Data-Driven Political Science

Ingmar Weber
Qatar Computing Research Institute
Doha, Qatar
ingmarweber@acm.org

Ana-Maria Popescu
Pinfluencer
San Francisco, California
anamaria.popescug@gmail.com

Marco Pennacchiotti
eBay Inc.
San Francisco, California
mpennacchiotti@ebay.com

ABSTRACT
The tutorial will summarize the state-of-the art in the growing area of computational political science. Like many others, this research domain is being revolutionized by the availability of open, big data and the increasing reach and importance of social media. The surging interest on the part of the academic community is matched by intense efforts on the part of political campaigns to use online data in order to learn how to best disseminate information and reach the right potential donors or voters. In this context, a tutorial can summarize existing methods in a fascinating, high-interest area and allow participants with diverse backgrounds to get inspiration from the methods and problems studied. The tutorial will feature seminal research concerning (i) political polarization, (ii) election prediction and polling, and (iii) political campaigning and influence propagation. The goal is not only to familiarize attendees with ideas from related conferences such as WWW, ICWSM or CIKM, but also to present ideas and quantitative methods closer to political science such as Poole's and Rosenthal's NOMINATE score for a politician's political orientation.

Categories and Subject Descriptors
J.4 [**Social and Behavioral Sciences**]: Sociology; H.3.5. [**Online Information Services**]: Web-based services

General Terms
Experimentation, Human Factors, Theory

Keywords
political science, big data, election prediction

1. TARGET AUDIENCE AND OBJECTIVE
This half-day tutorial targets researchers from the growing area of computational social science as well as, more generally, researchers with interests in web and data mining or its applications to political science. Attendees will be familiarized with the problems studied, the methods applied and the datasets used in a range of research work in computational political science both before but mostly after the start of the Web era. After the tutorial, participants will have an in-depth knowledge of the current state-of-the-art concerning (i) quantifying political polarization, (ii) election predictions and (iii) quantitative analysis of political campaigning.

WSDM'13, February 4–8, 2013, Rome, Italy.
ACM 978-1-4503-1869-3/13/02.

2. OUTLINE
The tutorial will cover seminal and important academic papers and techniques in the area of Data-Driven Political Science [1]. It will be organized in three topically coherent parts with a focus on (i) political polarization, (ii) election predictions and polling, and (iii) political campaigning and influence propagation. Each part will roughly last 1 hour.

2.1 Political polarization:
We will discuss studies and measures of political polarization where the object of polarization can be a politician, a social media user, a web site, a web search query, or text such as a political speech. Focus will be given to (i) how to define and validate political polarization and political orientation, (ii) how to compute it algorithmically, and (iii) the datasets being used in studies. The methods presented will include some well-known in political science, but not traditionally used in computer science, such as the NOMINATE score and its variants. While we will focus on the U.S. political system due to its simple left/right structure, we will discuss ways to generalize existing methods to multi-dimensional political systems [2,3,4,5,6,7,8,9,10,11,12,13].

2.2 Election predictions and polling:
Recently, many new approaches to using social media for election predictions have been presented. Some studies claim good results even using mere per-party tweet counting but others have cast serious doubts on those results and the methodology. We will give a survey of these different approaches and their results. We will also feature prediction market-based approaches and other novel methods to political forecasting and polling [14,15,16,17,18,19,20,21].

2.3 Political campaigning and influence propagation
Social media and the Web in general can be used to observe political activity but also to shape politics, in particular through political campaigning and attempts to start grassroot movements among enthusiastic political followers. We will describe studies covering either centralized political campaigning or the effects of "viral" propagation on political opinions. The focus here will be on methods that quantify the effects of such campaigns [22,23,24,25,26,27,28].

3. ORGANIZERS' BIOS
Ingmar Weber is a senior scientist at the Qatar Computing Research Institute. His research covers a wide subject area from classical information retrieval, to sponsored search, to interdisciplinary studies in web science. While at Yahoo! Research Barcelona, he led the development of the Political Search Trends and Political Hashtag Trends demos which assign political leanings to search queries and Twitter hashtags respectively. In August 2012 he taught the course "An

Introduction to Web Science" at the RuSSIR summer school, featuring elements from computational political science. Together with Ana-Maria and Marco he organized the PLEAD (Politics, Elections and Data) workshop at CIKM 2012.

Ana-Maria Popescu is a research scientist working with early stage social startups. Previously, she was a text mining scientist at Yahoo! Labs where she most recently worked on social media mining and user modeling. While at Yahoo!, she co-led the Twitter analytics effort for the Yahoo! News Election 2012 coverage. Ana obtained her Ph.D. in 2007 from University of Washington with a thesis in information extraction from the Web.

Marco Pennacchiotti is an applied scientist and principal engineer at eBay, where he leads the science on social media. Before joining eBay in June 2012, he was a research scientist at Yahoo! Labs, focusing on text mining, information extraction applied to big data and social media. At Yahoo!, he wrote papers focusing on user-profiling also applied to the political domain. Marco obtained a Ph.D. in natural language processing and information extraction in 2007 from the University of Rome, Italy.

4. REFERENCES

[1] Weber, I., Popescu, A.-M., and Pennacchiotti, M. 2012. PLEAD 2012: politics, elections and data. In *CIKM'12*.

[2] Pennacchiotti, M., and Popescu, A.-M. 2011. Democrats, republicans and Starbucks afficionados: user classification in twitter. In *KDD'11*.

[3] Weber, I., Garimella, V. R. K., and Borra, E. 2012. Political Search Trends. In *SIGIR'12*.

[4] Zhou, D. X., Resnick, P., and Mei, Q. 2011. Classifying the political leaning of news articles and users from user votes. In *ICWSM'11*.

[5] Borra, E., and Weber, I. 2012. Political Insights: Exploring partisanship in Web search queries. *First Monday* 17, 7.

[6] Laver, M., Benoit, K., and Garry, J. 2003. Extracting policy positions from political texts using words as data. *APSR*, 97, 2.

[7] Weber, I., Garimella, V. R. K., and Borra, E. 2012. Mining Web Query Logs to Analyze Political Issues. In *WebSci'12*.

[8] Weber, I., Garimella, V. R. K., and Teka, A. 2013. Political Hashtag Trends. In *ECIR'13*.

[9] Malouf, R., and Mullen, T. 2008. Taking sides: user classification for informal online political discourse. *Internet Research*, 18.

[10] Conover, M., Ratkiewicz, J., Francisco, M., Goncalves, B., Flammini, A., and Menczer. F. 2011. Political polarization on twitter. In *ICWSM'11*.

[11] Efron. M. 2004. The liberal media and right-wing conspiracies. In *CIKM'04*.

[12] Slapin, J., and Proksch. S. 2008. A scaling model for estimating time-series party positions from texts. *AJPS*, 52, 3.

[13] Poole, K. T., and Rosenthal, H. 1985. A spatial model for legislative roll call analysis. *AJPS*, 29, 2.

[14] Jungherr, A., Jürgens, P., and Schoen. H. 2012. Why the pirate party won the German election of 2009 or the trouble with predictions. *SSCR*, 30, 2.

[15] Tumasjan, A., Sprenger, T., Sandner, P., and Welpe, I. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *ICWSM'10*.

[16] Rothschild, D. 2009. Forecasting Elections Comparing Prediction Markets, Polls, and Their Biases. *Public Opinion Quarterly*, 73, 5.

[17] Gayo-Avello, D., Metaxas, P. T., and Mustafaraj. E. 2011. Limits of electoral predictions using twitter. In *ICWSM'11*.

[18] Metaxas, P., Mustafaraj, E., and Gayo-Avello, D. 2011. How (not) to predict elections. In *PASSAT'11*.

[19] Lui, C., Metaxas, P. T., and Mustafaraj, E. 2011. On the predictability of the US elections through search volume activity. In *IADIS'11*.

[20] Jin, X., Gallagher, A. C., Cao, L., Luo, J., and Jiawei, H. 2010. The wisdom of social multimedia: using Flickr for prediction and forecast. In *MM'10*.

[21] Hummel, P., and Rothschild, D. 2013. Fundamental Models for Forecasting Elections. In *AEA'13*.

[22] Cogburn, D. L., and Espinoza-Vasquez, F. K. 2011. From networked nominee to networked nation: Examining the impact of Web 2.0 and social media on political participation and civic engagement in the 2008 Obama campaign. *Journal of Political Marketing*, 10, 1-2.

[23] Small, T. 2008. The facebook effect? On-line campaigning in the 2008 Canadian and U.S. elections. *Policy Options*, November.

[24] Smith, A. 2009. The internet's role in campaign 2008. *Pew Internet & American Life Project*, 15.

[25] Kushin, M. J., and Yamamoto, M. 2010. Did social media really matter? College students' use of online media and political decision making in the 2008 election. *Mass Communication and Society*, 13, 5.

[26] Dimitrova, D., Shehata, A., Strömbäck, J., and Nord, L. W. 2011. The effects of digital media on political knowledge and participation in election campaigns: Evidence from panel data. *Communication Research*, November.

[27] Hong, S., and Nadler. D. 2012. Which candidates do the public discuss online in an election campaign? The use of social media by 2012 presidential candidates and its impact on candidate salience. *Government Information Quarterly,* 29, 4.

[28] Vitak, J., Zube, P., Smock, A., Carr, C. T., Ellison, N, and Lampe, C. 2011. It's complicated: Facebook users' political participation in the 2008 election. *CyberPsychology, behavior, and social networking*, 14, 3.