Thesis: Influence Level Prediction on Social Media Through Multi-Task and Sociolinguistic User Characteristics Modeling

Abstract: Prediction of a user's influence level on social networks has attracted a lot of attention as human interactions move online. Influential users have the ability to influence others' behavior to achieve their own agenda. As a result, predicting users' level of influence online can help to understand social networks, forecast trends, prevent misinformation, etc. The research on user influence in social networks has attracted much attention across multiple disciplines, from social sciences to mathematics, yet it is still not well understood. One of the difficulties is that the definition of influence is specific to a particular problem or a domain, and it does not generalize well. Another challenge arises from the fact that all user interactions occur through text. Textual data limits access to non-verbal communication such as voice. These facts make the problem challenging.

In this work, we define user influence level as a function of community endorsement, create a strong baseline, and develop new methods that significantly outperform our baseline by leveraging demographic and personality data. This dissertation is divided into three parts. In part one, we introduce the problem of influence level prediction, review influential research across different disciplines, and introduce our hypothesis that leverages user-centric information to improve user influence level prediction on social media. In part two, we answer the question of whether the language provides sufficient information to predict user-related information. We develop new methods that achieve good results on three tasks: relationship prediction, demographic prediction, and hedge sentence detection. In part three, we introduce our dataset, a new ranking algorithm, RankDCG, to assess the performance of ranking problems, and develop new user-centric models for user influence level prediction. These models show significant improvements across eight different domains ranging from politics and news to fitness.

Committee:

- Professor Rebecca Levitan, Mentor, Brooklyn College
- Professor Elena Filatova, City Tech College
- Professor Michael Mandel, Brooklyn College

Outside member:

- Professor Andrew Rosenberg, Google