EXperimental Learning

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Big Data and Social Analytics certificate course

MODULE 4 UNIT 1
Video 4 Transcript

MIT BDA Module 4 Unit 1 Video 4 Transcript

Speaker key

XD: Xiaowen Dong

HY: Hapyak

XD: So in the previous videos I mentioned that we use networks to capture relationships or structures between data entities therefore it is important to understand the structures of networks which is the subject of the following two videos. In this particular video we are going to study vertex centrality measure which basically answers the following question: How to find important vertices in a graph?

The most straightforward way for us to measure the importance of a vertex is by its degrees which we also call degree centrality. As we studied in the previous videos, a node with high degree means that it is connected to the most number of neighbors which usually indicates that it is the important vertex in the network.

00:01:01

As an example, we see a graph here where the colors on the vertices indicate their degree centrality, namely the brighter the color the higher the degree centrality. Degree centrality can be useful, for example, to identify individuals who have most connections in a social network.

The second measure of centrality is the closeness centrality which measures, on average, how close one vertex is to all the rest vertices in the graph. Specifically the closeness of a vertex in the graph is defined as the inverse of the mean of the distances between this vertex and all the rest vertices where the distances between a pair of vertices i and j is the length of the shortest path between them, namely d_ij.

In our example we see that the, loosely speaking, vertices that are in the center of the graph has a higher closeness centrality and such a centrality decays smoothly towards the boundary of the graph. Closeness centrality is a very natural centrality measure in the network. Practically, an individual in a social network with a very high closeness centrality might find his opinions reach all the other users very quickly in the network.

00:02:12

The third centrality measure is called betweenness centrality which measures the extent to which a particular vertex lies on the path between all other vertices. Specifically, the betweenness of vertex i is computed as follows. For every pair of other vertices, such as j and k, we first compute the ratio of the number of shortest paths between j and k that goes through i, to the total number of shortest paths between j and k. The betweenness centrality of vertex i is therefore defined as the sum of such ratios for all possible pairs of j and k.

The betweenness is, in one sense, a slightly different centrality measure because instead of measuring how well the vertex is connected, it measures how much it falls between the others. Such a measure can therefore be used to identify vertices that have most control over information propagation in the network; a practical example being individuals who are key to communications to the others in a social network.

00:03:18

Finally, we introduce the concept of eigenvector centrality. The idea behind eigenvector centrality is that a vertex is considered important if it is connected to vertices that are themselves important. The eigenvector centrality of vertex i is the i^{th} entry of the eigenvector that corresponds to the largest eigenvalue of the adjacency matrix.

The technical explanations of such definitions can be found in the additional notes but such a definition gives a nice property of the eigenvector centrality, that it can be high either because a vertex is connected to many other vertices or it is connected to only a few, but important ones. In this example we see that vertices with high eigenvector centrality are themselves close, forming a group of important vertices in the graph.

Practically, eigenvector centrality can be used to identify individuals who know important people in the social network or webpages that are linked to important webpages on the Internet. Actually, Google's PageRank algorithm is a variant of the eigenvector centrality measure.

HY: What comparisons can you draw about the four centrality measures?

Continue watching for a comparison of the four centrality measures.

00:04:33

XD: In this video we have introduced four different kinds of vertex centrality measures. Here we see a comparison between them on the same graph. In the additional notes we will talk more about their respective advantages and disadvantages. So, notice that vertex centrality measure quantifies the importance of individual nodes. In the next video we will take a more global view of the structure of networks.

Tel: +1 224 249 3522 | **Email:** info@getsmarter.ac | **Website:** getsmarter.mit.edu