

# DEEPAYAN CHAKRABARTI

---

701 1st Ave, Sunnyvale, CA 94089

(408) 349 2950; deepay@yahoo-inc.com

## Personal Information

- Work address: 701 1st Ave, Sunnyvale, CA 94089. (408) 349 – 2950
- Home address: 1604 Hope Dr., Apt 123, Santa Clara, CA 95054. (408) 235 – 8917
- Email: deepay@yahoo-inc.com
- Citizenship: Indian; Visa status: H1-B
- Birth year: 1979

## Professional Preparation

### Employment

- Research Scientist, Yahoo! Inc. (08/2005 to present)

### 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. GPA: 4.0
- B.Tech. in Computer Science and Engineering, (07/1996 to 05/2000)  
Indian Institute of Technology (IIT), Kanpur, India. GPA: 9.7 (out of 10)

### Research Experience

I have worked on algorithmic and statistical challenges arising in a broad range of problems primarily derived from Web Search, Computational Advertising, and Graph Mining.

**Web Search:** The Web is characterized by multiple sources of information: searches, anchor-texts, user tags, etc. Combining these to extract the “true signal” has been the underlying theme of my work. Consider, for instance, the problem of segmenting a webpage into semantically distinct sections. There are several cues available: the visual rendering of the webpage, textual features such as the fraction of text within link anchor-texts, and the DOM-tree hierarchy of HTML components comprising the webpage. All of these contain information, mostly relevant but occasionally conflicting. We designed an algorithm to properly combine all of these, while *learning* the relative importance of each cue. Again, generating titles for webpages and search result “quicklinks” requires merging data from searches, anchor-texts, webpage content, etc.

**Computational Advertising:** As a foundation of many Web-based businesses, computational advertising has attracted increasing attention in recent years. The aim is to pick the best matching ads for any given query, which typically requires an estimate of the click-through rate (CTR) of the (query, ad) pair. The key issues here are the high dimensionality of the feature space size and data sparsity, both of which are seen in extremes typically not found in other domains. In the same spirit as in Web Search, our work improved the traditional IR-based ad ranking methods by combining IR scores with click feedback from historical logs. To control sparsity, we merged data from coarser to finer resolutions of a domain-specific hierarchy, using the low-variance CTR estimates from the former to inform the estimates in the latter. In addition, we proposed multi-resolution explore-exploit strategies that use the hierarchy to speed up exploration in the space of ads and, essentially, make explore-exploit solutions work at “Web scales.”

## Research Experience (contd.)

**Graph Mining:** My thesis, titled *Tools for Large Graph Mining*, under Dr. Christos Faloutsos, looks at three problems in mining large graphs (such as the Internet, the Web, protein networks, etc.) First, we find common patterns (“laws”) that hold for most real-world graphs, and have developed the *R-MAT* graph generator which matches almost all these patterns much better than other published generators. Second, we find a closed-form formula for the *epidemic threshold* in viral/information propagation over an *arbitrary* graph; above this threshold, an infection may turn into an epidemic, whereas the infection dies out below this threshold. Third, we have developed techniques to find “*natural*” groups of nodes scalably and *automatically*; our algorithm figures out both the required number of groups as well as their memberships.

## Awards and Honors

- 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.

## Professional Service

- Local arrangements co-chair for KDD 2007.
- Student member of CMU departmental Ph.D. admissions committee for 2001-2003.
- Program Committee member of KDD 2008, WWW 2008 and 2009, ICDE 2008, PKDD 2008, ECDM 2008, MLG 2007, and LinkKDD 2006.
- Reviewer for ICDE 2006, JMLR 2005 and 2006, DMKD 2005, INFOCOM 2004, SIGMOD 2003, IEEE Communication Letters 2003, and VLDB 2002.

## Book Chapters

- 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.

## Tutorials

- D. Chakrabarti and D. Agarwal: *Algorithmic Challenges in Online Advertising*, in CIKM 2008, Napa, CA.

## Patents

- Granted: *Customization of information retrieval through user-supplied code*, Patent number 6,611,834, by G. Aggarwal, D. Chakrabarti, P. K. Dubey, N. P. Garg, S. Ghosal, A. K. Gupta, A. Kulshreshtha, Ashutosh and S. K. V. Murthy; assignee IBM Corp.
- Filed: Nine patent applications, all in the USA, by Yahoo! Inc.

## Released Software (at <http://www.cs.cmu.edu/~deepay/index.html#Sw>)

- The [NetMine](#) system extracts many patterns given a large graph as input, and has been used by the Northrop Grumman Corp. (Mark Hoy and Jayshree Ranka).
- The [CrossAssociations](#) system automatically “groups” nodes in a large graph.
- The [F4](#) system performs automatic time-series prediction using chaotic time series methods.

## Refereed Conference Papers (at <http://www.cs.cmu.edu/~deepay/#Pubs>)

1. D. Chakrabarti, R. Kumar, F. Radlinski, and E. Upfal: *Mortal Multi-Armed Bandits*, in NIPS 2008, Vancouver, Canada.
2. D. Chakrabarti, R. Kumar, and K. Punera: *Generating Succinct Titles for Web URLs*, in KDD 2008, Las Vegas, NV.
3. D. Chakrabarti, R. Kumar, and K. Punera: *A Graph-Theoretic Approach to Webpage Segmentation*, in WWW 2008, Beijing, China.
4. D. Chakrabarti, D. Agarwal, and V. Josifovski: *Contextual Advertising by Combining Relevance with Click Feedback*, in WWW 2008, Beijing, China.
5. D. Agarwal, A. Broder, D. Chakrabarti, D. Diklic, V. Josifovski, and M. Sayyadian: *Estimating Rates of Rare Events at Multiple Resolutions*, in KDD 2007, San Jose, CA.
6. S. Pandey, D. Chakrabarti, and D. Agarwal: *Multi-armed Bandit Problems with Dependent Arms*, in ICML 2007, Corvallis, OR.
7. D. Chakrabarti, R. Kumar, and K. Punera: *Page-level Template Detection via Isotonic Smoothing*, in WWW 2007, Banff, Canada.
8. S. Pandey, D. Agarwal, D. Chakrabarti, and V. Josifovski: *Bandits for Taxonomies: A Model-based Approach*, in SDM 2007, Minneapolis, MN.
9. 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, Anchorage, Alaska.
10. D. Chakrabarti, R. Kumar and A. Tomkins: *Evolutionary Clustering*, in KDD 2006, Philadelphia, PA.
11. J. Sun, H. Qu, D. Chakrabarti and C. Faloutsos: *Neighborhood Formation and Anomaly Detection in Bipartite Graphs*, in ICDM 2005, Houston, Texas.
12. J. Leskovec, D. Chakrabarti, J. Kleinberg and C. Faloutsos: *Realistic, Mathematically Tractable Graph Generation and Evolution, Using Kronecker Multiplication*, in PKDD 2005, Porto, Portugal.
13. D. Chakrabarti: *AutoPart: Parameter-Free Graph Partitioning and Outlier Detection*, in PKDD 2004.
14. D. Chakrabarti, S. Papadimitriou, D. Modha and C. Faloutsos: *Fully Automatic Cross-Associations*, in KDD 2004, Washington, USA; also a CMU Tech Report.
15. D. Chakrabarti, Y. Zhan, C. Faloutsos: *R-MAT: A Recursive Model for Graph Mining*, in SDM 2004, Lake Buena Vista, FL.
16. 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, Lake Buena Vista, FL.
17. Y. Wang, D. Chakrabarti, C. Wang and C. Faloutsos: *Epidemic Spreading in Real Networks: An Eigenvalue Viewpoint*, in SRDS 2003, Florence, Italy.
18. D. Chakrabarti and C. Faloutsos: *F4: Large Scale Automated Forecasting using Fractals*, in CIKM 2002, McLean, Virginia, USA; also a CMU Tech Report.
19. 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, Williamstown, USA.

## Refereed Journal Papers (at <http://www.cs.cmu.edu/~deepay/#Pubs>)

1. D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos: *Epidemic Thresholds in Real Networks*, in ACM TISSEC, 10(4), 2008.
2. D. Chakrabarti and C. Faloutsos: *Graph Mining: Laws, Generators and Algorithms*, in ACM Computing Surveys, 38(1), 2006.
3. 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.
4. 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.