HERITAGE INSTITUTE OF TECHNOLOGY

Diffusion of information using contagion models in real life small world networks

Authors
ANKITA PANDEY
DEBOPRIYO BHATTACHARYA
HARSHITA PRIYADARSHINI

Mentor
PARTHA BASUCHOWDHURI
ASSISTANT PROFESSOR
Heritage Institute of Technology

December 9, 2017

Abstract:

The study of epidemic disease has always been a topic that mixes biological issues with social ones. When we talk about epidemic diseases we will be thinking of contagious diseases caused by biological pathogens which spread from person to person. The patterns by which epidemics spread through groups of people is highly determined by network structures within the population it is affecting. To develop this idea more fully, we relate it to another basic structural issue- the fact that groups of people can be connected by very short paths through the social network. The fact that social networks are so rich in short paths is known as the small-world phenomenon or the "six degrees of separation."

Introduction:

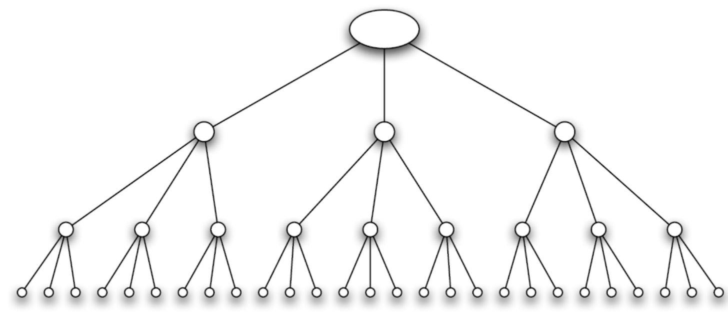
There are clear connections between epidemic disease and the diffusion of ideas through social networks. Both diseases and ideas can spread from person to person, across similar kinds of network that connect people, and in this respect, they exhibit very similar structural mechanism to the extent that the spread of ideas is often referred to as social contagion.

Add to this, the small world effect could have substantial implications for networked systems. A small world network is a type of graph in which both nodes are not neighbors of one another but most of the nodes can be reached from the other by a small number of connections.

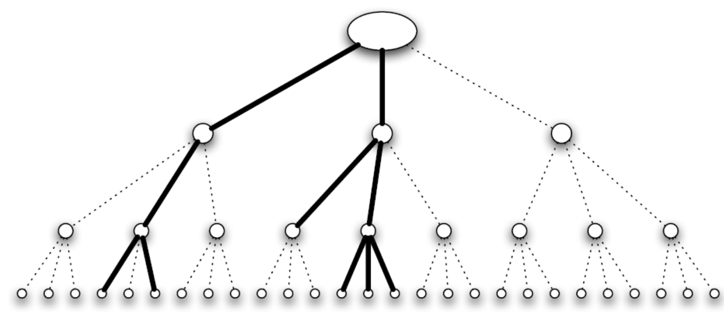
In this work, we are concerned with modelling of the contagion model for spread of information in a social network, while correlating its characteristics with the properties of a small world network. The data under consideration is from the social media platform, Twitter.

Connections to the Diffusion of Ideas and Behaviors:

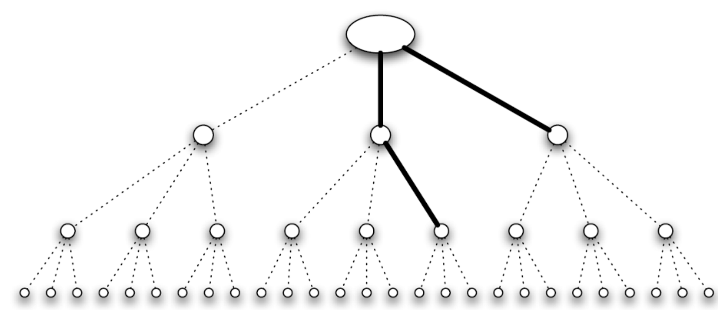
In the context of discussion about networks, the biggest difference between biological and social contagion lies in the process by which one person "infects" another. With social contagion, people are making decisions to adopt a new idea or innovation. With diseases, on the other hand, not only is there a lack of decision-making in the transmission of the disease from one person to another, but the process is sufficiently complex and observable at the person-to-person level that it is most useful to model it as random. That is, we will generally assume that when two people are directly linked in the contact network, and one of them has the disease, there is a given probability that he or she will pass it to the other.



(a) *The contact network for a branching process*



(b) *With high contagion probability, the infection spreads widely*



(c) *With low contagion probability, the infection is likely to die out quickly*

Models of the spread of disease:

1. **THE SI MODEL**

   The within-host dynamics of the disease is reduced to changes between a few basic disease states.

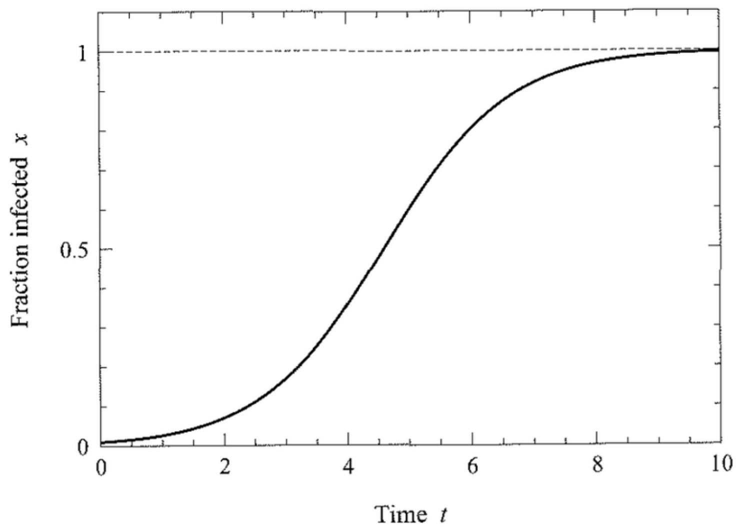
   In the simplest version there are just two states,

   (a) Susceptible: An individual in the susceptible state is someone who does not have the disease yet but could catch it if they come into contact with someone who does.

   (b) Infected: An individual in the infected state is someone who has the disease and can, potentially, pass it on if they come into contact with a susceptible individual.

   Here, we are focused more on what's happening at the level of networks and populations than on what's happening within the bodies of the individual population members. The allowed transitions between states can be represented by flow charts like this simple one for the SI model.

   An S-shaped "logistic growth curve" for the fraction of infected individuals is observed. The curve increases exponentially for short time, corresponding to the initial phase of the disease in which most of the population is susceptible, and then saturates as the number of susceptible dwindles and the disease has a harder and harder time finding new victims.



2. **THE SIR MODEL**

In the SI model individuals, once infected, are infected (and infectious) forever. For many real diseases, however, people recover from infection after a certain time because their immune system fights off the agent causing the disease. Furthermore, people often retain their immunity to the disease after such a recovery so that they cannot catch it again. To represent this behavior in our model we need a new third disease state, usually denoted R for recovered. The corresponding three-state model is called the susceptible-infected-recovered or SIR model.

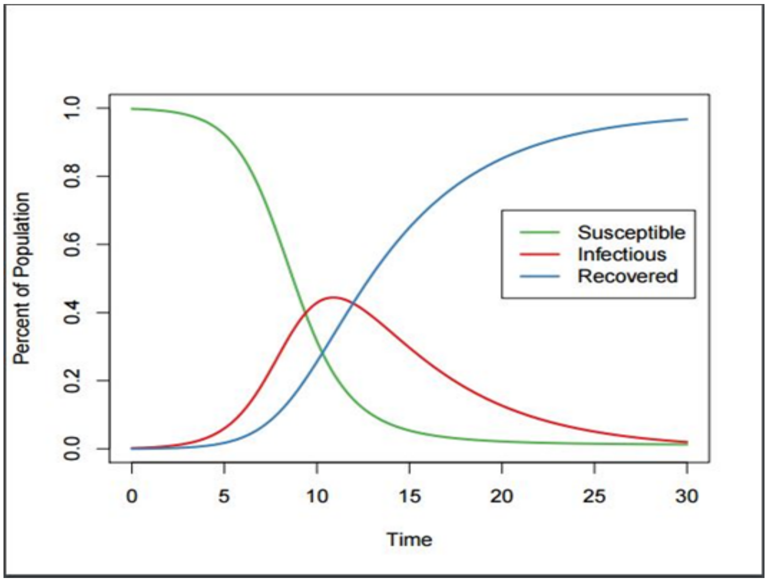
The flow chart for SIR model:

The dynamics of the fully mixed SIR model has two stages.

(a) In the first stage, susceptible individuals become infected when they have contact with infected individuals. Contacts between individuals are assumed to happen at an average rate β per person as before.

(b) In the second stage, infected individuals recover (or die) at some constant average rate γ.

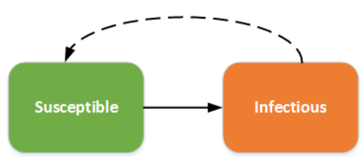
The three curves in this figure show the fractions of the population in the susceptible, infected, and recovered states as a function of time:



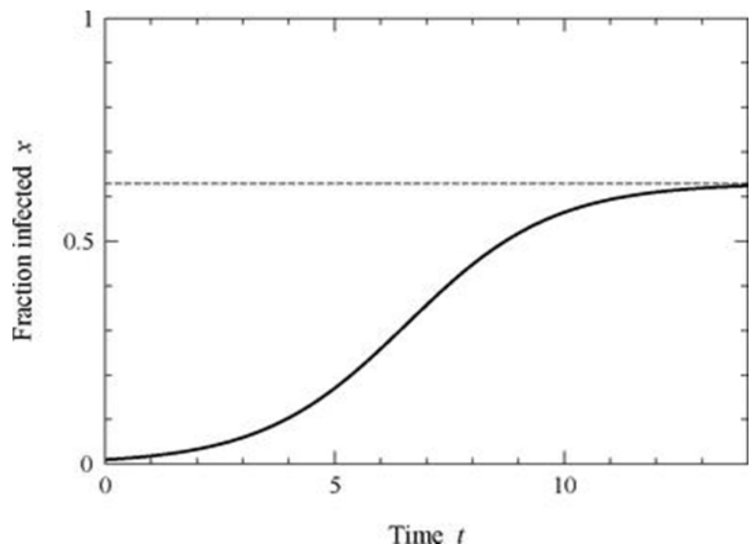
3. **THE SIS MODEL:**

   A different extension of the SI model is one that allows for reinfection, i.e., for diseases that don't confer immunity on their victims after recovery, or confer only limited immunity, so that individuals can be infected more than once. The simplest such model is the SIS model, in which there are just two states, susceptible and infected, and infected individuals move back into the susceptible state upon recovery.

   Flow chart for the SIS model:



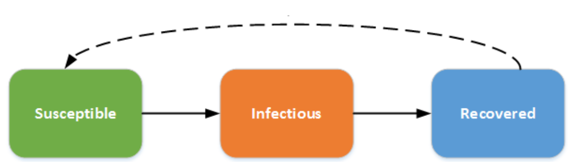
   Fraction of infected individuals in the SIS model. The fraction of infected individuals in the SIS model grows with time following a logistic curve, as in the SI model. Unlike the SI model, however, the fraction infected never reaches unity, tending instead to an intermediate value at which the rates of infection and recovery are balanced.



4. **THE SIRS MODEL:**

   In this model individuals recover from infection and gain immunity as in the SIR model, but that immunity is only temporary, and after a certain period of time individuals lose it and become susceptible again. A new parameter δ to represent the average rate at which individuals lose immunity.

   Flow chart for the SIRS model:



GOALS:

To determine the type of compartmental model present in a large social media platform like Twitter. This will help study the role of public opinion as well as its influence on the people connected to a particular person who expresses his or her opinion on any topical issue. This type of influence can be referred to as Behavioral Contagion. It will indicate the propensity for certain behavior exhibited by one person to be copied by others, who are somehow in contact with the former.

The small world phenomena have gone on to be popularized as "Six Degree of Separation" hypothesis which is a theory that everyone is just six or few steps away from any other person in the world. Considering the three main properties of small world, which are clustering coefficient, degree distribution and average path length, a mathematical model between the compartmental model determined and the small world network will be worked upon.

This will help study the diffusion and influence of information on social media platform, Twitter.

PROGRESS:

For extraction of data from Twitter, web crawling was carried out using a program written in Python. 'urllib' and 'BeautifulSoup' are the Python libraries used in the program. Currently, any hashtag can be searched for on Twitter, for a time/range during which one wants. A certain account handle can also be specified from which one wants to extract tweets.

The problems with using Twitter API is that since it is limited on a free account, it allows one to make only 15 requests every 15 minute window for every OAuth token (account one uses to authenticate with).

The problem with Selenium is that even though it performs as required, it being a browser, takes time to load all the images and execute the javascript code etc., which can be programmed. But, it gets complicated and difficult. 'urllib' when used correctly is found to be the fastest way to scrape tweets from Twitter as per one's needs. All the necessary tweets and data extracted can be stored in .csv files.

References

[1] Mark E. J. Newman. *Networks: An Introduction*. Oxford University Press, 2016.

[2] David Easley, Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, 2016.

[3] Frank Mittelbach, Michel Goossens, Johannes Braams. *The LaTeX Companion*. Addison-Wesley, 2013.