MODELING COMPLEX SYSTEMS

ASSIGNMENT 4

January 11, 2019

Maxfield Green and David Landay
University of Vermont
Comp. Systems and Data Sci.

0.1 Problem 1

CASCADING MODEL

(a) Why is this network acyclic?

For a graph to be acyclic, there will be no closed paths among any of the nodes in the network. In other words, no set of edges will complete a circuit, or, nodes will never be visited more than once when following edges.

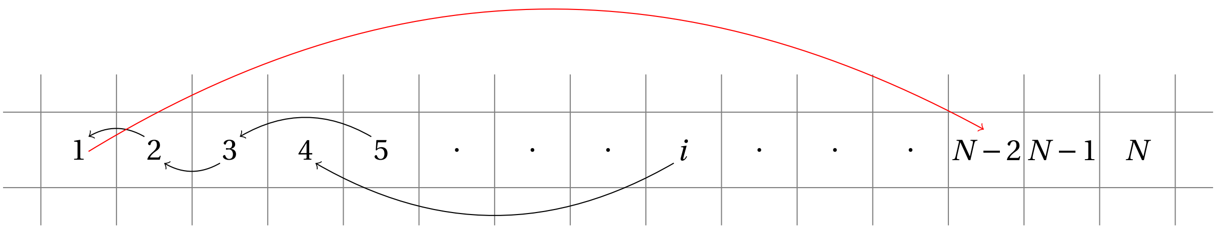


Figure 1: $1-d$ lattice network: nodes i may form edges to any $i' < i$. Arrows represent directed edges. The red arrow represents an edge that breaks the network construction.

In this example, nodes that are lower in the hierarchical structure of the network can never direct to nodes that are greater in the hierarchy by construction. If an edge directed from node i' to node i, where the value of the index represented by i' was less than that of i, then we would have a blatent contradiction of the network construction because we assume edges only from i to $i' < i$. Thus, edges can never complete a cycle, and the network is acyclic.

(b) What is the average in-degree of vertex i? What is the average out-degree?

If $i = 1$, the out degree is zero, the node cannot direct to any node with higher index. The in degree would be $p(N-1)$, where p is the probability that an edge connects two nodes.

If $i > 1$, the out-degree is $(i-1)p$ because there are $i-1$ potential directed connections available. The in-degree would be $(N-i)p$ as there are $N-i$ nodes left for the i^{th} node to direct to.

If $i = N$, the out-degree is $(N-1)p$ because the node cannot connect to itself and thus has $N - 1$ potential pairs. The in-degree is zero since no node in the network can direct to the N^{th} node.

(c) Show that the expected number of edges that run from nodes i' >i to nodes i' < i

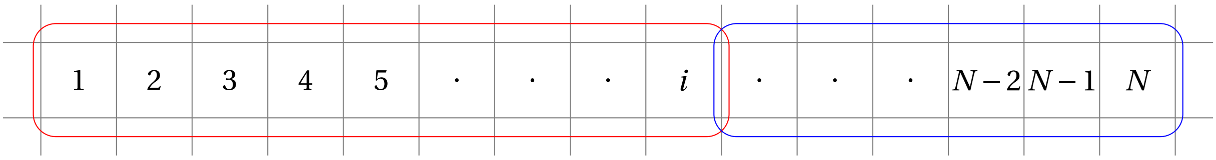


Figure 2: $1-d$ lattice network: nodes i' may form edges to any $i < i'$

The diagram above describes how nodes $i' > i$ link to every node $i' \leq i$. We can see that to for every $i' > i$, we expect that $N - i$ total components connect to i. As explained in the previous section, this is equivalent to the expected in-degree for i. Simultaneously, the expected number of nodes i' that link to $i' \leq i$ is equal to i. Therefore, the expected number of nodes $i' > i$ that connect to all $i' \leq i$ is the product of the two quantities:

$$(N - i)\, i$$

$$= N\, i - i^2 \;\therefore$$

(d) Assuming N is even, what are the largest and smallest values of the quantity calculated in (c) and where do they occur in i?

Assume that N is even.

$$\implies N = 2k \qquad \text{(by def.)}$$

The network is bounded by $i = 1$ and $i = N$

$$\text{if } i = 1, \implies N\,i - i^2 = N - 1 = 2k - 1 \qquad \text{(from c)}$$

$$\text{if } i = N, \implies N^2 - N^2 = 0$$

This means that $i = N$ is the smallest value of the quantity calculated by (c). Next, we would want to find the largest value of i calculated in (c). To do this, we want to maximize the difference between the quantity calculated at $i = 1$ and i.

Let $i = \frac{N}{2}$, then

$$i = \frac{N}{2} \implies N\,i - i^2 = N\,\frac{N}{2} - \left(\frac{N}{2}\right)^2$$

$$= \frac{4k^2}{2} - \frac{4k^2}{4}$$

$$= 2k^2 - k^2$$

$$= k^2$$

This means that $i = \frac{N}{2}$ yields more connections than $i = 1$, dictated by the calculation in c. We can treat $i = \frac{N}{2}$ as the new upper bounds. For the sake of argument, let $i = \frac{N}{2} + 1$. Then

$$N\,i - i^2 \implies = N\left(\frac{N}{2} + 1\right) - \left(\frac{N}{2} + 1\right)^2$$

$$= \frac{N^2}{2} - \frac{N^2}{4} - 1$$

$$= 2k^2 - k^2 - 1 = k^2 - 1$$

Thus, $i = \frac{N}{2}$ is an upper bound for all $i \geq \frac{N}{2} + 1$. Similarly, if we let $i = \frac{N}{2} - 1$, then

$$N\, i - i^2\, N \left(\frac{N}{2} - 1\right) - \left(\frac{N}{2} - 1\right)^2$$

$$= \frac{N^2}{2} - \left(\frac{N^2}{4} + 1\right)$$

$$= \frac{N^2}{2} - \frac{N^2}{4} - 1$$

$$= 2k^2 - k^2 - 1$$

$$= k^2 - 1$$

Thus, $i = \frac{N}{2}$ is an upper bound for all $i \leq \frac{N}{2} - 1$. Hence, the value of i that yields the largest number of connections for (c) is $i = \frac{N}{2}$ $\therefore$

0.2 Problem 2

(a) What is the degree distribution $\{p_k\}$? What are the generating functions $G_0(x)$ and $G_1(x)$ for the degree and excess degree distributions?

In a *degree-regular* network, every node will have the same number of edges. Hence, the degree at every node will be equal. In our case, we are generating a random network where every node has exactly four connections, thus, the expected degree for any node on the network is $< k > = 4$.

Because every node has the same degree, 4, the probability that a node is of degree k is

$$p_k = \begin{cases} 1 & \text{if } k = 4 \\ 0 & \text{otherwise} \end{cases}$$

because $p_k = \frac{n_k}{n}$, the proportion of nodes of degree k with respect to the total number of nodes in the network, equals $\frac{n}{n} = 1$ when $k = 4$, and 0 otherwise. The degree distribution of the entire network, therefore, can be denoted $\{p_k\}$, and is equal to $\{4, 4, 4, \ldots\}$.

We will call the function that generates the degree distribution G_0.

$$G_0(x) = p_0\,x^0 + p_1\,x^1 + p_2\,x^2 + p_3\,x^3 + p_4\,x^4 + \ldots + p_k\,x^k$$

$$= \sum_k p_k\,x^k \tag{1}$$

The expected value of k over the entire network, or first moment about k, can be expressed as the derivative with respect to the random variable x of $G_0(x)$ evaluated at $x = 1$. From **1**, we get

$$\frac{d}{dx} G_0(x) = \sum_k k\,p_k\,x^{k-1}$$

Evaluating the derivative at $x = 1$, we confirm our intuition of the expected degree the network:

$$\frac{d}{dx} G_0(x) |_1 = \sum_k k \, p_k \, (1)$$

$$= (1 \times 0 \times 1) + (2 \times 0 \times 1) + (3 \times 0 \times 1) + (4 \times 1 \times 1) + \ldots$$

$$= 4$$

The excess degree describes the number of nodes that are accessed by following an edge from a random node. In our case, the *degree-regular* graph has an expected degree of 4, so at any given node if we follow an edge, we should expect that neighbor to have 3 branching edges to different nodes. We can describe the excess degree distribution by another probability generating function. Let us denote the excess degree distribution as $G_1(x)$.

$$G_1(x) = \frac{G_0'(x)}{G_0'(1)}$$

$$= \frac{\sum_k^\infty kp^{k-1}}{4}$$

This confirms our previous knowledge of the networks excess degree to be 3. Evaluating the derivative of $G_1(x)$ at x = 1 resolves to 3.

(b) Find the minimal value of k that is still large enough to expect that the entire network will be connected in one single giant component.

In a *degree-regular* network of N nodes of degree K a given node has (N-K)K nodes to connect to. I the case of an infinite sized network with degree 2, there is only one possible way for the network to be composed of one giant component. This structure would be one big loop. In the case of degree 3, it is highly unlikely to achieve a structure other than a giant component. Therefor, it is possible when K =2 but much safer when k = 3.

(d) Write a pairwise approximation of an SIS model for this network.Explain the moment closure involved.

Derived from a compartmental model as well as the moment closure discussed in class, we came up with the following system that describes how the number of edges that connect infected-infected, susceptible-infected and susceptible-susceptible nodes change over time.

$$\frac{d[SS]}{d\widetilde{t}} = [SI] - \lambda[SI](\frac{3}{4})\frac{2[SS]}{S}$$
$$\frac{d[SI]}{d\widetilde{t}} = 2[II] - [SI] + \lambda[SI]((\frac{3}{4})\frac{2[SS]}{s} - (\frac{3}{4})\frac{[SI]}{S} - 1)$$

Where $\lambda = \frac{\beta}{\alpha}$ and α = recovery rate, β = infection rate.

The moment closure gives us an approximation of a higher order property given lower order information. We are getting an approximation about the number of triplets in the SSI configuration given the number of SS edges that lead to SI edges. The ratio of $\frac{<k_{ex}>}{<k>}$ allows us to approximate the effects of a heterogeneous degree distribution. However in this particular model, there is a homogeneous degree and excess degree distribution as every node has degree four.

0.3 Problem 3

Epidemic Spreading on the configuration model

(a) Write a heterogeneous mean field system for an SIS epidemic model on a configuration model network.

Our last mean field system for an SIS model assumed that all nodes in a compartment had the same degree. This leads us to assume that every infectious node observes the same transition to infection rate. This is however not true. We need to consider heterogeneity in transmission rate. To do this, we will place infectious nodes in the I_k compartment, where k is the degree of the node. This way we have an $\frac{\tilde{I}_k}{d\tilde{t}}$ for each regime of degree.

The resulting mean-field model adapts to be

$$\frac{\tilde{I}_k}{d\tilde{t}} = \lambda\, k\, (p_k - \tilde{I}_k)\, \theta - \tilde{I}_k$$

where θ is the moment closure. We interpret this as we normally do; that it is the probability that a random edge around a susceptible node leads to an infectious node.

(b) Modify the system by adding a new state Vaccinated nodes. Assume that around vaccinated nodes, the transmission rate (both to and from these nodes) is reduced a factor of f.

For our new model, we now have two different types of each compartment; infected and susceptible individuals who are vaccinated, and infected and susceptible individuals who are not. The resulting difference means that we require two sets of compartments. This lead us to the following mean-field equations for each proportion:

$$\frac{d\tilde{I}_{Nk}}{d\tilde{t}} = \lambda\, k \left(p_{Nk} - \tilde{I}_{Nk} \right) \left(\theta_N + f\, \theta_V \right) - \tilde{I}_{Nk} \tag{1}$$

$$\frac{d\tilde{I}_{Vk}}{d\tilde{t}} = \lambda\, k \left(p_{Vk} - \tilde{I}_{Vk} \right) \left(\theta_N + f\, \theta_V \right) - \tilde{I}_{Vk} \tag{2}$$

where

$$\theta_N = \frac{\sum_k k\, \tilde{I}_{Nk}}{\sum_k k\, p_{Nk}} \tag{3}$$

$$\theta_V = \frac{\sum_k k\, \tilde{I}_{Vk}}{\sum_k k\, p_{Vk}} \tag{4}$$

The moment closure is now a sum of the two probabilities that a random edge around a susceptible node leads to either a vaccinated or non vaccinated infected node. The proportion of infected nodes that are vaccinated can be thought of as vaccines which failed. The spread of the disease between vaccinated nodes is now controlled by a factor f. To calculate $\frac{d\tilde{I}_k}{d\tilde{t}}$, therefore, we get

$$\frac{\tilde{I}_k}{d\tilde{t}} = \frac{d\tilde{I}_{Nk}}{d\tilde{t}} + \frac{d\tilde{I}_{Vk}}{d\tilde{t}} \tag{5}$$

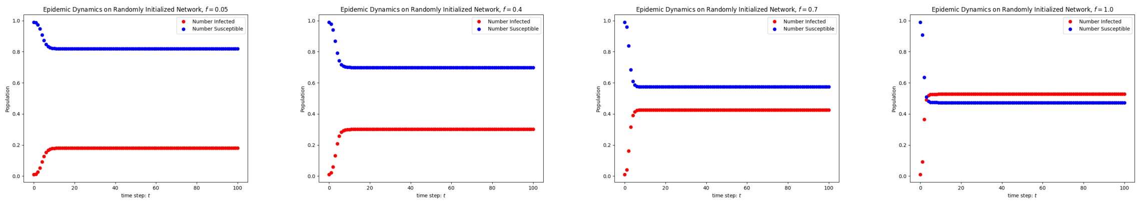
When we integrate each $\frac{\tilde{I}_k}{d\tilde{t}}$, we are approximating the proportion of infected individuals at time $t+1$ with degree k. To find the total proportion of the network that is infected, therefore, we can sum proportions for all $\tilde{I}_k$ to find the total proportion of infected individuals.

$$I_k = \int_t^{t+1} \frac{d\tilde{I}_k}{d\tilde{t}} \tag{6}$$

$$I = \sum_k I_k \tag{7}$$

The figures below are examples of the altered dynamics of the network, given the inclusion of the vaccinated population.

(c)



(a) $f = 0.05$ **(b)** $f = 0.4$ **(c)** $f = 0.7$ **(d)** $f = 1.0$

Figure 3: Change in proportion of Infected and Susceptible over time

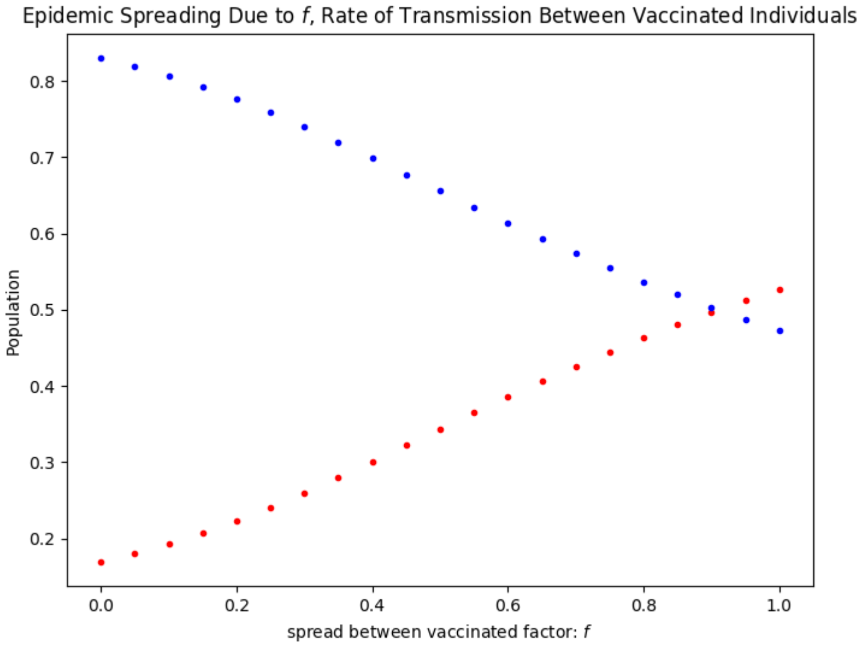
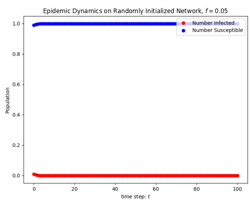
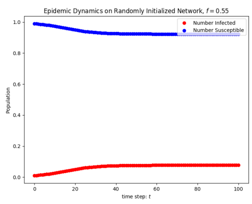


Figure 4: Effect of spreading factor f between vaccinated nodes

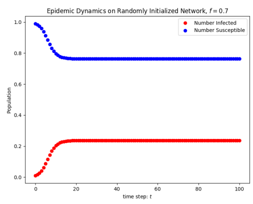
(d)



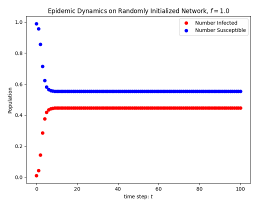
(a) $f = 0.1$



(b) $f = 0.5$



(c) $f = 0.7$



(d) $f = 1.0$

Figure 5: Change in proportion of Infected and Susceptible over time

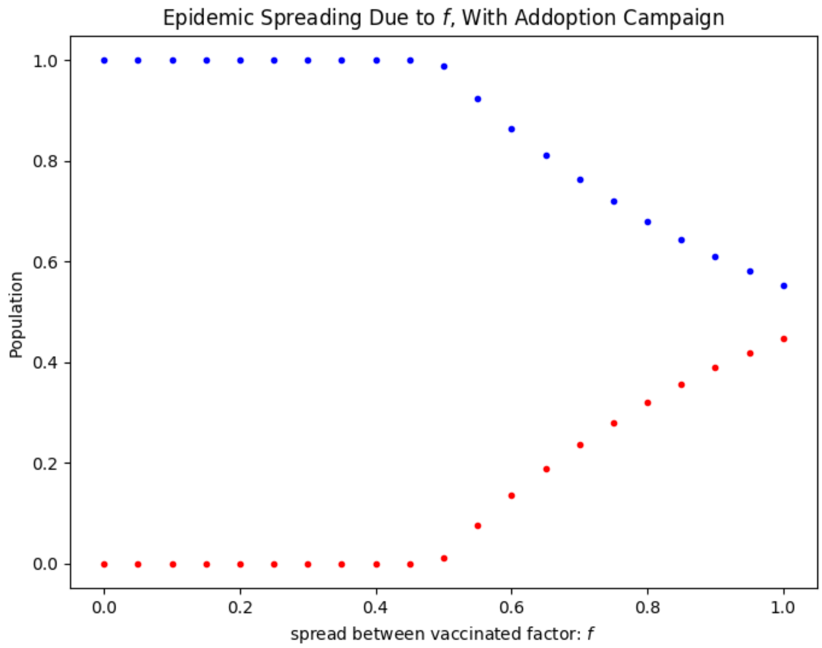


Figure 6: Effect of spreading factor f between vaccinated nodes

Links to two animations showing the dynamics of the network over time can be found here: No vaccine adoption, Upper 40% Vaccine adoption. The two examples show us that without adoption, the proportion of infected individuals does not die out over time. It also highlights how higher f will yield a higher proportion of infected individuals. This might simulate a vaccine that controls one infection, but makes individuals more susceptible to another.

0.4 Problem 4

The Voter Model

(a) A brief description of the network:

We choose was an interaction network from the hit HBO series "Game of Thrones". In this voter model, we are tracking the support of John Snow from the most powerful characters in Westeros. Our network contains 107 nodes, each representing a character from the series. Linking the nodes are 352 edges. The average degree of our network is 6.579. The degree distribution is displayed below:

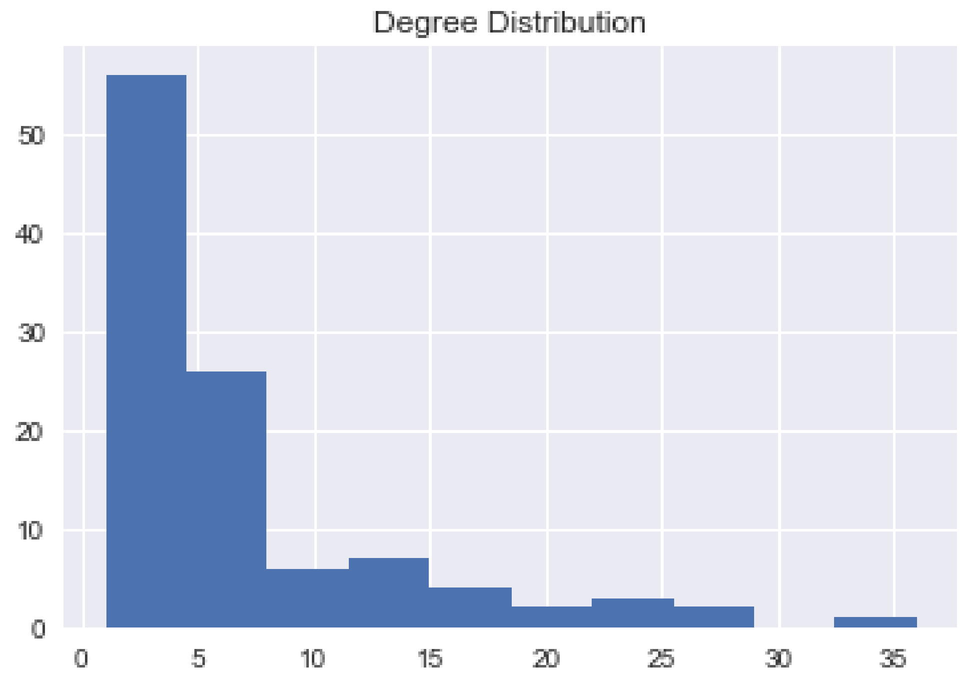
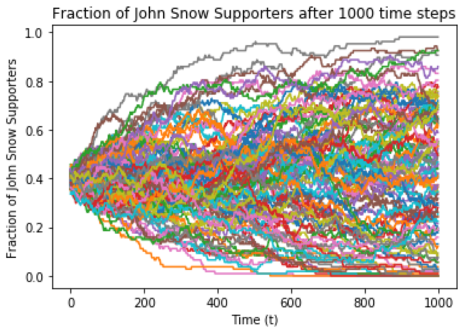


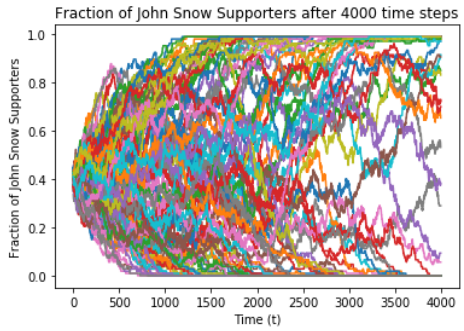
Figure 7: Degree Distribution of network

(b)How often do the characters form a consensus? After how long?

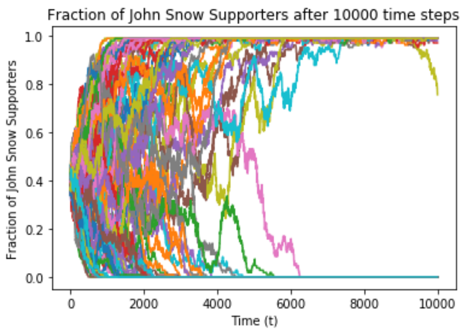
To see how communities come to consensus in our model, simulations from four time intervals are plotted below. Each line represents a different simulation where the fraction of John Snow supporters are plotted against time.



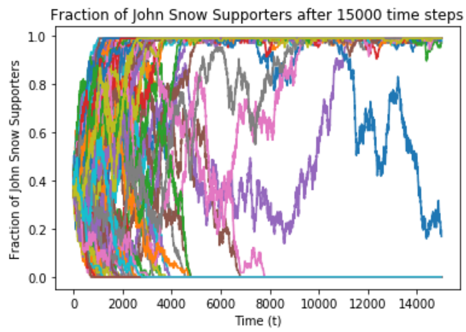
(a) Fraction of supporters over 1000 time steps



(b) Fraction of supporters over 2000 time steps



(c) Fraction of supporters over 10000 time steps



(d) Fraction of supporters over 15000 time steps

We see that over time, almost all simulated voting communities come to a consensus. Below we see the consensus time distribution with the average highlighted. The time distribution is skewed to right, some communities take a lot longer to come to a consensus.

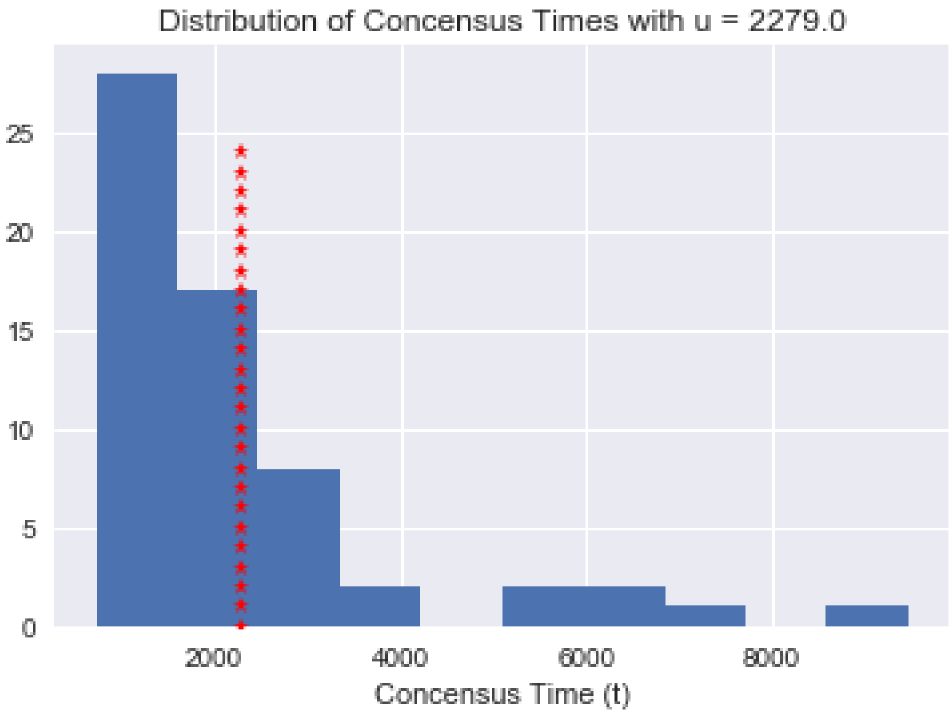


Figure 9: Degree Distribution of network

The degree distribution of the network is heavy tailed, meaning that there are a main characters like John Snow or Deaneries Targarion that have many more edges than a character that only interacts with one other person on set. Thus, these popular characters are much more likely to be selected as the neighbor that influences a nodes opinion at a given time step in the voter model. These nodes have a larger vote in this model, other people will change their own opinion with a high probability to match a well known characters. Thus, there may be a correlation with power law degree distributions and time before consensus.