EXperimental Learning

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Big Data and Social Analytics certificate course

MODULE 4 UNIT 1
Video 3 Transcript

MIT BDA Module 4 Unit 1 Video 3 Transcript

Speaker key

XD: Xiaowen Dong

XD: In this video we are going to look at basic definitions of undirected network, directed network and weighted networks and also important concepts such as the degree of the vertex, path and the shortest path in network and also component of the networks. Since we are going to use graphs to represent networks, from now on I will use the terminologies networks and graphs interchangeably.

00:00:34

A graph consists of a set of vertices or nodes which are connected by a set of edges or links. So in this case we have an example graph that has four vertices and four edges. The vertices are indexed by the numbers in the blue circles but such a ordering is usually arbitrary.

Usually vertices represent data entities and edges represent that there exists a certain type of relationships between such entities. For example, we can think of vertices as people and edges indicate whether they are family members or colleagues, etc. Graphs can usually be represented by an adjacency matrix where the ij-th entry of this matrix is 1 if vertices i and j are connected or 0 otherwise.

Here we have an example of a so-called undirected graph therefore the adjacency matrix is symmetric. A graph can also be directed, meaning that edges can also be directional. In this example we see that we have five directed edges between four different vertices. Directed graphs can be used to model relationships that are not necessarily reciprocal, for example, followers on Twitter.

Directed graphs can also be represented by an adjacency matrix where the ij-th entry is 1 if there is an edge from the vertex i to the vertex j. However such an adjacency matrix does not have to be symmetric anymore.

00:02:01

HY: What is the difference between a directed and an undirected graph with regards to the adjacency matrix?

The adjacency matrix of an undirected graph is symmetrical, while the adjacency matrix of a directed graph is not.

XD: In the previous examples edges are binary in the sense that they indicate whether the relationship exists or not, but they can also be associated with a non-negative weight that indicates the strength of the relationship such as the weighted graph that we see here. Examples of weighted graphs can be a network of scientists where the edges represent the number of papers that they have co-authored or a network of customers where the edges represent the number of shops that they co-visited.

Now, one important concept is the degree of the vertex. In an unweighted graph the degree of a vertex is simply the number of edges incident to it. In a weighted graph the degree of a vertex is defined as the sum of weights of all the incident edges. For example, in this weighted graph we see that vertex 1 has a degree 2 and vertex 2 has a degree 5. Intuitively, the degree of a vertex tells how well and strongly it is connected to the rest of the graph. For example, in a collaboration network, a vertex with high degree represent a scientist who has collaborated with many other people.

00:03:21

HY: The degree of vertex tells us how well and how strongly a vertex is connected to the rest of the graph.

True.

Correct, well done. The degree of vertex does actually indicate how strongly a vertex is connected to the rest of the graph.

False.

Incorrect. The degree of vertex does actually indicate how strongly a vertex is connected to the rest of the graph.

XD: Another basic concept in a graph is path. A path is a sequence of vertices such that every pair of consecutive vertices are connected by an edge. Here we show three paths in a graph. The shortest among all the paths that connecting i and j is called the shortest path with length d_ij. In this example, the shortest paths between vertices 1 and 2, is from 1 to 4 to 2, which has a length 2. Notice that shortest paths do not have to be unique as there might exist more than one path that have the same shortest length.

Shortest path is an important concept because it measures how fast one vertex can reach another vertex in the graph. In practical examples this may be used to study how fast information can be propagated along the network.

HY: Why is the shortest path concept important?

It's important because it measures how quickly one vertex can reach another in the graph.

XD: The last concept that I'm going to introduce in this video is component, and component, in an undirected graph, is the maximum subset of vertices such that every vertex can be reached by other vertices by some paths. In the example graph shown here we have three components. The first is vertices 1, 2, 3, 4, the second is vertices 5 and 6 and the third one is vertex 7.

00:04:46

Component of the largest size is usually called the largest connected component or giant component.

So, as a small recap, in this video we have introduced some basic definitions and concepts and we are going to use them to measure the structures of networks in the following videos.