Jayanta Mondal

3220 A. V. Williams Building
Department of Computer Science
University of Maryland
College Park, MD, 20742

Cell: (240) 463-7974
Email: jayanta@cs.umd.edu
Web: www.cs.umd.edu/~jayanta/
Google Scholar Profile

RESEARCH INTERESTS

My research interests lie in the broad area of big data analytics, network science, and cognitive analytics. My dissertation research focuses on performing real-time analysis of graph-structured data. Example applications include personalization, recommendation, and behavioral analysis in social networks; intrusion and advanced threat detection in file access/download graphs; anomaly detection, pattern mining in financial transaction networks; event detection and notification in spatio-temporal networks, and so on. Along side my thesis work, I am developing an online classifier to distinguish malware from benign software using real-world intrusion events. I have also worked on text analytics and recommender systems, as a core team member of Minekey, a machine learning-based startup. Additionally, during my two internships at Microsoft Research, I have developed tools and techniques for a database-as-a-service-platform. Specifically, I built a partitioning advisor for OLTP workloads, and developed online algorithms to maintain an optimal set of physical designs for evolving OLAP workloads.

EDUCATION

- **University of Maryland**, College Park, MD USA   **August 2010 – 2015 (Expected)**
  Masters without Thesis and PhD Candidate, Computer Science
  Advisor: Prof. Amol Deshpande
- **Indian Institute of Technology**, Kharagpur, INDIA   **August 2002 – June 2007**
  Bachelor of Technology (Honors) and Master of Technology, Computer Science and Information Technology
  Advisor: Prof. Sudeshna Sarkar

RELEVANT GRADUATE COURSEWORK

- **University of Maryland, USA:** Database Systems, Social Networking Databases, Probabilistic and Graph Data Management; Large-scale Analytics, Social Network Data Management, Mobility of Data In Network, Local Data And Privacy, Models for Socio-Technical Networks, High Performance Computing, Fundamentals of Software Testing, Computational Gene Finding and Gene Assembly.

- **Indian Institute of Technology, Kharagpur, India:** Machine Learning, Information Retrieval, Data Warehousing and Data Mining, Artificial Intelligence, Natural Language Processing, Principles of Distributed Systems, Advanced Graph Theory, Algorithms for Bio-informatics.

- **Completed Coursera Courses:** Machine Learning (Stanford), Introduction to Data Science (UW), Mining Massive Dataset (Stanford), Pattern Discovery in Data Mining (UIUC), Statistical Inference (JHU).

RESEARCH EXPERIENCE

- **University of Maryland, College Park, USA**.   **August 2010 – Present**
  Graduate Research Assistant, Advisor: Prof. Amol Deshpande, Collaborator: Prof. Tudor Dumitras

  Declarative Framework for Detection and Monitoring of Active Subgraphs: An active subgraphs definition consists of both structural predicates (i.e., how different nodes are connected) as well as conditions on how active (e.g., how many phone calls a node is making every 6 hours) the nodes in a subgraph are. Introduced *graph view*, and independence layer, to separate the query interface from the underlying dynamicities of the data. Designed a Datalog-based declarative interface to query the graph view. Currently exploring techniques rooted in sampling, probabilistic data structures, query predicate reordering, to implement the language efficiently.

Ego-centric Aggregation Framework For Large Dynamic Graphs: Designed and implemented a flexible, general-purpose, and user-extensible in-memory framework to execute large number of neighborhood-based aggregate queries (e.g., computing TOP-K hash tags tweeted by the followers of a node). The framework is built around the notion of an aggregation overlay graph, a pre-compiled data structure that encodes the computations to be performed when an update or a query is received. The overlay graph enables sharing of partial aggregates (by exploiting overlaps among neighborhoods) across different ego-centric queries, and also allows partial pre-computation of the aggregates to minimize the query latencies. Achieved a throughput of 500K queries per second for a graph of size 320 million.

A Replication Framework for Large Dynamic Graphs: Designed and implemented an in-memory distributed graph database with a special focus on supporting low latency query execution, while minimizing bandwidth consumption. Developed a hybrid replication policy that monitors the read-write frequencies of the nodes to dynamically decide what data to replicate, and whether to do eager or lazy replication. Evaluated a clustering-based approach to amortize the costs of making these replication decisions, and introduce a fairness criterion to control the tradeoff between eager and lazy replication. Proposed a fairness criteria to minimize latency skew that can appear in hybrid replication strategies. Implemented a prototype distributed graph using CouchDB, and evaluated our system on Amazon EC2. Our techniques saved up to 33% of network bandwidth.

Influence Graph-based Malware Detection: Introduced the notion of influence graph to model the download activity behavior of software on a host machine. By combining telemetry from anti-virus and intrusion-prevention systems, we reconstruct and analyze 19 million graphs from 5 million real hosts. We identify several strong indicators of malicious activity, such as the growth rate, the diameter and the Internet access patterns of downloader graphs. Designed a random forest classifier, building on these insights. Our system achieves 98.0% true-positive rate, with a 2.0% false-positive rate, and is able to detect malware an average of 9.24 days earlier than leading anti-virus products.

- **Microsoft Research, Redmond, USA**   **May 2014 – August 2014**
  Research Intern, Mentor: Sudipto Das

  Online Physical Design Tuning for Analytical Workloads: Worked on the online version of the physical design tuning problem, where the workload itself can change over time, in contrast to dealing with a static snapshot of the workload. Developed cost-based techniques (in close consultation with the query optimizers cost model) to decide what set of physical designs (primarily indexes) to maintain as time progresses with the goal of minimizing the overall cost of query execution in the presence of various system-level constraints.

- **Microsoft Research, Redmond, USA**   **May 2013 – August 2013**
  Research Intern, Mentor: Sudipto Das

  Automatic partitioning of Windows Azure SQL Databases: Worked on the problem of automatically scaling tenant databases in a multi-tenant database-as-a-service platform. Specifically, developed techniques to automatically shard SQL Server databases through careful analysis of the workload, database schema, and available resources. Additionally, developed a framework to generate insights about the database schema and workload that can empower users about why certain databases are hard to partition.

- **Indian Institute of Technology, Kharagpur, India**   **July 2006 – June 2007**
  Masters Student and Student Researcher, Minekey Inc; Supervisor: Prof. Sudeshna Sarkar

  Content Extraction from webpages: Worked on the problem of news extraction from HTML files crawled from the web. The key challenges were to filter out the garbage texts, advertisements, etc., and extract relevant textual information. Adapted techniques like wrapper induction and learning advertisement-placement models to achieve the goal.

  Automatic concept extraction from English text: Addressed the problem of organizing the results of a keyword search query, into various concept categories in which the keyword has occurred. Designed and implemented a co-occurrence based semantic clustering technique to automatically extract concepts associated with the search results of keyword queries. The technique worked by building a semantic co-occurrence graph and then finding densely connected words through graph clustering.

PUBLICATIONS

- Managing Large Dynamic Graphs Efficiently, *SIGMOD, 2012*
  Jayanta Mondal and Amol Deshpande
- Stream Querying and Reasoning on Social Data, Encyclopedia of Social Network Analysis and Mining, 2014
  Jayanta Mondal and Amol Deshpande
- EAGr: Supporting Continuous Ego-centric Aggregate Queries over Large Dynamic Graphs, *SIGMOD, 2014*

Jayanta Mondal and Amol Deshpande

- A Coarse-grained Partitioning Advisor for OLTP Workloads, *Under preparation, 2014*,
  Jayanta Mondal and Sudipto Das
- The Dropper Effect: Insights into Malware Distribution with Downloader Graph Analytics, *To appear in ACM SIGSAC (CCS), 2015*
  Bumjun Kwon, **Jayanta Mondal**, Jiyong Jang, Leyla Bilge, and Tudor Dumitras
- GSTREAM: Towards a Unified Streaming Graph Data Management System, *Submitted to Big-O(Q) VLDB Workshop on Big-Graphs Online Querying, 2015*
  Jayanta Mondal and Amol Deshpande
- A Declarative Framework to Detect and Monitor Active Subgraphs, *Under preparation, 2015*
  Jayanta Mondal and Amol Deshpande

WORK EXPERIENCE

- **Minekey Inc., Gurgaon, India**   **July 2007 – June 2010**
  Senior Software Engineer (2007 - 2009), Technical Manager (2009-2010). Reported to: Gaurav Bhatia

  One of the initial six members of **Minekey**, a startup which was incubated at **IIT Kharagpur**. Primarily worked on (a) automatic concept extraction from text data, (b) modeling user behavior, (c) generating recommendations, and (d) designing backend architecture of iThink and Twezr for performance and scalability.

  News Recommendation System (Minekey): Curated news from webpages, extracted topics from news using unsupervised techniques, and modeled users' reading habit to generate recommendations.

  Social Application on Facebook (iThink): Designed and implemented the Facebook application named iThink. Scaled our application to handle more than 10 million users using sharded MySQL servers and Memcache.

  People-centric Unified Inbox (Twezr): Designed and implemented a unified inbox to combine messages from various email services (Gmail, Yahoo mail, Microsoft Exchange) and social networks (Facebook, Twitter). Provided a user-centric and ranked view of the aggregated messages. Developed a smart crawler by modeling user behavior. Developed entity resolution and contact ranking algorithms. Used CouchDB (a document-oriented database) to scale our application.

TALKS

- Real-time query processing on Large Dynamic Graphs. *Database Seminar, Microsoft Research, Redmond*, July 2014.
- EAGr: Supporting Continuous Ego-centric Aggregate Queries on Large Dynamic Graphs. *Conference Presentation. Snowbird, SIGMOD*, 2014.
- Challenges in Managing and Querying Graph Databases in Real-time. *Invited Talk. IIT Kharagpur, India*, December 2013.
- Managing Large Dynamic Graphs. *Conference Presentation. Scottsdale, SIGMOD*, 2012.

SERVICE

- Reviewer: IEEE Transactions on Knowledge and Data Engineering (TKDE), 2015
- External Reviewer: ACM Special Interest Group on Management of Data (SIGMOD), 2015
- External Reviewer: International Journal on Very Large Databases (VLDBJ), 2014

TEACHING EXPERIENCE

- Teaching Assistant for CMSC216 – Introduction to Computer Systems in Spring 2012, University of Maryland
  Duties included: Tutorial lecture, grading and conducting TA office hours.
- Teaching Assistant for CS31001 – Computer Architecture and Operating System in Fall 2007, IIT Kharagpur
  Duties Included: Tutorial lectures, grading, and conducting laboratory hours.

SKILLS

- **Languages:** Java, C, C#, Perl, Shell scripting, PHP

- **Tools, Platforms, and Systems:** Weka, Map Reduce (Hadoop), Apache Storm, Apache Spark, Giraph, GraphX, CouchDB, Redis, Amazon EC2, Windows Azure, PostgreSQL, MySQL, Microsoft SQL Server.

HONORS AND MENTIONS

- Kulkarni Summer Research Fellowship at University of Maryland, College Park, 2015

- Travel Award, SIGMOD, 2012

- Dean's fellowship at University of Maryland, College Park, 2010–2012

- Nominated for the best B.Tech. project (titled *Design and Implementation of a Regional Language Search Engine*) award in the Computer Science and Engineering Department (6 out of 60 students were nominated), 2007

- Ministry of Human Resource Development (MHRD) Scholarship, IIT Kharagpur, 2006–2007

- Merit-cum-Means Scholarship, IIT Kharagpur, 2005–2006

- National Council Of Educational Research And Training (NCERT) Scholarship, 2000–2007

REFERENCES (More available on request)

Prof. Amol Deshpande	*Dr. Sudipto Das*	*Prof. Tudor Dumitras*
Department of Computer Science	Researcher, DMX Group	ECE Department
University of Maryland	Microsoft Reaserch	University of Maryland
College Park, Maryland, 20740	Seattle, WA, 98052	College Park, Maryland, 20740
Phone: (301) 405-2703	Phone: (425) 705-7819	Phone: (301) 405-7466
Email: amol@cs.umd.edu	Email: sudipto.das@microsoft.com	Email:tdumitra@umiacs.umd.edu