

Tutorial Program

Online Experiments for Computational Social Science

Presented by Eytan Bakshy and Sean Taylor

Taught by two researchers on the Facebook Data Science team, this tutorial teaches attendees how to design, plan, implement, and analyze online experiments. First, we review basic concepts in causal inference and motivate the need for experiments. Then we will discuss basic statistical tools to help plan experiments: exploratory analysis, power calculations, and the use of simulation in *R*. We then discuss statistical methods to estimate causal quantities of interest and construct appropriate confidence intervals. Particular attention will be given to scalable methods suitable for “big data,” including working with weighted data and clustered bootstrapping. We then discuss how to design and implement online experiments using PlanOut, an open-source toolkit for advanced online experimentation used at Facebook. We will show how basic A/B tests, within-subjects designs, as well as more sophisticated experiments can be implemented. We demonstrate how experimental designs from social computing literature can be implemented, and also review in detail two very large field experiments conducted at Facebook using PlanOut. Finally, we will discuss issues with logging and common errors in the deployment and analysis of experiments. Attendees will be given code examples and participate in the planning, implementation, and analysis of a web application using Python, PlanOut, and *R*.

Eytan Bakshy is a researcher and senior member of the Facebook Data Science Team. He has been conducting field experiments at Facebook for over three years, focusing peer effects in networks. Bakshy holds a Ph.D. in information from the University of Michigan and

a B.S. in mathematics from UIUC.

Sean J. Taylor is a research scientist on the Facebook Data Science Team specializing in field experiments on Web and mobile platforms. His research interests include causal inference, social influence, information credibility, and evaluation of predictions. Taylor holds a Ph.D. in information systems from New York University and a B.S. in economics from the University of Pennsylvania.

Social Media Threats and Countermeasures

*Presented by Kyumin Lee,
James Caverlee, and Calton Pu*

The past few years have seen the rapid rise of many successful social systems — from web-based social networks (such as Facebook, and LinkedIn) to online social media sites (such as Twitter and YouTube) to large-scale information sharing communities (such as reddit and Yahoo! Answers) to crowd-based funding services (such as Kickstarter and IndieGoGo) to web-scale crowdsourcing systems (such as Amazon MTurk and Crowdfunder).

However, with this success has come a commensurate wave of new threats, including bot-controlled accounts in social media systems for disseminating malware and commercial spam messages, adversarial propaganda campaigns designed to sway public opinion, collective attention spam targeting popular topics and memes, and propagate manipulated contents.

This tutorial will introduce peer-reviewed research work on social media threats and countermeasures. Specifically, we will address new threats such as social spam, campaigns, misinformation and crowdturfing, and overview countermeasures to mitigate and resolve these threats by revealing and detecting malicious participants (such as social spam-

mers, content polluters, and crowdturfers) and low quality contents. This tutorial will also overview available tools to detect these participants.

Kyumin Lee is an assistant professor in the Department of Computer Science, Utah State University, (kyumin.lee@usu.edu). Lee's primary research interests are in information quality and data analytics over large-scale networked information systems like the web, social media systems, and other emerging distributed systems. His current work focuses on both a negative and a positive dimension. On one hand, he focuses on threats to these systems and designs methods to mitigate negative behaviors; on the other, he looks for positive opportunities to mine and analyze these systems for developing next generation algorithms and architectures that can empower decision makers. He received a highly-competitive Google Faculty Research Award in 2013. He has published 30 peer-reviewed research papers in top journals and conferences such as TIST, SIGIR, WWW, CIKM, and ICWSM. His work was introduced by the MIT Technology review. Lee received his Ph.D. from Texas A&M in 2013.

James Caverlee is an associate professor in the Department of Computer Science and Engineering, Texas A&M University (caverlee@cse.tamu.edu). Caverlee's research focuses on web-scale information management, distributed data-intensive systems, and social computing. Most recently, he's been working on (1) spam and crowdturfing threats to social media and web systems; and (2) geosocial systems that leverage large-scale spatio-temporal footprints in social media. Caverlee is a recipient of the 2010 Defense Advanced Research Projects Agency (DARPA) Young Faculty Award, the 2012 Air Force Office of Scientific Research (AFOSR) Young Investigator Award, a 2012 NSF CAREER Award, and was named a Texas A&M Center for Teaching Excellence Montague-CTE Scholar for 2011-2012. Caverlee received his Ph.D. from Georgia Institute of Technology in 2007.

Calton Pu is a professor in the College of Computing, Georgia Institute of Technology (calton.pu@cc.gatech.edu). Pu's research inter-

ests are in the areas of distributed computing, Internet data management, and operating systems. He has published more than 250 papers in several system-related areas, including operating systems, transaction processing, systems reliability, security, and Internet data management. He worked on spam and denial of information (with several academic and industry partners), service computing (with IBM Research), and automated system management (with HP Labs). He has served on more than 100 program committees for more than 50 international conferences and workshops, including PC cochair of SRDS, ICDE, CoopIs, DOA, and general cochair for CIKM, ICDE, CEAS, and SCC. He is an affiliated faculty of Center for Experimental Research in Computer Systems (CERCS), Georgia Tech Information Security Center (GTISC), and Tennenbaum Institute. Pu received his Ph.D. from the University of Washington in 1986.

Route Planning and Visualization Using Geo-Social Media Data

*Presented by Hsun-Ping Hsieh,
Thomas Sandholm, and Cheng-Te Li*

Geosocial media data, produced by GPS-enabled devices, location-based services, and digital cameras, are ubiquitous thanks to the maturity of mobile and web technologies. Geographical activities of human beings are tracked in the form of trajectories. User-generated geosocial trajectory data enable a novel application, route planning, which aims to recommend travel routes satisfying trip requirements. In this tutorial, we aim to introduce two popular topics related to the analysis of geosocial media data: route planning and geodata visualization. The first part provides a broad review of recent advances on the route planning problem using GPS trajectories and uncertain trajectories that come from different sources and possess diverse properties and problems. Given geosocial query requirements depicting the desired routes, which are divided into three categories, that is, location, context, and social, we elaborate three mainstream approaches of route planning: graph search, pat-

tern mining, and inference or learning. The second part gives a technical introduction and practical advice on how to visualize geosocial data using various tools, including Google Maps, D3, Google Fusion Tables, Google Earth, Tableau Public, Open Street Map, Python Heatmap, Stamen, and Mongolabs. Hands-on examples are provided to elaborate techniques of cloud data storage, scalable geomarker positioning, and interactive maps for visualization.

Hsun-Ping Hsieh is a Ph.D. candidate at the National Taiwan University with research interests on geosocial and urban computing. He worked as a research intern at Microsoft Research Asia and received the Excellent Stars of Tomorrow award in 2013. His representative recognition includes the ACM KDD Cup 2010 First Prize, and Garmin Fellowship 2014.

Thomas Sandholm is a principal research scientist at HP Labs in Palo Alto, CA, USA. He holds a PhD in computer science from the Royal Institute of Technology in Sweden, and worked as research staff on distributed systems and geosocial media analysis at Argonne National Labs, Lund University and KAIST.

Cheng-Te Li is a postdoctoral researcher at the Institute of Information Science in Academia Sinica, with research interests on social networks, big data mining, and geosocial computing. He earned his Ph.D. in computer science from National Taiwan University. His representative international recognition includes Facebook Fellowship Finalist Award 2012, and ACM KDD Cup 2012 First Prize.

Large Scale Network Analytics with SNAP

Presented by Jure Leskovec and Rok Susic

Techniques for social media modeling, analysis and optimization are based on studies of large

scale networks, where a network can contain hundreds of millions of nodes and billions of edges. Network analysis tools must provide not only extensive functionality, but also high performance in processing these large networks.

The tutorial will present the Stanford Network Analysis Platform (SNAP), a general purpose, high performance system for analysis and manipulation of large networks. SNAP is being used widely in studies of web and social media. SNAP consists of open source software, which provides a rich set of functions for performing network analytics, and a popular repository of publicly available real world network datasets. SNAP software APIs are available in Python and C++.

The tutorial will cover all aspects of SNAP, including SNAP APIs and SNAP datasets. The tutorial is targeted toward entry level audience with some programming background, thus the Python API will be presented in more detail than the C++ API. The tutorial will include a hands-on component, where the participants will have the opportunity to use SNAP on their computers.

Jure Leskovec is an assistant professor of computer science at Stanford University. His research focuses on mining and modeling large social and information networks, their evolution, and diffusion of information and influence over them. Problems he investigates are motivated by large scale data, the Web and online media.

Rok Susic is a research associate in the Department of Computer Science at Stanford University.