MIT EXperimental Learning

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Big Data and Social Analytics certificate course

MODULE 4 UNIT 4
Video 7 Transcript

MIT BDA Module 4 Unit 4 Video 7 Transcript

Speaker key

AS: Arek Stopczynski

HY: Hapyak

AS: Networks of physical interactions can be very complex, extremely complex, and the same time we know they are not random. Think about your own physical interactions. It's not just that you are bumping randomly into people, you actually work in groups, you go out together with people. There is a structure there, but traditionally it has been extremely hard to discover because if we take a population and we take a network of physical interactions and we aggregate it even after an hour or two hours, it becomes a hairball. Everyone is almost connected to everyone else.

00:00:46

And understanding communities or discovering communities, understood as groups or teams of people in that network, has been very hard. And primarily it has been hard because those communities, they tend to be overlapping. So every individual might belong to several communities – think your family and your kayaking club and your colleagues at work – and those communities are dynamic. They change quite a bit, sometimes at the order of days, sometimes at the orders of weeks. Some are stable, but there's a lot of going on in terms of the change.

And finally, they have soft boundaries. So you belong to your kayaking club but you're kind of meh about it so you're only going to participate in 20% of their interactions and their actual activities. You still belong to this community but you are not the core member.

So, how can we actually attack this problem? How can we take this huge hairball of interactions, physical interactions, and slice it in a way that would reveal what is the underlying structure in this complex social system. Traditionally, adding time dimension to the networks, making them temporal networks, has been increasing complexity because suddenly you have a new dimension that you need to take care of, but something very interesting happens when we study networks of social interactions, physical interactions, as they actually happen.

00:02:11

If you plot the data, you can see that the network doesn't get constructed by random links appearing all over the network. There are actual small communities that are present at every single point in time. Just think about your last meeting or your last get-together with friends. You were a very tightly connected group of people that met for minutes or maybe hours, and then you dissolved and then you all went to another set of meetings. And this is the aggregation, those people moving between the meetings what creates this huge hairball very quickly. But if our data is good enough, if it has resolution that is high enough, at the order of minutes we can actually see the network not as a hairball but as the actual interactions that lead to the creation of these networks.

So we can look at the communities directly, rather than trying to disentangle this hairball; can actually see how those communities happen in time. So the problem of community detection really becomes the problem of tracking common communities or common gatherings across time. So what we are doing, we are literally taking those multiple little communities that we are seeing in a sliced... in a slice of time and we are matching them across multiple time slices, and we are saying, oh, this looks like the same community because the members are very similar. It was a single meeting. Great. Now we have a single meeting.

And now we go to a long periods of time and we are saying, this community that has certain members, some of that members might have participated for the entire duration of the meeting and some might have left after few minutes. We're going take this graded membership and match it to different communities across long periods of time, and when it matches we're going to say, oh, it looks like the same community.

00:04:06

HY: The movement of people from one interaction to the next is the cause behind the formation of hairball networks.

True

Correct, well done. Hairball networks are formed very quickly by the movement of individuals from one network or gathering to another.

False

Incorrect. Hairball networks are formed very quickly by the movement of individuals from one network or gathering to another.

AS: So you are getting together with your kayaking club. Even though the members might be slightly different every time because not everyone can join, or you might have extra guests, it will actually look very similar to this community, to other kayaking club meetings across time. And this is what we build. We basically build a hierarchical clustering dendrogram where we are seeing that those communities are the same. So now we have a unit, we have a community that we can track across time and it's embedded in time, so we can say exactly when people have met in this community or how long have been participating in that.

And from there, what we are seeing is commonly there is a structure. There is a core and periphery structure. So think about community that has met multiple times, and let's think how much time people has spent in that community. What we are seeing is that there's a graded membership. So some people have been at every single meeting of this community and some people have only joined once or twice. Once we plot it this way we can look what is the largest gap that we are seeing in the participation profile. And if this gap is larger than what we would randomly expect from a null model where we just randomly create participation profiles, we are saying there is a core and periphery structure. And the core is exactly those generators of the community, people that always meet, that are almost always present at the meetings of the community, and then the periphery, people that would sometimes join, maybe would... maybe have joined once or twice but are not really the ones that are generating or driving this community.

00:05:53

So this way we are arriving to the understanding of the complex social system, not by disentangling this huge hairball but actually looking at the interactions themselves, and this is only possible if we have data that has resolution that is high enough.

HY: Think of an example of a social community where you are a core member and where you are only an occasional participant.

Thank you for your reflection.

AS: The topic of temporal networks is of course a vast topic and there are many, many, many papers published. I hope this gave you a little bit of insight into how we can actually think about temporal networks, especially in context of social interactions. And, of course, there is the written material and reading material for you if you want to dive deeper.