

Huan Gui, **Jialu Liu**, Fangbo Tao, Meng Jiang, Brandon Norick, Lance Kaplan and Jiawei Han, "Embedding Learning with Events in Heterogeneous Information Networks", 2017. (under review)

Chi Wang, **Jialu Liu**, Nihit Desai, Marina Danilevsky and Jiawei Han, "Constructing Topical Hierarchies in Heterogeneous Information Networks", Knowledge and Information Systems (KAIS), 44(3), pp.529-558, Springer London, 2014.

### conference/proceedings

Jingbo Shang, **Jialu Liu**, Meng Jiang, Xiang Ren, Clare Voss and Jiawei Han, "Automated Phrase Mining from Massive Text Corpora", arXiv preprint arXiv:1702.04457, 2017. (under review)

Jingbo Shang, Meng Qu, **Jialu Liu**, Lance M. Kaplan, Jiawei Han and Jian Peng, "Meta-Path Guided Embedding for Similarity Search in Large-Scale Heterogeneous Information Networks", arXiv preprint arXiv:1610.09769, 2017. (under review)

Fernando Chirigati, **Jialu Liu**, Flip Korn, You Wu, Cong Yu and Hao Zhang, "Knowledge Exploration using Tables on the Web", Proc. 2017 Int. Conf. on Very Large Data Bases (VLDB'17), 10(3), pp.193-204, Munich, Germany, Aug 2017.

Huan Gui, **Jialu Liu**, Fangbo Tao, Meng Jiang, Brandon Norick and Jiawei Han, "Large-Scale Embedding Learning in Heterogeneous Event Data", Proc. of 2016 Int. Conf. on Data Mining (ICDM'16), pp.907-912, Barcelona, Spain, Dec 2016.

**Jialu Liu**, Xiang Ren, Jingbo Shang, Taylor Cassidy, Clare Voss and Jiawei Han, “Representing Documents via Latent Keyphrase Inference”, Proc. of the 25th Int. Conf. on World Wide Web (WWW’16), pp.1057-1067, Montreal, Canada, April 2016.

**Jialu Liu**, Jingbo Shang, Chi Wang, Xiang Ren and Jiawei Han, “Mining Quality Phrases from Massive Text Corpora”, Proc. of 2015 ACM SIGMOD Int. Conf. on Management of Data (SIGMOD’15), pp.1729-1744, Melbourne, Australia, May 2015.

**Jialu Liu**, Chi Wang, Jing Gao, Quanquan Gu, Charu Aggarwal, Lance Kaplan and Jiawei Han, “GIN: A Clustering Model for Capturing Dual Heterogeneity in Networked Data”, Proc. of 2015 SIAM Int. Conf. on Data Mining (SDM’15), pp.388-396, Vancouver, Canada, Apr. 2015.

**Jialu Liu**, Charu Aggarwal and Jiawei Han, “On Integrating Network and Community Discovery”, Proc. of 2015 Int. Conf. on Web Search and Data Mining (WSDM’15), pp.117-126, Shanghai, China, Feb. 2015.

Xiang Ren, **Jialu Liu**, Xiao Yu, Urvashi Khandelwal, Quanquan Gu, Lidan Wang and Jiawei Han, “ClusCite: Effective Citation Recommendation by Information Network-Based Clustering”, Proc. of 2014 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’14), pp.821-830, New York, NY, August 2014.

Chi Wang, Marina Danilevsky, **Jialu Liu**, Nihit Desai, Heng Ji and Jiawei Han, “Constructing Topical Hierarchies in Heterogeneous Information Networks”, Proc. of 2013 IEEE Int. Conf. on Data Mining (ICDM’13), pp.767-776, Dallas, TX, December 2013.

Quanquan Gu, Charu Aggarwal, **Jialu Liu**, and Jiawei Han, “Selective Sampling on Graphs for Classification”, Proc. of 2013 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’13), pp.131-139, Chicago, IL, August 2013.

**Jialu Liu**, Chi Wang, Marina Danilevsky and Jiawei Han, “Large-Scale Spectral Clustering on Graphs”, Proc. of 2013 Int. Joint Conf. on Artificial Intelligence (IJCAI’13), pp.1486-1492, Beijing, China, August 2013.

**Jialu Liu**, Chi Wang, Jing Gao and Jiawei Han, “Multi-View Clustering via Joint Nonnegative Matrix Factorization”, Proc. of 2013 SIAM Data Mining Conf., pp.252-260, (SDM’13), Austin, TX, May 2013.

Chi Wang, Hongning Wang, **Jialu Liu**, Ming Ji, Lu Su, Yuguo Chen and Jiawei Han, “On the Detectability of Node Grouping in Networks”, Proc. of 2013 SIAM Data Mining Conf., (SDM’13), pp.713-721, Austin, TX, May 2013.

**Jialu Liu**, Deng Cai and Xiaofei He, “Gaussian Mixture Model with Local Consistency”, Proc. of the Twenty-Fourth AAAI Conf. on Artificial Intelligence (AAAI’10), pp.512-517, Atlanta, GA, July 2010.

## workshop paper

**Jialu Liu**, Chi Wang, Jeffrey Yufei Liu, Kin Hou Lei and Jiawei Han, “Ranking-Based Name Matching for Author Disambiguation in Bibliographic Data”, Proc. of KDD Cup 2013 Workshop (KDDCUP’13), Chicago, IL, August 2013.

**Jialu Liu** and Jiawei Han, “HINMF: A Matrix Factorization Method for Clustering in Heterogeneous Information Networks”, Proc. of 2013 Int. Joint Conf. on Artificial Intelligence Workshop (IJCAI-HINA’13), Beijing, China, August 2013.

## demo paper

Fangbo Tao, George Brova, Jiawei Han, Heng Ji, Chi Wang, Brandon Norick, Ahmed El-Kishky, **Jialu Liu**,

Xiang Ren and Yizhou Sun, “NewsNetExplorer: Automatic Construction and Exploration of News Information Networks”, Proc. of 2014 ACM SIGMOD Int. Conf. on Management of Data (SIGMOD’14), pp.1091-1094, Snowbird, UT, June 2014.

Fangbo Tao, Kin Hou Lei, Jiawei Han, ChengXiang Zhai, Xiao Cheng, Marina Danilevsky, Nihit Desai, Bolin Ding, Jing Ge, Heng Ji, Rucha Kanade, Anne Kao, Qi Li, Yanen Li, Cindy Xide Lin, **Jialu Liu**, Nikunj Oza, Ashok Srivastava, Rod Tjoelker, Chi Wang, Duo Zhang, and Bo Zhao, “EventCube: Multi-Dimensional Search and Mining of Structured and Text Data”, Proc. of 2013 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’13), pp.1494-1497, Chicago, IL, August 2013.

## Awards

Grand prize of Yelp Dataset Challenge, 2015

SDM paper invited to SADM special issue “Best of SDM 2015”, 2015

Yahoo!-DAIS Research Excellence, 2015

Data Mining Research Award, Data Mining Group, UIUC, 2014

2<sup>nd</sup> place of KDD Cup: Author Disambiguation Challenge, 2013

Wing Kai Cheng Fellowship in Computer Science, UIUC, 2011

1<sup>st</sup> place of National Data Mining Contest for College Student in China, 2011

Scholarship for Outstanding Students, Zhejiang University, 2008 to 2010

1<sup>st</sup> class Award for Research and Innovation, Zhejiang University, 2010

Award for Excellent Student, Zhejiang University, 2008 and 2009