EXperimental Learning

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Big Data and Social Analytics certificate course

MODULE 4 UNIT 1
Video 2 Transcript

MIT BDA Module 4 Unit 1 Video 2 Transcript

Speaker key

XD: Xiaowen Dong

HY: Hapyak

XD: Hello, everyone. I hope you have enjoyed the class so far. I'm Xiaowen, I'm from the Human Dynamics Group of MIT Media Lab. In the next few videos, I'm going to talk about an important subject, namely network. I will introduce some basic definitions and concepts of networks, along with typical data processing techniques on networks. So, in this very first video, I will first briefly mention the motivations, why we are interested in studying network, and then I will give you a few examples on empirical networks, along with the mathematical representation of networks. So, in this course, we focus on big data. The data are big, obviously, because nowadays they came in overwhelming volume, but volume is not the only problem. Data are increasingly complex, with rich structures behind them.

00:00:59

For example, temperature records in a local regions are related to their geographical proximities; traffic volumes in a road network depends on the topology of the network; and the behavior of a group of people may be influenced by the friendships between them. We are interested in studying networks because, first, networks capture such structures or relationships between data entities. As I said before, it can be a geographical network, transportation network or social network. Secondly, we have developed effective ways to analyze and measure such networks. In the following videos, we are going to talk about various measures, and analysis tools on networks.

Here, we see four different kinds of empirical networks. The first is the United States transmission grid, which is an electrical network for delivering electricity in the country. We see that this network captures the structures of electricity transmission lines in the United States. The second is the road network in Minnesota, which is a transportation network. This network would obviously influence the mobility patterns in the region. The third example is a brain network, in which we see that different regions of the brain are linked to each other, if they communicate with each other. This network reveals how the different regions in the brain are functioning in a collective way.

00:02:27

Finally, we have a social network of Twitter users, where there is an edge between users A and B, whether A follows B or B follows A. It is clear, in this case, that this network reflects the social relationship between Twitter users, which usually drives their behavior and personal preferences. In this course, we represent networks by a mathematical object called graphs, which are subjects of a branch of mathematics called graph theory. Graph theory, as a field, dates back to the famous Königsberg bridge problem, almost 300 years ago, which asks whether there exists a walk through the Prussian city of Königsberg that would cross each of the seven bridges once and only once. In 1736,

Swiss mathematician, Leonhard Euler, published a paper to prove that such a solution does not exist, which marks the beginning of graph theory. In particular, Euler represented the four different parts of the city as four vertices and the seven bridges as seven edges that link these vertices. This gives rise to a graph structure based on which he was able to solve the problem.

HY: Graph theory will be expanded on in the following videos, what is your previous experience?

a. Extensive
b. Limited
c. None

Thank you.

XD: Ever since then, graphs have been widely adopted as mathematical tools to model pairwise relationships between entities, and we will see this in more details in the following videos.