rtweet: Collecting Twitter Data

A proposal for

POLS 904: Statistical Computing

by Michael W. Kearney

Fall 2016

1 Background

My research examines the intersection of new media environments and political communication using innovative quantitative research methods. My current work analyzes "big data" to better understand the relationship between selective exposure and social media. In working to systematically collect and wrangle large amounts of data, I created and now maintain *rtweet*, the successor package to *twitteR* for collecting Twitter data. Originally developed to collect data for my dissertation, *rtweet* now features numerous diverse methods for interacting with Twitter APIs and represents the future of open-source efforts in collecting Twitter data.

2 Proposal

Most of the work toward creating the package is done. I have already published *rtweet* on CRAN and have received a fair amount of feedback via email and Github. I still want to add a couple new features and implement at least a handful of improvements, but my next big goal is to advertise the package and have my work count toward something a little

more tangible. With this in mind, for my semester project in *Statistical Computing and Foundations*, I propose I write and submit an article about *rtweet* to the *R Journal*.

3 Details

As an introductory piece, the article will provide three different overviews. First, the article will provide an introduction to working with Twitter's APIs. This section will also include some basic discussion about the the nature and limitations of publically available Twitter data. Second, I will provide an overview of the major functions and features in *rtweet*. Third, I will provide quick examples of different statitical analyses, including sentiment [textual] analysis, social network analysis, and time series.

The R Journal encourages authors to discuss similar packages, so I would like to provide some points of comparison between *rtweet* and *twitteR*. In the third section described above, I would also demonstrate how to make *rtweet* compatible with several statistical packages. This may also feature integration with *data.table* or other ways to work with relational databases.