Using Social Network Analysis to Evaluate the Development of Professional Connectivity

Michael FitzGerald, The Rucks Group, LLC, mfitzgerald@therucksgroup.com
Emma Leeburg, EvaluATE, emma.leeburg@wmich.edu
Lana Rucks, The Rucks Group, LLC, lrucks@therucksgroup.com

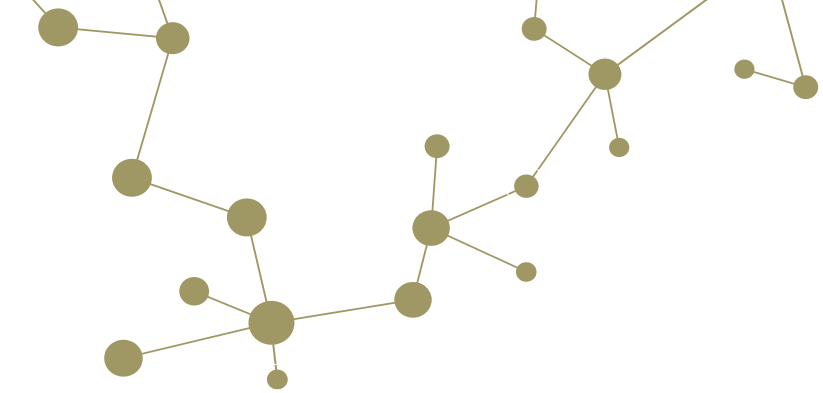
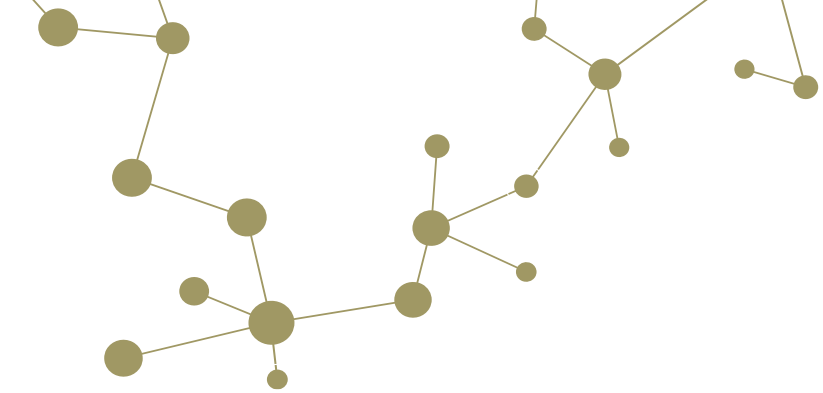


TABLE OF CONTENTS



INTRODUCTION

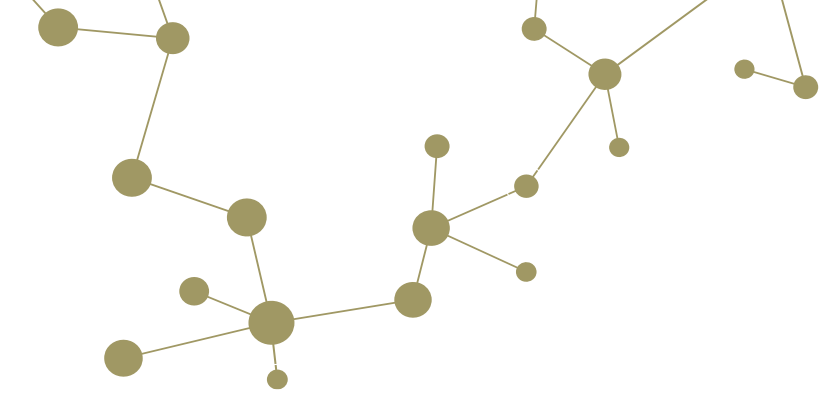
The success of initiatives that develop a skilled technical workforce increasingly depends on the growth and strength of professional interpersonal connections. These connections can be between education and industry professionals, or in the case of coordination networks they could be connections among the individuals collaborating on projects and research. Whether increasing professional connectivity is an explicit goal or a means of achieving a stated goal, project teams need a method to evaluate the growth and development of these professional connections over time. Social network analysis (SNA) provides a useful methodology for evaluating and describing both the structure and development of the interpersonal connections—or professional connectivity—within these contexts. SNA allows one to capture and illustrate the connections among individuals within a given defined population at a given point and to evaluate how those connections change over time (i.e., before and after the introduction of some activity or intervention).

Using a real illustrative example, this paper provides an explanation of the foundational concepts of SNA, practical guidelines for capturing survey data to conduct a SNA, and an overview for interpreting the basic measures that describe the interconnectedness of networks. It should be noted that SNA is frequently used to describe and evaluate the position of individuals within a network. This paper, however, will focus on SNA analysis and interpretation at the macro level.

OVERVIEW OF THE BASIC ELEMENTS & CONCEPTS OF SOCIAL NETWORK ANALYSIS (SNA)

Upon first hearing of the concept of SNA, people can frequently conflate it with social medial platforms in general and Facebook in particular thanks to the movie *The Social Network*. SNA has certainly been used to illustrate and analyze the interconnections among individuals on social media platforms, but SNA methodology is used to study interconnections among individuals—or collections of individuals—in many contexts that have nothing to do with social media platforms.

The very basic elements and concepts of SNA are not difficult to learn but, as with any methodology, the terminology and language can be less than intuitive. This section provides an overview of the basic concepts along with associated terminology that will be used throughout the paper.



Basic Elements of interest in SNA

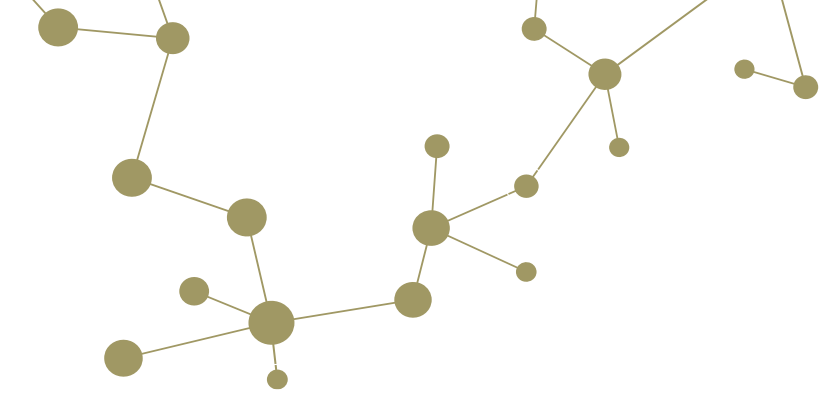
The two most basic elements of SNA are **individuals** (also called actors and *nodes*) that are connected via one or more types of **relational ties** (also called *edges*, *links*, or *arcs*). While those in the field tend to use the terms *nodes* and *edges*, this paper will use the more intuitive words *individuals* and *relational ties.*

Individuals can be a single person or a collection of individuals including groups, teams, organizations, institutions, or communities. It should be noted that, for research purposes or to help us better understand our network, we might also be interested in including *attributes* for each of the actors such as those listed below.

TARGET	POSSIBLE ATTRIBUTES OF INTEREST IN ANALYSIS
People	• Gender • Age • Race • Political affiliation
Institutions/organizations	• Profit vs non-profit • Educational vs industry • Numbers of employees
Communities	• Rural vs urban • Region • Population

Table 1. Types of targets and attributes that can be analyzed using SNA.

Networks are characterized by some type of **relational tie** among **individuals**. At the most basic level relational ties might be characterized by their nature—or type—such as romantic partners, friends, or professional connections. At another level we might characterize the relational tie by specific kinds of interactions as described in Table 2. In addition, it might be of interest to characterize these relational ties or interaction types by degrees or strength in terms of such things as frequency, closeness, warmth, and so on.



RELATIONAL TIES TYPES	TYPES OF POSSIBLE INTERACTIONS	WAYS TO CHARACTERIZE THE STRENGTH OF INTERACTIONS
Friends	• Leisure or recreational activities • Personal conversations	• Frequency of interactions • Perceived level of closeness
Professional connections	• Project collaborators • Co-committee members • Mentor/mentee • Ties by organizational structure (manager, direct report)	• Frequency of interactions • Affiliation for the other

Table 2. Examples of the types of relational ties, types of interactions, and ways to characterize interaction strength.

Individuals and relational ties in a network over time

SNA is a quantitative methodology. Much of its appeal, though, lies in the fact that it provides a powerful mechanism for illustrating the interconnections among individuals at a given moment or over multiple points in time. Each dot in Figure 1 represents an individual with the lines indicating relational ties. The figure shows the change in the relational ties within this group over a three-year period. In year 1 this group of individuals could be characterized as largely disconnected with only a few individuals having multiple connections and many with no connections. By year 3 most individuals had three or more connections and only one remained disconnected from the others in the group.

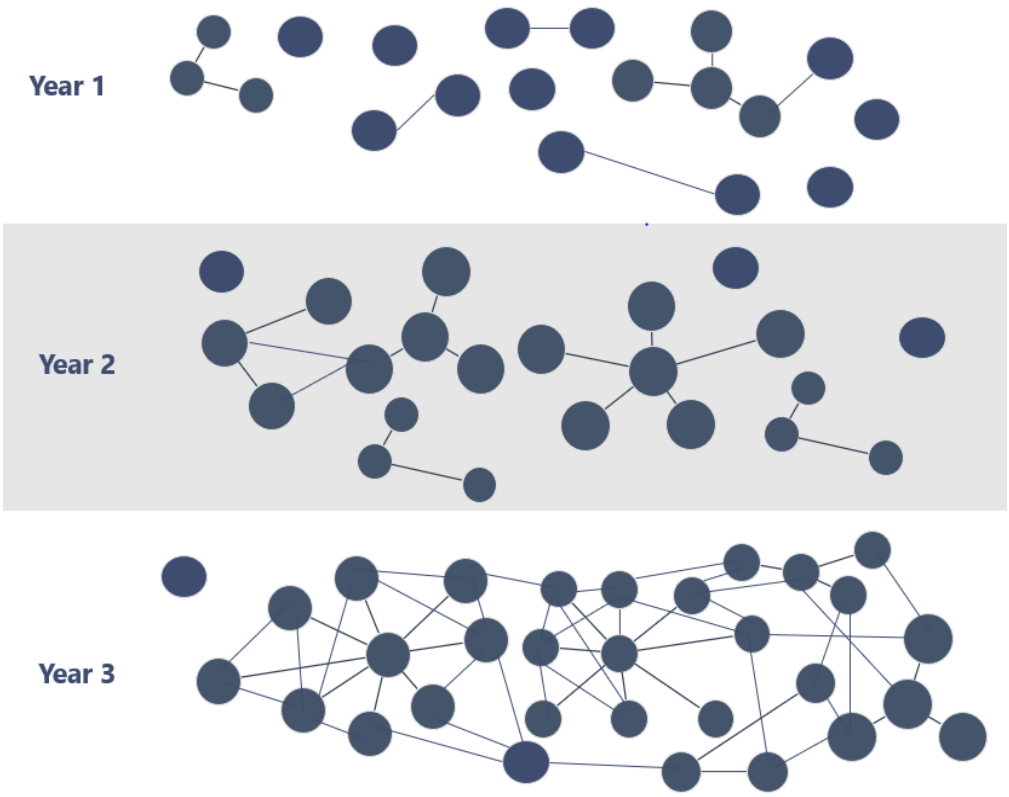
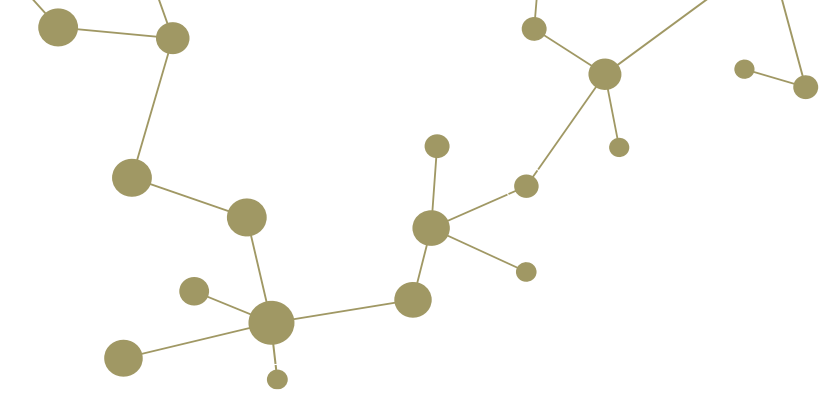


Figure 1. Illustration of a social network developing over a three-year period.

DATA COLLECTION

On the surface, this process might seem simple and straightforward. However, as with most survey data collection efforts the challenges are in the details. This section will include the various decisions that must be made (mostly sequentially) along with some potential issues that need to be addressed to ensure that the data collected provide the most accurate picture possible of the full network. The decisions made regarding the data collection process for the example presented in this document illustrates the types of decisions that must be made. The process for collecting full network data involve these four steps, which are described in greater detail in the remaining section:

1. Define the actor population
2. Define what constitutes are relational tie
3. Design a survey to capture the data
4. Administer a survey to capture the data



Defining the Individual Population

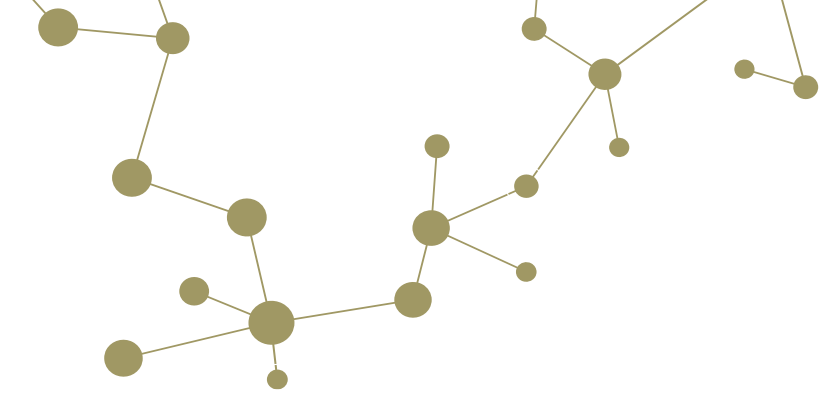
The first step is to define and bound the network population of interest. When conducting a full social network analysis, the name of every person in the population must be included in the survey. The goals of the project should be the main driver for defining the population, but there are feasibility issues that must also be considered. First, it must be possible to obtain a list of every individual within that population. Second, the list must not be so large that respondents quit the survey or submit incomplete responses.

Generating an accurate list of the population. The project team must invest the time up front to ensure that all the individuals on the list are part of the target population. Even the most carefully sourced list is likely to include individuals who are not part of the target population (e.g., administrative contacts) and individuals who were—but are no longer—part of the target population (e.g., those who have retired). Not doing so undermines the accuracy of the network description as generated through SNA. In addition, individuals who are part of the population will be missing from the list due to clerical errors or the fact that the list might not have been updated with the newest members of the target population.

The best way to address this issue is to allow respondents to write in the names of individuals with whom they have some type of relational tie. If the survey includes this option, someone will need to review each added name to determine if the individual is a member of the target population.

Defining the Relational Tie

In the context of professional connectivity, we are usually interested in understanding connections that go beyond a single one-time encounter either at a conference or after a meeting. These informal interactions are certainly important for initiating professional connections. However, asking respondents to recall each of these kinds of connections would be both burdensome and susceptible to recall error. Most people would be hard pressed to recall—even by description—each person that they met, much less recognize their name on a list. A relational tie should signify some type of professional connection that is characterized by more substantive interactions that involve exchanging ideas, providing guidance, collaborating on projects, participating together on committees, and the like.



Developing Survey Items

One of the most difficult aspects of SNA is to develop a survey tool that captures valid data. The survey tool must—at a minimum—be designed to capture each respondents' full set of relational ties within the population. In the context of professional connectivity, it might also be valuable to capture the types of interactions that are of interest (e.g., collaborated on a project) as well as some value or characterization that captures the strength of the relational tie (e.g., interaction frequency).

Survey items to capture relational ties and type of interactions. Respondents will need to consider each of the individuals in the target population and determine whether or not they have a relational tie with them. It is important to provide a limited timeframe for respondents as they select connections. The specific timeframe should be driven by the goals and context of the project; but asking people to recall relational ties from more than one to two years would be difficult for most and would be prone to recall error.

Consider that in addition to asking whether the respondent has a relational tie with an individual, it might also be useful to understand the type of interactions they have had with the person. The burden of this type of reporting can be reduced by first asking respondents to identify everyone with whom they have a professional connection based on the survey parameters of what constitutes a professional connection, and then asking the respondents to think about the specific kinds of interactions they have had with the connections they selected. This two-step method is not commonly utilized in SNA surveys, but the pretest conducted during the development of the survey in the subsequent example indicated that it was indeed easier on the respondents. It should be noted, however, that asking the questions in this fashion requires survey software functionality that is not available on all platforms.

Survey items to capture the quality or strength of the connection. When investigating the growth and development of social networks is important not only to determine who connects with whom, but also to capture aspects related to the strength and quality of those connections. The most typical aspect to capture is frequency. Other aspects that could be captured include friendliness, closeness, levels of trust, etc.

Example: Data Collection

A project being conducted by EvaluATE – the evaluation hub for the National Science Foundation's (NSF) Advanced Technological Education (ATE) program – aims to develop the evaluation capacity of ATE projects and centers. A key goal for increasing that evaluation capacity is to increase the professional exchanges among ATE evaluators. EvaluATE has implemented a number of activities and initiatives that are designed to increase professional connectivity including opportunities for ATE evaluators to meet through organized events (e.g., receptions at professional conferences) and providing a means for connecting and communicating with other ATE evaluators through a dedicated Slack channel.

The population of interest for this project is "current ATE evaluators." This focus is driven by the project goal to "increase professional exchanges among ATE evaluators." The project team was able to obtain a list of ATE evaluators from an internal database. This list had 145 names after duplicates and retirees were removed. This list is not immediately updated, so the project team knew it would miss individuals who had only recently become ATE evaluators. Consequently, a question was included on the survey to allow for write in names.

In the case of the work conducted by EvaluATE, respondents were asked to consider only those connections that included more substantive types of evaluation-related interactions which could include any of the following: Providing or receiving evaluation guidance, resources, or information; Working together on an evaluation; Collaborating on educational or outreach activities (e.g., article, presentation, committee).

The ATE evaluators were asked to review the full list and select each evaluator with whom they had at least one evaluation-related interaction during the past 12 months. After seeing each of the listed names, respondents were then provided with the opportunity to write in the names of any other ATE evaluators with whom they have had at least one evaluation-related interaction in the last 12 months. The next screen was a matrix question that included the name of each "selected" and "write in" ATE evaluator in the first column and asked the respondent to select the types of evaluation-related interactions they have had with each person on the last 12-months (see figure that follows).

	PROVIDED guidance, resources, or information	RECEIVED guidance, resources, or information	Worked together on an evaluation	Collaborated on an educational or outreach activity	Other
Name 1	☐	☐	☐	☐	☐
Name 2	☐	☐	☐	☐	☐
Name 3	☐	☐	☐	☐	☐
Name (write-in)	☐	☐	☐	☐	☐

Figure 2. Response options for the question, "What types of evaluation-related interactions have you had with each person in the last 12 months? (Select all that apply)" that was used for EvaluATE project example.

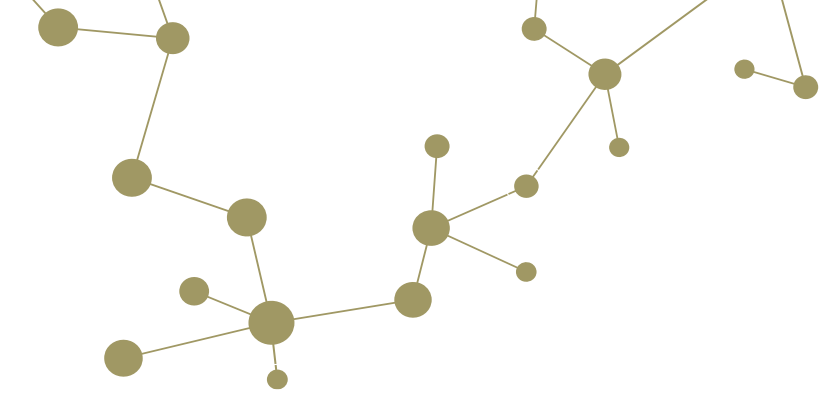
After identifying each of their ATE evaluator connections, respondents were asked to indicate how often they had interacted with that individual on evaluation-related matters (i.e., 1-2 times, 3-10, or more than 10) and then finally asked to select or identify the types of interactions they had. While respondents were provided with the four types of interactions as previously described (e.g., providing evaluation guidance, resources, or information, etc.), they also had the option to select "Other" and to describe the type of interaction. An abbreviated copy of the SNA survey items is provided in Appendix A.

	1 – 2 times	3 -10 times	More than 10 times
Name 1	☐	☐	☐
Name 2	☐	☐	☐
Name (write in)	☐	☐	☐

Figure 3. Response options for the question, "About how often have you had evaluation-related interactions with each person in the last 12 months?" that was used for EvaluATE project example.

ANALYZING SNA DATA

Unlike data used in other fields of statistical analysis, social network data always consist of at least two datasets: the **individual list** (or node in SNA terminology) and the **relational tie list** (or edge list in SNA terminology). The individual list includes the full set of individuals' names. The relational tie list describes the paired relationships among the individuals.



Individual List

An individual list (i.e., node list) will include one row for each actor and might also include one or more attributes (e.g., gender, race, age) for each (see Table 3).

Name	Gender	Age	Employment
John	M	35	Consulting, research, or evaluation firm
Terry	F	54	Consulting, research, or evaluation firm
Bob	M	33	Independent consulting practice
Cathryn	F	37	Independent consulting practice
David	M	68	Independent consulting practice
Evelyn	F	30	Higher education
Steven	M	76	Independent consulting practice
Marcus	F	41	Independent consulting practice
Maggie	F	38	Independent consulting practice
Elaine	F	51	Consulting, research, or evaluation firm
Frazer	M	42	Independent consulting practice
Jacinda	F	46	Higher education

Table 3. Partial example of a partial individual list (or node list) dataset.

Relational Tie List

The relational-tie list (i.e., edge list) will include a row for each of the paired relationships among the actors and might also include descriptions of the types of relations that exist between the dyads and some type of value (e.g., frequency of interaction, strength of tie; see Table 4).

Source	Target	Frequency	Received guidance from	Provided guidance to	Worked on project together
John	Terry	3	1	0	1
John	Jacinda	1	1	1	0
John	George	2	0	1	1
John	Maggie	3	1	0	0
Terry	Steven	2	0	1	1
Terry	Marcus	2	0	0	1
Terry	John	1	1	0	1
Bob	Terry	3	1	1	0
Bob	Cathryn	1	0	0	1
Cathryn	Terry	2	0	1	0
Cathryn	Bob	1	1	1	0

Table 4. Partial example of a relational tie list (i.e., edge list) dataset.

VISUAL AND QUANTITATIVE DESCRIPTIONS OF A NETWORK IN SNA

A key strength of SNA as a method for evaluating networks is the ability to generate visualizations of those networks which can provide rich and useful information about the network as a whole and the position of individuals within that network (see Figure 4). Software packages, such as the Gephi and UCINET file formats, are required to generate these visualizations. In addition to generating powerful visual depictions of networks, SNA can also be used to describe quantitatively the networks at a given point and then tracked to monitor changes over time. For example, SNA provides measures for analyzing the extent to which the individuals within a network are interconnected (i.e., density) and the extent to which connections are held relatively equally among individuals (i.e., centralization).

Density. The simplest SNA measure to describe connectedness at the network level is density, which measures the extent to which individuals in a network are interconnected. It is calculated as the total number of paired connections or ties in a network divided by the maximum number of ties possible.

Centralization. Degree of centrality is a common individual-level measure used in SNA to denote how influential a given individual is within the network and based purely on the number of connections associated with the individual. Centralization, however, is a network-level measure that provides an indication of how centralized a network is. In a highly decentralized network, most individuals within the group have similar numbers of connections. In a highly centralized network, most connections are held by a relatively small number of individuals.

The magnified portion of the network shown in Figure 5 provides a clearer picture of the relatively few highly connected or centralized individuals within the ATE Evaluator Network at baseline. Figure 6 shows the many individuals with relatively fewer connections that is more typical within the ATE Evaluator Network. However, it should be noted that some of the less connected individuals have developed strong connections with each other as indicated by the darker lines.

Example: Visual and Quantitative Descriptions

The survey data gathering effort for the EvaluATE project yielded 115 network individuals with a total of 422 relational ties. It should be noted that – typical with any survey – some of the 145 evaluators on the population list did not respond to the survey and a few additional individuals who were not included on the population list were written in by other respondents. Consequently, the illustration quantitative descriptions of the network depicted here does not represent the full true network.

Each dot in the figure that follows represents an *individual* (or "node" in SNA terminology) in the evaluator network. The larger and darker the dot, the more connections that individual has.

Each line represents a *relational tie* (or "edge" in SNA terminology) between two individuals. The darker the line, the more interactions reported for that connection over the prior year. An arrow indicates that one individual identified the other as a connection. Connections with an arrow on each end indicate that each had identified the other as a connection (see Figure 4).

The ATE Evaluator Network had a density of 4.9% at baseline. It should be noted that this is probably an underestimation of the true interconnectedness among the 115 individuals represented in the network. There were 54 individuals who did not complete the survey but were identified as a connection by one or more people who did. If each of these 54 individuals had provided connection information, the interconnectedness of the network – as measured by density – would most certainly be higher.

The ATE Evaluator Network had a centralization score of .49 at baseline. Centralization ranges from 0 (i.e., all individuals have the same number of ties to others) to 1 (i.e., all ties are held by a single individual in the network). The high degree of centralization of the ATE Evaluator Network at baseline can be clearly seen in Figure 5. Most of the ATE evaluators within the network had relatively few but sometimes strong relational ties with other ATE evaluators.

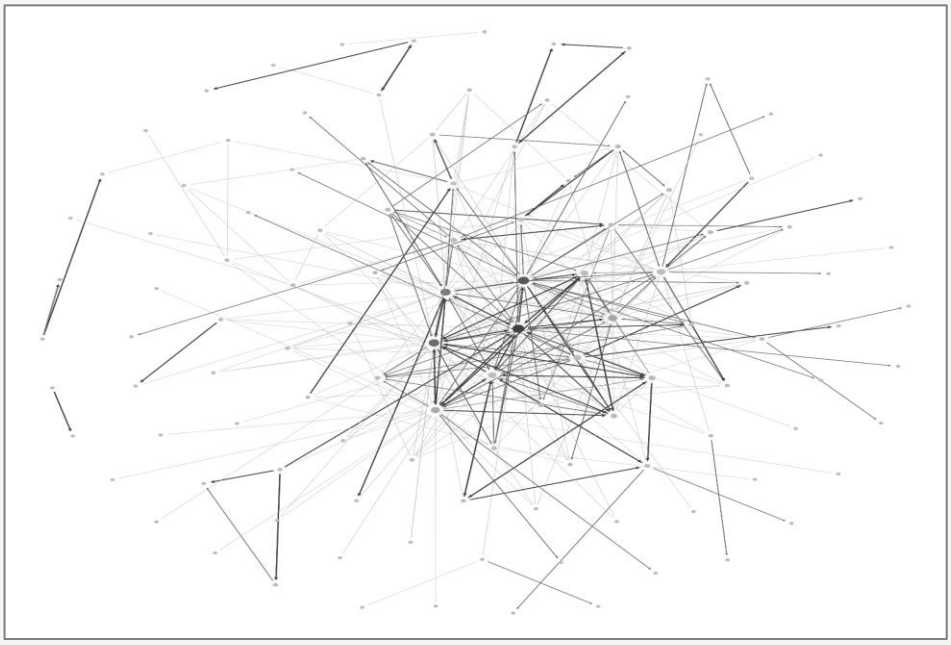


Figure 4. ATE Evaluator Network at baseline showing 422 connections among 115 individuals.

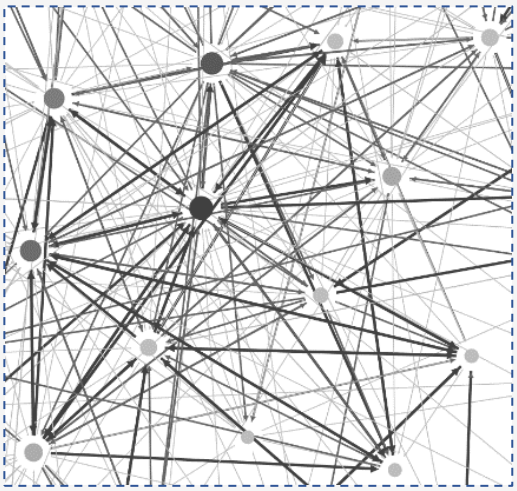


Figure 5. Magnified portion of the ATE Evaluator Network illustrating high degree of centralization within the network at baseline.

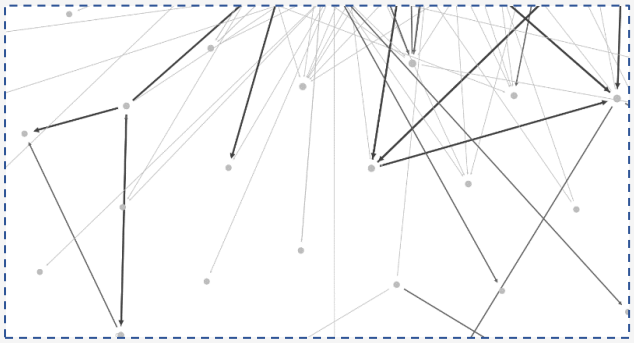
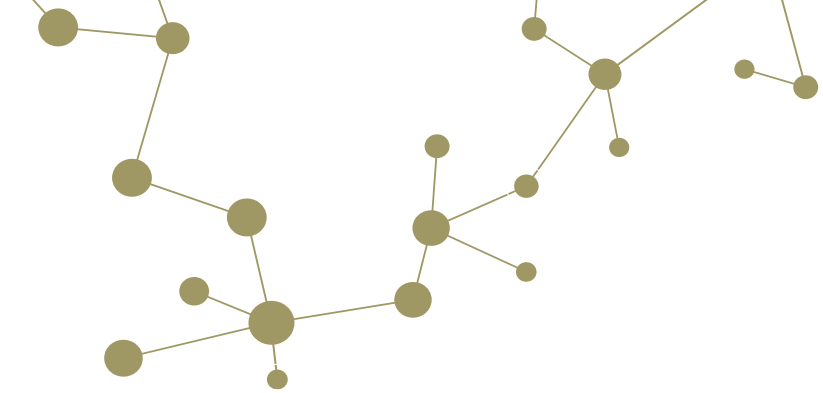


Figure 6. Magnified portion of the ATE Evaluator Network showing the more typical individuals within the network who have relatively few - but sometimes strong - relational ties with others in the network.



SUMMARY THOUGHTS

The purpose of this paper was to provide a basic understanding of the foundational concepts of SNA, practical guidelines for capturing the necessary survey data to conduct the SNA, and an overview regarding how to interpret the basic measures used to describe the interconnectedness of networks at a macro level.

The example baseline data shows that many ATE evaluators connect with each other on evaluation-related matters, but there are certainly opportunities for growth in terms of the number of connections, the frequency of interactions among connected individuals, and in the overall connectedness – or density - of the network. Ideally, as the network of ATE evaluators becomes more interconnected, there will be less of a reliance on a relatively small group of individuals for evaluation-related guidance, support, and collaboration.

SNA is a well-established and complex methodology that requires high levels of expertise and experience in practice. Readers who are new to the topic are advised to seek guidance from someone with expertise and experience in the methodology before conducting SNA in practice.

This work is supported by the National Science Foundation EvaluATE: Evaluation Learning and Resource Hub for Advanced Technological Education (DUE #1841783). The opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not reflect the position or policies of the National Science Foundation.