

# Week 8-1: Paper Summaries

## *CE-510 Seminar: Social Media Mining*

Student Name: Jiaqi Guo    NetID: JGR9647

### ■ Online Human–Bot Interactions: Detection, Estimation, and Characterization

#### ■ Paper's contribution:

1. Presents a framework for analyzing many features extracted from social media data and user metadata including friends, Twitter content and emotions, network patterns, and activity time series
2. Their detection system was tested using existing public datasets and collected manually annotated datasets. The newly annotated data was used to retrain the model trained from the existing data.
3. Using different models and score thresholds to distinguish between humans and bots, we predict the presence of 9-15% social bots on Twitter.
4. Analyze the differences in user behavior between social connections and news flows: Humans are more likely to interact with a human-like account. Some bots are more random about who they connect to

#### ■ Possible Improvement Directions:

1. There is no simple set of rules to assess whether an account is human or bot. Therefore, whether manual data annotation will be subject to a certain subjective influence, resulting in inaccurate final prediction results. Can we use some officially confirmed robot accounts as data sets to improve the accuracy of prediction

### ■ Everyone' s an Influencer: Quantifying Influence on Twitter

In this paper they investigate the attributes and relative influence of 1.6M Twitter users by tracking 74 million diffusion events that took place on the Twitter follower graph over a two-month interval in 2009.\

To calculate the influence score for a given URL post, they tracked the diffusion of the URL from its origin at a particular "seed" node through a series of reposts—by that user's followers, those users' followers, and so on—until the diffusion event, or cascade, terminated.

#### ■ Possible Improvement Directions:

1. The possibility that "ordinary influencers" are more cost-effective in many cases is a statistical model based on observed data and does not imply causation. It's quite possible that content streamed by an external source will be delivered in a completely different way than the content the user chooses for himself. Similarly, while we have considered various possible cost functions, other assumptions about cost are of course possible and may lead to different conclusions.