First network exploration

AI, Gabor – Mar 4, 2020

This assignment is due on the next class day. If you are unable to complete it, you will turn in a printout of your code along with an explanation of why you were unable to finish it. Remember that your name, date, and period go in the upper right hand corner of what you turn in.

The assignment is an exploration of what happens with networks. We will understand a network to mean large graph, and typically this means millions of nodes. Some examples are brain cells (some connected by synapses), servers on the internet, people (linked by friendships), actors (linked when they star in a common movie), web sites on the internet.

Up until about 20 years ago, the way that networks were modeled was the following: A graph was constructed with a certain number of vertices (call it n). Then, edges were added at random until the average degree of the vertices achieved a certain desired amount.

Part 1 of your this assignment is to implement the above paragraph for $n = 100{,}000$ where the average degree of each vertex in your network is to be 4, and then in a second example it should be 5. Your network will have only undirected edges, no self loops, and a pair of vertices can have at most a single edge between them. You will turn in a depiction of a of a relationship between two quantities (a precalculus type of graph), one graph for when the average degree is 4, and one for when the average degree is 5. The horizontal quantities, x, are integers going from 0 to the maximum degree over all vertices. The vertical axis represents the number of nodes that have degree x, call it the $degreeCount(x)$. Thus, if you sum all of the y values, you should get the total number of vertices, while if you take the sum of each x value times the corresponding $degreeCount(x)$, you should get twice the total number of edges. You should turn in Part 1 on one side of a single sheet of paper. The actual values you plot should be indicated, and of course which graph has which average vertex degree should be indicated, too.

Please complete this part before looking ahead to Part 2, on the other side of the paper.

First network assignment continued (part 2).

Part 1 should be finished before reading further

The amazing thing is that there are certain properties that almost all real work networks exhibit, whether biological in nature or not, including the examples on the opposite side of this page. The way you modeled it in Part 1, which was the standard way up until about the year 2000, was incorrect. When one first hears this, it really is quite surprising, that anyone could make such a statement with applicability to such disparate collections of many objects which have some randomness in their connections.

The explanation for the incorrectness is that most networks tend to grow from small to large, and the connections are not made in the random way that one initially supposes. In fact, based on observation and borne out by modeling, when a new node is adding edges, the attachments can be viewed as random, but the likelihood of attachment to an existing node is proportional to the number of attachments that the existing node already has! Qualitatively, one might say that the rich get richer. The person who has a lot of friends is more likely to be befriended by others than those persons with fewer friends.

Part 2: You are to model the above paragraph, again using $n = 100{,}000$ and where the average degree of each vertex is 4 in one example, then 5 in the second. Specifically, start from a small number of vertices, say 5 or 6, connected as you wish, and then add one new vertex at a time, deciding how many edges it will have so that the average degree stays at about what you decided for your example. Once you've decided how many edges a new node is to have, connect those edges up to the other existing vertices so that the likelihood of connection to another vertex is proportional to the degree of that vertex. Once your graph is completed, you should display the same type of diagram as on the previous page except that you will have to be more judicious in how many values of x you display. Make sure to indicate on your diagrams the maximum of all the vertex degrees in your graph.

Part 2B) Tweak the model for part 2 in one way and produce comparable graphs. A possible tweaking would be to take pity on the lonely singletons and say that the likelihood of connection to another vertex is proportional to 1 plus the degree of that vertex.