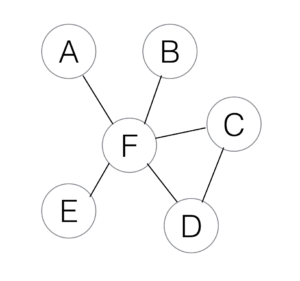
By representing phenomena as networks, we can study the mathematical properties of their structure. Typically, in network analysis, we quantify properties using network metrics. These network metrics are developed to answer specific questions. For example, we can ask which nodes are the most important in the network. The metrics which answer this type of question are called centrality metrics. The three most common centrality metrics are degree, closeness and betweenness. In the context of social networks, one question we can ask is whether our friends are friends of one another as well. The clustering coefficient was developed to answer this question.

*How many friends do you have?*

**Degree**. The degree of a node is the number of edges connected to it. It is a measure of how many relationships the node has.

In the case of Facebook, degree is the number of friends that a user has. Degree only considers the friends of a certain user and doesn’t really take into account the Facebook network as a whole. Because of this, degree is considered as a local metric. A user who has high degree we can think of as a local celebrity – he has lots of connections in his sphere of influence, but not necessarily outside of it.

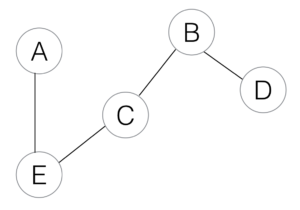


The node F is the node with the highest degree in the network, since it is connected to everyone else, whereas A, B, and E have the lowest, since each only has one connection, F.

*How near are you to everyone else?*

**Closeness**. The closeness of a node is defined as the inverse of its farness, where farness is defined as the sum of the distances of the shortest paths from the node to every other node in the network. A node which has high closeness requires very little travel time to get to other nodes in the network.

Since we need to calculate shortest paths between every node pair in the network, closeness is, as opposed to degree, considered a global metric. Every node in the network needs to be considered to compute it.

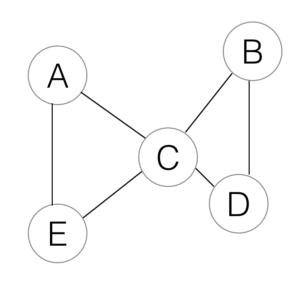


Node C has the highest closeness in the network. Since C is at the midpoint of the linear network, it has the smallest average distance to every node in the network. On the other hand, A and D are at the ends of the line, and so their average distance to the other nodes is at the maximum.

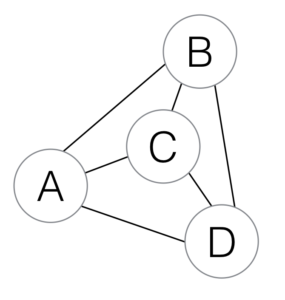
*How often do you connect other people?*

**Betweenness**. The betweenness of a node is defined as the number of shortest paths (of every other node in the network) that passes through it. Intuitively, we can think of a node with high betweenness as a bottleneck connecting two or more large collections of nodes with one another. Removing a node with high betweenness from the network will often divide the network into disjoint subnetworks.

Again, betweenness is a global metric since it requires the calculation of shortest paths.



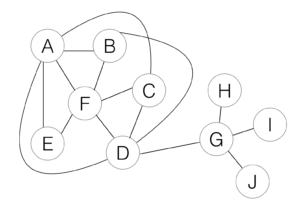
Node C has the highest betweenness here, since the shortest paths from {A,E} to {B,D} need to go through C. On the other hand, no shortest path passes through any other node in the network.



Node C has high closeness and low betweenness, since the shortest path from C to any other node is one, while the shortest path between nodes other than C do not pass through C.

*Are your friends acquainted with one another?*

**Clustering**. To calculate the clustering coefficient of a node, we look at all other nodes connected to the specified node, and count the number of edges which exist among the node’s neighbors. The clustering coefficient is a measure of how connected the node’s neighbors are – do they cluster around the node?



Node F has high clustering. The majority of F’s neighbors are connected amongst themselves. The opposite happens with node G, wherein none of his neighbors are connected.

All of the metrics defined above refer to a specific node in the network. We can also associate a metric to the whole network. How do we do this? By averaging the metric over all the nodes in the network.

Usually, the metrics presented above are normalized to be in the range 0 to 1, that is from weakest to strongest. A degree of 1 says that the node is connected to every other node in the network. A clustering coefficient of 1 means that all the neighbors of the node are connected to one another.

Taking it one step further, networks sometimes have added information in the form of edge weights. These edge weights quantify the strength of relationship represented by the edge. These edge weights can be included in the calculation of degree, closeness, and betweenness by setting the distance between two nodes as the corresponding edge weight.

For precise mathematical definitions, check out the documentation of NetworkX.

<https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/>