Analyzing the Impact of Exogenous Shocks on Collaborative Volunteering Systems

Daniel M. Romero, Assistant Professor, University of Michigan School of Information
Ceren Budak, Assistant Professor, University of Michigan School of Information
Lionel Robert, Associate Professor, University of Michigan School of Information

Background: Crowdsourcing has become a very popular way of organizing crowds to produce high quality work. In some settings, individuals in crowds not only make independent contributions to the work, but also coordinate their contributions with others in the crowd. Two examples of this are Wikipedia and GitHub, where millions of people collaborate to produce encyclopedic articles and software, respectively. An important research topic that has been considered by various disciplines is to identify the properties of the crowd, or of the technology that facilitates crowdsourcing, that best equip a crowd to produce high quality work. Much of the prior work assumes that the environment around the crowd is fixed and ignores changes that occur, not within the crowd, but in its environment. However, collaborative crowdsourcing often takes place in highly disruptive environments, where changes are substantial, frequent, and unpredictable. For example, world events such as the death of a celebrity can generate shocks to the corresponding Wikipedia articles that are characterized by a large influx of viewers and editors. We know very little about the dynamics of crowd collaboration during times of shocks. To address this, we combine empirical and theoretical analyses to develop an understanding of how the attributes of a crowd and their future performance are affected by exogenous events.

Description: Students will work under the supervision of Professors Daniel M Romero, Ceren Budak, and Lionel Robert. Research projects will involve large-scale analysis of collaborative crowdsourcing platforms such as Wikipedia and GitHub. Students will examine the impacts of external shocks in both Wikipedia and GitHub collaborative crowds. To accomplish this, the students will perform several tasks such as (i) developing of algorithms or other techniques to identify external shocks; (ii) creating and validating measures of crowd collaboration and (iii) collecting large data sets from both communities through APIs and/or web scraping. Finally, students will analyze large data sets to determine the effects of external shocks on both communities. Students will develop their skills in programming, data science, social network analysis, data mining, information retrieval, and social science theories of large collaborative crowds.

Mentoring Plan: The student will meet regularly throughout the project with the research team. Beyond discussion of the progress of the research project, the meetings will include discussion on developing personalized goals for academic, professional, and personal growth. These goals will include identifying research interests, outlining specific skills and/or knowledge the student hopes to gain through the experience, and determining strategies for reaching their goals.

Desired qualifications: The student should have strong programming skills (in Python, R, Java, or C++) and a strong interest in applying their skills to a research project. Experience with large
scale data gathering, manipulation, and exploration and knowledge in machine learning, data mining, or social network analysis is a plus.