

Matthew Saltz

saltzm@gmail.com · +1 225 205 7903

<http://www.matthewsaltz.com>

EDUCATION

Erasmus Mundus Master of Data Mining and Knowledge Management Sept 2013-June 2015

· **Université Lumière Lyon 2**, Lyon, France, *Data Mining & Complex System Modeling in Social Science* 2013-2014

· **Universitat Politècnica de Catalunya**, Barcelona, Spain, *Statistical Modeling & Data Mining* 2014-2015

Awarded the Category A scholarship of Erasmus Mundus, an organization sponsored by the E.U. to encourage international cooperation in a variety of cross-university master's programs. Includes full tuition, living stipend, personal travel budget

University of Georgia, B.S., M.S. in Computer Science, *GPA: 4.0/4.0 (B.S.), 3.96/4.0 (M.S.)* August 2009-2013

Oxford University, Course: Modernist Literature, *GPA: 4.0/4.0* May-June 2010

Honors & Awards: Foundation Fellowship (1 of 11, University of Georgia's highest academic scholarship), First Honor Graduate (1 of 14 students graduating with a 4.0), Outstanding Undergraduate Student Award for Computer Science (only recipient), Phi Beta Kappa, Computer Science representative on the Dean's Student Advisory Board

RELEVANT EXPERIENCE

Hewlett-Packard Enterprise Vertica, Cambridge, MA, *Software Engineer, Machine Learning* August 2015-Present

- Helped lead the initial development of Vertica's in-database machine learning capabilities
- Responsible for the design and development in C++ of distributed machine learning algorithms in Vertica, including kmeans and data normalization
- Designed and implemented in C++ a resource-managed infrastructure for caching tables and auxiliary data structures in memory that led to a speedup of 1-2 orders of magnitude for our iterative machine learning algorithms, allowing them to outperform Apache Spark in some cases tested
- Worked on ensemble gradient boosting machine training and prediction in Distributed R
- Served as an intern mentor for summer of 2016 and serving as an intern coordinator for summer of 2017

Master's Thesis Research, Universitat Politècnica de Catalunya February 2015-June 2015

- Kernelized the energy function of the bidirectional associative memory (BAM) and demonstrated that treating inference as an optimization problem on the kernelized energy function improves recall with more patterns and higher amounts of noise
- Publication (author): *A New Kernelized Associative Memory and Some of Its Applications*, accepted to the 2016 European Conference on Artificial Intelligence (ECAI)

Data Management (DAMA) Group, Universitat Politècnica de Catalunya, *Research intern* July-August 2014

- Designed and developed in Java a distributed vertex-centric community detection algorithm based on the WCC metric using Apache Giraph, taking advantage of advanced features such as custom aggregators, combiners, and message types
- Performed scalability tests on several real life graphs. The implemented algorithm scales best with the largest graphs and obtains all communities for a graph of 1.8 billion edges in just over an hour using 32 worker machines
- Publication (author): *Distributed Community Detection with the WCC Metric*, accepted to the 2015 ACM SIMPLEX workshop (a companion to the WWW 2015 conference)

Master's Thesis Research, University of Georgia May 2013-August 2013

- Created and implemented in Scala a novel algorithm for subgraph isomorphism that outperformed two well-known algorithms by up to several orders of magnitude on synthetic graphs of up to 10M nodes and 250M edges and on two real life datasets
- Publication (author): *DualIso: An Algorithm for Subgraph Pattern Matching on Very Large Labeled Graphs*, industrial track of 2014 IEEE International Congress on Big Data

Graph Research with Dr. John A. Miller, University of Georgia August 2012-August 2013

- Implemented in Java and Scala centralized sequential and parallel versions of strong simulation, an algorithm that uses various topological criteria to find matches of a query graph in a data graph
- Publication (co-author): *A Distributed Vertex-Centric Approach for Pattern Matching in Massive Graphs*, 2013 IEEE International Conference on Big Data

The Home Depot, Atlanta, GA, *IT Intern* May-August 2012

- Created and developed a novel technique for the segmentation of customers based on their transaction histories. Wrote SQL, bash scripts and custom Java code to implement the full processing pipeline using Apache Mahout on a Hadoop cluster
- Designed and implemented a novel security solution for connecting to Apache Hive via JDBC. Wrote Java code to create an authentication Tomcat service which spawned Hive servers, allowing the use of Hive's built-in authorization methods

SKILLS & INTERESTS

Languages (ordered by proficiency)

C++, Scala, Java, Python, R, Prolog, Bash

Platforms & Tools

Apache Hadoop, Giraph, Akka, Theano, Linux, googletest, perf, gdb, vim

Interests

Spanish (proficient), French (proficient), music (guitar, piano, singing), improv