AWARI
Vol. 4, 2023. 1-9, DOI: 10.47909/awari.625

Pro-Metrics

OPINION ARTICLE

Misunderstandings (and some criticisms) in social network analysis

Steve Borgatti

Department of Management, Gatton College of Business and Economics, University of Kentucky, USA.
Email: sborgatti@uky.edu

Transcription by: Francisca Ortiz Ruiz, Millennium Institute for Care Research (MICARE), Santiago, Chile.

Presentation moderator: Alejandro Espinosa-Rada, Social Network Lab, ETH Zürich, Switzerland.

ABSTRACT

The Chilean Society for Social Network Science (ChiSocNet) has played a pivotal role in the organization of the social networking community in South America, Ibero-America, and beyond in recent years. It recently spearheaded the inaugural Chilean Social Networking Conference (ChiSocNet), held in Santiago, Chile, from January 4-7, 2023. This conference brought together researchers and practitioners from diverse backgrounds, united by a shared interest in the social networking perspective. The program consisted of 13 sessions on nine topics, with more than 80 presentations. In addition, 180 individuals registered to attend the conference, including speakers and the general public. Furthermore, six free workshops were held, with registrations ranging from 40 to 100 people, depending on the topic. In this context, the international keynote address was delivered by Steve Borgatti, Professor and Chair of the Department of Management in the College of Business at the University of Kentucky. A transcript of this presentation is provided in this article.

Keywords: social network analysis; network science; network theory; challenges; Chile.

INTRODUCTION

The Chilean Society for Social Network Science (ChiSocNet) has helped organize the social networking community in South America, Ibero-America, and abroad in recent years. It recently organized the 1st Chilean Social Networking Conference (ChiSocNet) held in Santiago, Chile, January 4-7, 2023. This conference brought together researchers and practitioners interested in the social networking perspective. The program consisted of 13 sessions on nine topics, with more than 80 presentations. In addition, 180 people registered to attend the conference (between speakers and the general public). In addition, six free workshops were held, with registrations ranging from 40 to 100 people, depending on the topic. In this framework, Steve Borgatti, professor and chair at the University of Kentucky in the Department of Management in the Gatton College of Business and Economics gave the International Keynote talk. This article provides the revised transcript of this presentation, which can be found in video format on ChiSocNet's YouTube channel.

AUTHOR'S PRESENTATION

Alejandro Espinosa-Rada: I think it is difficult to introduce Steve since he is a very well-known

How to cite: Borgatti, S. (2023). Misunderstandings (and some criticisms) in social network analysis. *AWARI; 4,* 1-9. DOI: 10.47909/awari.625.

Received: 06-11-2023 / **Revised:** 12-10-2023 / **Accepted:** 15-11-2023 / **Published:** 21-11-2023
Editor: Francisca Ortiz

AWARI
Vol. 4, 1-9, DOI: 10.47909/awari.625

person, but I am going to say some formal things, and maybe later, I will add a couple of more informal things. In terms of the formal, Steve Borgatti is a professor and chair holder at the University of Kentucky, Gatton's Department of Management in the College of Business and Economics. His research focuses on social networks and at the University of Kentucky, well, he houses the LINKS Center research institute, which focuses on the study of social networks around the relational perspective. He is currently an associate editor at Computational and Mathematical Organizational Theory and the Journal of Supply Chain Management and serves on the board of directors of Field Methods. He is the former editor-in-chief of Connections, a senior editor of Organization Science, and has served as the editor of several journals. These include Administration Science Quarterly, Journal of Management, and Sociological Methodology, among several others. He is the author of several well-known programs in the networking world such as UCINET, which is a software package for social networks; Anthropac, which is for analyzing domain and cultural data; and E-NET, which seeks to analyze network data collected through personal network designs and research. He has served two terms as president of INSNA, the professional association for social network researchers, and founded the SOCNET listserv. Co-founding editor of Field Methods (when it was CAM - Cultural Anthropology Methods Bulletin). Directed the NSF Summer Institute for Research Methods in Anthropology for three years. That's the formal part, and the informal part, I think, if you look up and analyze networks, in the history of networks, I think Steve has been a super important person who has worked with Freeman, with Martin Everett, with many others, and I think it's hard to calculate the relevance he had, and still has, for many generations who got into the encyclopedia especially in the nineties in the two thousand even up to today, so much so that they recently adapted them for R as well. And if you are still nostalgic about wanting to use the things that were used in college but still want to use R, is already solved the problem. Well, I don't want to preamble anymore because I really think we all want to hear from you, Steve, so thank you very much from the organization side for being with us today.

KEYNOTE

Steve Borgatti: Thanks Alejandro, thanks to the organizers, and to the participants. My topic today is misunderstandings and some criticisms of social network analysis. In my opinion, there are three types of social network studies. Prototypical types I have named "links," "achievement," and "style." The "links" are studies that want to explain the presence and absence of a link between each dyad in the network, and of course, there are other network analyses as well, but it is a big idea and the mechanisms (of preference and opportunity) that are key variables. The second type, I have named it "achievement". These studies try to explain achievement, performance, performance, and value creation, and they do it based on a mechanism that interests us, which is the mechanism of social capital; it is the idea that, in some way, the links help to be successful. The third one, "style," these are the studies that try to explain why it is that each person has his style, his way of speaking, his way of eating, his aptitudes, his ideas, and why it is that certain pairs are similar and others are different. Here, the preferred mechanism of us networkers is diffusion, contagions, and interpersonal influence. I am going to use these three as a framework to organize my presentation; although I don't think we are going to get to number three, I am sure we will be able to talk about one and two.

These three types of studies are not all which we could place in a more complete and more complicated framework. For example, on the left, we have different levels of analysis that distinguish between studies that are interested in the antecedents of network variables and those that are interested in the consequences of the fact that we are all entangled in a network. The types of networks are influenced by the theoretical framework, which means different levels of complexity.

SOCIAL CAPITAL

Let's start with the concept of social capital. I think there are three sets of ideas in social capital.

1. First, the idea of social resources is associated with Nan Lin. His proposal is that it is useful to have friends and contacts with various

resources they can lend you or use to your advantage. They can bring certain ideas that can help you. So it is hypothesized that if you have a large and diverse network, there will be someone in that network who can help you when you need it. Thus, in this type of analysis we are mainly counting things, such as, for example, the number of women you know, the number of bankers, the number of people in positions of high power or economic resources.

2. The second group of ideas is that of structural holes, which I associate with Ronald Burt. It is almost the same idea as the previous one. The idea is that it is useful to have a large personal network because, in that network, you will see someone who has what you need at a given moment. For example, someone who has the money to invest in your project and knows about computers and can help you at that moment. It is also useful if all the contacts are not connected to each other. For example, we are interested in having access to different points of view, so if our contacts are all connected, they influence each other and reach a consensus, and with that, we lose access to new information without novelty or innovation. These gaps in access to new information are what we call structural holes in the literature. A small, high-density, small personal network like the one I am pointing out is not as useful as a larger network with fewer links. So much of the discussion around this topic focuses on poverty in very closed and dense networks. Thus highlighting the relevance of more open networks with structural holes.
3. The third group of ideas within the notion of social capital is the notion of centrality. This idea can be associated with Freeman, although many authors are involved in developing this literature. This literature argues that not only the direct links that matter (as we saw in the issue of structural holes) but that the node's position in the network is crucial. There is value in the position in the network, as it is related to proximity to others, or that one has access to certain information before others, or control over what circulates in the network, and so on.

Considering these three ideas that circulate social capital, I raise a criticism. Today, most studies of social capital, at least in my area of work, which is business, define this notion only from the perspective of Burt's structural holes. This is a testament to the great impact of the author and his theory, but perhaps it has also reduced diversity in the field, which is an unintended consequence. His ideas are great, but they have been so good that they have destroyed any competition, in my view.

CENTRALITY MEASURES

It's a shame because I like all the other centrality ideas and measures. These centrality measures have been a big part of my work and are very interesting. For example, we have the *betweenness* measure, which measures the nodes located on the network's best paths and therefore has the best control over the flow of information. The *closeness* measure measures the closeness to all the other nodes in the network, thus seeing the reception of everything that flows in it. To illustrate, these are the people who receive a piece of information before others and can therefore change it a little by demonstrating their control. The *eigenvector* is the measure of popularity in the sense that it does not measure the number of friends per se but the number of friends they have in turn. It is similar to degree centrality but in an improved form. *Fragmentation* measures how many pairs of nodes are not reached by any path if we remove a node, functioning as a determinant of global cohesion. It is part of the family of induced centrality measures, where you measure a specific property of an entire network, remove the node, and measure it again, and the difference is induced centrality. Centrality measures are adaptable, as they can be created according to what you want to measure in the contexts in which you work. This makes them very interesting and beautiful measures, and they are not being used as much as they could be. Maybe in other disciplines, such as political science, but not in business management, which is more my area.

STRUCTURAL HOLES

Now that we accept the popularity of structural holes, let's review them in more depth. There are two measures of structural holes:

AWARI
Vol. 4, 1-9, DOI: 10.47909/awari.625

effective size and constraint. Both measure a combination of degree and density, the lack of density within Ego's personal network. For example, we have a personal network down here; node "G" is Ego (see Figure 1).

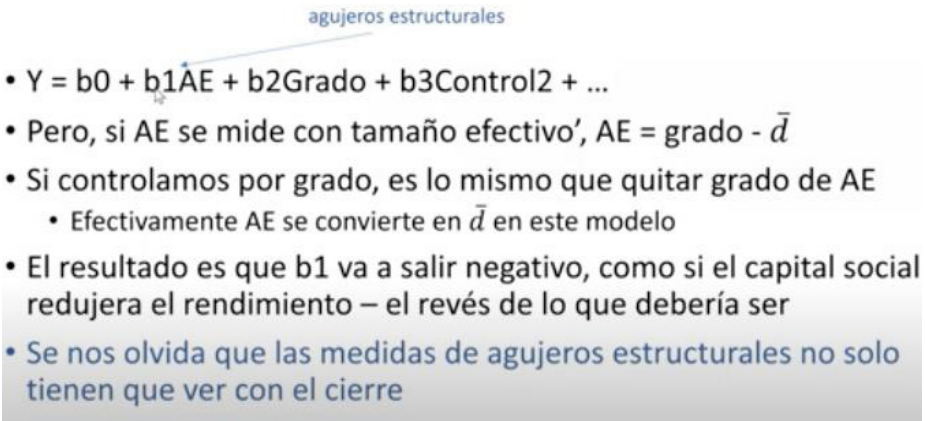

Figure 1. Examples of "effective size" and "*constraint*".
Source: Slide extracted from the presentation in question (in Spanish).

Here are the alters to which G is connected. The effective size measure is precisely *n-đ*. *n* is the number of alters; *đ* is the average degree in the personal network, so, for example, "E" is connected to one alter, not including the link to Ego. "A" is connected to three alters, and the average of all these degrees is 1.33. To calculate effective size is 6 minus 1.33, which gives us 4.67. This is a very logical measure. If Ego's friends, the alters, had no links between themselves, they would all have zero degrees, the average would be zero, and the measure would be six minus zero. In that case, the effective size would be equal to the size, but because some are connected to others, they are redundant and the measure takes away points for that. It is logical.

Now we run into a misunderstanding: what are we going to do with this measure? We're going to put it into a regression variable. The dependent variable is going to be something like performance, as always in Burt's case, and the independent variables are going to be various controls, and our variable of interest is the measure of structural holes. If we do the analysis we expect that the coefficient b1 is going to come out positive and significant. What happens is that we inevitably have to control for degrees in the regression. We do that because the degree is a very important variable affecting everything, so we have to put the degree in the regression. But if we do that, what happens? We have structural holes (see Figure 2), which we will measure with the effective size and added grade and the other controls.

But we already know that the effective size, the formula is grade minus "d" slash, so the grade is already twice. We are effectively removing the grade. Putting grade here as a control will remove the grade from the AE measurement and leave only what is effectively density. Density divided by "n" minus one is the closure. So, if we do the analysis, what will happen is that the coefficient b1, now that this variable is density, will be negative as if the capital stock were related to yield in a negative sense. If we interpret this, we would be saying that the higher the capital stock, the lower the yield. This is a mistake, a misunderstanding.

The same happens with *constraint*. The first thing is that this is an inverse measure of structural holes, i.e., the larger value corresponds to fewer structural holes, and of course, in regression, we have to expect a negative signal instead of a positive one, and this causes problems that we forget. It is believed that the largest maximum value of constraint is one, but it is not, since the largest value of constraint is 1.125; plus, these are binary and undirected links, so the minimum value is 0.5 and not zero. There are few who know these facts, and if one has a fairly large network that influences the result of the *constraint* measurement, generating another misunderstanding.

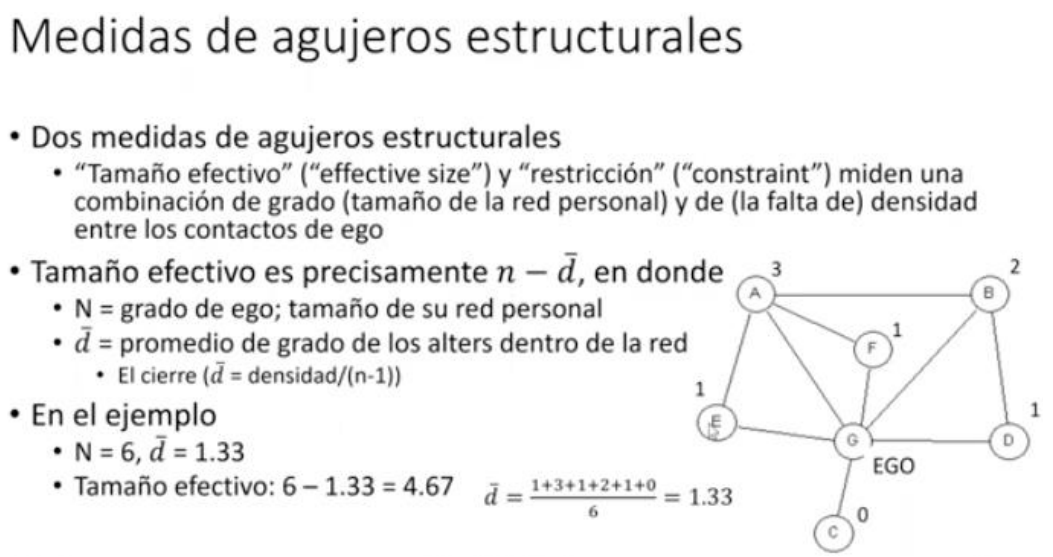

Figure 2. Misunderstanding of structural holes.
Source: Slide extracted from the presentation in question (in Spanish).

In general, it is assumed that the maximum value of *constraint* occurs when you have a personal network where there is perfect density, i.e., where all alters connect to all other alters. But in reality *constraint* has its maximum value in a personal network where there are fewer connections between all the alters (as in the network on the right of Figure 3).

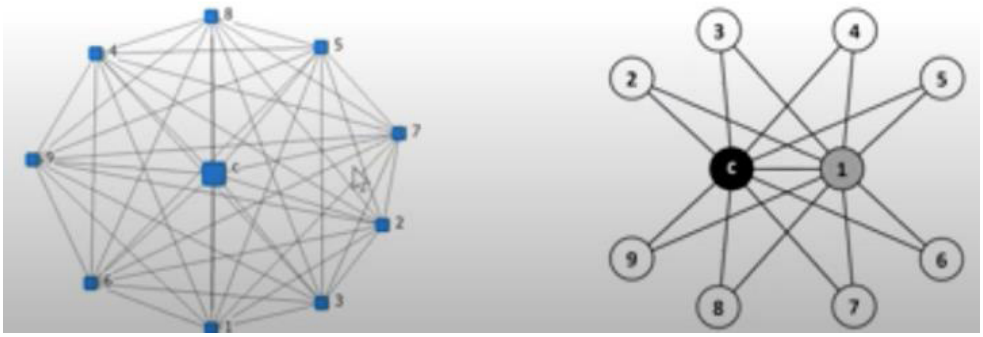

Figure 3. Example of networks with maximum constraint value.
Source: Slide extracted from the presentation in question.

This is strange. Because looking at the network on the right, it is obvious that if C is negotiating between different alters, it is intermediating between them. It has possibilities since two is not connected to four and can be an intermediary for those two; two is also not connected to three and five. There are chances to be an intermediary, but according to the measure, there are fewer chances than the network on the left. In this case, the misunderstanding is that this is unknown, but the cause is a problem with the measurement, which does not measure exactly what we wanted to measure.

LINKS IN THE NETWORKS

This section refers to research that studies how links are formed, who connects to whom and why, how centrality develops, measures at the node level, why certain people have more centrality than others, and what determines the shape of the network.

Mathematically, I believe in networks. I know that it is a mathematical object and it is very useful, but sociologically, I don't believe in the network as if it were something tangible. Because these concepts of social relationship,

AWARI
Vol. 4, 1-9, DOI: 10.47909/awari.625

a link, a friendship, are real, but within the sociological world. The network is often confused with the group, but they are two separate concepts. For me, the network is a way of describing the structure of a group. So, the group is the thing, and the network is something that is used to describe the structure of that group. I mention the group as an example, but you could apply the same thing to other sets. So the network is a pattern of links and nothing more. My thesis is that if one does not understand this, one may fall into certain misunderstandings.

The first example is the problem of limits. It is a problem that, in some cases, is quite big and causes a lot of anguish; every novice has this problem; he wants to know how I determine the limit of my network. It is a difficult problem, partly because networks do not really have limits. Obviously, the limit of the friendship network is the whole world or the human species, or maybe the human species and dogs and horses or maybe other animals, or cats, well maybe not cats (laughs). But groups do have boundaries, while networks don't: Groups are hard to define. They are always ambiguous and contested; some members are joining or leaving, and others want to be members, to mention a few examples. The limit exists in a certain sense, and the sense is that it depends on the very concept of the group. Every group has its members and partners, and some are not members, making it difficult to distinguish them. Those members have certain privileges and the identity of being a group member, but the network does not have that.

If we study networks, we can study a set of nodes that is defined by a natural group. In that case, we are fine, but the limit is not the limit of the network but rather the limit of the group we are studying. If we are studying a natural group, we have to pay a lot of attention to choosing the limit correctly. So, I wanted to say that it is possible to study networks that are not associated with a sociological group and are more artificial. For example, if I am interested, I can get the idea that maybe the pattern of links between the participants of a conference affects the quality of the conference or of the presentations within the conference. I could do a questionnaire and collect data on who is friends with whom at the conference, and of course, I understand that the boundary of that conference is very artificial, I understand that friendships do not stop in that order, but the truth is that there are more friendships outside the conference than inside. If I apply the same survey to many different conferences, then it's a more legitimate inquiry; because I might be interested in whether the pattern of friendships within this set somehow affects the quality of the conference performance, and you could analyze that, even if the boundaries are quite artificial. So you have to consider that it is important to distinguish between groups and networks, also considering the mathematical aspects, which are artificial but still useful.

Another point closely related to the boundary issue is the misunderstanding about which question I should incorporate into my questionnaire. There are several options: who are your friends? who do you look to for advice? who do you have the same hair color with? Every question one asks, every relational question one asks, creates a network at that moment. The researcher creates the network by asking the question. If I ask those who have the same hair color, we can build that network, but that network is probably not useful or helpful in predicting anything. The network is anything that one wants to understand or study that is relational. The point is always to consider what relationship we are interested in and which relationships have a connection to some dependent variable of interest.

NETWORK MODELING WITH SIENA

Let's look at a couple of more complicated and technical issues. We know that in the world and in the study of networks, there are models of network evolution, such as the SIENA models, but I would like to point out that these models do not focus on the development of a group because the evolution of a network is different from the development of a group. SIENA is not useful for studying how a group evolves from three people that grows to thirty, grows to three hundred, and becomes a political movement. This is a sociological phenomenon, and to study it, we have to do it from a sociological approach, which requires other models and theories.

What SIENA does is simpler. Siena models change between the existence (or not) in the dyad links. It is the structure that is being modeled, not the growth of the group. In a way, in SIENA, networks always exist because they are artificial; from a mathematical approach, they exist in the mind of the researcher. The network always exists. All the nodes exist at all times, although they cannot make certain links at certain times, so the only thing that changes is the pattern of relationships.

I hope I am already convincing you that the network is not true. According to how one sees everything, in a way, the whole world is mentally created. There is even a difference between the idea of a network and the idea of the group. The group and the link are constructed mentally and from a sociological reality, but I think that does not happen with the network.

NETWORK MODELING WITH ERGMs

Exponential Random Graph Models could be MEGA (laughs), but no, we have to call them ERGM. In these models, the network structure is modeled as a function of certain tendencies of the network to form certain microstructures. These microstructures cannot be seen in the observed network because they are hidden by other tendencies. There are several tendencies simultaneously; for example, if one is using a very open network and the other is very closed, the result will be a mixture of both. It is the job of the algorithm to discover these effects in what we can observe. In these models, it is imagined that the observed network was created by a process of evolution, and at each point in the process, there were opportunities to form a link or not to form links due to a function. The objective function, which we do not know, favors the creation of certain links, and here enters the different effects to form certain microstructures. For example, to create links that would create more closed triangles (transitivity), or more links for nodes that already have many links (this is the measurement of preferential attachment by degree), or to have more homophily (tendency to affiliate with people similar to oneself in terms of origins, language, personal style, activities, aptitudes, interests, demographic variables, etc.). Reciprocity and all the other effects that have been defined in programs that calculate ERGM. Using this kind of model is a good idea; it makes sense. We see observed data, we imagine that thing that we have seen that pattern of ties had a history, it evolved and we have to infer what were the forces that created it.

Now for the misunderstandings. These models define trends in a certain way and parameterize them, so there is a parameter for each trend. Sometimes, these trends are called social forces or mechanisms, and they are also characterized as endogenous, self-organizing, and emergent. This is where I have problems with the language used.

Starting with social mechanisms, I believe that what we see in ERGM models, for example, transitivity, is an effect. That is, it is a result of a social mechanism. Transitivity is not a force that changes networks, but it is a result of social forces. Of course, if there were a one-to-one relationship between a social mechanism and a parameter of the model, then nothing would happen, it is just a convenient way of speaking, but this is not the case because each of the effects we see in ERGM has several social mechanisms that can create it. For example, the transitivity effect has many potential mechanisms, such as the idea of a psychological mechanism, where we have Heider's cognitive equilibrium theory. The idea is that I want to align myself with my friends. If I like Michael and Michael likes John, then I would like to like John too because that way, I avoid cognitive dissonance. This psychological mental process results in a tendency to have closed triangles in the network, but that is a result of something that is in the minds of individuals. Another mechanism is opportunity. I become friends with John, and he probably has other friends, that he is going to introduce me to, and when he does that, the triangles close. Again, we have a tendency toward transitivity, and these are not the only mechanisms that can result in transitivity; there are others, including hemophilia.

Thus, these trends are often described as endogenous effects for purely statistical reasons. Any effect that can be calculated with data from the network itself, for example,

AWARI
Vol. 4, 1-9, DOI: 10.47909/awari.625

reciprocity, is called endogenous. If we are studying a network without any other data, we can calculate transitivity and preferential *attachment*. So, all those effects are, in a statistical sense, endogenous. But what happens is that we confuse social mechanisms with these tendencies or effects found in the model, and, therefore, we now believe that the social mechanisms are endogenous and inherent in the network because it is a network. That is as if these social mechanisms were an integral property of the networks. This is a mistake because it is not the network (the network does not even exist – remember?); the nodes act. It is the nodes who have agency, and who resent cognitive dissonance. It is the nodes who want to respect cultural norms, and who want to reciprocate interactions. When we look at effects as if they were endogenous to networks, we are making a small mistake (although it is understandable why the word 'endogenous' is used initially - it originates from a statistical term).

EFFECTS AS "SELF-ORGANIZED" AND "EMERGENT"

Other words that are often used in this context, and with which I disagree, are the notions of self-organized and emergent. Social mechanisms, i.e., effects (e.g., transitivity, reciprocity, homophily), are said to be self-organized and emergent; almost as if they are uncaused and magically appear. Of course, again, I understand why the words emerged. In all these models, the idea is that we have something that can be described as a macro phenomenon, which is the pattern of links in a network, and we are going to explain that macro phenomenon with some small rules or microstructures or micro forces, but none of those micro forces want to create the pattern that we see in the data. No actor has the notion that you have to reciprocate when someone gives you something or that you want your friend's friend to like you; you want to create a network with certain given statistical properties, such as a network with a core and a periphery. No actor is trying to do that, and no social mechanism wants to do that because the pattern we see results from many different forces acting simultaneously.

So I understand the idea that the pattern is self-organizing - there is no one who wanted to create that pattern, but it emerges from the combination of other things. Almost unexpected, almost unpredictable. But, what happens is, for some reason, and I don't think it's the statisticians who make this mistake if not the users. It is believed that it is the trends, like transitivity, that are self-organizing, that are the social mechanisms that emerge in a self-organizing chaotic way. This is a mistake because, of course, all these things are based on psychology, culture, and institutions, and I think what is happening is that we are forgetting that we are studying people, well in my case, in organizational studies, but the nodes always have certain abilities and certain limits that determine what is going to happen in a network. It is not just because it is a network that causes the structures that we see.

CONCLUSION

Obviously, the networking field is in crisis. I think you can hear it in my tone during the presentation, where I say that mistakes are being made, and I don't like all that, but the truth is that it's not all that way. I have had a career of more than thirty years, where I have done only one thing: study social networks. At the same time, I have said in this presentation that I don't believe in social networks. So obviously, there is a crisis, but it's not in the field; it's in me (laughs). I have tried to draw attention to some small misunderstandings, some ontological complaints about networks, because I think they are not real; about the effects of stochastic models, that I don't see as social mechanisms, which would be a mistake. The theme that ties my complaints together is that we forget a little about who we are. We are social scientists, and I see in the world now that network *science* is emerging, as well as big data and all that. I think they don't have a strong interest in people, groups, organizations, and social mechanisms. I also know that they very easily confuse their models with reality. They confuse the map with the territory. And I do not want us to make the same mistake. Of course these are opinions that I raise here, but I think that a Keynote is appropriate to raise these opinions. Thank you very much.

Acknowledgments

We gratefully acknowledge the support of the Chilean Society for Social Network Analysis (ChiSocNet) for this presentation and article. Francisca Ortiz Ruiz received support from ANID Millennium Science Initiative Program (ICS2019_024) and ANID FONDECYT REGULAR Nº1230437. In addition, thanks are due to Felipe García for his contribution as assistant in writing this paper.

Conflict of interest

The author(s) declare(s) that there is no conflict of interest.

Data Consent Statement

No data was generated during the development of this study. The full presentation can be found at the link below on YouTube:https://youtu.be/IZN8I6a27mc?si=ZG8DtZ9v2tSe0OZZ. ●