# Overview of Social Network Metrics

After we create a social network, we can start doing descriptive analytics. A common way of doing this is to describe the key characteristics of the network using a set of metrics.

In what follows, we will first introduce the geodesic concept. We will then discuss various network centrality measures such as the degree, the closeness, the betweenness, and the graph theoretic center. Centrality measures quantify the importance of a node in a social network from various perspectives. They can be computed based on the whole network structure, or a subgraph thereof.

The geodesic represents the shortest path between two nodes. In this example network, the geodesic between nodes A and I is represented by the red edges. Note that the weights can be taken into account when you calculate the geodesic. The geodesic is a very interesting concept. Consider a fraud detection setting, for example. You can calculate the geodesic between a node representing a currently legitimate customer and a node representing a fraudulent customer. When the fraudulent node is closer, it has a greater influence and impact.

The degree of a node represents the number of edges or connections.

In directed graphs, a further distinction can be made between in-degree (representing the incoming connections) and out-degree (representing the outgoing connections).

In the example, node A has two connections, node B has one connection, node C has three connections, and node D has two connections. For this network, the maximum number of connections and degree equals 3.

More generally, in a network with n nodes, the maximum degree equals n - 1. The normalized degree can be obtained by dividing the degree by the maximum degree. In this example, you can see that node C has a normalized degree of 1 because it has the maximum number of connections.

The closeness measures the extent to which a node is near to all other nodes in the network. In other words, it measures the distance of a node to all other nodes in the network. The closeness of a node i is computed as the inverse of the sum for j going from 1 to g of the distance between ni and nj where g represents the number of nodes in the network. Note that the distances are calculated by using the geodesic or shortest path.

For a network with g nodes, the maximum closeness is obtained when a particular node is connected to all other g - 1 nodes, resulting in a closeness of 1 divided by (g - 1).

Hence, we can calculate the normalized closeness by dividing the distances by g -1, or in this case, 3. We can again see that node C has maximum closeness. The closeness is a very important measure. If a fraudulent node has a high value for closeness, then fraud might easily spread through the network and contaminate the other nodes. Finally, note that the farness is defined as the reciprocal of the closeness.

The betweenness counts the number of times that a node or edge occurs in the geodesics of the network. The betweenness for node i can be computed as the sum for j less than k, of gjk(ni) divided by gjk, where gjk(ni) represents the number of geodesics connecting node j and k and passing through node i, and gjk is the number of geodesics between node j and k. The betweenness can be normalized by dividing by (n minus 1) times (n minus 2) divided by 2, which is the number of pairs of nodes excluding the node itself.

Here you can see a simple example of how the betweenness can be calculated. Nodes A and E are between no two other nodes. Hence, their betweenness is 0. Node B is between A and C, A and D, and A and E. This results in a betweenness of 3. Node C has the highest betweenness because it is between A and D, A and E, B and D, and B and E. Finally, Node D is between E and C, E and B, and E and A. This results in a betweenness of 3.

Here you can see a more complex network. The red node has the highest betweenness because it connects two communities. When a node in one community needs to connect with a node in the other community, the

connection needs to pass through the red node. If the red node is infected by fraud from one community, it can easily pass this on to the other community. Hence, if this node can be disabled, the fraud contamination can be controlled.

The graph theoretic center is the node with the smallest, maximum distance to all other nodes. This node is the most central node in the network. In terms of information diffusion, the graph theoretic center is the node that influences other nodes the fastest. In the network shown here, the red node is the graph theoretic center. All other nodes can be reached within two steps. If the graph theoretic center is contaminated with fraud, then fraud might rapidly spread throughout the rest of the network. Hence, special attention should be paid to this node.





Close