

# Assignment # 3

## Modeling Complex Systems (CS/CSYS 302)

Lily Shapiro, Mahalia Clark, & Thomas O'Leary

### Question 1: The cascade model

#### Part A

This network will always be acyclic because if we pick a random node,  $i$ , and step along one of its outward edges chosen at random to another node  $i'$ , and keep stepping along random outward edges from node to node, we can only step to nodes whose value of  $i$  is smaller, never larger. So wherever we start, we will always end up at a smaller value of  $i$ . We can never step from a node with a small value of  $i$  back to a node with a larger value of  $i$ , so we can never return to our starting point to complete a loop.

#### Part B

To calculate the average in-degree of vertex  $i$  we can think about all the nodes that have an index  $> i$ , of which there are  $N - i$ . Each of these could be connected to node  $i$  with a probability  $p$ , so the average in-degree of vertex  $i$  is  $(N - i) * p$ . Conversely, to get the average out-degree of vertex  $i$  we can think about the number of nodes with index  $< i$ , of which there are  $i - 1$ , so the average out-degree of vertex  $i$  is  $(i - 1) * p$ .

#### Part C

To calculate the expected number of edges that run from nodes  $i' > i$  ('big nodes') to nodes  $i' \leq i$  ('small nodes') we can begin by counting the number of big nodes and the number of small nodes. Then we can think about the expected number of connections between them if nodes in one group are connected to nodes in the other group with probability  $p$ . There are  $N - i$  nodes with  $i' > i$  (big nodes) and  $i$  nodes with  $i' \leq i$  (small nodes). Any individual node among the  $N - i$  big nodes could have up to  $i$  edges running to small nodes, one to each of them, and if each of these edges exists with probability  $p$ , we would expect any single big node to have  $i * p$  edges to small nodes. If we sum these up over all the big nodes, we get a total of  $(N - i) * i * p$  edges from big nodes to small nodes, or  $(N * i - i^2) * p$  edges.

#### Part D

Assuming  $N$  is even, what are the largest and smallest values of the quantity calculated in c) and where do they occur (in terms of  $i$ )? If we plotted  $(N * i - i^2) * p$  over  $i$ , we would get a downward-facing parabola with roots at  $i = 0$  and  $i = N$ . Since  $i$  can only have values from 1 to  $N$ , it cannot equal zero, so the smallest value of the function is 0 at  $i = N$ . The largest value of the function occurs at the top of the parabola, which is found halfway between the roots, at  $i = \frac{N}{2}$ , where the function has a value of  $\frac{N^2}{4} * p$ .

### Question 2

### Question 3

Our network represents the 48 contiguous US states in addition to the District of Columbia (49 nodes total). Knowing that question 4 revolved around analyzing the voter model and the capacity of a network to reach consensus (i.e. the same state of all nodes), we thought it pertinent to have our chosen network directly correspond to this, thereby choosing a basic U.S. map. The appearance of the network itself loosely matches with the geographic orientation of

states and subregions within the U.S and a view may decipher patterns corresponding to known locations of states in such a network structure, although there are obvious departures as the network needs not necessarily exactly correspond. Clearly the nodes in this network are representative of US states, as labeled, with the edges representing the geographic borders shared between states. Maine is the only U.S. state connected by a single edge, sharing a land border only with New Hampshire (the beautiful seacoast of NH prevents it from coming into contact with Massachusetts). Other states are far more connected, particularly Missouri and Tennessee which have 8 edges apiece. The average degree (or average number of edges shared between states) is 4.28, a logical output as most states share about 4 to 5 borders with other states.

In Fig() we have displayed the nodes according to betweenness-centrality, where it is obvious that Missouri is the most central state within the network, that is, it most often acts as a connection between shortest path lengths amongst other nodes. New York also serves as an apparently important central state, acting as an apparent bottleneck, cutting off New England from the rest of the country. States that have high betweenness centrality likely have larger influence on network dynamics, possibly including how and if states will reach a consensus within the voter model.

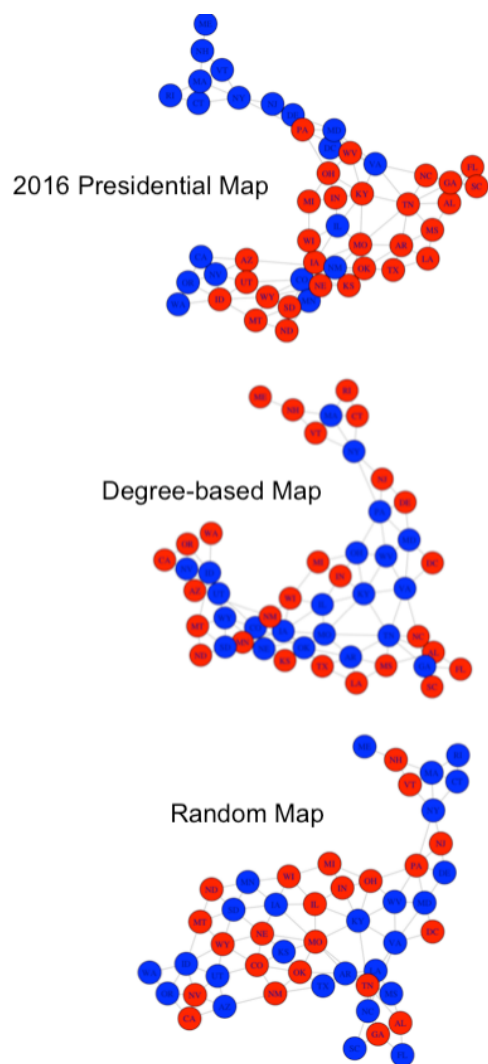
The graph is moderately clustered, with a clustering coefficient of 0.507, which is in accordance with the relatively regular geographic structure of states on the map. Assessing the modularity of the network reveals 6 distinct communities (Fig:) with an associated value of 0.579. These more densely connected communities may influence each other more than those outside of these communities.

## Question 4

### Part A

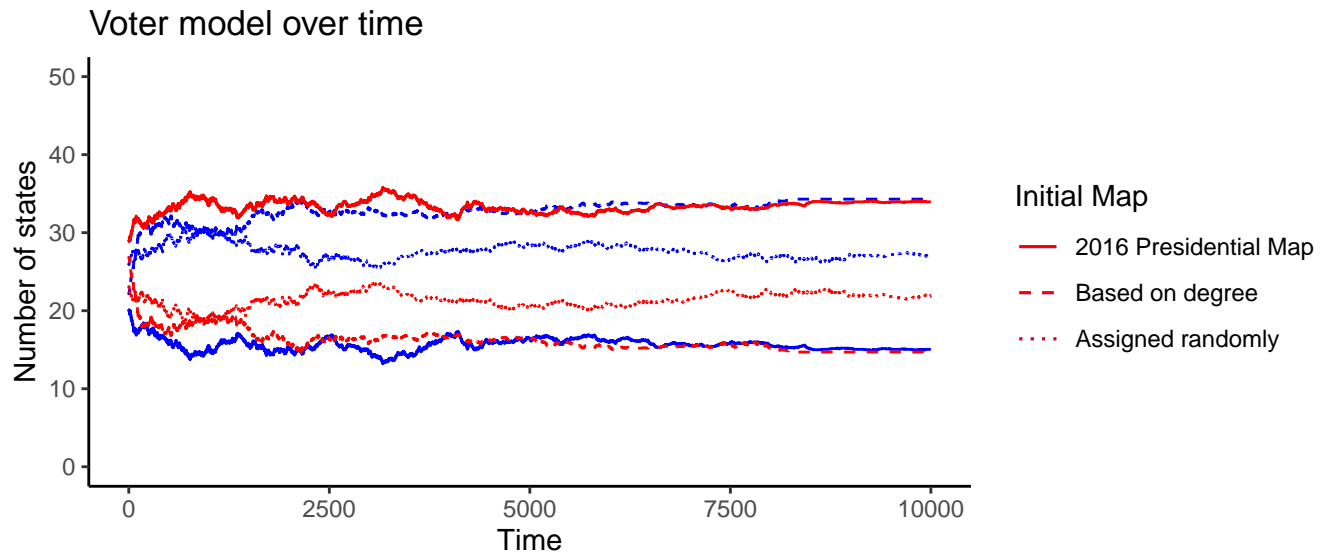
Using the network you chose in the previous problem, implement the voter model and run it on your network using some initial conditions (i.e. 50% red and 50% blue randomly distributed). Averaged over a few runs, does the network tend reach consensus? If so, how fast? Try a few different initial conditions.

## Initial Conditions



**Fig. XXX:** Graphical representation of the initial conditions for the networks for each of the initial conditions that we tested, 2016 Presidential Map (top), Degree-based map (middle), and a Random Map (bottom).

## Voter Model Runs



### Part B

Now, find the degree distribution of your network and run (or imagine) the voter model using a resulting configuration model network. How do you expect the dynamics to compare to the real network? Explain with simulation results or in words.

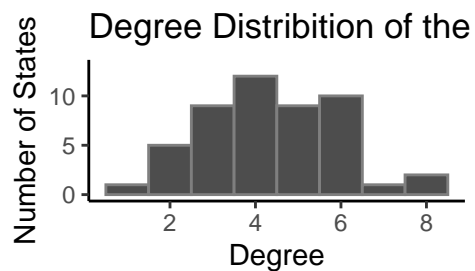


Fig. XXX: Degree distribution of the United States Map Network