CSE 8803 EPI: Data Science for Epidemiology, Fall 2020

Lecturer: B. Aditya Prakash
Scribe: Jiaming Cui

Lecture # 5
September 9, 2020

---

1 Summary of Lecture Content

This lecture introduces some network-based models of contagions, which can be applied to describe many common phenomena in people's daily life.

The lecture starts with the Independent Cascade Model (IC Model). Then we discussed how the post popularity drops off over time, which follows the power law with the slope close to -1.5.

Next, we discussed the SpikeM model, which captures the rise and fall patterns in social media. The main idea of the SpikeM is about un-informed bloggers, external shock at time n_b, and infection. Besides, the infectiveness of a blog-post at age n also follows the power law. Then, we discussed the complex contagion, in which a node will act on the contagion exposed by this node's neighbor. For different types of contagions, the exposure curve can also be different.

Next, we briefly mentioned the models that are not covered in the lecture, including decision-based models, voter models, hybrid models, 2-mode models, and so on. We also analyzed the complexity of the models. Then, we discussed the taxonomy of the mathematical models for epidemiology, which can be intuitively classified into two main categories: ODE based models and Stochastic ODE based models.

Finally, we discussed models for multiple contagions, which includes the predator-prey Lotka-Volterra model and competing contagions model. In the basic competing contagions model, it has three states: Susceptible, Infected1, Infected2. Besides, we can also extend such a basic model with an interaction factor ϵ to capture the interaction between two competing objects.

2 Independent Cascade Model

Independent cascade model is one of the most fundamental contagion models [1]. In Figure 1, each edge (u, v) has weight p_{uv}. When one node becomes active, it will activate each of its neighbors with the probability on the edge.

The advantage of the IC model is that it can be intuitive and has some nice properties. However, the disadvantage is that it may involve too many parameters since for each edge in the graph, we need a weight parameter p_{uv}.

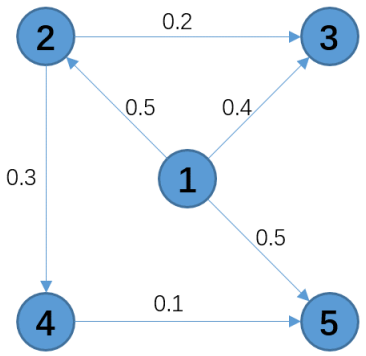

Figure 1: Independent cascade model

Next, we discussed the properties that real world cascades have if they are following the IC model. The analysis [2] on blog data of 45,000 blogs participating in cascades shows how the popularity (number of in-links) changes to lags (days after post). The analysis result is in Figure 2, which demonstrates that the relationship between the popularity of the post over time in log-scale and the lags in log-scale has an exponent (the slope for the log-scale curve) of -1.6, which is close to -1.5 in Barabasi's stack model [5]. Here popularity denotes the number of in-links to the post after some time after the post was initially published.

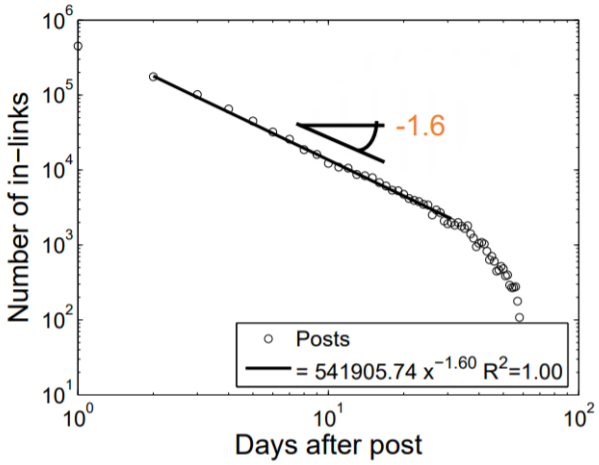

Figure 2: The popularity of the post follows power law

3 "Rise and Fall" (SpikeM)

This seems to be different that the standard IC model (where Figure 2 would be expected to be exponential). How can we extend the IC model to capture such patterns? SpikeM model [4] is one model which leverages the above observation and captures the rise and fall patterns in social media. The main idea of the SpikeM are the following three points:

1. Un-informed bloggers
2. External shock at time t_b

3. Infection

Here, the contagion model is that all bloggers are un-informed at time t_0. At time t_b, the external shock was exposed to one of the bloggers, then this shock will infect the others via the edges following the IC model from t_{b+1}.

The infectiveness of a blog-post at age n in SpikeM model is assumed to be

$$f(n) = \beta \times n^{-1.5} \tag{1}$$

where β is the strength of the infection (quality of news) and $f(n)$ is the decay function (how infective a blog posting is). Here, the reason why we are assuming the $f(n)$ in the form is that this form satisfies the power law discussed in Section 2 with an exponent of -1.5.

One of the advantages of the SpikeM model is its ability in "What-if" forecasting on upcoming spikes shown in Figure 3, which means that given (1)first spike, (2)release date of two sequel movies, (3)access volume before the release date, SpikeM can forecast not only the tail-part but also the rise-part [4].

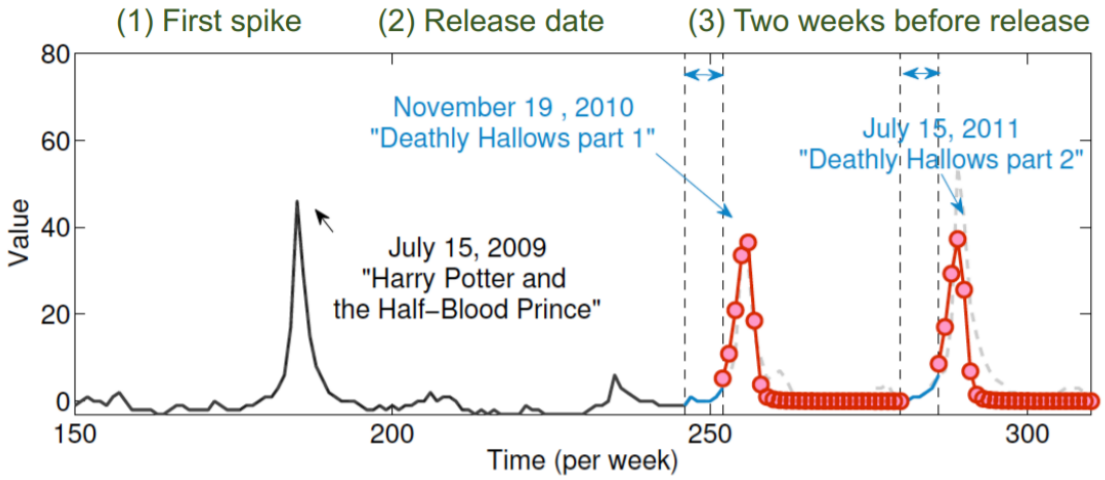

Figure 3: The forecast of SpikeM on upcoming spikes

4 Complex Contagions

Complex contagion focuses on the process from exposure to adoptions. To be more specific, it focuses on how the node acts on the contagion when the node's neighbor exposes the node to contagion.

With the intuition that the probability of adopting a new behavior depends on the number of friends who have already adopted, the exposure curves are used to demonstrate the relationship between the probability of adopting and the number of friends adopting.

Specifically, the exposure curves may show different patterns for different objects. For example in Figure 4, the objects are the virus, then the more one's neighbors are infected with the virus, the higher probability that he will be also infected with the virus. However, the marginal effect of the number of infected neighbors will decrease. SIR/IC type models have this kind of behavior.

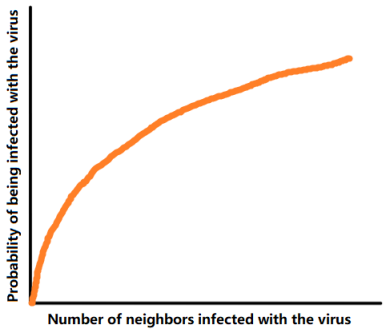

Figure 4: Different kinds of exposure curves

Another example is in Figure 5, where the objects are iPhone, then the situation is that one may not choose to buy iPhone when seldom of his friends are using iPhone. However, when more of his friends are starting to use iPhone, then the probability that he will also choose an iPhone will increase fast. However, if he has decided to not use the iPhone, then any further increase in the numbers of his friends using iPhone will make little difference. Decision type models have this kind of behavior.

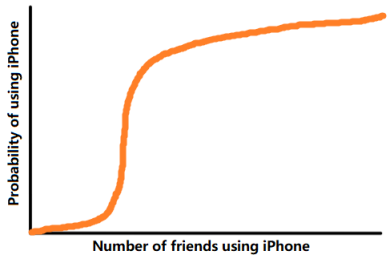

Figure 5: Different kinds of exposure curves

The application of the exposure curves is common in the marketing agency, where they estimate the adoption curve for different products [6]. For example in Figure 6, when the number of recommendations received is 3, the customer will have the highest probability of purchasing the product.

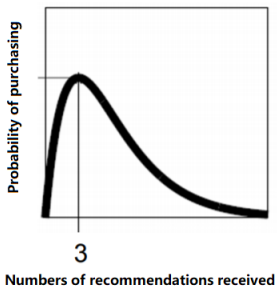

Figure 6: Adoption curve

To model the shape of the curve, the conception of persistence and stickiness of the curve are proposed [6]:

1. Persistence of P is the ratio of the area under the curve P and the area of the rectangle of $length(max(P)), width(D(P))$, which measures the decay of exposure curves.

2. Stickiness of P is $max(P)$, which measures the probability of usage at the most effective exposure.

Here, persistence describes how fast the curve goes up. For example in Figure 7, idioms and music have lower persistence than that of a random subset of hashtags of the same size while politics and sports have higher persistence than that of a random subset of the same size. The reason is that politics and sports can be more general and acceptable for people than idioms and music.

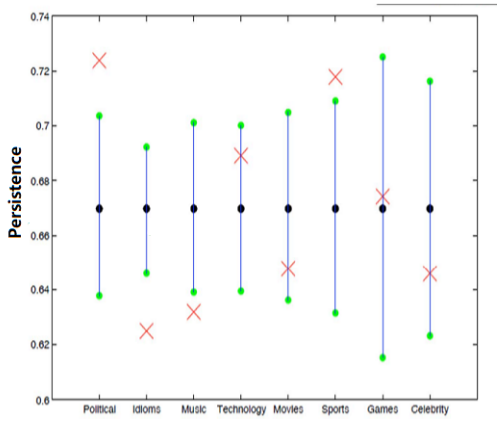

Figure 7: Persistence

As for stickiness, higher stickiness means that people will have a higher probability to accept it at the peck. For example in Figure 8, music has higher stickiness than random while tech, and movies have lower stickiness than random.

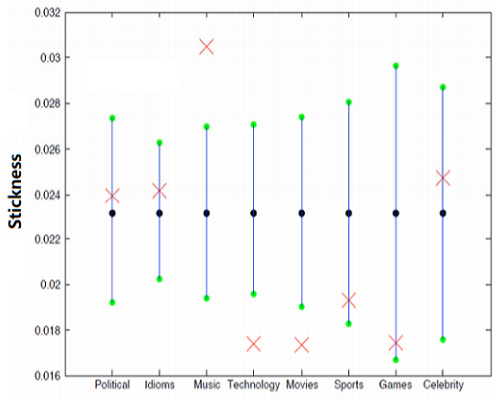

Figure 8: Stickiness

5 Model Taxonomy

Some of the models that haven't been covered in the lecture in detail include decision-based models, voter models, hybrid models, and 2-model models on heterogeneous networks.

Linear threshold model [1] is one of the decision-based models, where weights (thresholds) are attached to nodes. In linear threshold model, the propagation happens if the number of the neighbors exceeds the threshold, which models the peer pressure situation. In voter models, each node picks the 'ideology' of a random neighbor.

Next, we discussed the complexity of the models. Generally speaking, the mathematical models are the least complex ones while the multiagent models are the most complex since the objects are heterogeneous. Social network models and other simulation models have medium complexity compared with mathematical models and multiagent models.

Besides, the mathematical models of the epidemiology can be classified into two main categories shown in Figure 9: Differential equation based models and network based models. Differential equation based models include ODE based models and Stochastic ODE based models. Network based models include Random net models, template based models, and realistic social network models. Besides, there also exist some spatially explicit models like patch based models and cellular automata models.

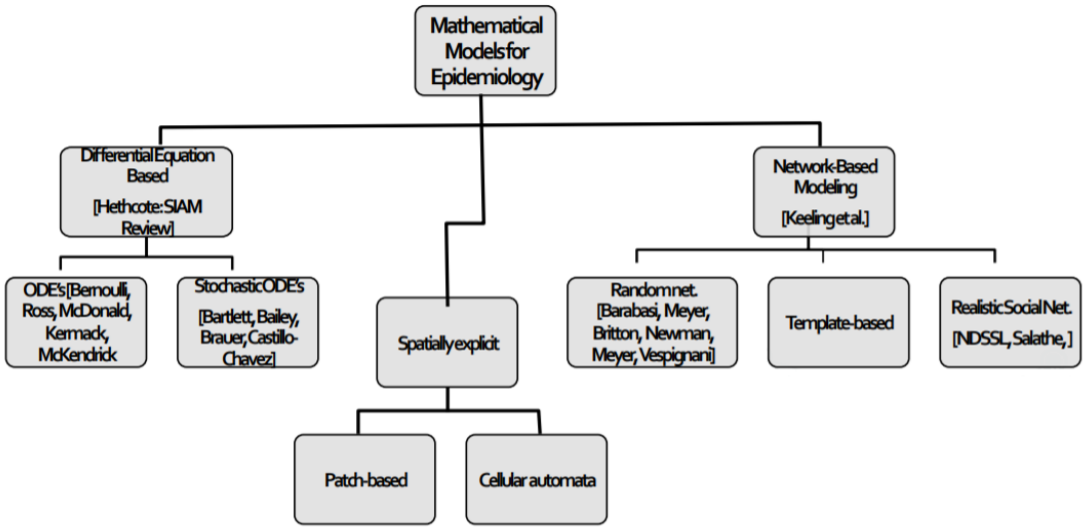

Figure 9: The continuum of complexity of the models [3]

6 Multiple Viruses

In this section, we will focus on the situation that two different kinds of viruses are spreading in one network at the same time.

6.1 "Predator-Prey" Model: Lotka-Volterra Model

First, the lecture demonstrates the 1910 "Predator-Prey" model, which has the following assumptions:

1. The prey population finds ample food at all times.
2. The food supply of the predator population depends entirely on the prey populations.
3. The rate of change of population is proportional to its size.
4. During the process, the environment does not change in favor of one species.

Therefore we will have the following equations for the prey and predator:

$$\begin{aligned} \frac{dx}{dt} &= \alpha x - \beta xy \\ \frac{dy}{dt} &= \delta xy - \gamma y \end{aligned} \tag{2}$$

The fixed points for the equations above are:

$$\begin{aligned} x &= 0 \\ y &= 0 \end{aligned} \tag{3}$$

which is an unstable point that represents the extinction of the prey and predator.

$$\begin{aligned} x &= \frac{\gamma}{\delta} \\ y &= \frac{\alpha}{\beta} \end{aligned} \tag{4}$$

which is a non-hyperbolic fixed point that represents the co-occurrence of the prey and predator. Specifically, although not obvious, it's inferred that the predator and prey populations cycle will oscillate around this fixed point.

6.2 Competing Contagions

The competing contagion is also common in our real life: For example, the competition between iPhone and Android, Blu-ray and HD-DVD... In biological area, the examples are like the common flu and avian flu.

An example model for competing contagion is a modified flu-like model, which has the mutual immunity that one person can only be infected by one virus at one time. We have 3 states: infected1, infected2, and susceptible.

Here, the susceptible people will be infected with the virus 1/2 with an attack rate β_1 and β_2. Besides, infected1/infected2 people will recover to susceptible state with cure rate δ_1 and δ_2. Therefore, we have the following ODEs:

$$\begin{aligned} \frac{dS}{dt} &= -(\beta_1 + \beta_2)S + \delta_1 I_1 + \delta_2 I_2 \\ \frac{dI_1}{dt} &= \beta_1 S - \delta_1 I_1 \\ \frac{dI_2}{dt} &= \beta_2 S - \delta_2 I_2 \end{aligned} \tag{5}$$

Besides, we can also extend the models above to allow the interaction of virus 1 and virus 2 by introducing the interaction factor ϵ. Different ϵ corresponds to different real-life examples.

1. $\epsilon = 0$: Full mutual immunity. For example: Hepatitis vs. Hepatitis vaccine.

2. $\epsilon < 0$: Partial mutual immunity. For example: iPhone and Android.

3. $\epsilon > 0$: Cooperation. For example: iPhone and iPad.

The extended model satisfies the following ODEs:

$$\begin{aligned}
\frac{dS}{dt} &= -(\beta_1 + \beta_2)S + \delta_1 I_1 + \delta_2 I_2 \\
\frac{dI_1}{dt} &= \beta_1 S - (\delta_1 + \epsilon\beta_2)I_1 + \delta_2 I_{12} \\
\frac{dI_2}{dt} &= \beta_2 S - (\delta_2 + \epsilon\beta_1)I_2 + \delta_1 I_{12} \\
\frac{dI_{12}}{dt} &= \epsilon\beta_1 I_2 + \epsilon\beta_2 I_1 - (\delta_1 + \delta_2)I_{12}
\end{aligned} \tag{6}$$

which can be intuitively described by the following diagram shown in Figure 10.

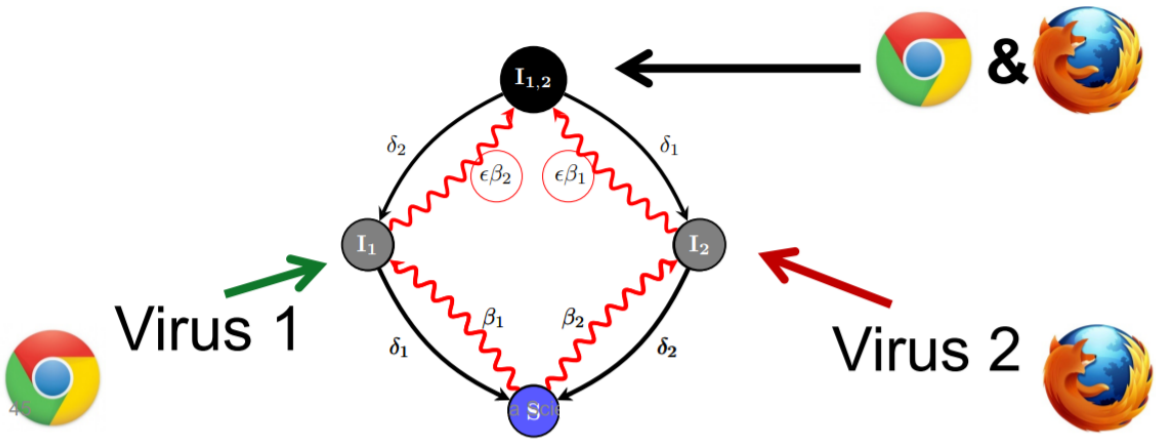

Figure 10: Diagram for the extended model

Besides, this ODE can be easily extended to network bared models if we make S, I_1, I_2, and I_{12} as the fraction of the people and β_1, β_2, δ_1, δ_2 as the probability.

References

[1] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, 2003.

[2] Jure Leskovec, Mary McGlohon, Christos Faloutsos, Natalie Glance, and Matthew Hurst. Patterns of cascading behavior in large blog graphs. In *Proceedings of the 2007 SIAM international conference on data mining*, pages 551–556. SIAM, 2007.

[3] Madhav Marathe, Naren Ramakrishnan, and Anil Kumar S. Vullikanti. Computational epidemiology. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014.

[4] Yasuko Matsubara, Yasushi Sakurai, B Aditya Prakash, Lei Li, and Christos Faloutsos. Rise and fall patterns of information diffusion: model and implications. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 6–14, 2012.

[5] Joao Gama Oliveira and Albert-László Barabási. Darwin and einstein correspondence patterns. *Nature*, 437(7063):1251–1251, 2005.

[6] Daniel M Romero, Brendan Meeder, and Jon Kleinberg. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 695–704, 2011.