To Follow or to Un-follow:

Simulating the Effects of Tweet Patterns on User Following

CSCI 8980: Simulation

Kristin Dahl

10/24/2011

I. Introduction

Over the past several years, Twitter has rapidly grown into one of the most popular means by which people can share and discuss current events, ideas, and personal thoughts. One of the hallmarks of Twitter is that it allows users to "follow" each other. Following provides an easy way for users to receive up-to-the-minute "tweets" from any other Twitter user, including friends, celebrities, politicians, and even brands.

Oftentimes, celebrities will use Twitter to advertise their latest movie or hit record. To politicians, it is a popular campaign tool, and for brands, it is an inexpensive means of advertisement. For these and many other Twitter users, it stands to reason that reaching as many followers as possible is important.

The problem explored by this simulation is to determine whether different tweet characteristics have an effect on the number of followers for a particular user. Is it possible for a user's follower base to continue to grow and never shrink? Behavior of the tweeter can create numerous results, and in this study we will attempt to discover the probabilistic attributes that make a user more likely to un-follow somebody.

This paper is divided into five sections. In Section II, we will further explore the research question that we aim to answer with this simulation. In Section III, the approach for the methodology of this simulation will be discussed based on multiple probabilistic aspects of establishing social connections between twitter users. In Section IV, we will discuss the results of the simulation as individual models and some additional discussion in Section V.

II. The Research Question

The specific question that is discussed in this paper is: what qualities will have the greatest influence on follower retention? The answers to this question are unclear because users follow each other for a variety of reasons, including interests, social influence, tweet quality, and whether they maintain an offline relationship. Because of the complexity associated with these social patterns, we have developed a simple model that focuses on the following three attributes: user influence, tweet quality, and tweet frequency.

We expect highly influential users to have a higher number of followers than users with little influence. However, does this change if a highly influential user has very poor quality tweets (or never tweets)? Using this simulation, we attempt to present which probabilistic independent behaviors affect the overall outcome of users' desire to un-follow somebody.

III. Methodology

In this simulation we attempt to generate independent probabilistic variables based on user interaction with Twitter followers to determine which attributes result in the maximum number of followers for a particular tweeter.

The tool base used in this simulation was Repast Simphony 2.0. The simulation was used to generate data based on three independent adjustable variables: User Influence, Tweet Quality, and Tweet Frequency. The simulation was executed separately for each independent variable to determine which of these attributes would contribute to the probability of "un-following" a tweeter. To determine initial values we construct a standard probabilistic neutral

value for each static variable. For each adjustable variable, we attempt to find the optimal value such that any given user would choose to un-follow a tweeter they are currently following.

Generating Followers

Within the Repast tool base, we generated 99 user agents and 1 "tweeter" agent. Each agent is initially placed at a random position in a grid of 25x25. At each simulation iteration, users are relocated to a random location adjacent to their current location. If ever one of the user agents is located on the same grid location as an allocated tweeter agent, the user agent automatically follows the tweeter agent and subscribes to their tweets.

Simulating Un-following

Again, our simulation of following is based on three independent variables: User Influence, Tweet Quality, and Tweet Frequency. At each iteration, the tweeter agent has a chance of tweeting (based on an assigned tweet frequency). If the tweeter agent tweets, all following user agents are immediately notified.

Each user agent keeps track of the length of their "relationship" with the tweeter agent by maintaining a count of the time units ("ticks") that have passed since they began to follow the tweeter. In addition, the total number of tweets a user agent has received is also maintained. The tweet frequency is then calculated by dividing the number of tweets received by the length of the relationship. If the tweet frequency for a specific user exceeds a randomly assigned tweet threshold, the probability that the user is going to continue following the tweeter decreases.

However, the probability of an un-follow is also affected by the tweeter's influence value and tweet quality value. At each tick, the user calculates a "follow probability," which is a value between 0 and 1 that is a combination of the calculated tweet frequency, the tweeter's tweet quality, and the tweeter's influence. Then, an un-follow probability is created by generating a random number between 0 and 1. If this random un-follow probability is greater than the follow probability, the user will un-follow the tweeter.

IV. Results

At each test for each independent variable, we assign neutral values for all other variables such that they do not affect the outcome of the current simulation. Each simulation was run for 5000 ticks with 99 user agents and 1 tweeter agent. A data point was generated at every 100 ticks.

Figure A is a graph that shows the change in followers based on the tweet frequency of the tweeter. The quality and influence values were each held constant at 1.0. It is apparent that a low frequency of 0 and a high frequency of 1 results in low followership. These frequencies represent inactive or overactive users who may deter current followers. The maximum number of followers peaks at a frequency of 0.5, simulating the optimum balance of tweeter updates. Frequencies of 0.25 and 0.75 share similar results, stabilizing at approximately 35 followers.

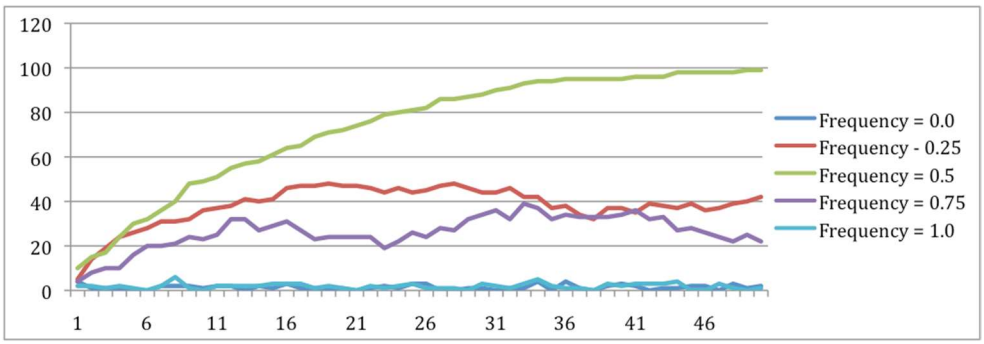

Figure A: Tweet Frequency

Figure B is a graph that shows the change in followers based on the tweet quality of the tweeter. The frequency and influence values were each held constant at 0.5 and 1.0, respectively. As excepted, higher quality tweets retained more followers and lower quality tweets deterred them. Interestingly, there does not exist a linear one-to-one correlation between user following and quality. At a quality value of 0.5, the average number of followers stabilizes at 20-30 followers. As quality increases, the percentage of followers increases to a maximum number of followers of 96.

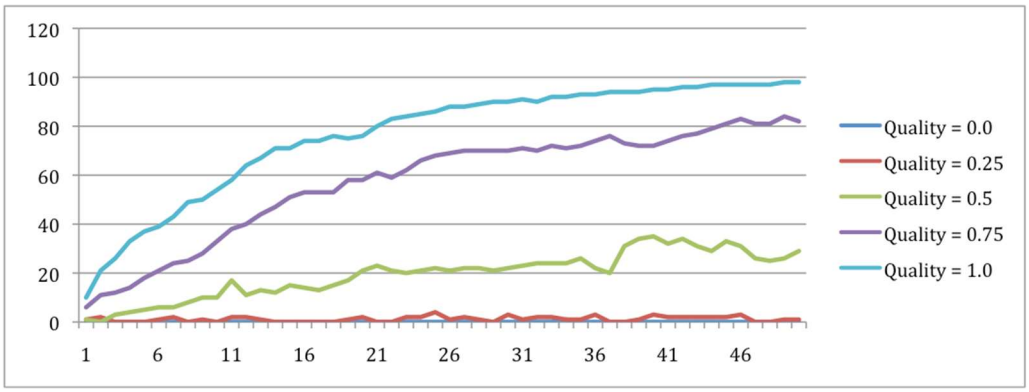

Figure B: Tweet Quality

Figure C is a graph that shows the change in followers based on the influence of the tweeter. The frequency and quality values were each held constant at 0.5 and 1.0, respectively. Similar to tweet quality, as predicted, higher tweeter influence resulted in a higher following and lower influence resulted in lower following. Again, it appears that there does not exists a linear correlation to the tweeter influence and the number of followers; generating a large gap between an influence of 0.5 and 0.75.

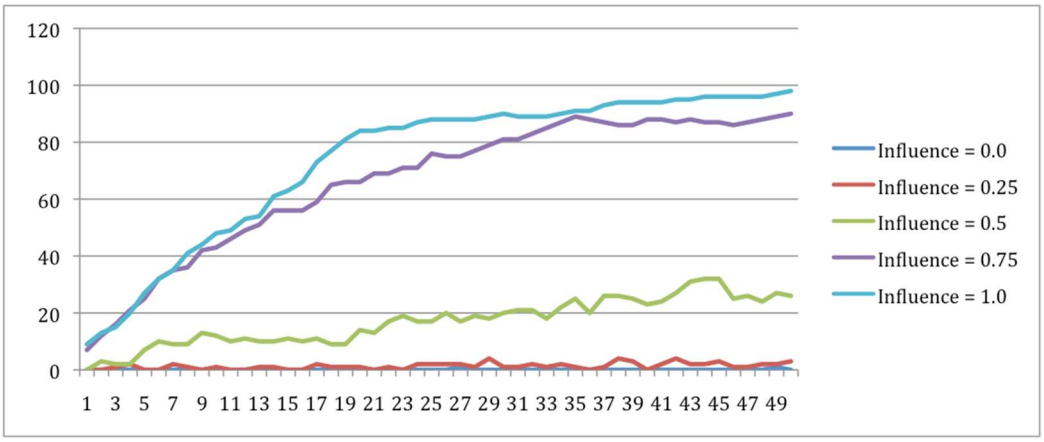

Figure C: Tweeter Influence

At each iteration each user determines if they wish to un-follow the tweeter. For tweet frequency, it is apparent that optimal frequency displays less variation in un-following the given tweeter. Frequencies with a lower following trend towards a more erratic behavior. Compared to frequency, tweet quality and influence display increased erratic behavior at optimum values, but are much more stable at their lowest values.

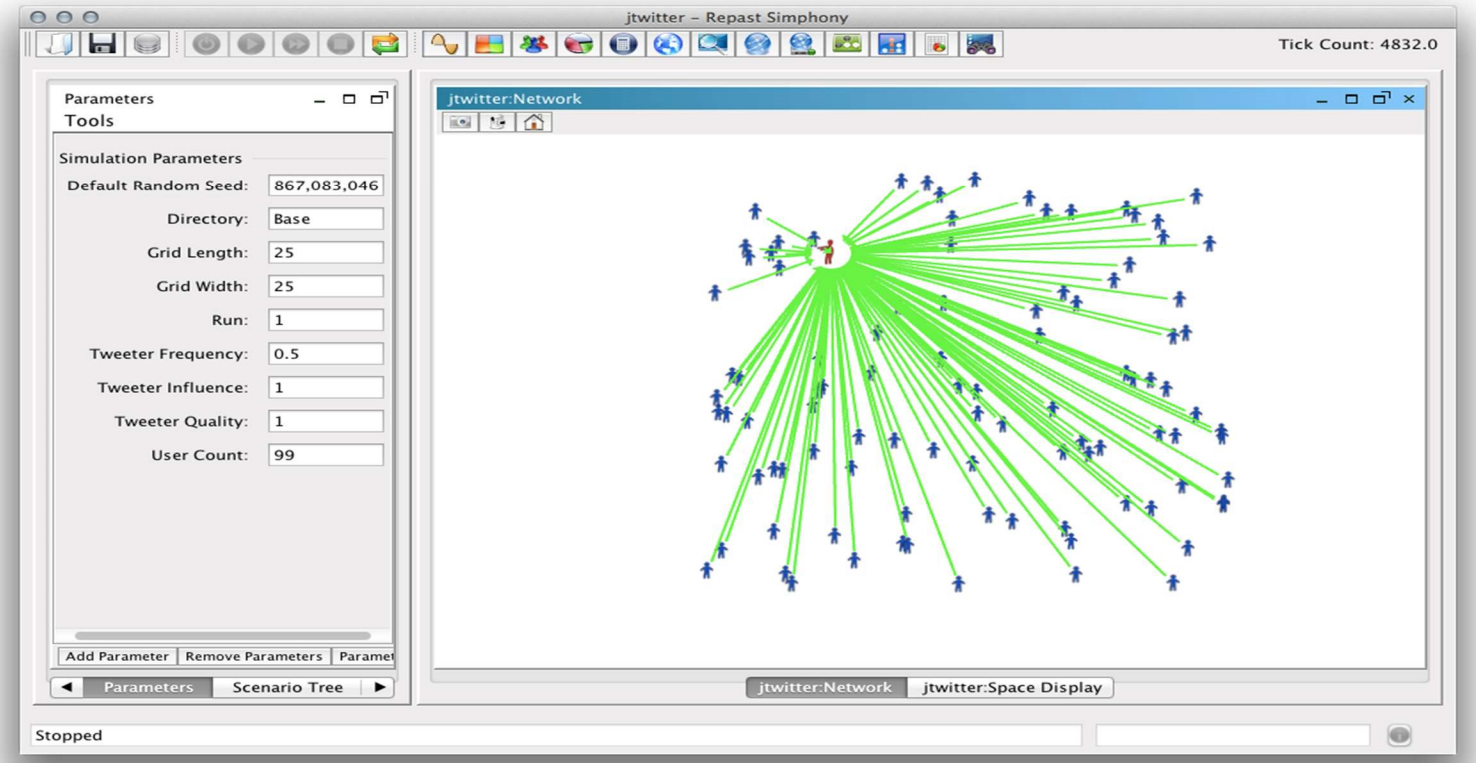

Figure D: Simulation Interface

V. Discussion

Of the three characteristics presented in this paper, we show that tweet quality and user influence are big factors in follower retention. However, even the most influential users with the highest tweet qualities still have to be mindful of the tweet frequencies, as this was found to be the most volatile characteristic.

In this simulation, tweet quality and tweeter influence were both assumed to have equal importance to the user. However, there still exist many external factors that may alter these results. We can never fully understand every permutation of every reasoning why un-following occurs, and it remains a question as to whether we can fully discover a correlation of motives for un-following Twitter users.

The complexity associated with this research topic allows for many opportunities of further research. Additional factors could be taken into account, including the length of the relationship, social ties based on an initial meeting, relationships on other social networks, or personality constraints. There are many different avenues that can be ventured down.