Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data

PABLO BARBERÁ | Social Media and Political Participation Lab | Department of Politics, New York University

Introduction
- 200M active Twitter users, 500M tweets per day. Can we use this massive amount of data to estimate quantities of substantive interest for social scientists?
- My contribution: a new method that generates ideology for any Twitter user on the basis of which politicians, journalists, and organizations they decide to "follow."
- Relevance: better understanding of the representativeness of Twitter population; precise ideal points are required to test hypotheses about political behavior and party competition.
- Key assumption: Twitter users prefer to follow accounts whose ideology is similar to theirs. Why? a) Homophily: clustering is a social network along common traits.
- b) Selective exposure: preference for opinion-reinforcing information.

Method
- Users and politicians’ ideology (θ, φ) are defined as latent parameters.
- Data: “following” decisions as a series of binary choices (Yij).
- Spatial following model: for n users, indexed by i, and m politicians, indexed by j:
  \[ P(Yij = 1 | \theta_i, \phi_j, \gamma) = \logit^{-1}(\alpha_j + \beta_i - \gamma_i) \]
  where:
  - \( \alpha_j \) measures popularity of politician j
  - \( \beta_i \) measures political interest of user i
  - \( \gamma_i \) is a normalizing constant
- Assuming independence, likelihood function is:
  \[ L(\theta, \phi, \gamma) = \prod_{i=1}^{n} \prod_{j=1}^{m} \logit^{-1}(\alpha_j + \beta_i - \gamma_i) \]
- Model is identified with unit variance restriction on \( \gamma \).
- The posterior density of each set of parameters is sampled using a Hamiltonian Monte Carlo algorithm (Hoffman and Gelman, 2011).

Data
- n = list of ~1200 highly visible political accounts from US, UK, Spain, Netherlands.
- m = followers of 33 accounts in each country, collected using Twitter REST API.
- Sample size: 300K in the US, 125K in the UK, 125K in Spain, 96K in the Netherlands.

Validation
- This method is able to correctly classify and scale Twitter users along the left-right dimension. Ideology estimates are highly correlated with:
  1. Ideal points for members of Congress based on their roll-call votes (DW-NOMINATE).
  3. Voter registration history in Ohio, after matching with Twitter accounts by full name and county.
  4. Left-right positions in Europe according to expert surveys (Chapel Hill – Bakker et al., 2012).

Application
- Social media and political polarization.
- To what extent is information diffusion through social media bounded by ideology?
- Is Twitter an “echo chamber” where individuals only receive information that reinforces their existing beliefs or, on the contrary, are cross-ideological interactions frequent?
- Previous studies find high levels of clustering along party lines (e.g., Conover et al., 2012)
- Data: 75 million tweets mentioning Obama or Romney from Aug. 20, 2012 to Nov. 6, 2012 collected using stream package for R (Barberá, 2013), available on CRAN.
- Analysis focuses on retweets (indication of information diffusion) sent by a sample of 300K users for which ideology was estimated.
- Results (see figure below): 1. Structure of conversations on Twitter supports idea of “online echo chamber”: most pairs of users involved in retweets mentioning Obama or Romney are ideologically similar.
- 2. Differences across ideological groups: liberals tend to engage more often in cross-ideological interactions, while right-leaning Twitter users show more polarized behavior.

Conclusions
- Structure of Twitter networks can be informative about policy positions: an “expert survey” with millions of respondents reporting how ideologically close they perceive politicians.
- Necessary to improve our understanding of how representative Twitter users are.
- Twitter presents several important advantages: vast amount of data, low cost, easy access, real-time nature. Unexplored source of information for old and new research questions.
- Meaningful comparisons of all types of political actors on same ideological scale: useful for wide range of empirical tests that require ideology estimates in Political Science.

References