MAT482: Rough Plan

Fin Vermehr, Binderiya Adishaa, Quentin Vilchez

March 3, 2018

1   Introduction

We would like to study how information spreads (or goes viral) on a social media.
For simplicity, we will focus on Twitter data.
We will use a modified version of the SEIZ model. Since, we would like to separate the susceptible
population into two. The idea being that we think that the virality of certain 'events' is due to the
fact that there are on-going discussion about topics around the event. We should look at 2 cases,
one with obvious discussion around it and one without, to see the difference.
(MAKE SEIZ diagram)

Figure 1: Example of Network Graph
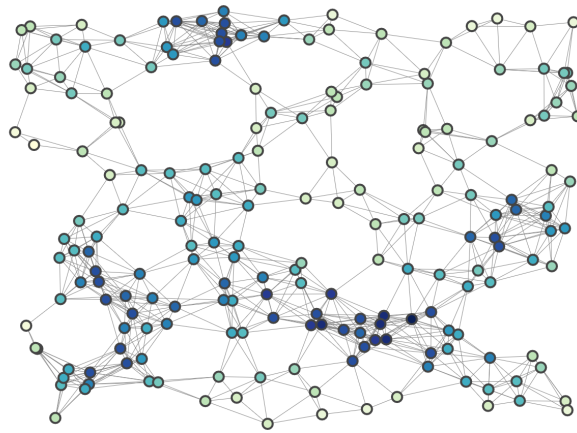
Then using a clustering algorithm, find major clusters. These clusters represent ideas. The
hashtags that are most in the centre of these clusters describe the purest form of the idea. Label
each cluster by this.

From the clusters of ideas found, ideas or trends that reach their peak popularity among twit-
ter users in a relatively short amount of time and slump in popularity soon after will be the focus
of the first part of the project. The reason is that ideas or trends of this kind might resemble
a disease outbreak making it possible to use SIR like model to model spread dynamic in time.
Moreover, it might be possible to extract contagiousness or contact rate for a given idea or trend.
Further analysis can be done to compare contagiousness in different geographical areas and across
political spectrum or to break down the features make up this type of ideas.

Next we must try to describe the underlying diffusion between these ideas. The SIR model for describing the spread of disease is ideal for this situation. The only two major adjustment that have to be made is that often ideas compete, and also that adopting an idea is a gradual process. You may not start spreading a newly acquired idea immediately as vocally as a long-time supporter.

There would for example be the idea of pro-abortion and anti-abortion. Therefore we would have an SIResque model that has the S, which are the susceptible, I which are the infected, and D which are the disbelievers. One problem that has to be resolved, is how to find opposing ideas.

$$
\begin{cases}
\dot{S} & = \Lambda - \beta S \frac{1}{N} - bS \frac{Z}{N} - \mu S, \\
\dot{I} & = \beta S \frac{I}{N} - \mu I \\
\dot{Z} & = bS \frac{Z}{N} - \mu Z
\end{cases}
$$

where b and β denote the per capita rates of idea rejection and adoption by susceptible respectively. Λ is the recruitment rate,