Engineering Probability and Statistics Course Project
Flow of information in social networks
Part II:interference Between News

Seyed Iman Hosseini Zavaraki*

May 18, 2018

1 Terms and conditions

By submitting a report on this course project you accept the following terms:
1) You waive any right to object to the grading procedure and final grade.
2) The grading system is not final and can be subject to change without notice.
3) Your submission may be forfeited in case of objection to the grading system.
4) Any submission after the deadline, WILL BE DISCARDED.
5) It is the duty of the student to submit a detailed and readable report including the supplementary materials (e.g. source code of programs he has made) to illustrate his/her work.
6) What you hand in should be your own work. The consequences of cheating are severe and you are strongly advised that you do not cheat.

2 Introduction

In the first phase, you were introduced to the problem of percolation and learned about the importance of Probability and Statistics as the language with which we can investigate and analyze the phenomena around us. As we saw, the solid mathematical foundation of probability makes it possible to prove precise results but more important to engineers, is the ability to take a problem from the real world, and make abstract models for it so that we can then predict and control the problem via these models.

But how can we be sure that a model "works"? What limits the discrepancy of a mathematical model with reality? The means of evaluation of models is mostly via applying it to real data from the past. For example, if a financial model works on data from 1990 to 2000 then we can be more optimistic about its performance in the future.

*hosseini.iman@yahoo.com

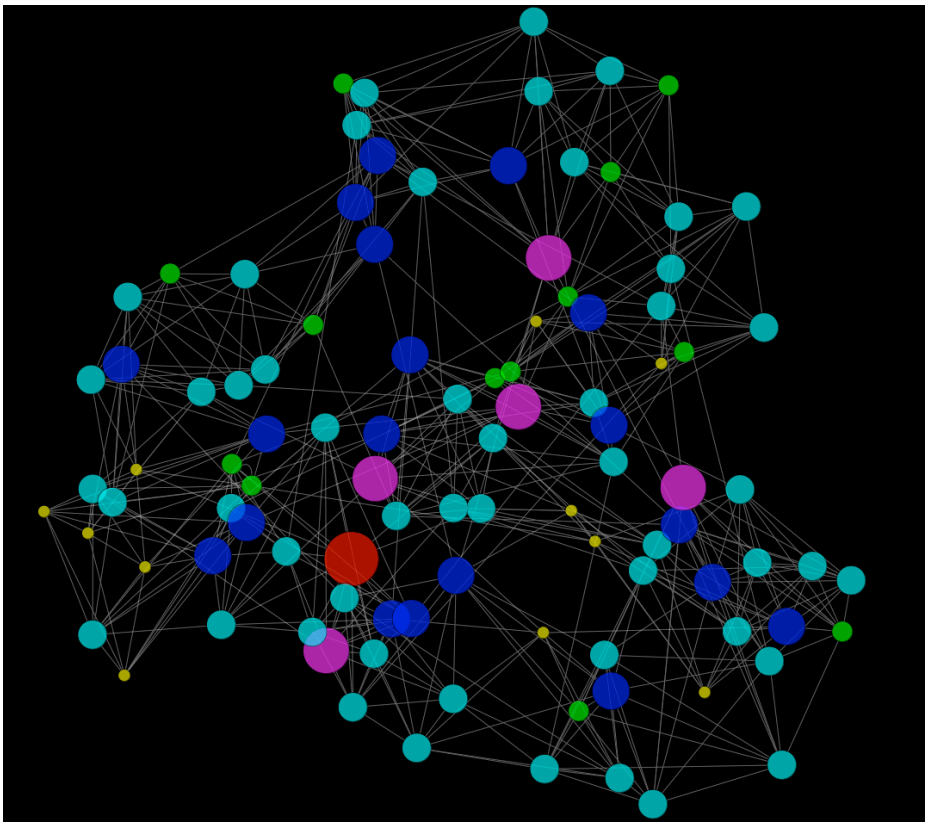

Figure 1: Watts-Strogatz graph with 100 nodes

In this part of the project, we will be investigating the phenomena of interference between flow of different news. First we present a model for flow of a single news, then we extend that model to take into account the effect of competing news. The intuition is that when there are various important news, the spread of them is hindered because attention is divided between them. These models are variously studied by companies to then influence opinions by means like timely leaks of information and thus have lead to data of social networks becoming very expensive. A recent example of use of such methods is the case of Cambridge Analytica during US presidential election.[1]

3 Synthesis Of Graphs

The first ingredient of our model would be a graph encapsulating the relation between different players. The nodes, can be twitter profiles, with an edge from X to Y implying that X follows Y. Or they can be people, connected according to their connections: a edge

[1] https://gizmodo.com/now-facebook-says-it-shared-the-data-of-up-to-87-milli-1824990995

would imply the two people have a probability of coming in close contact with each other. This case could be a basis to analyze the breakout of an epidemic virus.

For many cases, there is no decent real data available so there are synthetic methods to generate graphs and use those as basis for further work. Each of these models aim to incorporate specific feature of real world graphs, to be reflective of real use-cases.

Task 2.1 (36 points)

Investigate the following graph generation models (the models are not complicated, and a Wikipedia look-up suffices to understand how to make them) and implement a program in the language of your choice, to generate graphs based on these models and to output the result both as a file, and as a graphical visualization. Also for each model, plot the degree distribution and the average path length [as a function of size].

+) Barabasi-Albert model

+) WattsStrogatz model

+) Mediation-driven attachment model

4 Cat Working With A Computer

What is more viral than a cute cat working with a computer? An image of such cat is being shared worldwide and Tommy the Cat is now an internet celebrity. So how does that happen?

We have a graph modeling profiles and their follower relationships. To model time, imagine a turn-based scheme. At each turn, nodes view other profiles, and share posts. Initially some initial nodes post the image.

Now we also assign to each node, two boolean values $informed$: which represents if that node has seen the image of the cat and *shared* which represents if that node has shared the image and a positive real number less than 1 and call this *conductance*. This is the probability that a given person, after seeing Tommy, would share the image. Also, for each edge, representing X following Y, assign a weight w that is the probability that X at each turn, looks up Y page. If Y has shared the image, then X would become $informed$ -If a node becomes $informed$ it would stay informed forever- then according to it's conductance it would choose whether to share the image or not. -If it chooses not to share it, it would not share it forever- Then, if X shares the image, then starting from next turn, the followers of X would also be $informed$ of the image if they look up X profile.

Task 2.2 (16 points)

You are given two files, $Nodes.txt$ and $Edges.txt$ storing the data of a social network. On each line of $Nodes.txt$, the first number is the node id, the third number is the *conductance*. (ignore other fields) And in $Edges.txt$ each line represents an edge, the

Figure 2: Cat becomes worldwide sensation

first number is the follower, the second number is the profile being followed, and the last number is the weight.

At turn 0, nodes 0, 2, 4 are $informed$ and are *sharing*. Simulate the system for 240 turns, and plot the number of $informed$ nodes as a function of time. Do the same simulation on a Barabasi-Albert graph with 30000 nodes. (* For each test, you must run the simulation at least 10 times, and what you would plot must be the average of those 10 runs.)

5 Cat Vs Spinner

This time, a video of a fidget spinner is stealing the spotlight from the cat.

Task 2.3 (14 points)

To extend our model, now if a node decides to share one of the news, it cannot not share the other one for 10 turns. Besides that, everything else is the same: now we have two $informed$ and *shared*, one for each news. At the first turn, nodes 0, 2, 4 are

sharing news 1 and nodes 1, 3, 5 are sharing news 2. For each graph type (the given graph, and BA graph), plot the number of $informed$ as a function of type for the case of single and double news in one graph to compare the effect of interference.