MIT | EXperimental Learning

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Big Data and Social Analytics certificate course

MODULE 4 UNIT 1
Video 1 Transcript

MIT BDA Module 4 Unit 1 Video 1 Transcript

Speaker key

AS: Arek Stopczynski

HY: Hapyak

00:00:09

AS: Hello again. You're almost half way through the course. I hope it's going great so far. I'm having a blast actually talking to you. Right now, let's talk about networks and living in networks. In the world of today we are living in an increasing number of networks. By networks, we mean networks of physical interactions, but also online social networks, Facebook, Twitter and Google+, networks of telecommunication calls and texts, networks that you might not even think about that much but every time you are collaborating with some project on GitHub you're also part of the network of people creating the code and working together.

And in the world today it's the number of networks that increases, but also our ability to capture this data and to analyze it is growing. And this fact that the number of networks is growing and our ability to actually understand them is growing as well, it's extremely exciting because those networks, as we'll see in a moment, they start giving us the insight into complex social systems. They enable us to open them up and literally look how things are interconnected and work together and how processes unfold on these networks. And we've never been able to do it before. 20, 30, 40 years ago it was almost impossible to have his kind of resolution and this variety of data that really enables us to understand people that work together and live together.

HY: Which of the networks that you form part of would you be most interested in analyzing?

Thank you for your thoughts, please continue watching to see an analysis of a few different social networks.

AS: Something that's extremely important to remember is networks are different. Even if we capture networks that describe the same population, they actually differ in their structure and dynamics and they also signify different things, and they're important to understand different things about this population.

So, for example, networks of physical interactions, they tend to be random and very dense, especially if we consider long-range interactions versus short-range. Strangers, all the people that you bump into the corridor, it generates so many, so many interactions, so the network itself is extremely dense and quite random.

00:02:30

IN COLLABORATION WITH getsmarter

When we look at social networks such as Facebook, for example, it has much more structure, but it's still to some extent interactions between individuals that may not be really strong ties. Think about all those likes that you put in the posts of people that you don't really know. You kind of know them but not necessarily.

And then we go to telecommunication networks such as texts or calls. People relatively rarely text or call someone they don't really know, especially calls. They really provide a very strong signal of the social structure. They may not capture everything, you may not call every single of your friends, but when we actually see that you are calling someone, it usually means that there is some relationship in there, there is some link that is much stronger than, for example, Facebook link.

And now, looking at those different networks, it's actually interesting for studying different phenomena. For example, physical networks are extremely important for understanding of epidemic spreading because this is how pathogens that are infectious actually transfer from human to human through physical contact.

00:03:43

But to understand spread of rumors, probably Facebook might be more interesting or to understand social relationships, maybe the call or text network is actually the most important one.

HY: The spreading of rumors is most effective through proximity networks.

True.

Incorrect. The spreading of rumors is actually most effective through social networks such as Facebook, while the spreading of disease most effective through proximity networks.

False.

Correct, well done. The spreading of rumors is actually most effective through social networks such as Facebook, while the spreading of disease most effective through proximity networks.

AS: Networks are fascinating to study, not just because they describe the entire social complex systems, but also because our position, our own position in these networks may have a huge impact on our lives. So, for example, how we are positioned in the network may have an influence how well our team is doing at the university or at work. In a study called The Strength of the Strongest Ties, we took a course at technical university and we split people into teams and they were all working on the same assignments, what we were really interested in is how is the connectivity of the group to the rest of the class actually correlates with how well is this group doing.

So what we measured was exactly that. We have groups and we have 20 of them, let's say, and they are all composed of people attending the same class, the same course. And we looked at what is the correlation between how well this group is connected to the rest of the population with this group's performance. We actually saw that this connectivity of the group to the rest of the population dominates simple measures of who was in the group. So we measured things such as technical proficiency by asking people to fill out a little technical test at the beginning of the course or their

personality traits, big five, and we really showed that the connectedness of this group was the most important factor.

But interestingly, it wasn't any link that was important. It was only the strongest links that proved to be important. Basically we saw a phased transition, a curve that looked like that and at that point it was only the strongest links that actually correlate and told us something about the performance of the group.

00:05:59

And you can think about it as having people that you would be willing to ask to help you move. Unless you have people like that, even though you might kind of know that this is this person, this is that person, they're not going to actively engage with them until they are strong enough in terms of the ties that you are actually feeling comfortable to talk to them and this opens up the entire world of information to your actual team. So it's not just the quantity, it's also the quality of the links and the idea that cultivating strong links is important for the team performance.

But there's also another side of the story of your position in the network. In another study we are seeing that people who are very central in their social networks also tend to get infected early if there is an outbreak of epidemics such as flu and we kind of understand intuitively why this is happening. If you're a very social person, if you are very central in the network, there is a very good chance that if anyone in your network becomes sick, this is going to reach you very quickly.

00:07:13

So, of course, being a central person might have downsides. On the other hand, once we know that we can actually use this information to do targeted vaccination or targeted monitoring of this network to contain this epidemic spread by simply picking the right people, the most central people, and making sure that they are vaccinated… of course, the topics we are talking about here are pretty complex and there is lots of equations and lots of data processing that goes into that, so don't worry if this video is only a teaser or gives you a very high level overview of what's coming, there's the entire written material that covers exactly what I was talking about but in much greater detail.

00:07:54