



THE UNIVERSITY OF CHICAGO

Department of Statistics

MASTER'S THESIS PRESENTATION

ZHIYUAN GUAN

Department of Statistics
The University of Chicago

Topic Similarity between Twitter Users and Members of their Social
Networks

MONDAY, November 9, 2015, at 3:00 PM
Eckhart 110, 5734 S. University Avenue

ABSTRACT

The popularity of social network platforms such as Twitter has given rise to the availability of large amounts of user-generated texts. Much research has been devoted to effective clustering and topic discovery of Twitter users and their tweets. Hong and Davison (2010), among others, have implemented Latent Dirichlet Allocation (LDA) to topic model Twitter feeds. In this paper, we use their aggregation scheme to investigate the topic similarity of Twitter users and their *neighbors*, which we define to be users who are simultaneously followers and followees of a user. The dataset we collected contains 33 million tweets from 128,000 users. We demonstrate that Twitter users are indeed significantly more similar to their neighbors than strangers. We quantify the amount of information encapsulated by the topic proportions of a user's neighbors in the estimation of that user's topic proportions. Finally, we propose a synthetic estimate that leads to a more accurate recovery of the topic proportions of a user by incorporating the information of the neighbors.

For information about building access for persons with disabilities, please contact Laura Rigazzi at 773.702-0541 or send an email to lrigazzi@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following web site: <https://lists.uchicago.edu/web/arc/statseminars>.