Project 5 : Virus Propagation
UnityId : akagrawa

Introduction: This document contains the description of the virus propagation[1] implementation over the static contact network. The document illustrates few of the tasks and its outcome in the virus propagation model including checking for epidemic threshold, simulation for virus spread for different transmission probability(beta) and different healing probability(delta), applying different Immunization policies given k vaccines, estimating optimal number of vaccines required for different immunization policies in order to stop/control virus propagation. Please refer **vprop.R** for the implementation code and **README.md** file for the code description and usage. For more detail on the output plots, please refer the folder "**plots**".

Graph Description: The virus propagation is performed on the graph data-set of **static.network** graph data. These data files contain the vertex-vertex connection separated by the new line which is parsed and loaded into graph-like data structures for algorithm implementation. From the given dataset, we can assume that the graph is **simple(no parallel and self-loop), static, undirected, unweighted and unlabeled** for the virus propagation model.

Problems and Outputs

Option 1: Virus Propagation on Static Networks

Q1. For the SIS (susceptible, infected, susceptible) Virus Propagation Model (VPM), with transmission probability β = β1, and healing probability δ = δ1, calculate the effective strength (s) of the virus on the static contact network provided (static.network).

a. Will the infection spread across the network (i.e., result on an epidemic), or will it die quickly?
β1 = 0.2 , δ1 = 0.7

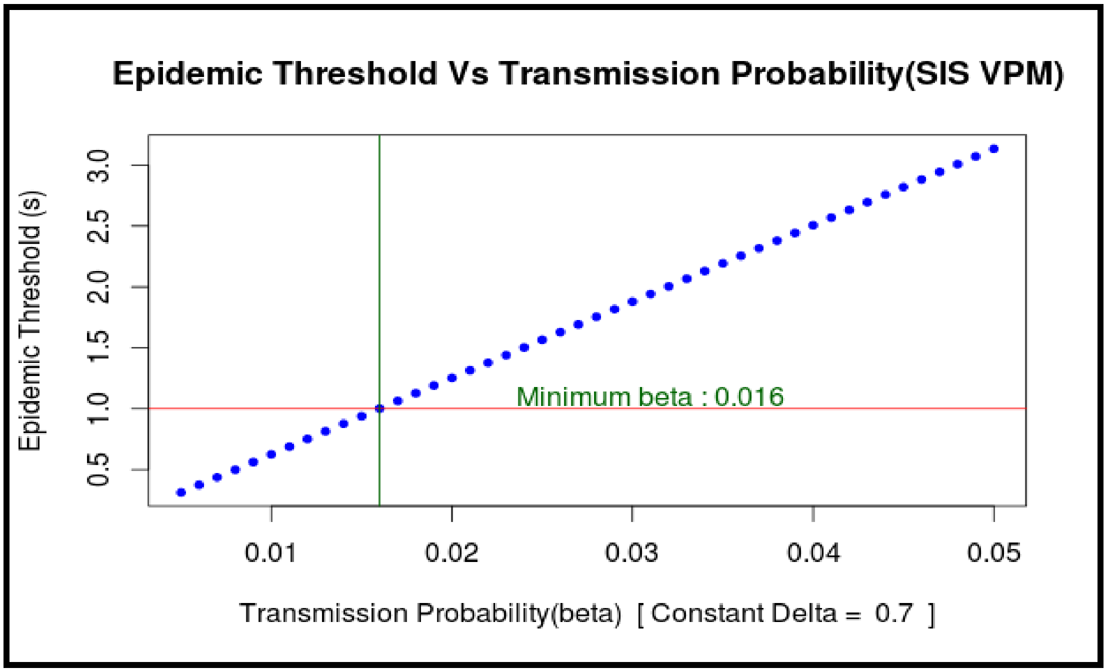
Output :

--- Epidemic Threshold : 12.52991 ----
Epidemic Threshold reached.. Virus will result in an epidemic

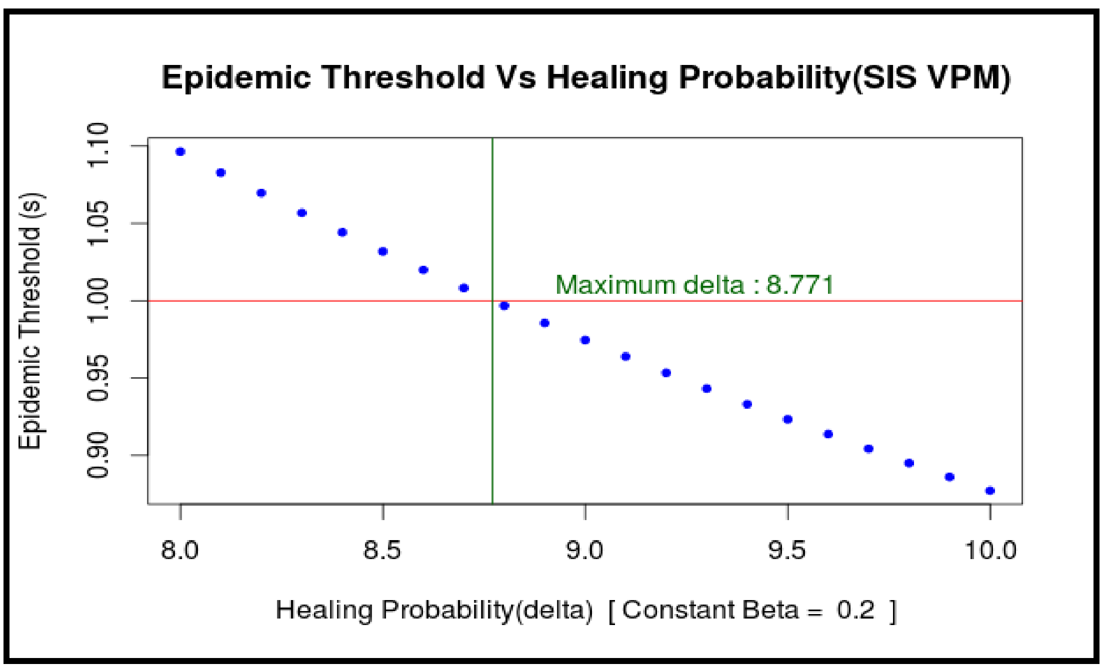
b. Keeping δ fixed, analyze how the value of β affects the effective strength of the virus. What is the minimum transmission probability (β) that results in a network-wide epidemic?

Output:

Keeping δ fixed, the minimum transmission probability (β) that results in a network-wide epidemic is 0.016 . The effect is linear if we keep increasing beta. The transmission probability is be very low given the contact network over fixed healing probability. Below plot reveals that increment in very small value of transmission probability will result in a network-wide epidemic.



c. Keeping β fixed, analyze how the value of δ affects the effective strength of the virus. . What is the maximum healing probability (δ) that results in a network-wide epidemic?



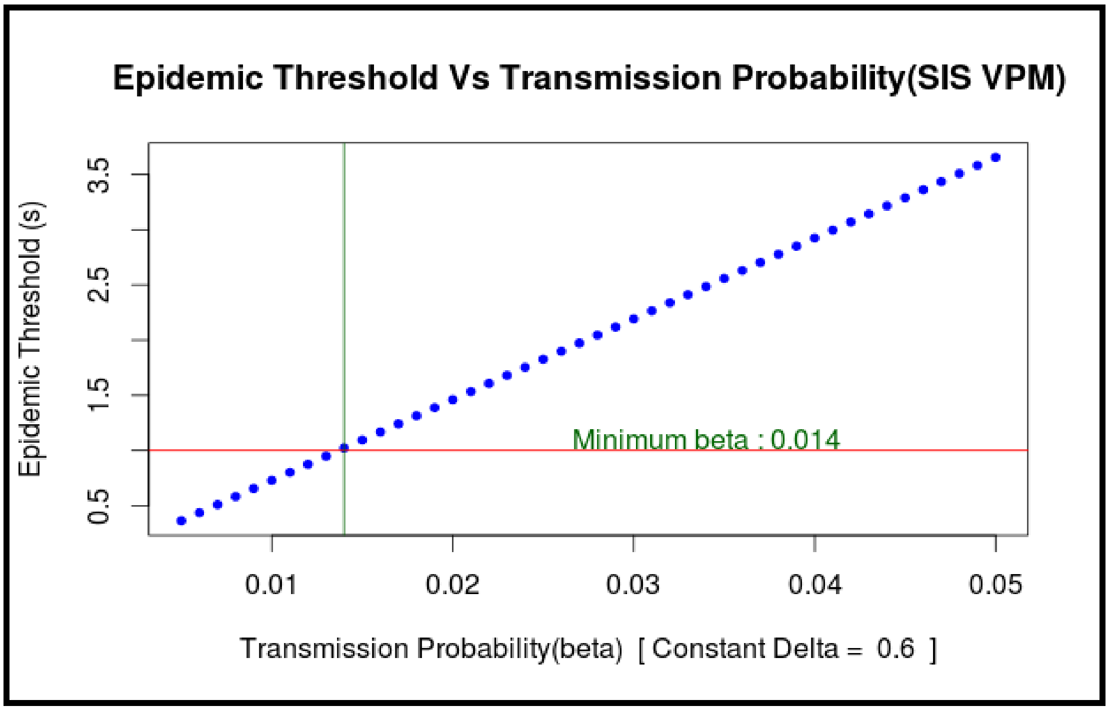
Keeping β fixed, the maximum healing probability (δ) that results(theoretically) in a network-wide epidemic is 8.771 which is impractical. Since probability can be maximum upto 1 but even the healing probability is high, there is no chance that the virus may die given the contact network over fixed transmission probability.

d. Repeat (1), (1a), (1b) and (1c) with β = β2, and δ = δ2. β2 = 0.01 , δ2 = 0.6

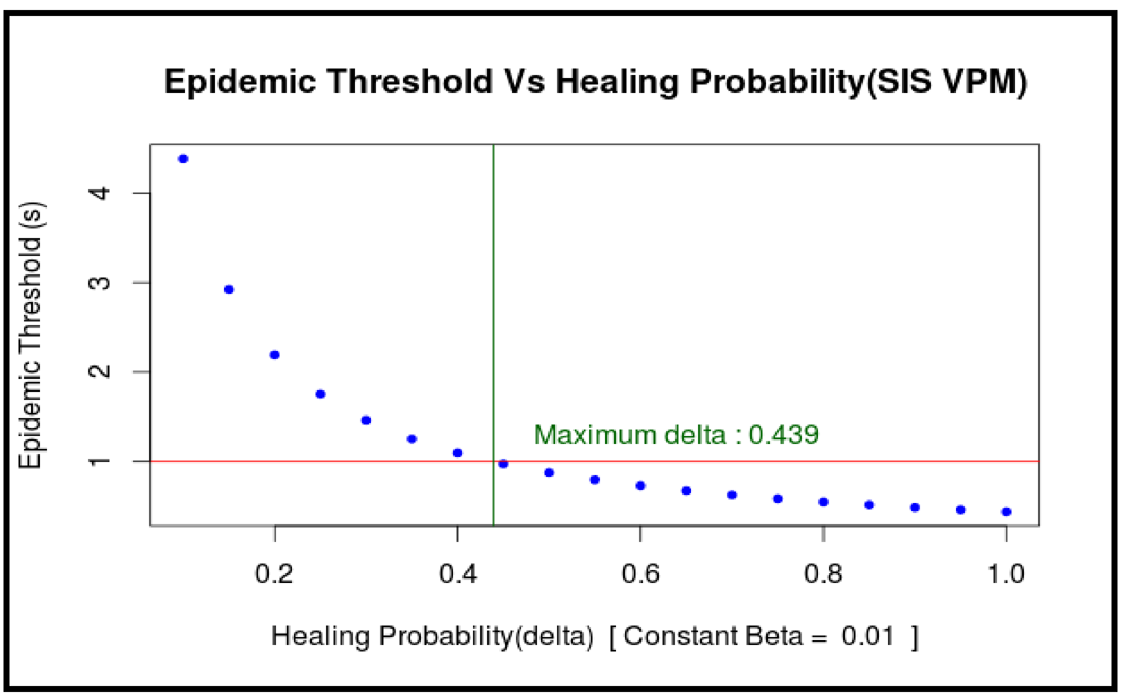
--- Epidemic Threshold : 0.7309116 ----

Epidemic Threshold not reached.. Virus will die quickly

Keeping delta fixed :



Keeping beta fixed :



From the above plots, it can be concluded that keep delta (healing probability) fixed, the minimum beta (transmission probability) will have a linear relation on epidemic threshold with positive correlation. The minimum beta that results(theoretically) in a network-wide epidemic is 0.014 i.e. also low. This means given such contact network, even the slight increment in transmission probability will increase the risk of network-wide epidemic.

While keeping beta(transmission probability) fixed, the maximum delta (healing probability) will have a negative correlation with epidemic threshold. The maximum delta that results(theoretically) in a network-wide epidemic is 0.439. This means given such contact network, even the healing probability must be good i.e. above 43.9% in order to make virus die quickly (i.e. no further spread in the network).

Q: 2 Write a program that simulates the propagation of a virus with the SIS VPM, given a static contact network, a transmission probability (β), a healing probability (δ), a number of initially infected nodes (c), and a number of time steps to run the simulation (t). The initially infected nodes should be chosen from a random uniform probability distribution. At each time step, every susceptible (i.e., non-infected) node has a β probability of being infected by neighboring infected nodes, and every infected node has a δ probability of healing and becoming susceptible again.

a . Run the simulation program 10 times for the static contact network provided (static.network), with β = β1, δ = δ1, c = n/10 (n is the number of nodes in the network), and t = 100.

<u>Virus Propagation Simulation :</u> The virus propagation model follows the below heuristics and assumptions:

1- Select the random nodes n/10 from the graph vertices using random distribution

2- These infected nodes will spread virus to their non-infected neighbour nodes with beta transmission probability.

3- In the same time duration, among all the infected nodes, few nodes are chosen for healing with delta healing probability and made susceptible again.

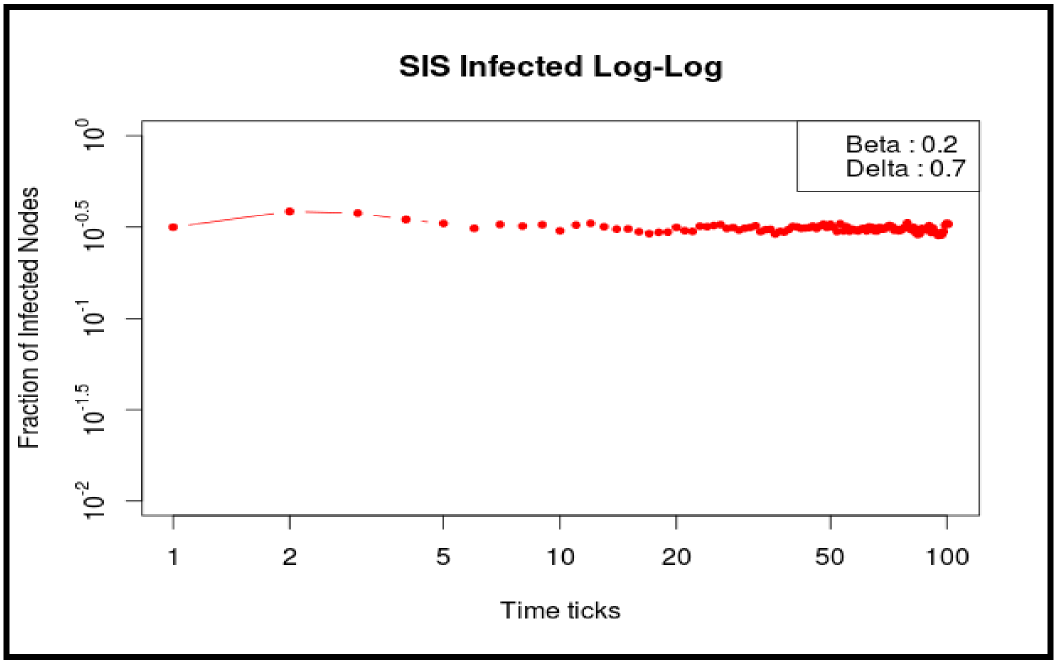
4- The above process is repeated for 100 time ticks and fraction of nodes infected were captured.

5- Process 1 to 4 in repeated for 10 times as number of simulations and average infected nodes over the timeticks were captured in a matrix.

A log-log plot is then sketched for time ticks and fraction of infected nodes.
Please refer the vprop.R file for the code and README.me file for the instructions to run the same.

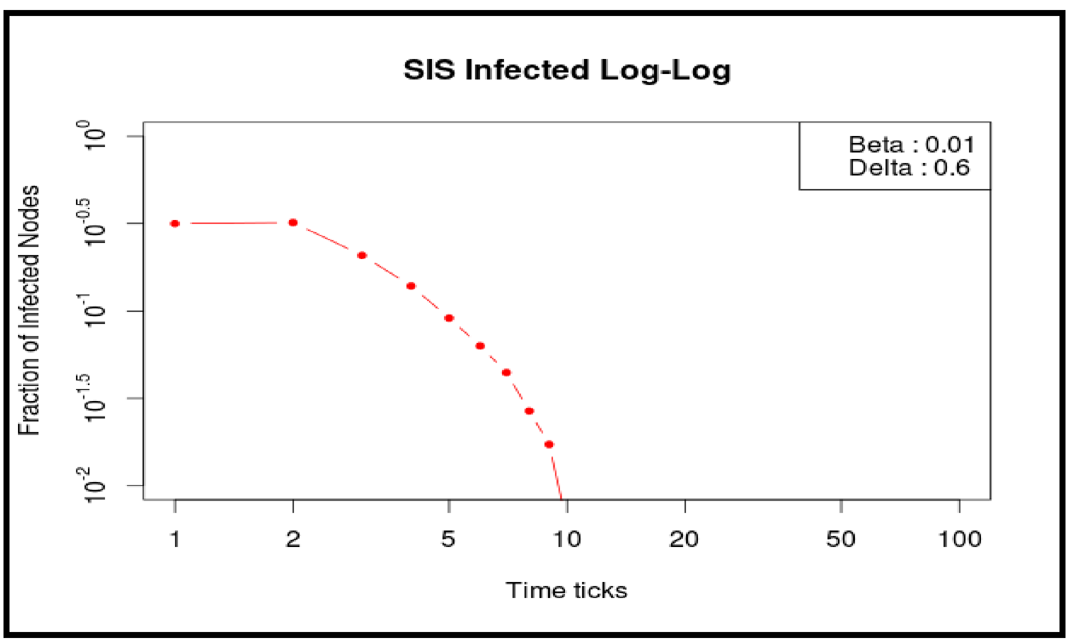
b. Plot the average (over the 10 simulations) fraction of infected nodes at each time step. Did the infection spread across the network, or did it die quickly? Do the results of the simulation agree with your conclusions in (1a)?

Please find the log-log plot below for the simulation produced as the outcome of the above mentioned processes. Running the simulations for 10 time over 100 time ticks for beta=0.2 and delta=0.7 produces the results shown in the below log-log plot. This results of the simulation **very much agree** with the epidemic threshold conclusion obtained earlier. For this network, given this transmission and healing probability the virus is not dying and there are always significant amount of infected nodes over the network.



c. Repeat (2a) and (2b) with β = β2, and δ = δ2.

Please refer vprop.R file for code implementation. The steps for simulation is same as mention above for different set of transmission and healing probability. Running the simulations for 10 time over 100 time ticks for beta=0.01 and delta=0.6 produces the results shown in the plot below. This results of the simulation **very much agree** with the epidemic threshold conclusion obtained earlier. For this network, given this transmission and healing probability the virus dies quickly which is clearly revealed in the plot below.



3. Write a program that implements an immunization policy to prevent the virus from spreading across the network. Given a number of available vaccines (k) and a contact network, your program should select k nodes to immunize. The immunized nodes (and their incident edges) are then removed from the contact network.
a. What do you think would be the optimal immunization policy? What would be its time complexity? Would it be reasonable to implement this policy? Justify.

In my view, the optimal policy would be to immunize the nodes with highest degree. If the adjacency matrix is given then the time complexity would be KN where K is the number of vaccines given and N is the number of nodes in the graph.

For your program, use the following heuristic immunization policies:
- Policy A: Select k random nodes for immunization.
- Policy B: Select the k nodes with highest degree for immunization.
- Policy C: Select the node with the highest degree for immunization. Remove this node (and its incident edges) from the contact network. Repeat until all vaccines are administered.
- Policy D: Find the eigenvector corresponding to the largest eigenvalue of the contact network's adjacency matrix. Find the k largest (absolute) values in the eigenvector. Select the k nodes at the corresponding positions in the eigenvector.
For each heuristic immunization policy (A, B, C, and D) and for the static contact network provided (static.network), answer the following questions:
b. What do you think is the intuition behind this heuristic?

For policy A : Random nodes in the graph is the most basic policy. It has a very less probability of selecting optimal or best nodes for immunization.

For policy B : The intuition of this policy is to select the nodes with highest connectivity, so that the overall density of graph can be reduce.

For policy C : Immunizing the node in the order of the degree of node one at a time will help selecting nodes that belongs to separate subgraphs.

For policy D : Using spectral property of graph in the immunization policy can help in selecting those nodes whose removal(immunization) can have greatest change in the variability of the graph.

c. Write a pseudocode for this heuristic immunization policy. What is its time complexity?

For policy A : Random nodes in the graph is the most basic policy. It has a very less probability of selecting optimal or best nodes for immunization.
Pseudo Code :*Let G be the Graph Object (Adjacency Matrix)*
Policy_A(G)
1- Select random nodes from set of vertices
2- In adjacency matrix, make the value 0 for the adjacent edges for which a random node is selected and delete the corresponding row and column for all the random nodes.

Time Complexity : N^2 (Cost of creating new adjacency matrix without those k random nodes)
Lets assume cost of creating new Adjacency matrix after deleting node = C , it will be directly used for the explanation of the time complexity of the rest of the policies.

For policy B :

Pseudo Code: Let G be the Graph Object

POLICY_B(G)

1- List<Node> nodes = K_HIGHEST_DEGREE(G)

3- G = DELETE_NODE(nodes)

Here the Graph object can be adjacency matrix. And Highest Degree of the node can be maintained in MAX-HEAP(NlogN heap building and maintaining cost). And cost of deleting node will be the cost of creating new adjacency matrix.

Time Complexity : logN * C = O(C)

For policy C : Immunizing the node in the order of the degree of node one at a time will help selecting nodes that belongs to separate subgraphs.

Pseudo Code: Let G be the Graph Object

POLICY_C(G)

1- for i =1 to K

2- Node N = HIGHEST_DEGREE(G)

3- G = DELETE_NODE(N)

Here the Graph object can be adjacency matrix. And Highest Degree of the node can be maintained in MAX-HEAP(NlogN heap building and maintaining cost). And cost of deleting node will be the cost of creating new adjacency matrix.

Time Complexity : KC + NlogN = O(KC)

For policy D : Using spectral property of graph in the immunization policy can help in selecting those nodes whose removal(immunization) can have greatest change in the variability of the graph.

Pseudo Code :

POLICY_D(G)

1- Get the first leading eigenvector of a graph O(m) where m is the no of edges[3]

2- Find K largest values from the eigenvector

3- Delete the corresponding indexes nodes from the graph

Time Complexity : O(m)

d. Given k = k1, β = β1, and δ = δ1, calculate the effective strength (s) of the virus on the immunized contact network (i.e., contact network without immunized nodes). Did the immunization policy prevented a network-wide epidemic?

<u>Policy A :</u>

--- Epidemic Threshold : 12.31393 ----

Epidemic Threshold reached. It is very high than the standard epidemic threshold. Virus will definitely result in an epidemic even applying this immunization policy. Even though the node selected are random, but for a great number of trials the average epidemic threshold will be around 10-12 which is also very high from standard epidemic threshold.

<u>Policy B :</u>

--- Epidemic Threshold : 1.080275 ----

Epidemic Threshold reached. According to the theorem[1] the virus will spread. But the immunization policy is much better than the previous policy. Virus may in this case eventually die since the epidemic threshold is slightly above the standard threshold. If the same policy is continued for immunization then only few more vaccines can prevent spreading the virus.

Policy C :

--- Epidemic Threshold : 1.084521 ----

Epidemic Threshold reached.. The epidemic threshold is almost the same as obtained by applying policy B. Here also according to the theorem[1] the virus will spread. But the immunization policy is much better than policy A. Virus may in this case eventually die since the epidemic threshold is slightly above the standard threshold. If the same policy is continued for immunization then only few more vaccines can prevent spreading the virus.

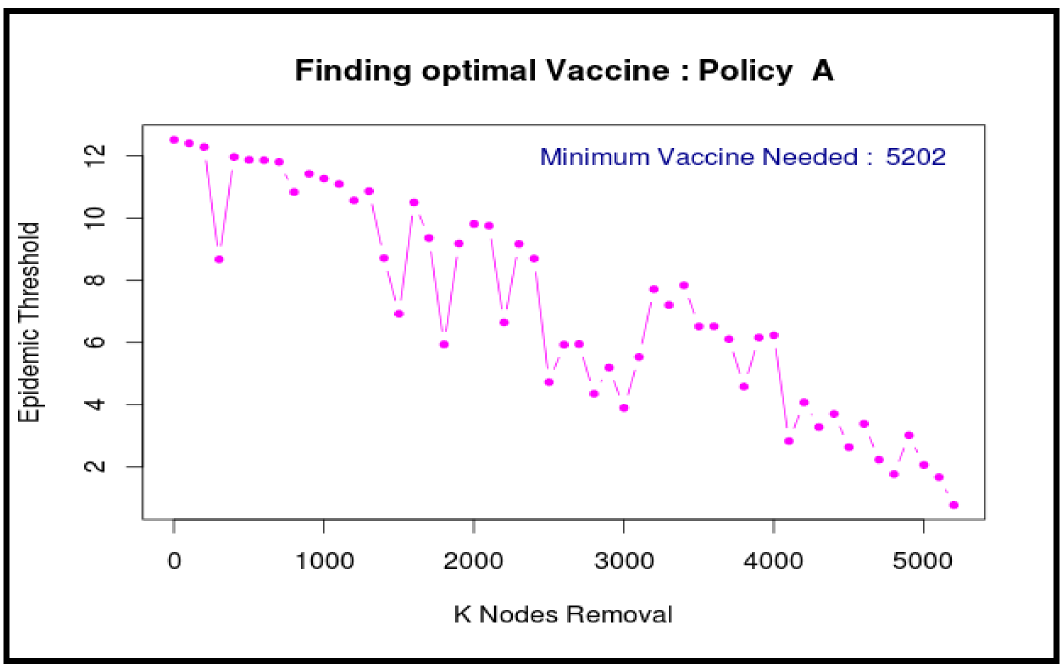
Policy D :

--- Epidemic Threshold : 3.070528 ----

Epidemic Threshold reached.. Virus will result in an epidemic. Since the epidemic threshold is much greater than the standard epidemic threshold. But the immunization policy is better than policy A but not as good as policy B and C in this case.

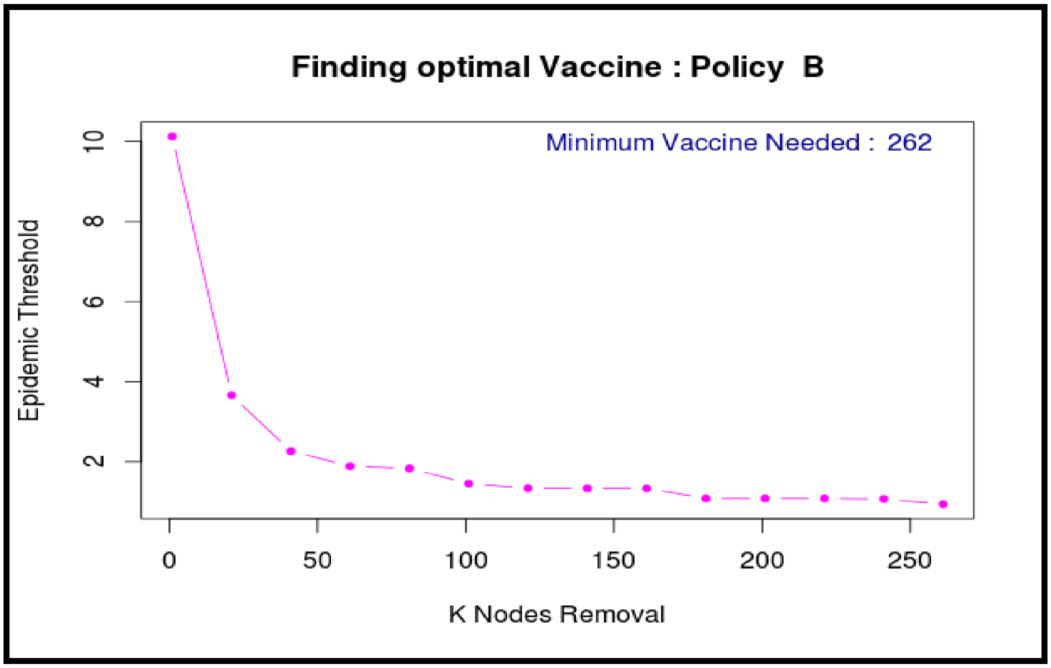
e. Keeping β and δ fixed, analyze how the value of k affects the effective strength of the virus on the immunized contact network (suggestion: plot your results). Estimate the minimum number of vaccines necessary to prevent a network-wide epidemic.

Policy A : The policy is to select random nodes, so the number of vaccines needed also depends on the properties of the nodes that are randomly get selected i.e. if high degree nodes are selected then lesser number of vaccines might be needed as compare to the nodes of lesser degree.

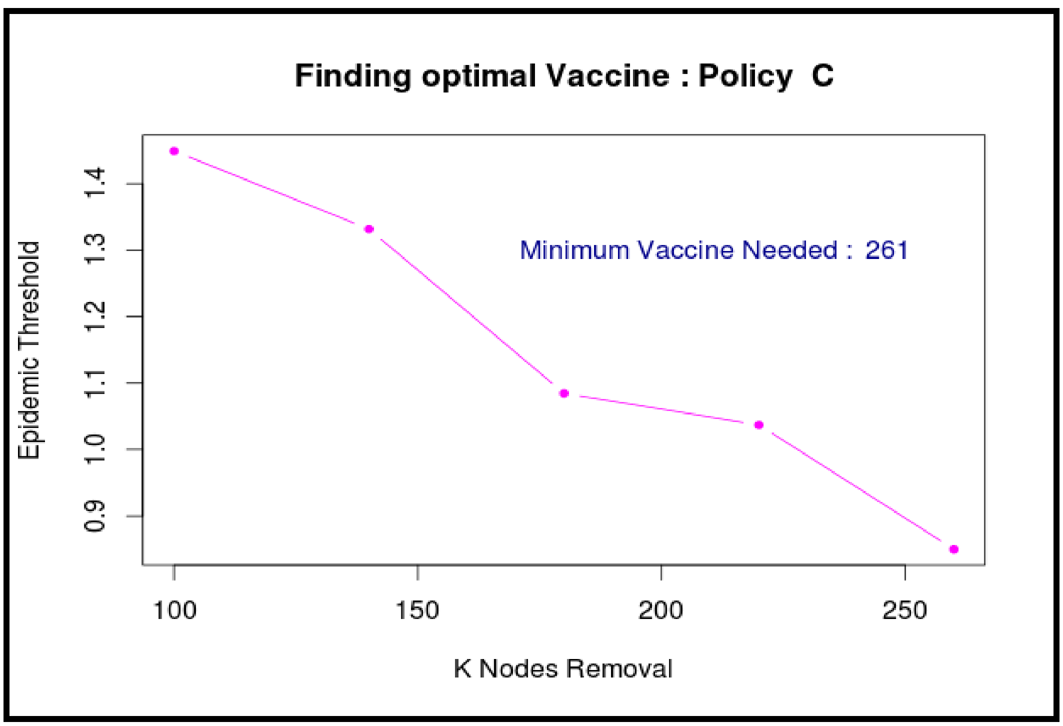


For a sample run, the vaccine needed = 5202. But in average the number of vaccines will not deviate much with this figure. The vaccines needed in this policy is close to the total number of nodes in the network which is 5715.

Policy B : The optimal number of vaccines needed in this case = 262
We can see the steep drop in the epidemic threshold when we increase number of vaccines. This observation reveals that policy is much effective as the number of vaccines in roughly 5% of the total nodes.

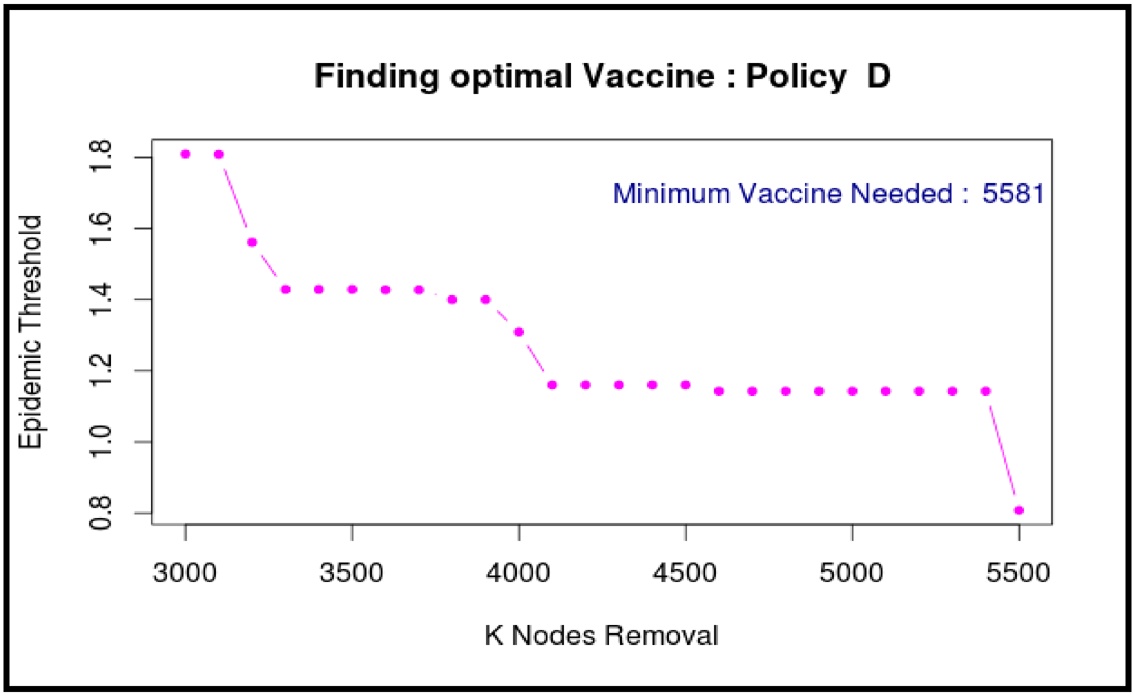


Policy C : Please find the detail analysis below for the policy C.



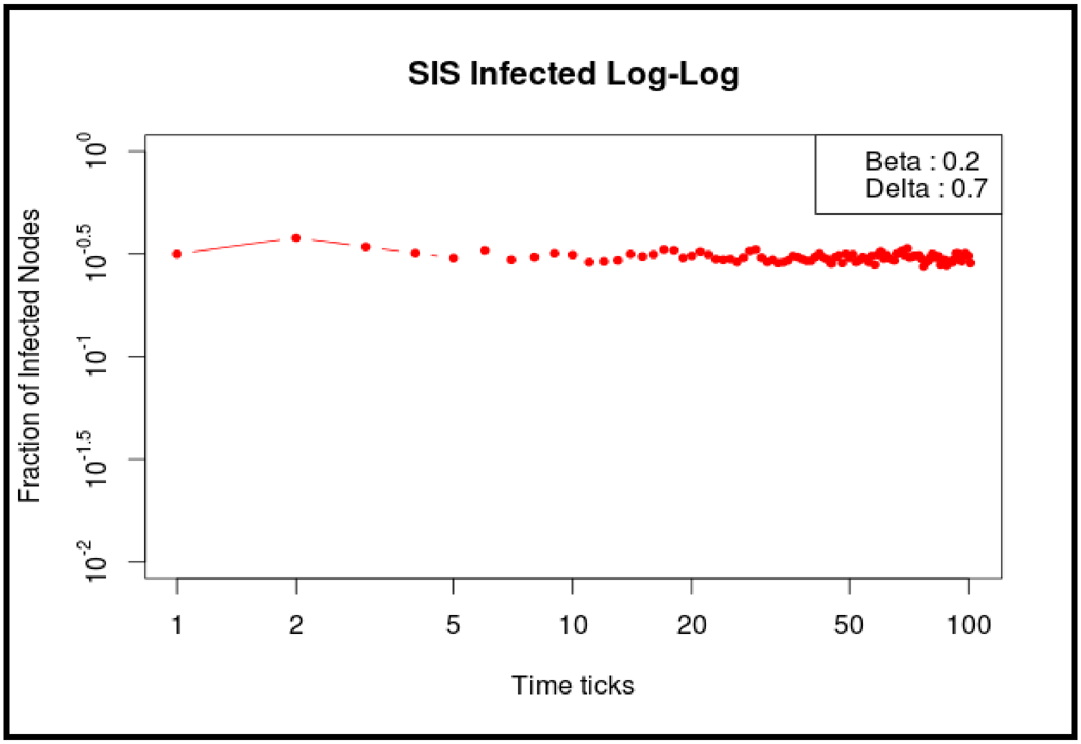
The optimal number of vaccines needed = 261. For policy C, the results are almost the same as policy B, which means this policy is equally good but it is a costly operation than policy B in terms of finding such nodes over the network.

Policy D : The number of vaccines required = 5581 in this case. These observations reveal that this policy is not at all a good policy as the vaccines required is nearly equal to the total number of nodes means it is like immunizing everyone in order to prevent from the virus which is by default the worst yet reliable policy.

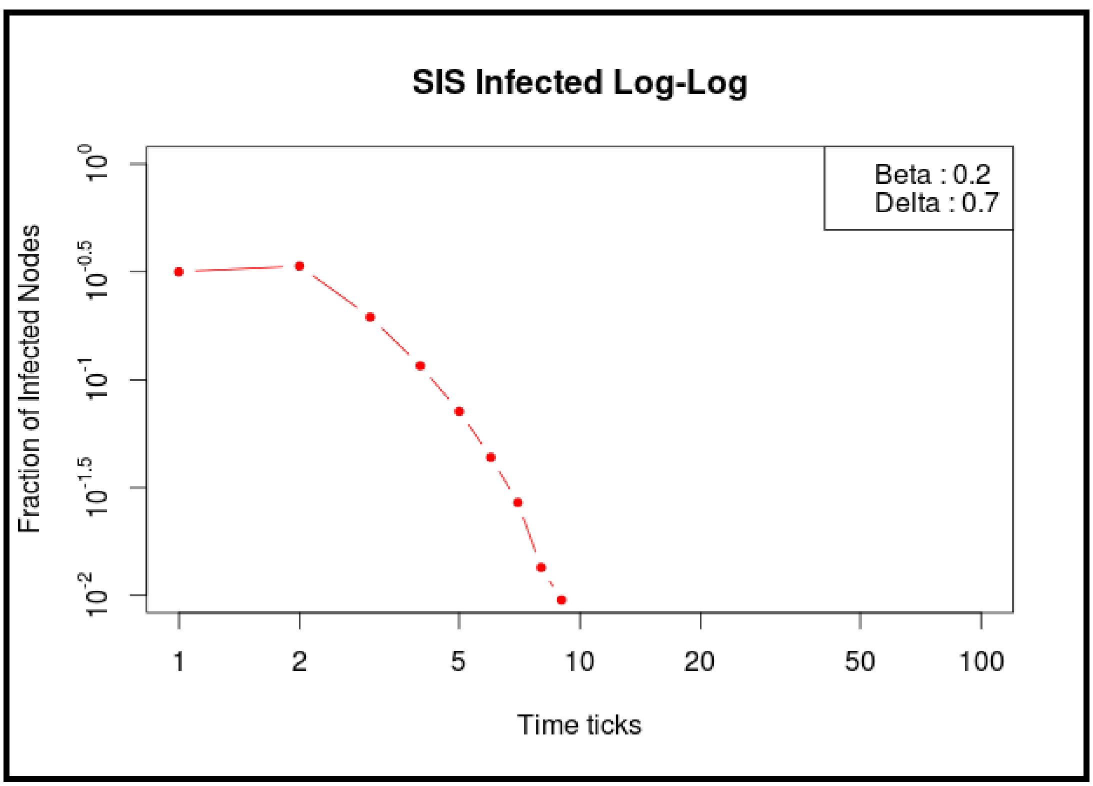


f. Given k = k1, β = β1, δ = δ1, c = n/10, and t = 100, run the simulation from problem (2) for the immunized contact network 10 times. Plot the average fraction of infected nodes at each time step. Do the results of the simulation agree with your conclusions in (3d)?

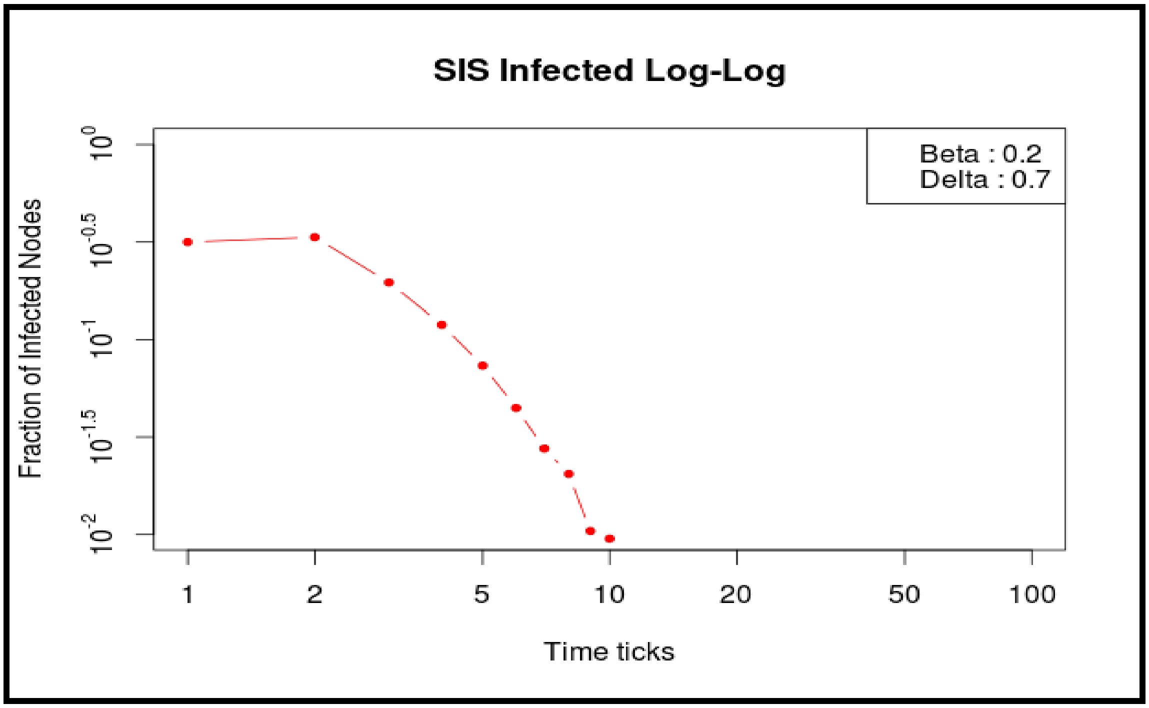
Policy A : Please find the plot below for the average fraction of nodes at each time step for the immunized network under policy A. Even after the immunization the average infected nodes are still non-zero after 100 time ticks. This means that the virus will not die after applying this immunization policy. The results in this case completely agree with the conclusion in 3d.



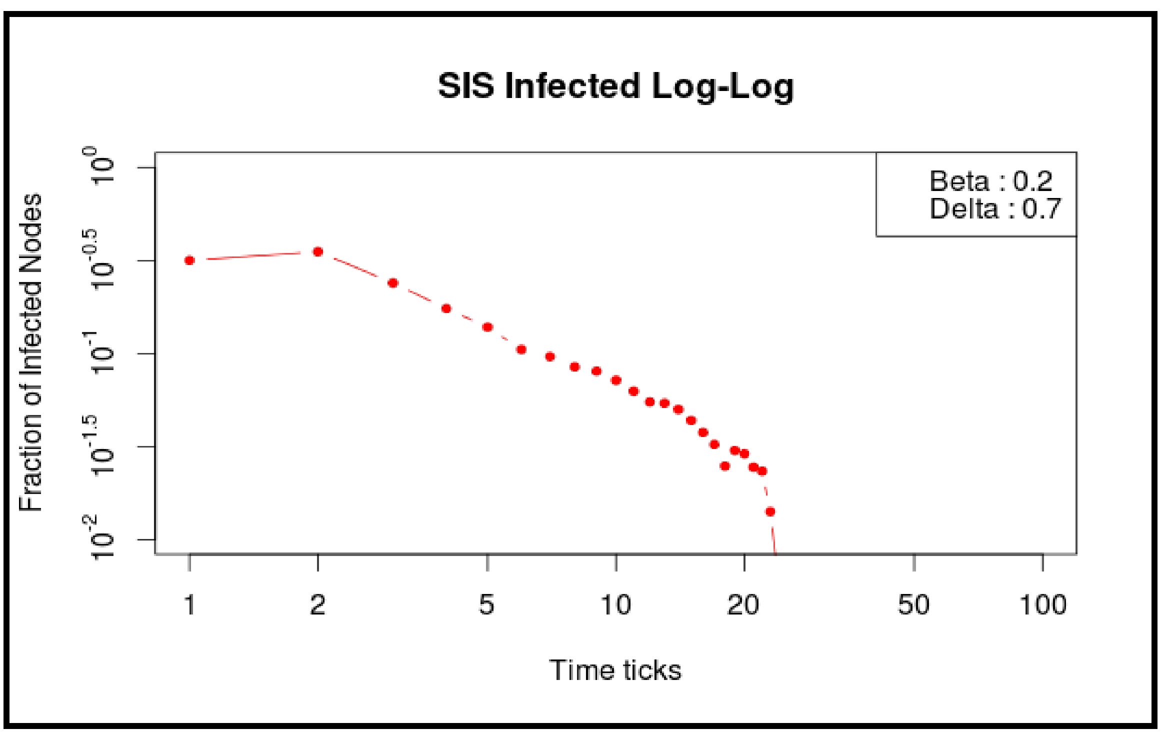
Policy B : We can see that in the immunized network of policy B, the virus is dying quickly. This does not agree with the 3d results theoritically. But in the inference is the same, since the epidemic threshold was very close to 1, it is subject to die eventually which we can see in the simulation plot.



Policy C : This simulation result is similar to that of policy C. Here also the virus will die eventually. The result also doesn't agree with the simulation theoretically but since the epidemic threshold is in the border line, it can be claimed that the virus will die eventually.



Policy D : In this policy, the virus is dying but after certain time ticks. The results doesn't agree with the solution in 3 d.



References :

[1] Prakash, A., D. Chakrabarti, M. Faloutsos, N. Valler, and C. Faloutsos. "Got the Flu (or Mumps)? Check the Eigenvalue!" Got the Flu (or Mumps)? Check the Eigenvalue! N.p., n.d. Web. 13 Nov. 2014.
[2] For igraph eigen value functions, http://igraph.org/r/doc/graph.eigen.html
[3] Eigen computation costs http://en.wikipedia.org/wiki/Eigenvalue_algorithm#Iterative_algorithms