

# MAT482: Rough Plan

Fin Vermehr, Binderya Adishaa, Quentin Vilchez

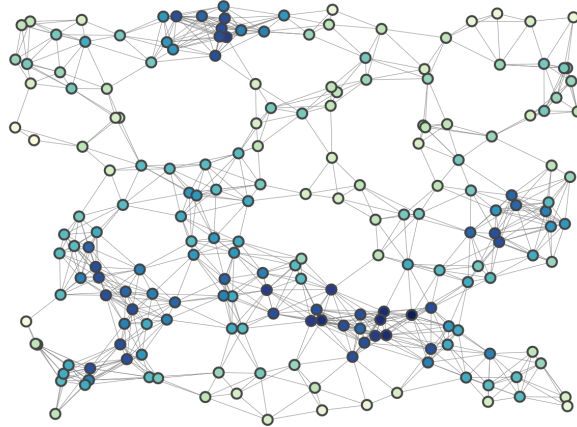
February 2, 2018

## 1 Introduction

Define vectors  $\mathbf{h}$ ,  $\mathbf{p}$  and matrix  $\mathbf{A}$ . Let all  $h_i \in \mathbf{h}$  be popular hashtags of 2017, so  $\mathbf{h}$  contains all major hashtags. Let all  $p_i \in \mathbf{p}$  describe the popularity of its corresponding hashtag  $h_i$ , so  $\mathbf{p}$  contains the popularity of the hashtags. Then let all vectors  $\vec{a}_i \in \mathbf{A}$  contain the accounts that retweeted hashtag  $h_i$ . Thus  $a_{if}$  contains a single account that retweeted  $h_i$ .

For each  $h_i$  find a coefficient  $r_{ij} = r_{ji}$  describing the relatedness between the  $h_i, h_j$ . This coefficient should be calculated from how often accounts in  $\vec{a}_i$  retweet  $h_j$  and how often accounts in  $\vec{a}_j$  retweet  $h_i$ . In other words, the coefficient  $r_{ij}$  is determined by looking at the overlap between  $\vec{a}_j$  and  $\vec{a}_i$ . Using this coefficient, plot a network graph like the following:

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 twitter 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.