Social Network Components

As I already explained, the two fundamental building blocks of a social network are nodes and edges. It is very important that both are unambiguously defined and optimized for the analytical task, such as churn prediction or fraud detection.

To construct the network, you typically start from a transactional data source such as a CDR log. You can see an example here. For each call, the log includes information about the caller, callee, duration, from location, to location, from provider, and to provider. The transactional data source can then be used to construct the network shown below. Depending on the business setting and application, the nodes or vertices can be customers, companies, products, credit cards, accounts, web pages, and so on. The edges or links define the connection between the nodes. The edge definition depends on the context that you are working in (for example, churn, fraud, and email). The edges can be weighted based on factors such as interaction frequency, importance of information exchange, intimacy, and emotional intensity. In our churn prediction example, the edge between customers 1234 and 9876 has a higher weight than the edge between customers 9876 and 6543, because customers 1234 and 9876 call each other three times a day whereas customers 9876 and 6543 call each other only once a month. The weights are usually positive values. Negative weights can be used to represent an animosity relationship, although this is very rare.

Networks can be visually represented in several ways. A first way to represent a social network is by using a sociogram, as you can see illustrated here. Our network has four nodes, A, B, C, and D. The nodes can also be colored according to their status such as fraud or non-fraud. Sociograms are convenient for small networks or subsets of networks, but not practical to represent large-scale networks. A second, more useful representation is an adjacency or connectivity matrix. This is a square matrix with the nodes in the rows and columns. The cells represent the connections, with 1 indicating a connection and 0 no connection. For undirected connections or edges, the matrix is symmetric. It will obviously be a sparse matrix because it will contain many zeros. Here you can see the adjacency matrix for the example network. An adjacency list is another way of representing the network by providing a list of all its connections. Here you can see the adjacency list for this network. Both the adjacency matrix and the adjacency list can be extended to also account for the weights of the network connections. Here you can see the weighted adjacency matrix and weighted adjacency list for the example network.

Close