DEEPAYAN CHAKRABARTI

2110 Speedway Stop B6500, CBA 6.462, Univ. of Texas, Austin TX 78712 (512) 232-5716; *deepay@utexas.edu*

Personal Information

- Work address: 2110 Speedway Stop B6500, CBA 6.462, Univ. of Texas, Austin TX 78712
- Email: deepay@utexas.edu; Office phone: (512) 232-5716
- Website: https://faculty.mccombs.utexas.edu/deepayan.chakrabarti

Professional Preparation

Current Employment

- Associate Professor, IROM, McCombs School of Business (since 09/2022)
- Amazon Scholar (since 11/2022)

Previous Employment

- Assistant Professor, IROM, McCombs School of Business (08/2014 to 09/2022)
- Research Scientist, Facebook Inc. (Ads and Data Science) (05/2012 to 08/2014)
- Senior Research Scientist, Yahoo! Research (08/2005 to 05/2012)

Education

- Ph.D. in Computational and Statistical Learning, (08/2002 to 06/2005)
  School of Computer Science, CMU, Pittsburgh PA.
  Thesis Title: Tools for Large Graph Mining (*Advisor: Dr. Christos Faloutsos*)
- M.S. in Knowledge Discovery and Data Mining, (09/2000 to 06/2002)
  School of Computer Science, CMU, Pittsburgh PA.
- B.Tech. in Computer Science and Engineering, (07/1996 to 05/2000)
  Indian Institute of Technology (IIT), Kanpur, India.

Research Experience

I have worked on a broad range of problems in large-scale graph mining, machine learning, and optimization. My work has applications in social network analytics, financial portfolio construction, computational advertising, and related fields.

Large-scale Graph Mining: We developed new algorithms and proof techniques to identify communities of people from the structure of their social network. We also used the network to find topical experts, "auto-complete" partial user profiles, and infer people's interests. In other work, we analyzed dynamic networks, calculated "epidemic thresholds" for viral contagion, and designed widely-used network generators.

Optimization: We developed new robust optimization methods for portfolio optimization under limited data. For instance, we may want to construct a portfolio of assets with a high reward-to-risk ratio. Since financial conditions change quickly, only the most recent data about these assets is relevant. Errors in this limited data can impact portfolio quality. Our robust methods work well even in such conditions, both provably and empirically.

Machine Learning: One stream of my work combines network analysis with traditional machine learning. For example, we merged network-based classification algorithms with deep learning. As another example, we constructed real-time storylines of ongoing events from tweets. Other work includes predicting rates for rare events that occur in computational advertising, segmenting webpages, web content analysis, new algorithms for multi-armed bandits, and optimal web traffic control.

Refereed Journal Papers

1. D. Chakrabarti: *Robust Linear Classification from Limited Training Data*, in the Machine Learning Journal, 111(5), 2022.
2. D. Chakrabarti: *Parameter-free Robust Optimization for the Maximum-Sharpe Portfolio Problem*, in the European Journal of Operational Research (EJOR), 293(1), 2021.
3. X. Mao, P. Sarkar, and D. Chakrabarti: *Estimating Mixed Memberships with Sharp Eigenvector Deviations*, in the Journal of the American Statistical Association (JASA), 116(536), 2021.
4. L. Zhao, D. Chakrabarti, and K. Muthuraman: *Portfolio Construction by Mitigating Error Amplification: The Bounded-Noise Portfolio*, in Operations Research, 67(4), 2019.
5. D. Chakrabarti: *Modeling Node Incentives in Directed Networks*, in the Annals of Applied Statistics, 11(4), 2017.
6. D. Chakrabarti, S. Funiak, J. Chang, and S. A. Macskassy: *Joint Label Inference in Networks*, in JMLR, 18(59), 2017.
7. D. Chakrabarti, and E. Vee: *Traffic Shaping to Optimize Ad Delivery*, in ACM Transactions on Economics and Computation, 3(2), 2015.
8. P. Sarkar, D. Chakrabarti, and M. Jordan: *Nonparametric Link Prediction in Large Scale Dynamic Networks*, in the Electronic Journal of Statistics, 8(2), 2014.
9. B. A. Prakash, D. Chakrabarti, N. Valler, M. Faloutsos, and C. Faloutsos: *Threshold Conditions for Arbitrary Cascade Models on Arbitrary Networks*, in Knowledge and Information Systems, 33(3), 2012.
10. J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani: *Kronecker Graphs: An Approach to Modeling Networks*, in JMLR, 11, 2010.
11. D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos: *Epidemic Thresholds in Real Networks*, in ACM TISSEC, 10(4), 2008.
12. D. Chakrabarti, C. Faloutsos and Y. Zhan: *Visualization of Large Networks with Min-cut Plots, A-plots and R-MAT*, in the Intl. Journal of Human-Computer Studies, $65(5), 2007$.
13. D. Chakrabarti and C. Faloutsos: *Graph Mining: Laws, Generators and Algorithms*, in ACM Computing Surveys, 38(1), 2006.
14. J. Sun, H. Qu, D. Chakrabarti, and C. Faloutsos: *Relevance search and anomaly detection in bipartite graphs*, in SIGKDD Explorations, 7(2), 2005.
15. S. Thrun, C. Martin, Y. Liu, D. Hahnel, R. Emery-Montemerlo, D. Chakrabarti, and W. Burgard: *A Real-Time Expectation Maximization Algorithm for Acquiring Multi-Planar Maps of Indoor Environments with Mobile Robots*, in IEEE Transactions on Robotics and Automation, 20(3), 2003.

Refereed Conference Papers

1. D. Chakrabarti: *SURE: Robust, Explainable, and Fair Classification without Sensitive Attributes*, in KDD 2023.
2. D. Chakrabarti: *Avoiding Biases due to Similarity Assumptions in Node Embeddings*, in KDD 2022.
3. D. Chakrabarti, and B. Fauber: *Robust High-Dimensional Classification From Few Positive Examples*, in IJCAI 2022 (long presentation).
4. X. Mao, D. Chakrabarti, and P. Sarkar: *Consistent Nonparametric Methods for Network Assisted Covariate Estimation*, in ICML 2021.
5. X. Mao, P. Sarkar, and D. Chakrabarti: *Overlapping Clustering, and One (class) SVM to Bind Them All*, in NIPS 2018 (spotlight presentation).
6. A. Pal, and D. Chakrabarti: *Label Propagation with Neural Networks*, in CIKM 2018.
7. C. Yang, D. Chakrabarti, A. Agarwal, and P. Konana: *Does Market Respond to Information in News Articles beyond Sentiments?*, in the Conference on Information Systems and Technology (CIST) 2018.
8. D. Chakrabarti, S. Funiak, J. Chang, and S. A. Macskassy: *Joint Label Inference in Networks*, invited to the WWW Journal Track, 2018.

9. X. Mao, P. Sarkar, and D. Chakrabarti: *On Mixed Memberships and Symmetric Nonnegative Matrix Factorizations*, in ICML 2017.
10. A. Pal, A. Herdagdelen, S. Chatterji, S. Taank, and D. Chakrabarti: *Discovery of Topical Authorities in Instagram*, in WWW 2016.
11. P. Sarkar, D. Chakrabarti, and P. Bickel: *The Consistency of Common Neighbors for Link Prediction in Stochastic Blockmodels*, in NIPS 2015.
12. D. Chakrabarti, S. Funiak, J. Chang, and S. Macskassy: *Joint Inference of Multiple Label Types in Large Networks*, in ICML 2014.
13. D. Chakrabarti, and R. Herbrich: *Speeding up Large-Scale Learning with a Social Prior*, in KDD 2013.
14. P. Sarkar, D. Chakrabarti, and M. Jordan: *Nonparametric Link Prediction in Dynamic Networks*, in ICML 2012; extended version published in the Electronic Journal of Statistics.
15. D. Chakrabarti, and E. Vee: *Traffic Shaping to Optimize Ad Delivery*, in EC 2012; invited to ACM Transactions on Economics and Computation.
16. B. Aditya Prakash, D. Chakrabarti, M. Faloutsos, N. Valler, and C. Faloutsos: *Threshold Conditions for Arbitrary Cascade Models on Arbitrary Networks*, in ICDM 2011; invited to KAIS Journal Special Issue (ICDM best papers).
17. A. Vattani, M. Gurevich, and D. Chakrabarti: *Preserving Pairwise Relationships in Subgraphs*, in ICML 2011.
18. D. Chakrabarti, and K. Punera: *Event Summarization using Tweets*, in ICWSM 2011.
19. P. Sarkar, D. Chakrabarti, and A. W. Moore: *Theoretical Justification of Popular Link Prediction Heuristics*, in COLT 2010 (best student paper) and invited to IJCAI 2011 (best paper track).
20. D. Chakrabarti, and R. Mehta: *The Paths More Taken: Matching DOM Trees to Search Logs for Accurate Webpage Clustering*, in WWW 2010.
21. D. Chakrabarti, R. Kumar, and K. Punera: *Quicklink Selection for Navigational Query Results*, in WWW 2009.
22. X. Wang, D. Chakrabarti, and K. Punera: *Mining Broad Latent Query Aspects from Search Sessions*, in KDD 2009.
23. A. P. Appel, D. Chakrabarti, C. Faloutsos, R. Kumar, J. Leskovec, and A. Tomkins: *ShatterPlots: Fast Tools for Mining Large Graphs*, in SDM 2009.
24. D. Chakrabarti, R. Kumar, F. Radlinski, and E. Upfal: *Mortal Multi-Armed Bandits*, in NIPS 2008.
25. D. Chakrabarti, R. Kumar, and K. Punera: *Generating Succinct Titles for Web URLs*, in KDD 2008.
26. D. Chakrabarti, R. Kumar, and K. Punera: *A Graph-Theoretic Approach to Webpage Segmentation*, in WWW 2008.
27. D. Chakrabarti, D. Agarwal, and V. Josifovski: *Contextual Advertising by Combining Relevance with Click Feedback*, in WWW 2008.
28. D. Agarwal, A. Broder, D. Chakrabarti, D. Diklic, V. Josifovski, and M. Sayyadian: *Estimating Rates of Rare Events at Multiple Resolutions*, in KDD 2007.
29. S. Pandey, D. Chakrabarti, and D. Agarwal: *Multi-armed Bandit Problems with Dependent Arms*, in ICML 2007.
30. D. Chakrabarti, R. Kumar, and K. Punera: *Page-level Template Detection via Isotonic Smoothing*, in WWW 2007.
31. S. Pandey, D. Agarwal, D. Chakrabarti, and V. Josifovski: *Bandits for Taxonomies: A Model-based Approach*, in SDM 2007.
32. J. Leskovec, D. Chakrabarti, C. Faloutsos, S. Madden, C. Guestrin and M. Faloutsos: *Information Survival Threshold in Sensor and P2P Networks*, in IEEE INFOCOM 2007.
33. D. Chakrabarti, R. Kumar and A. Tomkins: *Evolutionary Clustering*, in KDD 2006 (Google Scholar Classic Paper).

34. J. Sun, H. Qu, D. Chakrabarti and C. Faloutsos: *Neighborhood Formation and Anomaly Detection in Bipartite Graphs*, in ICDM 2005.
35. J. Leskovec, D. Chakrabarti, J. Kleinberg and C. Faloutsos: *Realistic, Mathematically Tractable Graph Generation and Evolution, Using Kronecker Multiplication*, in PKDD 2005.
36. D. Chakrabarti: *AutoPart: Parameter-Free Graph Partitioning and Outlier Detection*, in PKDD 2004.
37. D. Chakrabarti, S. Papadimitriou, D. Modha and C. Faloutsos: *Fully Automatic Cross-Associations*, in KDD 2004; also a CMU Tech Report.
38. D. Chakrabarti, Y. Zhan, C. Faloutsos: *R-MAT: A Recursive Model for Graph Mining*, in SDM 2004 (basis for the Graph500 supercomputing benchmark).
39. D. Chakrabarti, Y. Zhan, D. Blandford, C. Faloutsos and G. Blelloch: *NetMine: New Mining Tools for Large Graphs*, in SDM 2004 Workshop on link analysis, counter-terrorism and privacy.
40. Y. Wang, D. Chakrabarti, C. Wang and C. Faloutsos: *Epidemic Spreading in Real Networks: An Eigenvalue Viewpoint*, in SRDS 2003.
41. D. Chakrabarti and C. Faloutsos: *F4: Large Scale Automated Forecasting using Fractals*, in CIKM 2002; also a CMU Tech Report.
42. Y. Liu, R. Emery, D. Chakrabarti, W. Burgard and S. Thrun: *Using EM to Learn 3D Models of Indoor Environments with Mobile Robots*, in ICML 2001.
43. S. Thrun, W. Burgard, D. Chakrabarti, R. Emery, and Y. Liu: *A Method for Acquiring Multi-Planar Volumetric Models with Mobile Robots based on the EM Algorithm*, in ISRR 2001.

Thesis

Title: Tools for Large Graph Mining
Advisor: Dr. Christos Faloutsos
Institution: School of Computer Science, Carnegie Mellon University
Date: June, 2005

Books

1. D. Chakrabarti, and C. Faloutsos: *Graph Mining: Laws, Tools, and Case Studies*, (Synthesis Lectures on Data Mining and Knowledge Discovery), published by Morgan Claypool in 2012.

Book Chapters

1. D. Chakrabarti: *Graph Mining*, in *Encyclopedia of Machine Learning*, 2010.
2. D. Chakrabarti, C. Faloutsos, and M. McGlohon: *Graph Mining: Laws and Generators*, in *Managing and Mining Graph Data*, 2010.
3. D. Chakrabarti and C. Faloutsos: *Graph Patterns and the R-MAT Generator*, in *Mining Graph Data*, editors L. Holder and D. Cook, published by Wiley in 2006.

Patents

1. D. Chakrabarti: *Selecting users to receive a recommendation to establish connection to an object in a social networking system*, Patent Number 10210458, issued February, 2019.
2. D. Chakrabarti, D. K. Agarwal, and V. Josifovski: *Method and Apparatus for Web Ad Matching*, Patent Number 9824124, issued November, 2017.
3. D. Chakrabarti, S. A. Macskassy, S. Funiak, and J. Chang: *Label inference in a social network*, Patent number 9552613, issued January 24, 2017.
4. D. Chakrabarti, J. Chang, A. L. K. Ching, and M. Kabiljo: *Striping of directed graphs and nodes with improved functionality*, Patent number 9330199, issued May 3, 2016.

5. K. Punera, D. Chakrabarti, and X. Wang: *Mining broad hidden query aspects from user search sessions*, Patent number 9305051, issued April 5, 2016.
6. D. Chakrabarti, S. Ravikumar, and A. Tomkins: *System and method for evolutionary clustering of sequential data sets*, Patent number 8930365, issued January 6, 2015.
7. L. Duan, F. Li, S. Vadrevu, E. Velipasaoglu, S. Hajela, and D. Chakrabarti: *Automatic classification of segmented portions of web pages*, Patent number 8849725, issued September 30, 2014.
8. D. Chakrabarti, D. Agarwal, and V. Josifovski: *Method, apparatus and computer readable medium for indexing advertisements to combine relevance with consumer click feedback*, Patent number 8725752, issued May 13, 2014.
9. K. Punera, and D. Chakrabarti: *Method for Summarizing Event-Related Texts To Answer Search Queries*, Patent number 8666916, issued June 19, 2014.
10. S. Ravikumar, D. Chakrabarti, and K. Punera: *Generating succinct titles for web URLs*, Patent number 8346754, issued January 1, 2013.
11. D. Chakrabarti, M. Mital, S. Hajela, and E. Velipasaoglu: *Automatic visual segmentation of webpages,* Patent number 8255793, issued August 28, 2012.
12. D. Chakrabarti, K. Punera, and S. Ravikumar: *System and method for detecting a web page template*, Patent number 7987417, issued July 26, 2011.
13. S. Ravikumar, D. Chakrabarti, and K. Punera: *Method for segmenting webpages by parsing webpages into document object modules (DOMs) and creating weighted graphs*, Patent number 7974934, issued July 5, 2011.
14. D. Agarwal, D. Diklic, D. Chakrabarti, A. Broder, and V. Josifovski: *System and method for determining impression volumes of content items in a taxonomy hierarchy*, Patent number 7921073, issued April 5, 2011.
15. D. Chakrabarti, K. Punera, and S. Ravikumar: *System and method for smoothing hierarchical data using isotonic regression*, Patent number 7870474, issued January 11, 2011.
16. D. Chakrabarti, S. Ravikumar, and A. Tomkins: *System and method using hierarchical clustering for evolutionary clustering of sequential data sets*, Patent number 7734629, issued June 8, 2010.
17. G. Aggarwal, D. Chakrabarti, P. K. Dubey, N. P. Garg, S. Ghosal, A. K. Gupta, A. Kulshreshtha, Ashutosh and S. K. V. Murthy: *Customization of information retrieval through user-supplied code*, Patent number 6611834, issued August 26, 2003.
18. A. Pal, D. Chakrabarti, K. Subbian, and A. Kannan: *Semi-Supervised Learning via Deep Label Propagation*, Patent Application US15597290, filed November 22, 2018.
19. A. Pal, A. Herdagdelen, S. Chatterji, S. Taank, and D. Chakrabarti: *Systems and Methods for Recommendation of Topical Authorities*, Patent Application US20170220577, filed January 29, 2016.
20. D. Agarwal, D. Diklic, D. Chakrabarti, A. Broder, and V. Josifovski: *System and method for determining an event occurrence rate*, Patent Application US20110153550, filed in February, 2011.
21. K. Punera, D. Chakrabarti, and S. Ravikumar: *Quicklink selection for navigational query*, Patent Application US20100250528, filed in March, 2009.
22. E. Vee, D. Chakrabarti, A. Dasgupta, A. Ghosh, S. Ravikumar, and A. Tomkins: *Hierarchical structure entropy measurement methods and systems*, Patent Application US20090112865, filed in October, 2007.
23. D. Agarwal, D. Chakrabarti, and S. Pandey: *System and method for matching objects using a cluster-dependent multi-armed bandit*, Patent Application US20090043597, filed in August, 2007.
24. D. Chakrabarti, S. Ravikumar, and A. Tomkins: *System and method using flat clustering for evolutionary clustering of sequential data sets*, Patent Application US20070255684, filed in April, 2006.
25. D. Agarwal, D. Chakrabarti, V. Josifovski, and S. Pandey: *System and method for matching objects belonging to hierarchies*, Patent Application US20080140591, filed in December, 2006.

Awards and Honors

- McCombs Amplify award (2022-2023), nominated and decided by students, given to faculty members who create inclusive classrooms and improve learning outcomes.
- Amazon Scholar (since 2022)
- Trammell/CBA Foundation Teaching Award for Assistant Professors (2021-2022) granted by the McCombs School of Business.
- Faculty of the Year award (2020) for "excellence in teaching and outstanding contributions to the MSBA Program."
- Our KDD 2006 paper is a Classic Paper in Data Mining and Analysis, according to Google Scholar. Classic Papers "are highly-cited papers in their area of research that have stood the test of time."
- Our PKDD 2005 paper received the 10 year test-of-time prize at ECML/PKDD 2015.
- McCombs Research Excellence Grant, 2014.
- Our R-MAT generator is the basis for the new Graph500 supercomputer benchmark.
- Our COLT 2010 paper received the student best paper award.
- One of only five *"Siebel Scholars"* in 2002 from the CMU School of Computer Science.
- "Certificate of Merit" for 1996-97 and 1997-98 in IIT-Kanpur, India.
- National Talent Search Scholarship in 1994 from the Govt. of India.

Invited Talks

- *On Scalable Estimation for Overlapping Clustering Models*,
  in Joint Statistical Meetings, 2021.
- *Overlapping Clustering, and One (class) SVM to Bind Them All*,
  in the INDSTATS Conference, 2019.
- *Nonparametric Link Prediction in Dynamic Graphs*,
  in the TRIPODS Southwest Summer Conference, 2018.
- *Joint Inference of Multiple Label Types in Large Networks*,
  in the Dagstuhl Seminar on High-performance Graph Algorithms and Applications in Computational Science, 2014.
- *Nonparametric Link Prediction in Large Scale Dynamic Networks*,
  in the Statistics Seminar, UT Austin, 2014.
- *Personalization in Social Networks*,
  in the WNCG Seminar in UT Austin, 2014.
- *Nonparametric Link Prediction in Dynamic Graphs*,
  in the Purdue Statistics Symposium, 2012.
- *Theoretical and Statistical Formulations of Link Prediction*,
  in the Graph Exploitation Symposium at MIT Lincoln Lab, 2012.
- *A Theoretical Justification of Link Prediction Heuristics*,
  in MLG 2012.
- *Clustering Applications at Yahoo!*,
  in the NIPS 2009 Workshop on Clustering.

Tutorials

- D. Chakrabarti and D. Agarwal:
  Statistical Challenges in Computational Advertising, in KDD 2009.
- D. Agarwal and D. Chakrabarti:
  Algorithmic Challenges in Online Advertising, in CIKM 2008.

Professional Service

- Co-chair of the Web Mining and Content Analysis track at WWW 2017.
- Senior PC on KDD (2010, 2014, 2023) and WSDM (2016 and 2020) research tracks.
- Posters chair for WWW 2010.
- PC member on several conferences including KDD, WWW, ICML, SIGMOD, VLDB, and ICDE.
- Reviewer for several journals including the SIAM Journal of Computing, DMKD, ORL, JSPI, TKDD, POMS, Annals of Applied Statistics, and MISQ.
- Served on several NSF Panels.
- Local arrangements co-chair for KDD 2007.

Teaching and Advising Experience

- I developed and teach undergraduate-level Social Media Analytics and Advanced Analytics Programming (Spring 2015 onwards).
- I developed and teach graduate-level Data Analytics Programming (Summer 2015 onwards).
- Co-advisor of Long Zhao (NUS Business School), Xueyu Mao (Amazon), and Akhil Jalan (current student).

Funding Support

- McCombs Research Excellence Grant in 2023 and 2014.
- Institute for Foundations of Machine Learning grant in 2022.
- Dell Faculty Award in 2022.
- Facebook Faculty Awards in 2014 and 2015.

Released Software

- SURE: Fair classification without knowledge of demographics.
- NEWS: Create node embeddings for sparse networks.
- DIRECT: Classification from few positive examples.
- ROLIN: Robust linear classification from limited training data.
- AlphaRob: Constructing robust portfolios with high out-of-sample Sharpe ratios.
- SVM-Cone: Overlapping community detection under all low-rank models.
- SPACL: Community detection under the Mixed Membership Stochastic Blockmodel.
- GeoNMF: Community detection using non-negative matrix factorization.
- CrossAssociations: Parameter-free community detection in large graphs.
- NetMine: Extract patterns given a large graph as input, and has been used by the Northrop Grumman Corp.
- F4: Automatic time-series prediction using chaotic time series methods.