



Episode 119: Part of the pattern: How networks explain social behavior

Part of the pattern: How networks explain social behavior

VOICEOVER

Welcome to Up Close, the research, opinion and analysis podcast from the University of Melbourne, Australia.

SHANE HUNTINGTON

I'm Shane Huntington. Thanks for joining us. The way in which networks behave has great bearing on many aspects of our society. Whether we are talking about interactions between large banks, the users of illicit drugs or the distribution of pathogens, understanding the complex social networks involved provides a basis for finding new and effective ways of solving problems in areas as diverse as economics, biology, linguistics and defence science. To tell us more about this field we are joined by Professor Pip Pattison from the Department of Psychological Sciences at the University of Melbourne, Australia.

Welcome to Up Close, Pip.

PROFESSOR PIP PATTISON

Thank you Shane.

SHANE HUNTINGTON

Pip, first of all, can you define for us what we mean by a complex social network? I assume we're not always talking about humans here.

PROFESSOR PIP PATTISON

No, we're not. Indeed we're often talking about social entities and the connections between them. But the social entities can be as variable as individual people, social organisations, commercial organisations, states. They can also refer, perhaps less with a social meaning, to biological entities. So people analyse these days protein-protein interaction networks or gene regulatory networks. So the idea of social networks has certainly emerged to deal with relationships amongst social entities but the applications are taken beyond the boundaries of social phenomena these days.

SHANE HUNTINGTON

With regards to modelling these networks, what type of modelling are we actually talking about?

PROFESSOR PIP PATTISON

We're very interested in statistical modelling and we think that's important because modelling social networks is inevitably a rather uncertain business. Network links change on a regular basis and we need to understand the ways in which connections between entities vary as well as the way in which they stay the same. So a stochastic approach is very important.

We've developed a class of models which is very general in its conception and which allows us to explore in detail the kinds of regularities, or statistical regularities, in networks that we see and it allows us to identify in empirical cases which of a broader class of models is the most appropriate.

SHANE HUNTINGTON

When you talk about a stochastic model, can you give us a bit more on what you mean there and I guess what, mathematically, we're talking about in terms of these models, what they look like mathematically?

PROFESSOR PIP PATTISON

Okay, we often think of networks as what are called mathematical graphs. Graphs are really simple entities in a way. They're just collections of nodes - the nodes often representing social entities in our case - and the connections between them are really what make up the network. They're the sets of relationships which track, for example, the propensity for interactions between banks, as you mentioned before, or the connection of sexual contacts between individuals if we're thinking about the spread of sexually transmitted diseases.

From a statistical point of view, we need to think of some of these entities as variable or random quantities.

So the class of models we've been exploring tends to assume that we know who the nodes in the network are, but we're trying to model are the connections between those nodes. So we treat the links between nodes as random variables.

SHANE HUNTINGTON

When we're dealing with relationships, I can imagine there are both formal and informal versions of these relationships. Can you speak to that fact and how we distinguish between them?

PROFESSOR PIP PATTISON

The question of which relationships matter in any network is a really important substantive question and much of the social sciences literature has really spent some time trying to hone in on what the key relationships are. So, for example, in every organisation I think we know that the formal relationships matter: who reports to whom, who works with whom in particular work groups. But in addition to that we know that certain sorts of informal connections between people are often the things that really make the difference between an effective and an ineffective organisation.

So I think we've developed a lot of folklore about what those relationships are as a result of a wide number of empirical studies. For example, we know that collaboration matters in this informal sense but we also know that mentoring type relationships - relationships of seeking advice - matter a great deal. Sometimes socialising outside of work is also very important.

SHANE HUNTINGTON

Let's talk about a slightly different type of example: that of the spread of disease. How does the knowledge of social networks in this case work with regards to effectively dealing with the spread of disease and again can you speak to the formal and informal aspects of that?

PROFESSOR PIP PATTISON

Yes, well the spread of disease is a very important example of the application of networks. Much of the early modelling of the spread of disease made the assumption that connections between individuals that might lead to the transmission of disease occurred completely at random. What we've found, through empirical study, is that that assumption is very far from the truth. There's been a search going on over a number of years for identifying the characteristics of interactions that actually matter for the spread of disease. This has become particularly prominent where the relationships of interest are more intimate. So, for example, the spread of sexually transmitted diseases is a good example, and HIV where the spread is through blood borne infections that can come from such things as needle sharing or sexual contact or blood transfusions. Where there are those rarer forms of social relationship it's particularly important to understand how they're structured in order to understand the flow of the disease through a population. So what we've managed to do, I think, over the last few years - and I'm referring to the work of many people around the world - is to develop some understanding of how the relevant networks, whether they're needle sharing amongst at-risk populations or sexual contact in the population generally, we've got some understanding of the structure of those networks and we've been able, through careful empirical work, to start to build statistical models that allow us to then simulate the spread of disease through a population. And that capacity is very helpful because it not only allows us to understand and predict the course of a disease transmission in a population but it also allows us to explore potential intervention strategies to curb the flow of the disease.

SHANE HUNTINGTON

On the topic of quantitative modelling of social networks we're joined today on Up Close by Professor Pip Pattison. I'm Shane Huntington and we're coming to you from the University of Melbourne, Australia.

Pip, I assume the goal here in terms of the modelling is to understand all the ties in the network that we have well enough to be able to go back and use that to understand the system itself and then start making predictions. Is that the case?

PROFESSOR PIP PATTISON

That's exactly right. I mentioned before that we regard the connections in the network as random variables. What I didn't say then but which is very important to this enterprise is to recognise that those random variables are not independent of one another and can't be treated as such. So the key to developing predictive models is actually to develop models that explain the ensemble of ties that we see, all of the ties at once, so to speak. So we're modelling the whole of the system of network ties.

SHANE HUNTINGTON

When modelling these relationships in these very large systems I understand you need to consider all these individual components. Can you model those and get a good understanding of those on their own or must you look at the ensemble as a whole?

PROFESSOR PIP PATTISON

We found that we certainly can't look at individual ties but probably we don't need to be able to understand any one tie as a function of the entire ensemble. Truth lies somewhere in between those two extremes. The work we've managed to do is to identify how complicated that the dependence of one tie on those ties that are in the neighbourhood of the tie, so to speak, need to be in order to build plausible models that reproduce many of the features of the networks that we actually see in practice.

SHANE HUNTINGTON

Now in these networks you have a range of things going on. I can imagine some of these factors have a very large impact on how the system works and others, while relevant, have less impact. How do you go about determining that kind of relevance of the particular factors?

PROFESSOR PIP PATTISON

Through careful empirical work. In many ways there's an iterative approach to answering that question. We propose particular models which make an assumption about how to understand the impact of one part of the network on the other. We then test that empirically and determine whether or not that was an appropriate assumption to make. Of course if it was appropriate, we're happy. If it wasn't we then need to change the model that we propose and so through that iterative of process we come to understand just how important the impact of one part of the network is on other parts of the network.

SHANE HUNTINGTON

In terms of the complexity of these models and we talk about these factors, how many parameters are you talking about in terms of some of the ones we've spoken about? Are we talking about thousands of factors and parameters for each part of the network or is it just a few?

PROFESSOR PIP PATTISON

Surprisingly we can get by with a relatively small number of parameters. In many ways the fact that we can build models for complex networks with relatively small number of parameters gives us some encouragement that we really are on the right track here. In models of networks of friendships in large US high schools with the order of 1500 students, we can build good models with, again, less than 10 parameters. Of those parameters many of them refer to characteristics of the nodes and the impact of that on the tendencies for ties to occur. So for example, if we're modelling friendships amongst high school students we know that boys differentially have friendships with boys and girls with girls and so we need to take account of those node characteristics in modelling the network.

Of the network specific parameters in the model there are often relatively few and they refer to differential propensities for individuals to have various number of ties so that we can explain the distribution of the number of ties of any particular node in the network. Differential propensities for students to be involved in large clusters of nodes in the case of large friendship groups. In the case of sexually transmitted diseases

the overlap of two particular individuals with a sexual contact to have other sexual contact partners in common. There are some variations on that but those sorts of parameters do a particularly effective job in being able to model quite large networks.

SHANE HUNTINGTON

When we talk about the effectiveness of these models how do you go about testing that and how do you get to the point where you're satisfied with the efficacy of the model for that particular network you're working on?

PROFESSOR PIP PATTISON

That's a really good question. And again the answer is through an iterative process. One of the interesting features about these models is that we can simulate an entire distribution of graphs that are implied by a particular network. We can look at that distribution of graphs and see whether a graph that we've observed resembles that entire population of graphs or not. We can do that for features that are directly impacted by the model, for example the number of partners that nodes have, or the number of shared contacts. But we can also do it for features that are not included in the model directly, for example, how many steps it takes to get from one node to another in the network and the distribution of those path links in the network. We find that through consideration of the kinds of features of networks that people have identified as being important, for example, the degree to which there's subsets of nodes that are tightly connected, the degree to which we see cycles in networks, the degree to which we see people connecting different parts of clustered regions of the network, the degree in general to which one node is connected to another. Those sorts of characteristics are the things that people have identified as being important in the network. And we can use each of those characteristics to explore how well the model for the network actually captures the data that we have in front of us. It's unusual in that sense because of the rich structure of networks we can actually hone in much more readily on how well a particular model fits the data.

SHANE HUNTINGTON

Pip this modelling sounds in many regards very similar to the type of ensemble modelling we find in other fields such as thermal physics. It's very quantitative psychological work. Is this based very much on those areas where very large complex systems are being modelled in statistical ways?

PROFESSOR PIP PATTISON

That's a very astute observation, Shane Huntington. In fact there's a very deep connection between the kind of work that occurred in the 1930s in statistical physics and these kinds of models. We can trace a direct connection, interestingly, through work done by a now US based statistician Julian Besag, on the development of spatial models for the distribution of disease in plants through to the development of models for social networks. There's a common mathematical conceptualisation that underpins both the modelling of behaviour in gases, the distribution of disease in plants and the structure of social networks. It's very beautiful mathematical theory.

SHANE HUNTINGTON

This episode of Up Close coming to you from the University of Melbourne, Australia. I'm Shane Huntington and our guest today is Professor Pip Pattison and we're talking about quantitative modelling of social networks.

Pip, Melbourne, Australia as is the case with many large cities around the world has a particular problem with disease transmission such as HIV via needle exchange. Can you give us a bit of an idea of this needle exchange network in Melbourne and how you go about modelling that, what sort of work you do and what it tells us.

PROFESSOR PIP PATTISON

Sure. In relation to diseases like HIV and indeed Hepatitis C the relevant networks tend to be those of sharing needles or sexual contact or as I say, some other means of blood transmission. The work that we've done has actually focused initially on needle sharing and we've done this in conjunction with colleagues at the Burnet Institute and they've done some wonderful and careful empirical studies, actually sending workers out into the field, talking to individuals who use intravenous drugs and they've asked

those individuals to identify their needle sharing partners. They've then gone on to interview those needle sharing partners and asked them to tell them about their needle sharing partners and through this iterative process working through the network, getting a picture of particular fragments of this network which really exists over the whole population of intravenous drug users in Melbourne.

Of course it's completely impractical to talk to every single intravenous drug user in Melbourne and what we've been able to do by working with this group is to develop methods of using these sampled fragments of a network to build a picture for the entire network of intravenous drug users in Melbourne. We've done this quite recently and just managed to identify a model that we think does a really good job of reproducing those characteristics of the network that we can see through the data that's been collected and we're about to begin simulating the likely distribution of this network across the whole set of intravenous drug users in Melbourne.

SHANE HUNTINGTON

When you refer to a subset what sort of percentage of the entire needle sharing network do you think you've actually managed to get data on?

PROFESSOR PIP PATTISON

It's probably of the order of five percent, no more.

SHANE HUNTINGTON

From five percent you're able to essentially determine the characteristics of the entire network?

PROFESSOR PIP PATTISON

We can build a model for the entire network. Of course how well that captures the entire network is an empirical question. But we've done a lot of simulation studies to show that although we lose some precision in our capacity to model the network by using only a fragment, on average we get it right. So if we take sufficient number of these samples we would expect, on average, to be able to build a pretty accurate picture of the network.

SHANE HUNTINGTON

In the case of the data we're talking about does this involve just the movement of individuals and who they are or do you also take into account genetic aspects of the virus itself, its transmission and those more biological factors?

PROFESSOR PIP PATTISON

That's a great question. For us this is very early days but one of the things I should have mentioned that our team of field workers do is take blood samples from the individuals with whom they speak. So we do have a quite detailed picture of the makeup of their particular infection: in this case their hepatitis C infections. So slowly we're starting to build an understanding of how the social interactions as measured by sharing needles has an impact on the spread of particular types of hepatitis C virus in the network. All I can say at this stage is there's a very kind of close connection between the distribution of the viral types in the population and the social connections that we've been able to measure.

SHANE HUNTINGTON

What is the ultimate goal of being able to model a network like the needle sharing network in Melbourne? What will be the outcome for that community and the people that support them?

PROFESSOR PIP PATTISON

The ultimate goal is really one of being able to determine how best to use knowledge of the structure of the network to effectively intervene in the transmission of diseases. Depending on the structure of the network different kinds of strategies for intervention might be important. Whether they're educational strategies or whether they're treatment strategies. It could be that there are certain sorts of characteristics of individuals who make them particular targets for treatment. And in treating that one individual you might be doing more to protect the health of the entire population than if you pick another individual.

So they're the kinds of strategies that will be important and that we can explore in a modelling way before we start to tamper with actual empirical studies which we would then do to verify that those are indeed good strategies.

SHANE HUNTINGTON

Speaking about the details of the model just for a moment I understand you use a technique called adaptive sampling to great effect in these models. Can you describe what that is and how it helps the models and their accuracy?

PROFESSOR PIP PATTISON

If we think about modelling the network of needle sharing amongst intravenous drug users a typical strategy from a sampling point of view is to take a random sample of individuals. If we're interested in network features we might then explore the degree to which those sampled individuals share network ties or needle sharing ties. Now if we only have five percent of the population and then we look to see whether there are ties between every pair of them we might actually come to the conclusion that the ties are very sparse, there's not much needle sharing going on. That's because we've looked at people in very different parts of the network.

Adaptive sampling takes a different approach. Rather than taking a random sample of individuals and simply looking to see whether those individuals are connected, we actually take a random sample of individuals and explore the network neighbourhood of each of those individuals. We find out who their partners are and who the partners of their partners are. So in that way we identify particular characteristics of the local network that we think might be important. It's adaptive in the sense that who we sample next depends on who we just sampled. So in that way we find information that's very relevant to the network. But of course we need to take the fact that we've traced particular paths in the network to identify those new sampled members and so we need an appropriate statistical mechanism for handling the fact that we've, in a way, biased our observations so that we see more ties.

SHANE HUNTINGTON

Pip just finally: this modelling is intensive, it takes on a lot of different factors, there are a lot of parameters involved, what sort of processing power is required? Is this something that's particularly challenging for this field of expertise?

PROFESSOR PIP PATTISON

That's a very good question and it is indeed very challenging because networks grow as the square of the number of nodes that you have. So the capacity to model networks every time we sample a graph from a distribution we need simulation techniques to actually do all of the calculations for the statistical modelling because there are no closed form or analytic solutions for anything that we do. So a huge amount of computation is involved in any case but as networks grow large that requirement grows exponentially so we do need very significant computing power.

SHANE HUNTINGTON

Professor Pip Pattison from the Department of Psychological Sciences at the University of Melbourne thank you for being our guest on Up Close today and giving us such a great understanding of modelling these complex social networks.

PROFESSOR PIP PATTISON

Thank you Shane.

SHANE HUNTINGTON

Relevant links, a full transcript and more info on this episode can be found at our website at upclose.unimelb.edu.au. Up Close is brought to you by marketing and communications of the University of Melbourne, Australia. This episode was recorded on 19 November 2010. Our producers for this episode were Kelvin Param and Eric van Bommel. Audio engineering by Gavin Nebauer. Background research by Christine Bailey. Up Close is created by Eric van Bommel and Kelvin Param. I'm Shane Huntington. Until next time, goodbye.

VOICEOVER

You've been listening to Up Close. For more information visit upclose.unimelb.edu.au. Copyright 2010. The University of Melbourne.

