

Shoshana (Bluma) Gelley

CONTACT INFORMATION	Department of Computer Science & Engineering New York University Polytechnic School of Engineering 2 MetroTech Center Brooklyn, NY 11201	Email: bgelley@nyu.edu Phone: (646) 596-5790 Website: http://cse.poly.edu/~gelley
RESEARCH INTERESTS	Data analysis, data mining, online communities, Human-Computer Interaction.	
EDUCATION	Ph.D. in Computer Science, School of Engineering, New York University - Dissertation Advisor: Torsten Suel - Dissertation title: User Behavior Analysis and Decision Support in Online Communities. <i>Expected May 2015</i>	
	M.S. in Computer Science, Polytechnic Institute of NYU	2012
	B.S. in Computer Science, Touro College - Honors: Summa Cum Laude, Service Award	2010
PROFESSIONAL RESEARCH EXPERIENCE	Yahoo Labs, Haifa, Israel. Intern Scientist. Internship with the Answers Research group. Collaborated with 5 colleagues to investigate the impact of a site policy change on user engagement. Used R, Python, and Unix scripting, awk, and sed to build a pipeline to transform raw data for analysis. Analyzed Yahoo Answers usage data using R, created exploratory graphs, and conducted statistical tests. We found that user engagement was negatively impacted by the change, and that users adopt a consistent ‘persona’ with which they present themselves on the site. I presented the results to 30 scientists at Yahoo Labs, then completed the internship remotely and collaborated on writing a paper using effective communication and tools such as ShareLatex and git. The paper has been submitted to a leading conference. <i>July – November 2014</i>	
	Avaya Labs, Basking Ridge, NJ. Research Scientist Intern. Created a robust, multithreaded, error-resistant crawler in Java (~ 4K lines of code) to scrape an Ajax-based website with no API (Pinterest). Made extensive use of Unix commands and scripts (including screen) to manage scarce server resources and ensure maximum uptime for the crawler. Ran crawler for over a month and collected millions of data points, stored in a graph database (Neo4j). Analyzed data using R to reveal insights about user behavior on the site that challenged accepted wisdom. In particular, we found that the heavily-promoted follow mechanism on the site drives only a small amount of content discovery. <i>May – August 2013</i>	
	CRA Distributed Research Experience for Undergraduates, Columbia University, New York, NY. One of 50 undergraduate students who received a grant to participate in an NSF-funded summer research program. Worked together with Dr. Haimonti Dutta of Columbia University to explore novel methods for predicting epileptic seizures. Used Java and Weka’s API to build and train SVM predictors on EEG data. We concluded that the combination of SVMs and Granger Causality for determining causation between channels on an EEG shows promise for seizure prediction. <i>June – August 2009</i>	
DISSERTATION RESEARCH EXPERIENCE	Detecting Notability on Wikipedia. This project is supported by an Individual Engagement Grant from the Wikimedia Foundation to develop decision-support tools for determining whether proposed articles meet Wikipedia’s notability guidelines. The work involves choosing and extracting features, analyzing and visualizing	<i>December 2014 – present</i>

data for feature selection using R with Shiny, then building machine learning predictors in Python with scikit-learn. I am collaborating with researchers from the Wikimedia Foundation and INRIA to collect data and design the tools that will use the predictions. Throughout the planning and implementation process, we have been interfacing with the Wikipedia community to solicit and respond to comments, suggestions, and functionality requests.

Wikipedia Articles for Creation (AfC) process. *March – November 2014*

Long-distance collaboration with other researchers to study the impact of Wikipedia’s AfC process. Used MySQL to write and optimize queries over tables with tens of millions of rows and determine the impact of user characteristics on the success of the articles they created.

Predicting Deletion on Wikipedia. *Sep. 2011 – April 2013*

Carried out an intensive study of the deletion process on Wikipedia, then created a classifier that would support user decisions by determining whether or not an article was likely to be deleted. Designed a process for collecting difficult-to-gather data. Cleaned the data and extracted features using natural language processing, web APIs, and text processing skills. Built a classifier that predicts with high accuracy whether or not a Wikipedia article will be deleted.

ADDITIONAL RESEARCH EXPERIENCE Member of an interdisciplinary team studying the search behavior of humanities researchers. *2013 – 2014*

Worked with NYU Proteus (NLP) group on temporal information extraction. *2010 – 2011*

PUBLICATIONS Bluma Gelley and Ajita John. Do I Need to Follow You? Examining the Utility of the Pinterest Follow Mechanism. To appear in *Proceedings of the ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)*, 2015.

Sergey Nepomnyachiy, Bluma Gelley, Wei Jiang, and Tehila Minkus. What, Where, and When: Geotemporal Text Search in Twitter. In *Proceedings of the ACM SIGSPATIAL Workshop on Geographic Information Retrieval (GIR)*, 2014.

Jodi Schneider, Bluma Gelley, and Aaron Halfaker. Accept, Decline, Postpone: How Newcomer Productivity is Reduced in English Wikipedia by Pre-publication Review. In *Proceedings of the International Symposium on Open Collaboration (OpenSym)*, 2014.

Bluma Gelley and Torsten Suel. Automated Decision Support for Human Tasks in a Collaborative System: The Case of Deletion in Wikipedia. In *Proceedings of the International Symposium on Wikis and Open Collaboration (WikiSym)*, 2013.

UNDER REVIEW Bluma Gelley, Ran Wolff, and Dan Pelleg, Modeling Answerers’ Social Personae. Under review.

Bluma Gelley. Investigating Deletion on Wikipedia. Undergoing requested revisions for *Internet Computing*.

GRANTS AND HONORS Individual Engagement Grant, Wikimedia Foundation. *December 2014*

Graduate Assistantships in Areas of National Need (GAANN) Fellowship, U.S. Department of Education. *2009 – 2012*

Computing Research Association Distributed Research Experience for Undergraduates. *2009*

Award for Exemplary Service, Dedication and Leadership, Touro College. *2009*

Dean’s List, Touro College. *2007 – 2009*

Lander Honors Full Tuition Scholarship, Touro College. *2007 – 2009*

TECHNICAL SKILLS	<ul style="list-style-type: none"> • Java, Python, R, SQL, VB.NET/ASP.NET. • Statistical analysis, hypothesis testing, machine learning, Natural Language Processing, scikit-learn, Weka, NLTK. • Data presentation in R — ggplot2, Shiny. • Excel – formulas, pivot tables, charting. • Unix/Linux (shell scripting, sed, awk). • Web scraping and crawling; data gathering, cleaning, and munging. • Graph databases: neo4j, neo4j-Cypher; RDBMS: MySQL, SQLServer. • Git, subversion. • Familiar with MapReduce paradigms and the Hadoop ecosystem. 		
TALKS	<p>Conference presentation on work on Pinterest. CSCW, March 2015. (Upcoming)</p> <p>A brief introduction to R. Yahoo Labs, 2014.</p> <p>Understanding answerer motivation on Yahoo Answers. Yahoo Labs, 2014.</p> <p>Research presentation about work on Pinterest. Avaya Labs, 2013.</p> <p>Conference presentation on work on deletion in Wikipedia. WikiSym, 2013.</p> <p>Wikipedia as a Social Network. Polytechnic Institute of NYU, 2012.</p>		
PROFESSIONAL ACTIVITIES	<p>Panelist, Girls Talk Tech, Google Anita Borg Birthday Celebration event to attract women to Computer Science. <i>2015</i></p> <p>Reviewed papers for CSCW 2015 and KDD 2014.</p> <p>President, Touro College student chapter of the ACM. <i>2008 – 2009</i></p> <p>Secretary, Touro College student chapter of the ACM. <i>2007 – 2008</i></p>		
ADDITIONAL WORK EXPERIENCE	<p>Computer Science Instructor, Touro College, New York, NY. <i>2010 – 2011</i></p> <ul style="list-style-type: none"> • Worked within an existing curriculum to create innovative and engaging classes. • Course materials developed are being used by instructors currently teaching the course. <p>Web Development Intern, Saw You at Sinai online dating service. <i>March – August 2009</i></p> <ul style="list-style-type: none"> • Developed functionality for new area of website in ASP.NET/VB.NET. <p>Database Intern, New York Jewish Herald. <i>June – August 2008</i></p> <p>Teaching Assistant and Math Instructor, Yeshiva of Flatbush Elementary School. <i>2008 - 2009</i></p> <p>Math and Science Teacher, 8th grade, Beth Jacob of Gur. <i>2007 – 2008</i></p>		
REFERENCES	<table border="0" style="width: 100%;"> <tr> <td style="width: 50%; vertical-align: top;"> <p>Torsten Suel, PhD Associate Professor of Computer Science Department of Computer Science and Engineering New York University Polytechnic School of Engineering 2 Metrotech Center, 10.046 Brooklyn, NY 11201 (718) 260-3354 torsten.suel@nyu.edu</p> </td> <td style="width: 50%; vertical-align: top;"> <p>Ajita John, PhD Research Scientist Avaya Labs 211 Mount Airy Road Basking Ridge, NJ 07920 (908) 848-5679 ajita@avaya.com</p> </td> </tr> </table>	<p>Torsten Suel, PhD Associate Professor of Computer Science Department of Computer Science and Engineering New York University Polytechnic School of Engineering 2 Metrotech Center, 10.046 Brooklyn, NY 11201 (718) 260-3354 torsten.suel@nyu.edu</p>	<p>Ajita John, PhD Research Scientist Avaya Labs 211 Mount Airy Road Basking Ridge, NJ 07920 (908) 848-5679 ajita@avaya.com</p>
<p>Torsten Suel, PhD Associate Professor of Computer Science Department of Computer Science and Engineering New York University Polytechnic School of Engineering 2 Metrotech Center, 10.046 Brooklyn, NY 11201 (718) 260-3354 torsten.suel@nyu.edu</p>	<p>Ajita John, PhD Research Scientist Avaya Labs 211 Mount Airy Road Basking Ridge, NJ 07920 (908) 848-5679 ajita@avaya.com</p>		

Dan Pelleg, PhD
Director, Answers Research
Yahoo Labs
MATAM Advanced Technology Park
Tower 3 - Floor 7
Haifa, 31905 Israel
dpelleg@yahoo-inc.com