**Reaction Paper # 4: Network Cascade**

Mohammad Elkholy  
Computer Science and Engineering  
 The American University in Cairo  
 New Cairo, Cairo, Egypt  
moelkholy@aucegypt.edu

**SUMMARY**

**Cascading Behavior in Large Blog Graphs**

The authors in this paper explore how blogs on the WWW influence eachother, study the decay in popularity and propose a model to simulate the cascading effects observed in these blog graphs.

The authors extracted their dataset from a much larger dataset, they were looking for blogs that actively interacted with eachother in order to analyze the differences across time and topological structure of these networks. They measured temporal differences across days and not hours as they didn’t normalize for time differences. Self-edges were removed as they do not add much value to studying information diffusion. They consider three networks: Blogosphere, blog network and post network. All 3 are different representations of the same dataset. They do not take into account trivial cascades (a single post).

They noticed that traffic time across the week were different, more posts were made on weekdays than weekends. They thus had to normalize the data by dividing by the percentage of a blog links on that certain day. They first start to measure the decay of popularity, they did so by calculating the number of new citations against the number of days after a post was made. They all followed a power-law distribution with exponent of around -1.5.

They then tackle the network topology, The dataset is a weighted directed graph. Half the blogs belonged to the largest connected component while the rest were isolated. In contrast, the posts network had 98% of its nodes being isolated and are very weakly connected. The in- and out-degree distributions followed a heavy tail distribution with exponent of -1.7. They noticed that in- and out-links of active blogs were not correlated at all. Similarly, the blog-to-blog links followed a heavy tailed distribution, as well as in- and out-degrees of the post network. The only difference was that the number of posts per blog showed a knee at around 40 posts per blog, this was due to their bias of using active blogs.

they then try to observe the cascades, their structure and their incidence. Almost all of them were trivial and the rest were trees or chains, with the exception of of G107 which contains posts that link to other posts inside the cascade. They then measured the in- and out-degrees of these cascade structures. The out-degree cascade distribution is truncated as they had to remove links to other urls, news media or images. In-degree exponent was stable suggesting that posts will be linked/cited even if they appear late in the cascade.Cascade size follows a heavy-tailed distribution for all types of cascades. Chains however have outliers which might be due to posts refering posts back and forth. The probability of obersing a cascade of size n follows a Zipfian distribution. They also notice that cascade sizes would have to grow exponentially in order to notice a linear increase in diameter, which is a property of balanced trees and very sparse graphs.

The authors then tackle the problem of modeling such a cascade, and follow the SIS model in their algorithm. Each node is infected for only one timestep and then becomes susceptible to infection again, this goes on till no new nodes are infected. They validate their model using numerical simulations. They simulated the model 10 times and reported on the average. The selection of first infected nodes was experimented on as well. They noticed that picking a starting blog in a biased way leads to very large cascades that do not follow the heavy tailed distributions.

**Contagion**

The author in this paper tries to tackle the idea of a contagiom: an action that when exhibited, causes other people to also perform it. The author provides bounds for local interaction systems where such a contagion would manage to infect everyone in system. The author focuses on Binary interaction systems where a player can choose action 1 or 0 based on the actions of players that are in his/her immediate surroundings. He also provides and proofs a threshold for which a contagion would be able to propagate throughout the system and infect the entire population. He provides proof that the threshold holds if two main properties are satsified:

* Low neighbor growth: number of players who can be reached in k steps grows less than exponentially
* Local interaction system must be sufficiently uniform: some number alpha exists where for all players very far from the core group, alpha of their neighbors are closer to the core group.

The local interaction system the author consists of an infinite population, whom all interact with a very small subset of the population. He then defines certain properties like bounds on their neighbors and connectedness. Each player in a finite subset X can have two possible actions. The best action is dependent on the probability that the other player chooses a specfiic action. Each player tries to maximize his/her payoff based on interactions with his/her neighbors. A configuration is the sum of of the payoffs from all the player’s interactions. However, due to the player baseing the best action on the probability, it is not hard to see that the best response depends on the proportion of its neighbors following that specific action.

The author then provides rational for the contagion threshold by using several examples with varying radii of interactions as well as dimensions. For all of them we can see that the threshold follows some pattern for each specific radius and dimension we use, but they all tend to ½ when the radii tend to infinity.

The author then specifies some measures of interactions between nodes in the population. A group is p-cohesive if it has proportion p of its interactions within the group, i.e. the number of intra-cluster edges is p for each node in the group. He then mentions how weak or strong links affect the growth of action adoption. Strong links mean that neighbors are more likely to interact with eachother which implies lesser growth in the long run. He then mentions Erdos distances and how low neighbor growth is satisfied if the sum of nodes with erdos distance n from finite group X multiplied by some factor gamma > 1 & gamma ^-n tends to 0 as n grows to infinity. He then explains how erdos distance grows very rapidly if neighbors know each other. He then talks about delta uniformity, which is satsified for some value delta where the difference in proportions of lower labels are maxmized and bounded by delta.

The author then talks about more properties of contagion threshold using the above defined measures. He places upper and lower bounds on the threshold and proportion p for p-cohesive groups. He then uses these propositions to prove that for all local interaction systems, the threshold is <= ½ and if it satisifes low neighbor growth and delta uniformity then threshold is >= ½-delta. The intuition is based on the idea that contagion must spread slowly, and if it spreads fast then it spreads to non interactive players and dies. Uniformity and low neighbor growth ensure slow spread.

The author then touches on the idea of randomness and how it can affect a local interactive system but then proves that it can still be bounded by some new threshold.

**COMMENTS**

**Cascading Behavior in Large Blog Graphs**

This paper was rather easy to read and follow as they just analyzed the graph properties and structure for the most part. The generative cascade model was also very intuitive and simple. They however made several assumptions in order for this model to work. Blogs about certain topics are less likely to behave in the manner they observed , as let's say blogs about hockey are much less cited than blogs about celebrities. they also did not provide any rational for choosing their probability threshold for infecting neighbors. They also did not discuss the results of starting choice of blogs in depth, which is something that is important particularly if we want to find out if algorithms that try to find optimal starting nodes will actually affect the entire network or not.

**Contagion**

Very hard read. Few problem arise, this is a binary interaction system, it is not generalizeable to more complex interaction systems with more actions to be taken. It also does not take into account systems where equilibra exists between the two actions. Moreover, it assumes we have an infinite player base, which is not possible.

**REFERENCES**

[1] Stephen Morris, Contagion, The Review of Economic Studies, Volume 67, Issue 1, January 2000, Pages 57–78, https://doi.org/10.1111/1467-937X.00121

[2] Leskovec, Jure & McGlohon, Mary & Faloutsos, Christos & Glance, Natalie & Hurst, Matthew. (2007). Cascading Behavior in Large Blog Graphs Patterns and a model.