**Interview with Matthew Sergant, RAND Research Methods Center on Social Network analysis, July 15, 2025**

**Tell me about the methods center?**

Work in methods center. Still trying to find the focus and greater methods. There is a very diverse set of people that use the method across RAND and people are not always in conversation with each other, calling it different things and not always in conversation with each other. So goal is to create a community of folks to cross pollenate.

A lot of the applications of network analysis are embedded in research not involved in network analysis. So part of the challenge is to identify the people that should be brought into the community. They are putting together, … trying to figure out how to become involved in network analysis. Putting together the training tools and lectures so people interested have a place to go. Other methods centers, we do relatively fewer talks as people are interested in showcasing methods or get a bigger audiences so we do talks jointly advertised with NSRD. And lastly we try to find ways to get groups of people in the same sort of methods, so networks in health care or other orgs so rather than talking at a general method.

The goal we are missing organizationally; met with a client and they brought in McCrystyal Group and a former general said they were brought on to do a network study with an HSOAC client and they said McCrystal is the only organization to do this but their work was undergraduate level. But they were able to do, to study an organization, one way to do it and to describe organization is to map the internal network. You can ask everyone questions, who do you go to for advice, who do you work with on a daily basis? Who you report to or who you go to for advice? What happens is you map the organization and show it naturally segments into different groups and you can see where the fault lines are. Like you can see Boston office is not integrated with the rest of RAND. Or see that someone is not the CEO but everyone goes to this group director for advice and consultation.

So fairly standard, every org has a certain amount of dysfunction and everyone likes to see what their org looks like. There is a standard way to get the data; send an email with a battery of questions and you select coworkers you work with, go to for advice and put it in a matrix. From a RAND process you might do that for a project but the expertise and all the code is not easily produced for the next project and RAND does not have an ability to depersonalize that expertise. But something you can do it with AI relatively easily bc so much of it is populating emails, sending emails, taking the data, ingest it and run it in the analysis.

I’m speaking from an organizational standpoint, an autonomous group in health care that does surveys, ask what their support group looks like so similar data gathering and processes. Within health, couple researchers such as, David Kennedy and colleague, they coauthor a lot and do similar data gathering and processing. If you talk to him he can give similar examples of processes to do that at scale.

**What other ways do researchers use social network analysis?**

1) First, Off the cuff, the big initial branch is option 1, which I just described where gathering information about network and asking about relationship between people and projects.

2) Second, The preponderance of what RAND does is going out and try to find data that already exists in the world that speaks to your problem. Taking the data or the processs system and converting into a structured network. So example is 5 or more years ago Jon Fujimora took process of the air ops center, group in the Air Force that takes all the inputs and turns cranks and gives planes air tasking orders.

This set of military doctrine and procedure and he coded it all by hand into network data. A second tranche would be taking data and coding it in a systematic way to produce structured relational data for network. That is what Jon did.

3) A third would be taking written data and transforming it, data that has already been gathered for some other purpose and recasting it into data you need. Don Snider, Systems engineering where they had … for next gen nuclear arsenal… all of these steps, they already had the data but… Publication? I can check.

4) Fourth, a data fairy has what you need that some else has collected.

**Is there a consistent procedure for analyzing data?**

Some visualization that is done in Geffe, a lot of the heavy duty analysis, 70 percent in R and 30 percent in python. And for Don Sniders work there is mathematical explorations. But the same set of software programs that people are using in r or python.

**What is the long poll in the tent?**

Depends on the topic. Ultimately once you commit to doing network analysis the biggest issue is collecting and coding the data. Having a researcher frame a problem in network terms… usually when you have this complex process, at an early stage of thinking how do we work this through a network lens? You can often think of network stuff and not get data to do it.

The second method above (see Jon Fumimora’s example above) is semi automatable. The challenge is telling AI what the signal looks like and what it should look for and how to code it. That is a step that emerging AI and LLM is making possible and doable. This is the new thing that AI allows.

What jon did, the same level of interaction, you dip into areas with some depth. He used a 3-30 document that described the procedure for the air operations center. The first step is figuring out who all the actors are, who are the ppl, what data are they passing and he did it just entirely by looking at what was written in doctrine. The commander produces this, this group reads it, they go the ISR team and with that they did surveys to add more data, what is the classification, how do you access this? He did the first step from text documents which presumably you could automate. That is a powerful addition, figuring out how to leverage that and letting people know that is an..

The first method described above is automatable because it involves sending out emails and getting responses in structured format.

**Can you use a a general LLM or could you train a specific model?**

You will get lots of different datasets that are roughly similar, the language is in a formulaic structure of command relationships will make them standard but I don’t know whether you can use off the shelf model or whether there is a subset that you’d want to train your own model to do that specifically. Need to get into it to see how accurate the models were in the coding.

**How often are network analysis studies done at RAND?**

Short answer is I don’t know. T here is probably a hand full in health like David kennedy cranking things out. There are five or so projects a year, 5-10 using it in a major way where the major analysis is network analysis. And a lot more where network analysis used in more casual way, 30-40 projects?

**How large is community at RAND interested in social network analysis?**

About 100.

The question is with additional tools if you lowered the barrier whether you could move that (more casual use of social network analysis into a more focused approach). could take a large number of projects using network analysis in casual way and convert them into projects where it is the central part of the analysis. Usually, the limitation is getting good data.

The balancing act is figuring out what you want to show to the sponsor. Only sometimes is there an appetite for seeing hard core analysis moving beyond the pictures and few statistical measures and most don’t want to engage in deep academic discussions of what is going on in the network. So matching availability of data and what you can show your sponsor.

Is there value in using AI to manipulate the data in R and Python?

The issue is there is such variability in the sorts of questions you are asking that it is not immediately obvious to me what steps you’d automate. Getting things set up in a network. The parts of the process that are repeatable. Coud do so if a standard set of R libraries, where first cut in graphical representation in publication quality, that is snippets of code you can use. But people do so many different things that once they get to R it is hard to think of what parts you’d want to automate… so a lot of variability in the types of analyses. [Consider the comparison of the ISIS Twitter networks study and Fujimora’s study, they used very different social network analyses.] it is not clear what would be shared.

**Other ideas for AI?**

Internally it would be useful to be able to identify which RAND products are using network analysis and who is doing it. If you are a researcher, or lead a methods center, to look across RAND products to try to identify how the method is being used, who is doing it. Say we get a contract to do supply chain analysis, how has RAND employed network analysis to study supply chains for the military. Reviewing past studies that used this method would be useful .

**Other ideas?**

Nothing that springs to mind.

**Are there risks with adopting AI?**

Casual sense, if using ai to code things, ensuring the data is robust and valid and things are not getting miscoded and caste. For HSPC, a little hard to ensure you are following all the promises you made in your HSPC document by using AI to code data. De-identifying data and… harder to do if AI might be able to go through and unravel the signals and identities. I don’t se that as a risk of using ai but of ai being out there. I don’t see any other substantial risks.

**What about adoption at RAND?**

I think we are making responsible use of the tools and there are a lot of things that are about to make more resources available. It is a question, RAND chat is useful when I have government data that I want to analyze but for general questions and lit reviews for non rand things, the array of for-profit providers, open ai or anthropic get better results. So figuring out what the use cases for using rand products. So what can I use for anthropic and what can I use for randchat. Proprietary data is best for rand chat but either I am underutilizing randchat bc I don’t understand its value enough or there is a difference … how do you use one vs. the other.

Things everyone struggles with, citation management but everyone agrees ai is not there yet. I outsource… things like that I can probably do a better job of leveraging ai for. Could do similar things with template management but I don’t.