

# 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: 1600 Villa St, Apt 287, Mountain View, CA 94041. (650) 964 – 7720
- 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*
- 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

Thesis title: **Tools for Large Graph Mining (Advisor: Dr. Christos Faloutsos)**

I have worked on 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.

### Teaching and Graph Application Experience

- TA for 15–415 (Database Applications, Fall 2002) and 15–826 (Multimedia Databases and Datamining, Spring 2005).
- Mentor for undergraduates Mr. J. Leskovec (summer '03) and Miss M. Towne.
- Helped in the preparation of (accepted) grants *NSF-IIS-0209107* and *PITA*, 2004.

### Internships

- IBM Almaden Research Lab (Summer 2004)
- IBM T. J. Watson Research Lab (Summer 2002)
- AT&T Research Labs (Summer 2001)
- IBM India Research Lab (Summer 1999)

### Professional Service

- Local arrangements co-chair for KDD 2007.
- Student member of the CALD Ph.D. admissions committee for 2001 – 2003.
- Program Committee member of ICDE 2007, PKDD 2006, and LinkKDD 2006.
- Reviewer for ICDE 2006, JMLR 2005 and 2006, DMKD 2005, INFOCOM 2004, SIGMOD 2003, IEEE Communication Letters 2003, and VLDB 2002.

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

1. D. Chakrabarti, R. Kumar, and K. Punera: *Page-level Template Detection via Isotonic Smoothing*, to appear in WWW, 2007.
2. S. Pandey, D. Agarwal, D. Chakrabarti, and V. Josifovski: *Bandits for Taxonomies: A Model-based Approach*, to appear in SDM, 2007.
3. J. Leskovec, D. Chakrabarti, C. Faloutsos, S. Madden, C. Guestrin and M. Faloutsos: *Information Survival Threshold in Sensor and P2P Networks*, to appear in IEEE INFOCOM, 2007.
4. D. Chakrabarti and C. Faloutsos: *Graph Mining: Laws, Generators and Algorithms*, in ACM Computing Surveys, 38(1), 2006.
5. D. Chakrabarti, R. Kumar and A. Tomkins: *Evolutionary Clustering*, in KDD 2006 (pages 554 – 560), Philadelphia, PA.
6. 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 (pages 434 – 445).
7. J. Sun, H. Qu, D. Chakrabarti and C. Faloutsos: *Neighborhood Formation and Anomaly Detection in Bipartite Graphs*, in ICDM 2005 (pages 418 – 425), Houston, Texas.
8. J. Leskovec, D. Chakrabarti, J. Kleinberg and C. Faloutsos: *Realistic, Mathematically Tractable Graph Generation and Evolution, Using Kronecker Multiplication*, in PKDD 2005 (pages 133 – 145), Porto, Portugal.
9. D. Chakrabarti: *AutoPart: Parameter-Free Graph Partitioning and Outlier Detection*, in PKDD 2004 (pages 112 – 124), Pisa, Italy.
10. D. Chakrabarti, S. Papadimitriou, D. Modha and C. Faloutsos: *Fully Automatic Cross-Associations*, in KDD 2004 (pages 79-88), Washington, USA; also a CMU Tech Report.
11. D. Chakrabarti, Y. Zhan and C. Faloutsos: *R-MAT: A Recursive Model for Graph Mining*, in SIAM Data Mining 2004, Orlando, Florida, USA.
12. 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.
13. Y. Wang, D. Chakrabarti, C. Wang and C. Faloutsos: *Epidemic Spreading in Real Networks: An Eigenvalue Viewpoint*, in SRDS 2003 (pages 25-34), Florence, Italy.
14. D. Chakrabarti and C. Faloutsos: *F4: Large Scale Automated Forecasting using Fractals*, in CIKM 2002 (pages 2-9), McLean, Virginia, USA; also a CMU Tech Report.
15. 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 (pages 329-336), Williamstown, USA.

## Book Chapters

1. 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. 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.
2. Filed: (a) *System and Method Using Hierarchical Clustering for Evolutionary Clustering of Sequential Data Sets*, (b) *System and Method Using Flat Clustering for Evolutionary Clustering of Sequential Data Sets*, (c) *System and Method for Evolutionary Clustering of Sequential Data Sets*, filed by Yahoo! in 2006.

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

1. 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).
2. The [CrossAssociations](#) system automatically “groups” nodes in a large graph.
3. The [F4](#) system performs automatic time-series prediction using chaotic time series methods.