

Study of Diffusion in Networks

Priya Balachandran Mary

Abstract

This report introduces a tool which predicts the spread of a behavior in a network once contagions are introduced to the network. The tool helps choose contagions randomly in a network or choose nodes with higher degree centralities as contagions. The tool helps compare the spread of a contagion's behavior in Erdos-Renyi graph and graphs that follow a power law degree distribution. We discuss the working of this tool and the underlying assumptions through the remainder of this paper.

1 Introduction

1.1 Motivation

The spread of a behavior in a network is directly linked to characteristics of nodes involved in the network and their neighborhood. A study by Coleman, Herbert Menzel and Elihu Katz[1966] showed how the use of an antibiotic called 'Tertracycline' spread, when it introduced by a major drug manufacturer, 'Phizer'. The observations from this study was that in major cities, the initial prescriptions were by a few doctors and the doctors having many physician contacts in their network started prescribing the medicine quickly thereafter. This could be because doctors in the cities have more direct contacts and thus a wider neighborhood. The physician contacts of the doctor who adopted the drug could have had enough motivation from a physician's perspective to adopt it. We capture this neighborhood (degree) of an influencing node and level of willingness (threshold) and try to find reasons behind the spread of a behavior.

1.2 Problem Statement

Through the developed tool we study the spread of a contagion's behavior through an Erdos-Renyi graph and a graph with a power law degree distribution. The goal is to study the role of a network's structure in the spread of the contagion's behavior. We compare the spread of a behavior in giant components for both graphs by infecting random nodes in each of these giant components and through simulation find the number of infected nodes at the end of the simulation. We repeat the same process for nodes in both giant components with the higher degree centrality values. In both cases, we provide a threshold for the nodes indicating their willingness to adopt an infected node's behavior.

2 Implementation

The tool used for the study of diffusion in the giant components of the Erdos- Renyi and the power law distribution graphs was developed using R language.

Packages used:

Shiny: A package by RStudio to help generate interactive web applications.

igraph: An open source package used for network analysis and visualization. Packages for igraph are available in R, Python and C/C++.

Assumptions: Graphs are undirected and progressive.

The user interface and the server instance was developed using the package 'Shiny'. It takes the following input parameters as shown in Figure 1:

Study of Diffusion in Networks

Figure 1: Inputs to the tool.

1. Number of nodes for both graphs
2. Gnp probability value for the ER graph
3. Power exponent value to create a Barabasi-Albert model graph
4. Threshold for all nodes in the ER graph to switch to the behavior of the infected seed
5. Threshold for for all nodes in the power law graph to switch to the behavior of the infected seed
6. Number of seeds for ER Graph
7. Number of seeds for power law graph
8. Number of runs

The Shiny server instance creates an ER and a power law graph based on the default number of nodes, Gnp probability and the power exponent. Based on these default parameters, a giant component is extracted from both graphs. The server then dynamically creates an UI widget to limit the selection of the number of seeds to the size of the giant component for both graphs. Similarly, the threshold is limited to the highest degree centrality value in the giant components retrieved. The user can change the default inputs and click submit to update the threshold and seeds limits for both graphs.

Later, the user can set the number of runs to study the spread of the seed's behavior.

Run	Threshold	Gnp Probability	Number of nodes infected					
			Seed=1		Seeds=5		Seeds=10	
			Random	Max Degree	Random	Max Degree	Random	Max Degree
1	1	1	100	100	100	100	100	100
		0.5	46	61	100	100	100	100
		0.1	14	18	41	63	73	78
		0.05	4	13	25	40	50	61
		0.01	4	5	9	14	10	15
	2	1	1	1	100	100	100	100
		0.5	1	1	87	91	99	100
		0.1	1	1	9	28	37	46
		0.05	1	1	7	10	13	34
		0.01	1	1	5	6	10	11
2	1	1	100	100	100	100	100	100
		0.5	100	100	100	100	100	100
		0.1	69	86	100	100	100	100
		0.05	26	39	70	88	74	92
		0.01	4	8	23	27	35	37
	2	1	1	1	100	100	100	100
		0.5	1	1	100	100	100	100
		0.1	1	1	18	70	85	89
		0.05	1	1	5	28	26	44
		0.01	1	1	4	8	10	10

Figure 2: ER Model: Results

The original graph generated is shown in the panels:

- View ER Model graph and degree distribution
- View Power law Model graph and degree distribution

The infected network and the number of nodes infected along with supplemental information on the parameters chosen are provided the following panels in the UI.

- Results - ER Model graph
- Results - Power law graph
- Graph Information

3 Results

A simulation was conducted to study the spread of randomly selected contagions and maximum degree contagions with the parameter values mentioned in Table 1.

The results shown in figures 2 and 3 were obtained from the simulation.

3.1 ER Model graphs

As shown in Figure 4, when the threshold was set to 1 for all the nodes in the network, the number of nodes infected increases as the gnp probability of the network increases.

It was also observed that the number of nodes infected increased as the number of seeds chosen increased. Another interesting observation is that for the same number of chosen seeds, the seeds with

Run	Threshold	Power exponent	Number of nodes infected					
			Seed=1		Seeds=5		Seeds=10	
			Random	Max Degree	Random	Max Degree	Random	Max Degree
1	1	0.5	2	9	13	31	27	50
		1	3	14	10	39	25	61
		1.5	3	49	9	74	20	75
		2	2	85	7	92	14	100
		2.5	2	91	7	100	11	100
	2	0.5	1	1	5	6	11	13
		1	1	1	5	6	12	12
		1.5	1	1	6	5	11	10
		2	1	1	6	5	11	10
		2.5	1	1	6	5	11	10
2	1	0.5	8	18	36	58	54	76
		1	6	29	25	72	53	88
		1.5	50	81	65	92	56	92
		2	85	97	83	100	92	100
		2.5	91	99	97	100	100	100
	2	0.5	1	1	5	6	13	13
		1	1	1	5	6	14	11
		1.5	1	1	6	5	15	10
		2	1	1	7	5	12	10
		2.5	1	1	6	5	11	10

Figure 3: Power law graph: Results

Parameter	Values
Number of nodes	100
Number of seeds	1,5,10
Number of runs	1,2
Threshold	1,2
Gnp Probability	1,0.5,0.1,0.05,0.01
Power exponents	0.5,1,1.5,2,2.5

Table 1: ER and Power law parameters for simulation study

the maximum degree infected more nodes over the seeds chosen randomly.

In figure 5, the threshold is set to 2. It observes a pattern similar to figure except that when the threshold is set to 2, the ability of a contagion to affect its neighbors decreases. Also, when number of seeds is set to 1, the number of nodes infected is 1, the seed itself.

3.2 Power law Model graphs

In Figure 6, we observe that for different exponents, as the power exponent increased, the maximum degree seeds managed to infect more nodes compared to the randomly selected ones when the threshold was set to 1. In fact, the randomly chosen nodes seemed to show a decline in number of infected nodes as the power exponent for the network increased. However, when the threshold was set to 2 (Figure 7), the ability of the maximum degree node seeds decreased as the power exponent exceeded 1.

The number of nodes infected increased when the power exponent was less than 1. In this case, it was also observed that the nodes infected by the maximum degree seeds was higher than the ones infected by the random seeds which was the exact

opposite case when the power exceeded one. This is because the igraph package's barabasi-albert graph generator function created nodes with different degree distribution for default power exponent 1 and as the power increased beyond 1 its generated a less diverse degree distribution.

4 Discussion

It was observed that the number of nodes infected increased with the gnp probability of the network. Also in the ER graphs, for the same number of seeds, the maximum degree seeds infected more nodes than the seeds chosen randomly. This indicates that a more central node has greater influence over the spread of a contagion's behavior across the network than the rest. A smaller threshold helps nodes in a network easily adopt a behavior compared to higher thresholds. Also, when the diffusion process was run twice for the ER graphs, the number of nodes infected almost doubled from the first run. If we were to compare this observation to the study by Coleman, Herbert Menzel and Elihu Katz[1966], threshold may be defined as the level of unwillingness to adopt a behavior and the physicians could be viewed to have less threshold, less reluctance in adopting Tetracycline.

The power law graphs center the distribution of most nodes in a network around a set of nodes as the power of the network increased. This allowed for higher infection rates when the threshold was 1 for the maximum degree nodes and dropped drastically when the threshold was set to 2 since there weren't many nodes with two infected neighbors in the network.

Also when the power fell below 1 it decreased the number of nodes attached to the more central nodes and fairly decreased the number of infected nodes too when the threshold was 1.

When the threshold was set to 2 the random node seeds seemed to infect more nodes than the maximum degree node seeds and this was because almost all the nodes in the network were mainly centered around one node and these connected node only had a degree of 1. Any two of these could have been chosen randomly and had the central node as the common connection ending up infecting a few extra nodes when compared to the maximum degree seed nodes as show in figure 3.

From the study of the two different graph models mentioned, it is evident that the way a network is structured and the way a contagion is chosen affects the spread of a behavior in a social network.

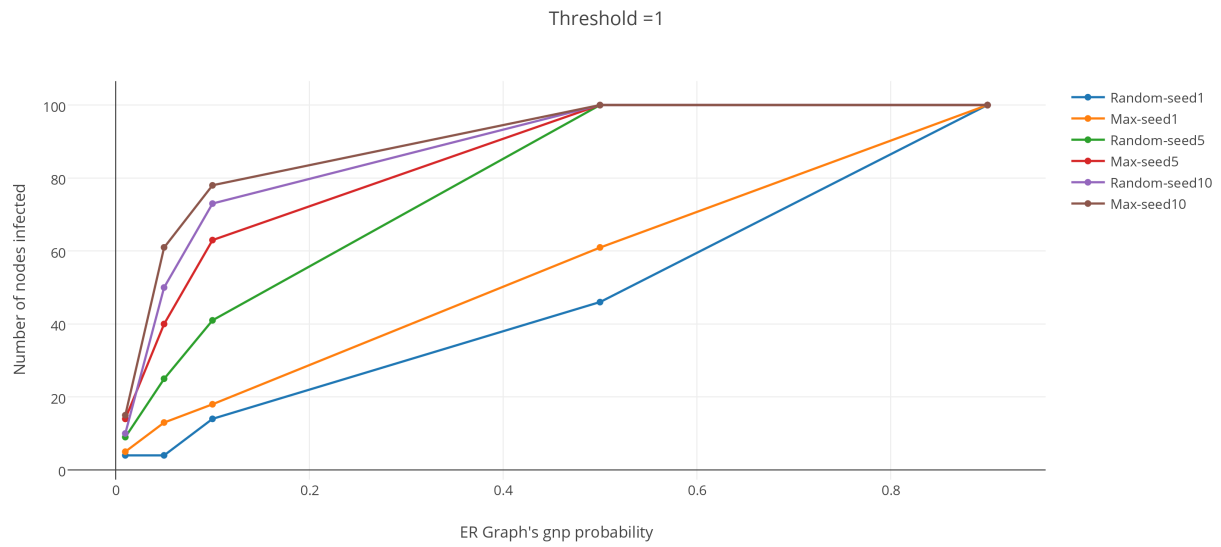


Figure 4: ER Model: Nodes infected when threshold=1

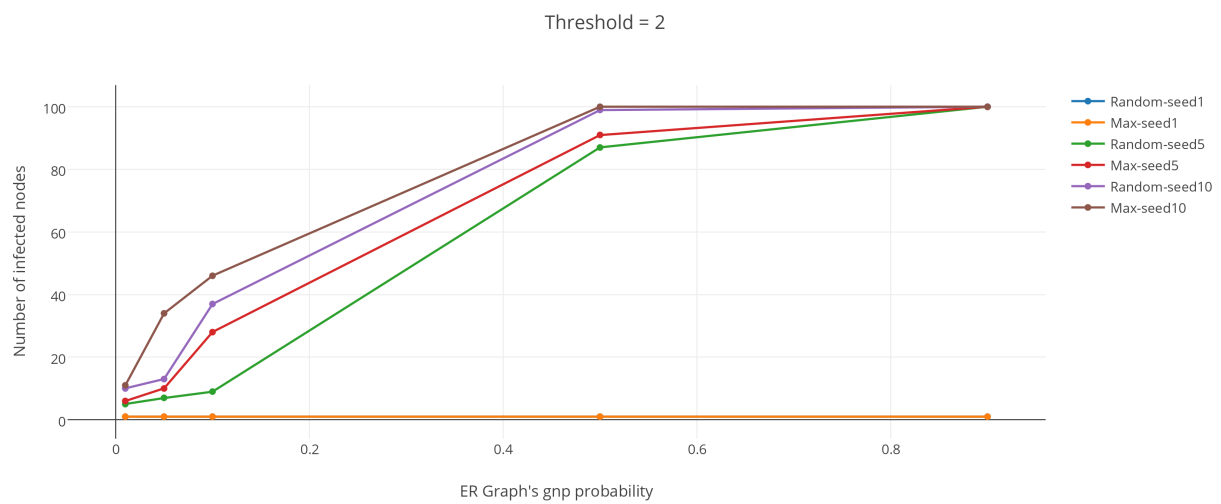


Figure 5: ER Model: Nodes infected when threshold=2

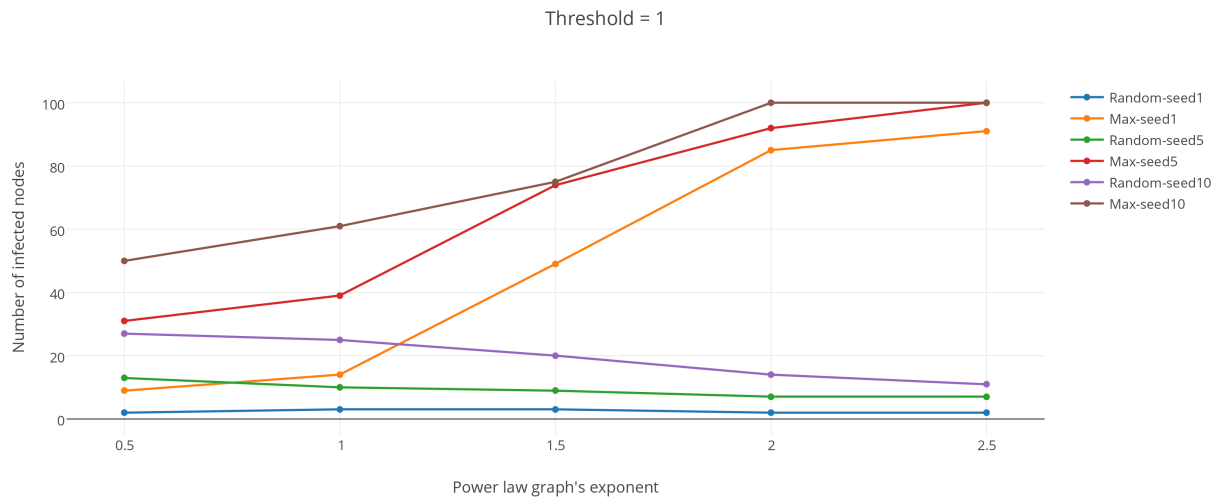


Figure 6: Power law Model: Nodes infected when threshold=1

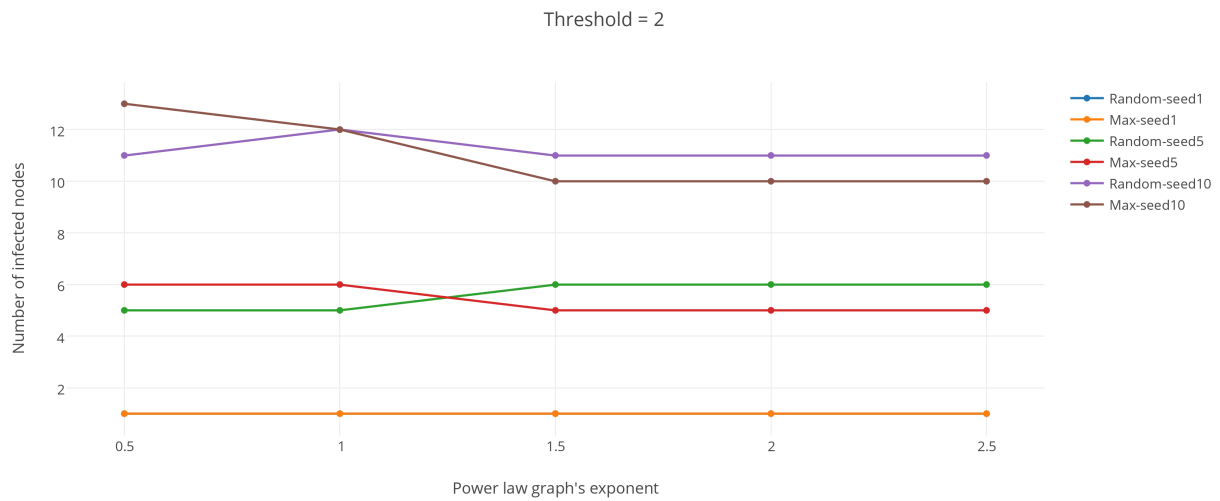


Figure 7: Power law Model: Nodes infected when threshold=2

References

- [1] James S. Coleman, Elihu Katz, and Herbert Menzel. *Medical innovation: A diffusion study*. Indianapolis: The Bobbs-Merrill Company, 1966.