Understanding the Chequamegon Bay Regional Food System Through Social Network Analysis

Deanna L. Schneider

University of Wisconsin – Eau Claire

DS785: Capstone

Dr. Mary Tripp

August 5, 2020

Abstract

The Chequamegon Bay Regional Food System in northern Wisconsin is a network of approximately 100 food growers, value-added producers, food system educators, retailers, and market managers. The network is supported by the educators employed by the University of Wisconsin-Madison, Division of Extension. The network, while active, has failed to produce the desired economic and social benefits expected. A review of the literature found many instances of social network analysis being used to assess regional food systems. Based on theories of social network analysis, the network was surveyed about their roles, affiliations, attitudes, and relationships. Survey data was used to do an exploratory social network analysis, including visualizations, actor prominence discovery, community detection, and simulation comparisons. Additionally, exponential random graph modeling was used to identify factors contributing to collaborative tie formation. Key findings include low density and high diameter in the Coordinate and Collaborate networks, the presence of Extension educators as central actors in the network, and a higher likelihood of food systems educators collaborating with each other than with non-food systems educators. The findings will be used as baseline data for future longitudinal studies of the network. Additionally, Extension educators will utilize the findings to develop targeted initiatives aimed at building network strength and resiliency.

Keywords: social network analysis, ERGM, regional food system, Extension

Contents

List of Tables

List of Figures

Understanding the Chequamegon Bay Regional Food System Through Social Network Analysis

The following project was completed as a client-based project in partial fulfillment of the requirements for the degree of Master of Science in Data Science. The primary client was Kellie Pederson, University of Wisconsin-Madison, Division of Extension Community Development Educator for Bayfield County, Wisconsin. Pederson's professional focus has been on small business management with an emphasis on sustainability. Her work has incorporated elements from the arts, health and wellness, and natural resources (University of Wisconsin-Madison, Division of Extension, n.d.). There were two additional clients. Sara DeGraff, the University of Wisconsin-Madison, Division of Extension Agriculture Educator for Bayfield County has focused her work on community development and farm management. Her background has been in Seed-to-Kitchen experiments, focusing on the connection between farmers, chefs, and communities (University of Wisconsin-Madison, Division of Extension, n.d.). The final client was Jason Fischbach, Extension Agriculture Educator for Bayfield County & statewide Food and Energy Woody Crops Specialist, University of Wisconsin Madison Division of Extension. Fischbach's research interests included hazelnuts, fast-growing tree crops, vegetables, fruit crops, and winter high tunnel greenhouse production (University of Wisconsin-Madison, Division of Extension, n.d.).

Chapter 1: Background

The Setting

The Chequamegon Bay is an inlet on the southern shore of Lake Superior. It is bordered to the southeast by Ashland County, Wisconsin and to the southwest by Bayfield County, Wisconsin. It is a beautiful and relatively remote area. Ashland County encompasses 1045 square miles and has a population of 15.5 people per square mile ("U.S. Census Bureau QuickFacts: Ashland County, Wisconsin," n.d.). Bayfield County is more populous, with 1478 square miles of land and 87.4 residents per square mile ("U.S. Census Bureau QuickFacts: Bayfield County, Wisconsin," n.d.). Home to the Apostle Islands National Lakeshore, numerous natural areas, and the Big Top Chautauqua, it has become a popular tourist destination.

The Food System

Historically, the land was deemed unsuitable for farming, with heavy timber stands and rocky soil. Large-scale agriculture never took hold here (Larson, 2005). In more recent years, small-scale food and agriculture have become integral to the economic, social, and cultural fabric of Ashland and Bayfield counties. A 2018 Local Food Survey conducted by the Center for Rural Communities at Northland College noted strongly held beliefs around and in support of a local food economy. Specifically, 92% of households wanted more of the food that they purchase to be grown in Ashland and Bayfield Counties; most were willing to pay more for locally-produced food; and 90% of households felt that small farms are important to the regional culture (Kemkes, Hofsted, & Tochterman, 2018). Additionally, even though farming accounted for just 1.9% of Ashland County's 2018 economy, Radke & Deller (2018) identified it as a growth segment worthy of future investment.

The University of Wisconsin-Madison Division of Extension educators in Ashland and Bayfield counties have been working in tandem with many local stakeholders to support and grow the local food system. In personal communication with Kellie Pederson (January 2020), she shared extensive information about the situation in the area. She emphasized their goal is to build regional wealth, vitality, and social equity through the development of the regional food system.

Within the last 10 years, there have been several new additions and expansions within the local food system. These include: new small diversified vegetable and fruit farms; new wineries, distilleries, and cideries; new restaurants focused on buying locally and offering local food on their menus; a new food producers cooperative that has branched out into a broader regional markets; new local food-related events for both community members and visitors; consistency and growth at farmers markets; presence of multiple CSAs; growth in programs providing access to local fresh fruit and vegetables for low-income community members; a growing presence of food sovereignty in the region; growth in capacity of organizations providing emergency food to populations in need; a new food center with commercial kitchen and co-packing capacity; growth in micro-loans available to producers; increase in quantity of local foods stocked at grocery stores; growth in commitments to purchase local food from cornerstone institutions; and a new degree program at the local college for sustainable farming and food systems.

The Problem

Despite these numerous recent developments and significant public support, Pederson noted that the food system of the greater Chequamegon Bay Area has not yet matured to become economically sustainable or self-perpetuating. CSA sales have declined, value added producers are facing continued challenges with growth and distribution to broader markets, and local food is still out of reach for many residents due to both real and perceived barriers. While great effort has been put forth, success has been modest. Many of the expected benefits, including sustainable livelihood creation, increased entrepreneurship, increased access to local food, and other useful “spillover effects” have yet to be realized. This has left key stakeholders questioning how best to continue to support the local food system. What is next for local food in the area?

When Pederson posed this question to colleagues and stakeholders, the discussion turned to considerations around the interconnectedness of the community and how businesses work together. Pederson noted that in the winter of 2018-2019, she and colleague Jason Fischbach conducted a series of networking events for value-added producers, farmers, and restauranteurs in the region. They called these “Farmily” events. What they learned was that many of the people in attendance were not familiar with one another. This led them to consider the possibility that what was needed was not more stakeholders but stronger connections between existing stakeholders.

The Network as Possible Solution

The impetus for focusing on the social and business relationships within the network is based on the idea that a network made up of strong ties will have greater economic success than a network made up of loose ties. Pederson’s general hypothesis was that if she could increase the density of the network, allowing for a more fluid transfer of knowledge and information, there would be a corollary increase in economic activity and success among the members of the network. She based this notion on the economic phenomena known as a dynamic cluster. Porter (2008) defined clusters as “geographic concentrations of interconnected companies, specialized suppliers, service providers, firms in related industries, and associated institutions (…) in particular fields that compete but also cooperate” (p. 213-214). In other words, in a dynamic cluster, firms are not just geographically co-located, but also often share knowledge and resources and build social capital through intentional and organic networks (Falck & Heblich, 2007; Porter, 2014). It was her hope that by developing a greater understanding of the social networks within the local food system, she could devise plans to increase the density of the network. The goal was a fully realized dynamic cluster in the Chequamegon Bay Region with its associated economic benefits.

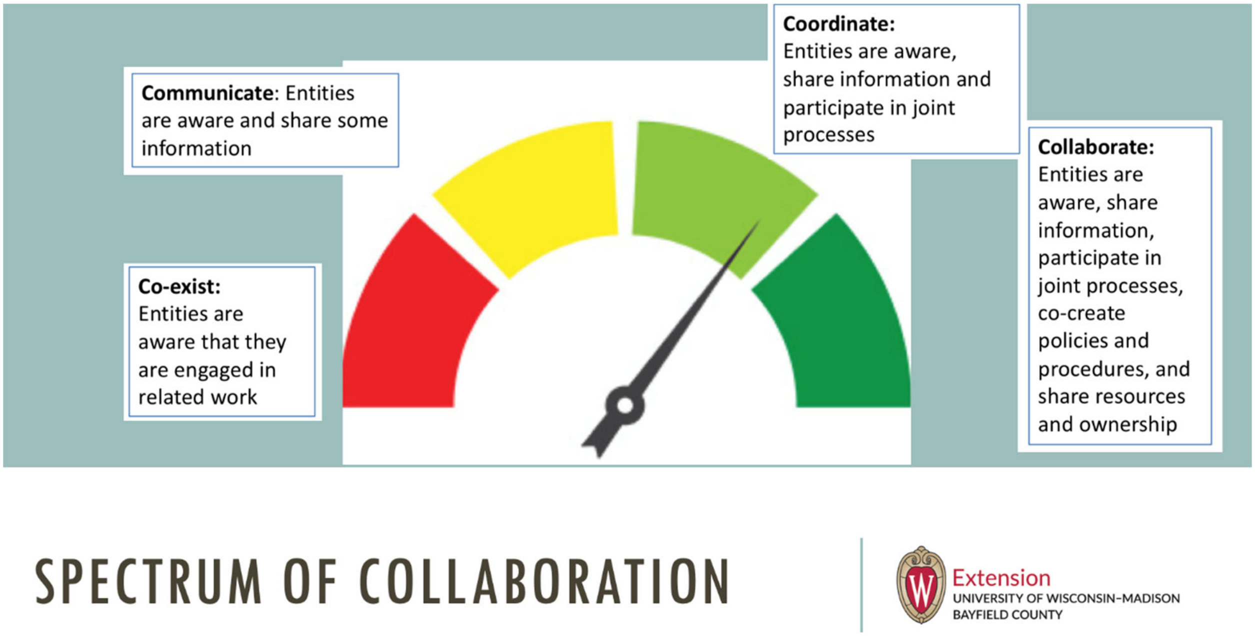
Defining Relationships Through the "Spectrum of Collaboration"

During the Farmily events, Pederson and Fischbach introduced the notion of the "Spectrum of Collaboration" (Figure 1). The Spectrum of Collaboration was a tool for identifying the nature of the relationship between individuals or organizations (Tamarack Institute, 2017). It encompassed a range of connection types - from the weakest connection (coexist) to the strongest connection (collaborate). Explicitly noted in the spectrum was the notion that weaker connections are a component of stronger connections. Collaboration encompassed coexisting, communicating, and coordinating, for instance. Each level added a new aspect to the relationship. For example, to move from coexisting (being aware of one another) to communicating, individuals must both be aware of one another and share some information.

Pederson and Fischbach realized that many of the stakeholders at their Farmily events were merely coexisting with one another, and very few stakeholders were approaching coordination or collaboration. A limitation of this realization was that their knowledge up to this point was anecdotal. They had not formally studied the relationships in the local food system.

Figure 1

Spectrum of Collaboration



Assessing the Network Through Social Analysis

The most scientifically valid way to assess the state of the relationships within the Chequamegon Bay regional food system was to conduct a formal social network analysis. Social network analysis is a methodological approach built on graph theory – the mathematical study of points and the lines which connect those points (Scott & Carrington, 2011). When applied to people and their relationships with one another, the people are represented by points, and the relationships are represented by lines connecting those points, constructing what is known as a sociogram. These points and lines can also be represented as a matrix, and then through matrix algebra, researchers can derive various numerical statistics that describe the network. These statistics can be used to understand how one network compares to another or how a single network changes over time.

For the Chequamegon Bay Regional Food System, the clients and I collected data and performed a baseline social network analysis such that I could answer the following questions:

1. What did the network look like, when presented visually?
2. What were the core statistics (density, transitivity, diameter, average degree, etc.) for the entire network and sub-networks?
3. Who were the important actors within the network?
4. Was the network made up of sub-communities and if so, around what factors were the sub-communities formed?
5. Could I identify factors that were predictive of strong network bonding (those relationships that fall on the far-right side of the Spectrum of Collaboration)? If so, what were those factors?

The intention was to use this baseline analysis to help the Extension educators form plans for future interventions that can strengthen the network.

This work was the first phase of a multi-year plan to continuously assess the network and determine if there is a significant relationship between network density and regional economic development, or between an individual's network position and personal/individual business prosperity.

Chapter 2: Literature Review

When considering a literature review for this project, three potential avenues of inquiry presented themselves: SNA as applied to food systems research, SNA as applied to Extension programming, and SNA as applied to dynamic business clusters. All three were relevant to the project and the client's objectives. As background, social network analysis (SNA) as a methodology to understand the shape and nature of social structures was first employed in the 1930s (Scott & Carrington, 2011). SNA began with simple visualizations and gradually added whole-network statistical measures and actor prominence measures (Furht, 2010). The advent of computers capable of fast calculations rocketed the use of social network analysis (and the literature about it) into a new dimension. SNA has been used to study everything from MicoRNA-Disease Associations (Zou, Li, Hong, Lin, Wu, Shi, & Ju, 2015) to social structures of captive chimpanzees (Clark, 2010). There is both great depth and breadth in the SNA literature. While it would be impossible to cover all of it, I highlighted some key relevant work.

SNA in Food Systems Research

The United States Department of Agriculture (USDA) has been studying the economic and social impact of local food systems since the late 1990s, but as of 2019 had yet to publish any studies using social network analysis as a methodology, though they did recognize it as a tool with great potential to understand the social capital of local food systems (Tropp, 2019). That does not mean SNA has not been applied to food systems. Social network analysis has been employed as an evaluative tool for food systems for the last couple of decades (Jarosz, 2000; Brinkley, 2018).

Despite broad use of SNA in food systems, there has been little agreement on the best way to conduct social network analysis. Instead, researchers have been exploring a myriad of ways to apply SNA to food systems work and doing so in a relatively scattershot way. For example, some studies focused on more informal relationships, such as advice networks and friendship networks (Chiffoleau, 2009; Chiffoleau & Touzard, 2014) while other studies focused on formal relationships, like joint membership in an organization (Crespo et al., 2014). Studies ranged from relatively simplistic, focusing on visualizations and descriptive statistics (Carboni et al., 2017) to advanced use of algorithms, such as the use of multinomial regression (Crespo et al., 2014). There has been a heavy emphasis on supply-chain networks and the implied social network that arises between producers and consumers (Trivette, 2018; Nicolosi et al., 2019). Longitudinal studies of regional food networks appeared less common than cross-sectional studies (Herrera & Dimitri, 2019; Christensen & O'Sullivan, 2015). Finally, while most studies relied on survey data or published databases, the use of simulated network data also appeared (Herrera & Dimitri, 2019).

The conclusions drawn varied as well. Social network analysis has been the basis for asserting that collaborations increased over time (Christensen & O'Sullivan, 2015); retailers were essential to regional food systems (Trivette, 2018); alternative food supply chains (like those found in regional food systems) renewed ties between producers who would be disconnected in more traditional food supply chains (Chiffoleau, 2009); and desire for social status within a food system could drive network tie formation (Chiffoleau & Touzard, 2014). Each of these conclusions, within the context of the research, appeared to be reasonably well-grounded.

Some other conclusions appeared to be less well-grounded. Herrera & Dimitri's 2019 study of organic dairy farms, for instance, made some questionable assumptions. They based their analysis on the assumption that a 50-mile radius denoted geographic proximity and limited their simulated network ties to that radius. Bell & Zaheer (2007), on the other hand, noted that individual-level friendship ties (like those found between individual farmers) span geographic holes, potentially reaching outside Herrera & Dimitri's artificial boundary. Further, Morgan

(2004) noted that we, unfortunately, know very little about how organizations learn and how important geographic proximity is to knowledge transfer. The weakness in Herrera & Dimitri's approach provided an opportunity for this study to contribute to the knowledge base by explicitly testing the effect of geographical proximity on relationship ties.

Despite the variance in approach to SNA with food systems, there was one commonality. Formal or informal relationships, organizations or individuals as nodes, simulated or genuine networks, advice-seeking or supply-chain networks, simple SNA or paired with additional modeling algorithms – despite the methodology or subject, each author asserted that SNA is a viable approach to building understanding about some aspect of local food systems.

SNA with Extension

While the literature on SNA and food systems has been developing, so has the literature on SNA applied to Extension programming. Extension, sometimes referred to as Cooperative Extension or Agriculture Extension, is an extension of the land grant university system. Federally supported by the United States Extension Department of Agriculture and the National Institute of Food and Agriculture, Extension aims to bring the knowledge gained through university research directly to citizens throughout the country (United States Department of Agriculture, n.d.).

Extension programming has been experiencing an on-going shift from the Extension educator as the sole expert (Rogers, 2010) to a model in which Extension educators are just one source of knowledge, embedded in a community of knowledge-makers (Bastos et al., 2018; Carr & Wilkinson, 2005). However, prior SNA research with Extension assumed Extension's role was to be the central node in a centralized star network (Christensen & O'Sullivan, 2015).

Alternatively, Lubell et al. (2014) proposed the concept of "Extension 3.0" – an approach to Extension programming that capitalizes on the decentralized nature of modern knowledge transfer, emphasizing the importance of multiple individuals sharing knowledge and resources. They noted that:

> …changes in knowledge systems have decreased the relevance of the traditional top-down continuum model of agricultural extension. Modern knowledge systems feature a diverse network of experts that includes universities and Cooperative Extension, but also many other types of actors who distribute information through multiple pathways including traditional outreach programs and new tools…. (p. 1092)

Lubell et al. (2014) focused primarily on the loose networks that can be developed using what they call information and communication technologies (ICTs) – online tools such as social media, mobile apps, and webinars. They went on to state that to understand Extension's current role and transition to more strategic roles, it was important to map existing networks and identify central and disconnected actors (Lubell et al., 2014).

While Lubell et al. (2014) implied that Extension should seek to decentralize their networks, Bastos et al. (2018) noted that core-periphery networks continue to proliferate, particularly where highly specialized information is concerned. They studied agriculture-related twitter data and found core-periphery models prominent, with government agents (like those that work for Extension) and news outlets at the center, and much less communication directly among producers. They offered the advice that Extension educators should position themselves as brokers in social networks to "…facilitate a more distributed, horizontal, and efficient flow of information among the periphery…" (p. 291).

Given that the Chequamegon Bay Food System study involves Extension educators as nodes in the network, the analysis could benefit from a focus on the core-periphery nature of the network and where Extension educators fit. The central question is, can Extension educators in the Chequamegon Bay region act as the brokers to connect the periphery (if, in fact, a core-periphery network exists)?

SNA for Business Clusters

Pederson was particularly interested in assessing the efficacy of the regional food system as a local business cluster. Giuliani & Pietrobelli (2011) asserted that SNA is a novel approach to the evaluation of cluster development programs. They defined cluster development programs as the efforts of the government to spur economic growth through supporting and developing a group of actors (business owners, firms, entrepreneurs) who are seeking more efficiency in production and higher innovation in a specific geographic area. This was most typically done by strengthening the relationships within the network.

Giuliani & Pietrobelli provided valuable background knowledge about the various types of networks seen in business clusters. They articulated when network weavers might strive to have each type of network. They noted that the approaches network facilitators could take to strengthen a network are highly dependent on the type of network sought. Their work was invaluable as a framework for identifying the next steps after the initial analysis of our network. However, they presented their work as a roadmap and did not apply it to any real-world data. This analysis, in part, attempted to apply their roadmap to our real-world data.

Final Thoughts

While there was a plethora of research on social network analysis with regional or local food systems and growing research on Extension's role in networks, none of the current studies were directly analogous to our study of the Chequamegon Bay food producer's network. Like some of the studies, our study incorporated both producers and suppliers, but our study also incorporated tangential supporters of the network, who Pederson hypothesized would be key players. Pederson did not wish to focus on supply chain mechanisms, or advice networks. Instead, she wished to focus on the spectrum of collaboration. Finally, our study used primary source data, not derived or simulated relationship data.

Chapter 3: Data Collection/Methodology

Data Collection

Data collection in social network analysis begins with identifying the nodes in the network. Nodes in the Chequamegon Bay Regional Food Network could have arguably been either organizations/businesses or individuals. After discussing the pros and cons of each approach, the client and I decided to focus on individuals as nodes. This decision was made for several reasons. First, in a rural community people matter, and name recognition is important. Social prestige among farmers has long been linked to the flow of information and knowledge (Lionberger, 1959), and social prestige is primarily affiliated with the individual. Second, while this study deals with a broad range of organization and business types, from non-profit agencies and educational organizations to restaurants and retail establishments, the prior educational efforts have focused not on the organizations/businesses, but on the individuals associated with them (K. Pederson, personal communication, January 2020). Finally, individuals in the network were often affiliated with multiple organizations or businesses. One individual may be affiliated with both a farm and a farmer’s market, for example. Additionally, due to economic necessity, many of the individuals in the network held secondary jobs, often in related fields such as restaurants. While social network analysis can be done on two-mode affiliation networks where individuals are tied to organizations and organizations are tied to each other (Borgatti & Everett, 1997; Carboni et al., 2017), two-mode analysis is more complicated, with fewer potential avenues of analysis. Finally, Pederson was less interested in the collaborations between organizations and more interested in the collaborations between individuals.

Once the decision was made to study individuals, the next decision to be made was which individuals to include in the analysis. Again, there are multiple approaches to solving this problem. One approach focuses on identifying just a few key individuals in the network, querying them about their important relationships, and subsequently querying the individuals they identified as important to them. This happens iteratively and is called snowball sampling (Doreian & Woodard, 1992; Naderifar et al., 2017). This approach is particularly effective when researchers are interested in a sub-network around some prominent actors. It can also be useful when researchers are interested in the entire network, but do not have reliable information on who all is in the network (McLeod & Vaughan, 2014). Alternatively, when the network is well-known, a viable technique to use is the fixed list technique, in which a fixed list of network actors is identified at the start of the research project. In the fixed list approach (sometimes also called the roster-recall approach), actors are given a list of other actors and asked to identify their relationship with those actors (Carolan, 2013; Reid et al., 2008).

Pederson identified 108 individuals whom she thought made up the bulk of the network. However, she was concerned that she may be missing some key individuals. Further, she wanted to make sure that everyone in the network consented to be a part of the research. To accommodate both of those needs, Pederson, DeGraff and I designed a two-stage data collection approach that incorporated elements of the snowball technique but resulted in generating a fixed list. The first stage was a 23-question online survey (Appendix A). While the survey was focused primarily on node characteristics, the final question allowed respondents to nominate up to three individuals who were important to them in their food system work. Pederson reviewed these nominations but did not identify any new candidate actors.

Other questions in the initial survey focused on characteristics that Pederson hypothesized might be related to their tendency to collaborate with others. Some of these questions focused on affiliations, such as the organized groups individuals belong to, the farmers' markets at which they vend or shop, and the social media outlets in which they participate. Other questions focused on individual characteristics, such as the degree to which respondents identified as an introvert or extrovert or their opinions on the value of collaboration. These questions were loosely based on the notions of horizontal and vertical individualism and collectivism as described by Triandis and Gelfand (1998). Categorizing people by their affinity for individualism or collectivism can help articulate when individuals are more focused on societal benefit (collectivism) or personal benefit (individualism). Collectivist people and cultures tend to be more oriented towards collaboration than individualistic people and cultures.

Of the 108 initially surveyed individuals, Pederson identified a final roster of 99 individuals. Via a second online survey, each person in the roster was asked to identify where they are on the Spectrum of Collaboration with each other individual if they were aware of or had some type of relationship with that individual. If a respondent did not know the person, they were directed to leave that survey item at the default choice of "Unknown." The survey was initially delivered via email on May 15, 2020. One additional person was added to the list after the publication and initial emailing of the roster. Two auto-generated follow-ups were sent to non-respondents on May 21, 2020, and June 8, 2020. Due to a low response rate, the Agriculture Extension Educator from Bayfield County, Sara DeGraff, personally contacted all non-respondents during the weeks of June 7 and June 14. The survey was closed on June 19, 2020 with 54 responses.

Data Cleaning/Preparation

Directed or Undirected Network

Networks can be either directed or undirected. A directed network is one in which the directionality of the ties matter (A → B ≠ B → A), whereas in an undirected network directionality is irrelevant (A → B = B → A) (Luke, 2015). Directedness is usually implied in the data collection question asked. For instance, if the survey asks, “from whom do you get advice” the implication is that the network is directed, because you are asking, “from” whom. In contrast, undirected network questions usually revolve around “with” types of questions – “with whom do you have sex” or “do you partner with any of these agencies.” While it is acceptable to treat a directed network as undirected, the reverse is not true (Scott, 2017).

The relationship question used in this study was an undirected question, so the analysis treated the network as undirected as well. However, for each relationship in which both respondents answered the survey, we have two answers about where those two individuals are on the spectrum of collaboration. In an ideal world, both halves of the dyad would consistently agree about their relationship. Unfortunately, within the complete cases, the two members of each dyad agreed only 51% of the time. Informant inaccuracy is common in social network analysis. Often, survey respondents are answering based on their cognitive network, as opposed to the actual relationships that exist (Killworth & Bernard, 1976). In other words, the respondents think they have a certain type of relationship, but the feeling is not mutual or not understood in the same way by both parties.

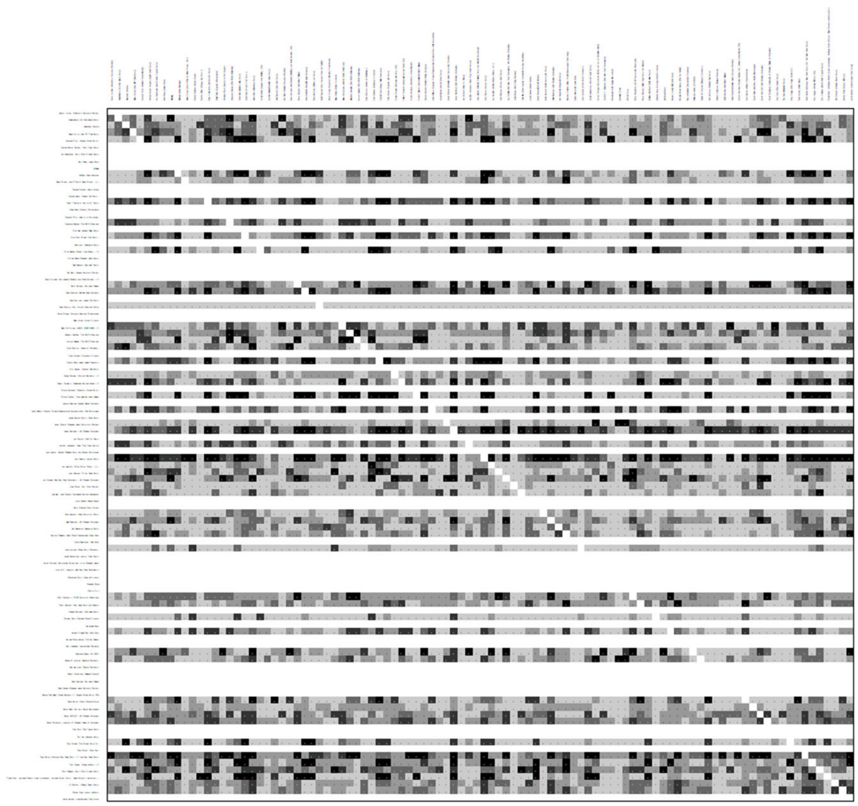
This low level of reliability limited our options. Fortunately, another viable method of simplifying a complex graph is to generate multiple binary graphs (Scott, 2017). The complex data in this study can be broken into five simple, undirected graphs – All Possible Relationships (for comparison only), Coexist, Communicate, Coordinate, and Collaborate. If either respondent in the dyad identified the relationship with at least that level, the dyad was included in the graph. For example, respondent A identified a "coordinate" relationship with respondent B. Respondent B identified a "communicate" relationship with Respondent A. The binary dyadic relationship A-B was included in the Coexist, Communicate, and Coordinate graphs, but not in the Collaborate graph. The dyad was included in the Coexist graph because coexisting is assumed as part of both coordinating and communicating (the Spectrum of Collaboration is cumulative).

Dealing with Missing Data

The relationship survey had a response rate of just 54%, with only 54 responses out of 100 possible responses. However, because of the how social network analysis is constructed, this means that we had complete information (both individuals in the relationship reporting on the relationship) for just 28% of the relationships in the network. If we include partial information, we had data for 79% of the network relationships. (See Figure 2 for visualization. See Appendix B for an explanation of calculations.)

Figure 2

Matrix of Missing Data



Note. The pattern of missing data is shown in white. Non-respondents are indicated by rows which are entirely white. The diagonal missing data is intentional, as individuals cannot have relationships with themselves.

Missing data is a significant issue in network analysis, impacting many of the statistics commonly used to describe networks (Stork & Richards, 1992; Huisman, 2009; Huang et al., 2019). Methods for dealing with missing data are still debated and developing, but boil down to three main categories: using only complete cases (those in which both halves of the dyad responded), using all available data (those in which either half of the dyad responded), or imputing missing data (Huang et al., 2019). Stork & Richards (1992) and Huisman (2014) both suggested using reconstruction, a form of "available data" analysis (which in our data would be using the partial information that we have to complete the reciprocal missing dyads). Reconstruction is only viable when interrater reliability is high, though (Stork & Richards, 1992). We have already seen that reliability was not high in our network data.

Huang et al. (2019) found reconstruction increases bias and instead supported unconditional mean imputation. Unconditional mean imputation can be computed three ways: using the overall density of the network; using the average of incoming ties; or using the average of outgoing ties (Huisman, 2009). Žnidaršič et al. (2017) supported imputing with the median of the k-nearest neighbors, though acknowledged that unconditional mean imputation of incoming ties is a viable alternative in most cases. Of these authors, the only study that dealt with valued networks like ours (in which coexist through collaboration were treated as ordinal values) was Žnidaršič et al. (2017). Therefore, I used their method of imputing the value based on the k-nearest neighbors to impute the missing data. I computed several values of k (three, five, and seven). None of the k-values were sufficient for imputing all the missing data, but k of five successfully imputed the most. For the remainder, I used the mean of the incoming ties.

Missing Data Methodology Caveats. One component of this project was a publicly available visualization with named nodes. The client and I notified respondents before participating that this network visualization would exist, and we gave them a link to the visualization for their reference. For this visualization alone, we used only available data. Using available data only circumvents the issue of trying to explain imputation to respondents and is an accepted practice within the SNA literature. Imputation, which was used for the private analysis, gave us a richer and more complete analysis, but the complexities outweighed the benefits for the public visualization.

Additionally, remember that we conducted two surveys – one that collected information about the people in the network and one that collected information about their relationships. Pederson included 27 individuals in the relationship survey that did not respond to the personal attributes survey. The missing personal attribute data could not be reliably imputed. Having that much missing data would have negatively impacted the models if those individuals were included in ERGM. As a workaround, I created a sixth network (Collaborate Complete) to visualize and assess the differences between the network with and without the non-responders. The individuals with missing personal attribute data were not highly central actors. Therefore, the Collaborate Complete network was used for the modeling component of the project.

Weighted or Unweighted Network

Networks can be either weighted or unweighted. In our data collection, we assigned an ordinal variable to each of the 5 levels of connectedness (unknown = 1, coexist = 2, communicate = 3, coordinate = 4, collaborate = 5). These values could be considered weights of a sort. However, while we know that relationship strengthens as we move through the ordinal scale, it is unlikely that each step is evenly distributed. Moving from coexisting to communicating may represent a smaller step than moving from coordinating to collaborating, for instance. While I could analyze the network without weights, disregarding the strength of the relationship entirely also discards important information in the lower-level networks. If we discarded weights, in the lower level Coexist network, a coexist-only relationship would seem identical to a collaboration relationship (which assumes coexistence is also occurring). The loss of this important information outweighed the concerns about the unevenness of steps, so weights were retained and utilized in those network statistics where weights are relevant.

When retaining weights, a choice needed to be made about how to handle simplifying two weights to one when we go from the complete matrix to the undirected network. We had already seen that we lack interrater reliability. Given that we had no additional information on which to base a decision, we had two viable choices. We could have randomly sampled from the 2 weights, or we could have summed the two weights (Csardi & Nepusz, 2006). Arguments could be made for either approach but summing provides the most information. For instance, if two respondents agreed that they collaborate, the weight attached to that relationship would be ten. If one reported collaboration while the other reported coordination, the weight attached to that relationship would be nine, indicating less confidence in the strength of the relationship, considering the assessment of it was not mutual. Summing was performed after the overall network was divided into the four component networks, so no strength information was lost prior to the simplification of networks.

Dummy Encoding

Many of the questions included in the personal attributes survey were categorical questions. Some of those categorical questions allowed for multiple answers. For both the single and multiple-choice categorical questions, using dummy encoding simplifies interpretation of exponential random graph models (ERGMs). EGRMs, like logistic regression, use a reference category when a categorical term is added to the model. The reference category is not explicitly included in the model, but instead is interpreted as the base state, from which all other categories vary (Luke 2015). Imagine a categorical question with three possible answers: red, blue, green. If this question were included in an ERGM as a categorical variable, it would result in two terms – one for blue and one for green. If the blue coefficient were positive, the interpretation of that term would be that individuals who selected blue were more likely to form ties than individuals who selected red (the reference variable). If the reference variable is meaningful as a point of comparison, this makes sense. However, if the researcher would prefer to be able to say that the model indicated that people who choose blue were more likely to form ties than people who did not choose blue, then dummy encoding the variables is appropriate. Dummy encoding creates a new variable for each category within a categorical variable and assigns a one if the respondent selected that category and a zero if the respondent did not select that category. The zero, or "off" state, becomes the new reference category. For this analysis, all categorical variables were dummy encoded.

Analysis

Network analysis typically follows the same overarching structure of most data analysis. It begins with exploration of the data, both through examination of numbers and through visualizations. Along the way, researchers develop testable hypotheses. The final step is testing those hypotheses through statistical modeling. My exploration phase encompassed several approaches – the Five Number Summary, centrality exploration, important actor visualization, and community detection. I used exponential random graph modeling to test the hypotheses.

The Five Number Summary

Most social network analysis begins with the examination of some whole network statistics. These are quantitative values which describe an aspect of the entire network, as opposed to aspects of specific sub-groups, nodes, or edges. Luke (2015) referred to these as the "Five Number Summary" (p. 11). The five numbers include: size, density, number of components, diameter, and clustering coefficient. The Five Number Summary is a bit of a misnomer, though, as several of the five should be expressed using multiple numbers. Perhaps a better name would be the five categories of descriptive statistics. Luke (2015) detailed each category.

Size. The size of the network primarily refers to the number of nodes in the network (Luke, 2015). The size is important because many network statistics are size-dependent. Density of a small-sized network is not equivalent to density in a large-sized network, for instance. For example, if someone lives in a small town, it is easy for them to know everyone. Their network size is small, and the density is high. However, if that same person lives in a large city, they may know just as many people, but it is a fraction of the overall population, leading to a much lower density. Understanding the difference size makes matters.

In addition to the number of nodes, my definition of size also included the number of edges (also called ties). Edges in an undirected network indicate any tie between two nodes.

Density. Density is the ratio of the number of observed ties compared to the number of possible ties (Luke, 2015). It is a number that varies between zero (no ties observed) and one (all ties observed). Density is a measure of connectedness among the individuals in our network (Freedman & Bess, 2011). While density should not be used as a comparison for different-sized networks, we can use it to compare equal-sized networks. I used density to compare the connectedness across our five sub-networks, each of which have the same number of nodes.

Number of Components. Components are sub-groups within the network in which all the nodes are connected to each other, either directly or indirectly (Luke, 2015). When an analyzed network has multiple components, it means that there are multiple sub-networks within the overall network with no connection between them. I compared the number of components across our five networks. I also included the size of the largest component and the total number of isolates. Isolates are nodes which are completely unconnected from all other nodes. These statistics again give us some sense of how well-connected the network is. A network with many components is lacking important connectors. Likewise, a network with many isolates tells us that there are people that are considered a part of the network, but who are not actually engaged with the network in meaningful ways.

Diameter. Diameter is a measure of compactness of the network (Luke, 2015). To move from one node to another, whatever is moving through the network must follow a path made of ties or edges. For any two nodes in a network, the total number of ties that must be followed to traverse between them is called a path. Each pair of nodes may have multiple routes (or paths) between them. The shortest path is the most direct route between two nodes, and is known as the geodesic (Freeman, 1978). To calculate diameter, researchers first calculate the geodesic between every pair of nodes in the network. The length of longest geodesic is the diameter of the network (Luke, 2015). From a practical perspective, if someone in the network wanted to disseminate information throughout a network of people by giving the information to one person and asking them to pass it on, the diameter is the number of people that the information would need to pass through to get to the most distantly connected member of the network. A network with a high diameter may have difficulty accurately disseminating information or transferring knowledge (Stoltenberg et al., 2019).

Clustering Coefficient. Finally, the clustering coefficient is also referred to as transitivity. Transitivity quantifies the tendency for people to form closed triangles – situations in which a friend of my friend is also a friend. Measured as a ratio between open and closed triangles, transitivity varies between zero and one (Luke, 2015). Transitivity is another indicator of the level of cohesion in a network, with higher clustering coefficients indicating a more tightly connected network (Giuliani & Pietrobelli, 2011).

Simulation Comparisons

Statistics about an observed network, without context, can be misleading. It can be difficult to interpret when, for example, transitivity is high or low, particularly if the researcher is unfamiliar with network analysis. One approach to putting networks in context is to compare them to simulated networks of similar size and density. Researchers have identified several commonly occurring network types that can be simulated (Luke, 2015; Harris, 2013). The networks in this analysis were compared with three types of standard networks: random, small-world, and scale-free.

Random networks were one of the first mathematical models of networks but remain important still today (Luke, 2015). In this model, each edge appears in the network with a researcher-chosen probability. To mimic each observed network, I created random networks using the density of the observed networks as the probability of tie formation within the simulated networks. Truly random networks are not often found in the real world, but they can still be used as comparisons for observed networks.

Small-world networks tend to be more typical of those found in real social networks, particularly where transitivity is concerned (Luke, 2015). Small-world networks are simulated by "re-wiring" nodes – randomly removing a tie and using it to connect two random nodes, using a researcher-chosen probability and a researcher-chosen number of initial neighbors. Luke (2015), noted that it only takes a few ties being rewired to reduce the diameter of the network substantially. In the simulations for our networks, I used the density of the observed network as the probability for tie creation and the observed network's mean degree for the initial number of neighbors.

Scale-free networks tend to generate degree distributions that more closely reflect real world networks. For instance, it is common in a real world network to have one or two individuals that are much more well-known than everyone else, so it is typical to see a large number of people with a low degree of connectivity, and a gradual decrease across the degree distribution, resulting in a small number of people with a very high degree of connectivity. The tendency for a small number of nodes to collect a high number of relationships is sometimes called preferential attachment (Luke, 2015). Consider a middle school classroom. A student that is considered "popular" is one that has more relationships. As new students enter the picture, they gravitate towards the popular people. The new students prefer to attach to those that are already known and liked, because there is the assumption that being affiliated with those that have power and prestige will provide some advantage. There are two ways to simulate scale-free networks – via a growing network or a non-growing network. The growing network approach iteratively adds nodes to the network, while the non-growing approach relies on a power law exponent for the degree distribution (Csardi & Nepusz, 2006). I generated both types of simulated scale-free networks, using the power of the observed network for the non-growing network and the degree distribution of the observed network to seed the growing network.

For each type of theoretical network, I generated 1000 simulated networks and statistics were compared to observed networks via boxplot distribution graphs. The statistics I compared included number of components, size of the largest component, number of isolates, diameter, and transitivity.

Centrality

Freeman's 1978 seminal work on network centrality noted that centrality can be calculated at the individual node level or at the level of the network. Centrality at the individual node level identifies which nodes hold positions of power within the network. Centrality at the network level describes the degree to which centralized powerful individuals exist within the network. A network where everyone is connected or the paths form a circle would have a low centrality score, whereas a star-shaped or linear network would have a high centrality score. For the purposes of our analysis, both individual and network centrality were measured using three types of centrality: degree, betweenness, and closeness. Networks or individuals with each type of centrality have benefits and limitations.

Degree Centrality. Degree centrality is the simplest type of centrality – referring to the total number of connections each node has (Luke, 2015). The idea is that a node with more ties is more powerful, as they have access to more information, resources, or whatever is being passed through the network. However, having too many ties can result in information overload, decreasing the node's functionality (Giuliani & Pietrobelli, 2011).

Closeness Centrality. Closeness centrality relies not on how many ties an individual node has, but on how short the geodesic distance is to all other nodes (Luke, 2015). An individual with high closeness can quickly send or receive information across the entire network, so these individuals are sometimes referred to as spreaders or sensors (Okamoto et al., n.d.). They can be key informants for a network organizer, as they often have their finger on the pulse of the network. Networks with high closeness have high trust and are good at joint problem-solving, but networks with too much closeness may lack innovation and get stuck in their ways (Giuliani & Pietrobelli, 2011). Unfortunately, there is no clear-cut line between when closeness is valuable and when it is detrimental. Monitoring closeness over time and pairing it with other evaluative measures of network success might be necessary to determine when a network has become too close.

Betweenness Centrality. Betweenness centrality is a measure of a node's position between other nodes, typically measured based on the number of times a node sits along other nodes' geodesics (Luke, 2015). An individual with a high betweenness centrality is important because they can be either a broker or a blocker between different groups. Consider a network with two clusters and one node that connects them. The connecting node would have a high betweenness score. To pass information from one of the clusters to the other, it would need to go through that node. Removal of a node with high betweenness could cause a disruption in the overall network functionality (Giuliani & Pietrobelli, 2011).

There is a special kind of broker that connects across nodes with differing affiliations. In our network, for instance, we may have brokers that have the role of value-added producer that bridges a gap between growers and wholesalers/distributors. Playing this special kind of brokering role has the potential to be beneficial for both the actor and the network (Giuliani & Pietrobelli, 2011). I reviewed actors with high betweenness scores to determine if they were playing this special kind of brokering role.

Important Actor Visualization

Visualizations of important actors provide information both about who the important actors are and where they are situated in the network. For each of the preceding types of centrality, I generated visualizations of who the important actors are in each network. This was done by first assigning every actor a centrality score. The score spread varied for each type of centrality. To normalize the data, each score was rescaled to a number between one and five, using the following rescaling algorithm, as recommended by Luke (2015):

> *rescaledVector = ((5-1) * (OriginalVector-MinimumofOriginalVector))/(MaximumofOriginalVector-MinimumofOriginalVector)+1*

In this algorithm "OriginalVector" was the vector of initial centrality scores. "Important" actors were identified as those actors which have a rescaled centrality metric equaling five (the highest value possible). To prevent information overload, the three types of centrality were first run through a principal components analysis to determine which of the three types of centrality provided the most information. In the findings, the single most important centrality visualization for each network is presented, though visualizations for all types of centrality are available in Appendix D.

One additional type of important actor was also visualized – cutpoint actors. These are the actors that, if removed from the network, would result in an increase in the number of components within the network (Luke, 2015). In other words, they are important actors in the network because they hold tenuous parts of the network together. It was important to identify these nodes because one role Extension educators might play is becoming additional bridges between these parts of the network, eliminating existing cutpoints and more tightly tying the network together.

Community Detection

Community detection is the practice of finding groups of nodes within a network that share a similar connectivity pattern, typically a pattern that is denser within the group than with other nodes outside the group (Geng et al., 2018; Wang et al., 2015; Giuliani & Pietrobelli, 2011; Luke, 2015). Community detection is used for a variety of purposes. For example, it can be used to identify cliques (groups of nodes which are entirely connected to one another which tend to have high trust and cooperation), to understand the overall network structure (such as differentiating between core-periphery, small world, and scale free structures), and to identify areas of the network that may be more prone to fragmentation (Giuliani & Pietrobelli, 2011). Community detection algorithms have proliferated in recent years (Wang et al., 2015) and articulating all the possible approaches to community detection is outside the scope of this project. However, a variety of community detections approaches as described by Luke (2015) were performed. These included clique detection, k-core detection, and automated community detection using the following algorithms: Edge Betweenness, Fast Greedy, Louvain, and Infomap. I chose these algorithms because they are appropriate for undirected, weighted networks (Luke, 2015).

Exponential Random Graph Modeling (ERGM)

In most statistical modeling, there is an underlying assumption that the observations are independent of one another. However, in networks, that assumption is violated. Network observations are explicitly dependent on one another. If actor A has a relationship with actor B, Actor B automatically has a relationship with Actor A. The two observations are dependent on one another. Network researchers needed an alternative way to conduct statistical tests of network data. ERGM fulfills that need. ERGM can do many things – estimation, simulation, and assessing goodness of fit. Hunter et al. (2018) stated that ERGM's main purpose is to simply describe the local factors that contribute to the overall structure of the network. Pederson was particularly interested in what factors might contribute to individuals forming collaborative ties in the Chequamegon Bay Regional Food System. ERGM has the capacity to determine those factors.

I used the implementation of ERGM found in the Statnet software package in R (Handcock et al., 2019) and applied it to the variation of the collaboration network that had the complete attribute data.

The general process of modeling involves conducting exploratory analysis, making hypotheses, and testing those hypotheses by building a model. The model includes "terms" which are variables added to the model. With ERGM, there is an infinite number of term combinations that could be tried iteratively, making modeling with ERGM quite complicated (Goodreau, 2007). Nevertheless, Luke (2015) and Harris (2103) both suggest variations on the following iterative process for selecting terms:

1. Start with a null model. A null model is one that uses the observed network as the dependent variable and the edges as the independent variable. Starting with the null model gives the researcher a basis for comparison for all other models that are explored.
2. Consider node-based terms. These are terms that contribute to the model when either one or both nodes in a tied pair have that attribute.

3. Consider dyadic terms. These are terms that contribute to the model when either both dyad partners have the same attribute, both do not have the same attribute, or there is a combination of one having it and the other not having it. The important factor here is that this category of terms looks at the importance of the attribute across pairs of nodes, as opposed to for an individual node.
4. Consider relational terms. These are terms that describe the relationship itself. (The only relational variable we tracked in our network was the geographic distance between two people.)
5. Consider structural terms. Structural terms add great complexity to the model, but they are often required to truly represent the non-linear nature of real-world networks (Harris, 2013). Structural terms can represent things such as the transitivity of the network, the degree distribution of the network, or the edgewise shared partners in the network.

I followed this general practice by iteratively adding terms to the initial null model based on both Pederson's initial hypotheses and the hypotheses that were generated during exploratory data analysis. I started with the simplest node-based terms and progressed through to structural terms.

Like other modeling practices, all the statistics included in an ERGM model should be statistically significant. As I iteratively added parameters to the model, I tested for statistical significance and removed those that had a p-value greater than .05. This can be complicated and can result in cascade effects, because ERGMs often require specific combinations of terms to converge. Insertion or deletion of one term can have a profound impact on the model (Goodreau, 2007). Interim models that included non-significant terms were not stored.

ERGMs also can have difficulty converging (Harris, 2013; Goodreau, 2007). A model that does not converge will have no coefficients – no information of any use to a researcher trying to understand the network. Models that did not converge were not considered.

I compared models that converged and that had statistically significant terms based on their Akaike information criterion (AIC). The AIC is a useful statistic to compare models that share the same base null model. When comparing models, a lower AIC indicates that the model is a statistically significant improvement on the prior model (Luke, 2015). In all, 25 models were worthy of AIC comparison. Of those, one model had the lowest AIC score.

Finally, I ran goodness of fit tests on the model with the lowest AIC score. Goodness of fit is built in to the ERGM package. It is performed by generating simulated networks based on the model and comparing the simulated networks statistics to the observed network. A well-fit model produces simulations that are largely in line with the observed network across most statistics (Goodreau, 2007).

If a model is well-fitting, we can make assumptions and predictions based on the network (Luke, 2015). For instance, the model can be used to predict the odds that a new member of the network will form a collaborative tie, based on some attribute of that new member. It can also be used to predict the likelihood that two members of the network will form a tie. The client was not interested in predicting the likelihood that two members of the network would form a tie but did request an in-depth explanation of the odds ratios for each term.

ERGMs are complicated models. There is no guarantee that a well-fitting model can be generated. Even so, they are still one of the best approaches to network modeling, and they do the best job of reflecting the complexities of real-world networks. Given that Pederson was interested in what personal attributes might contribute to tie formation, an ERGM model was the best method for answering that question.

Data Collection/Methodology Final Thoughts

While the data collected for this project has some limitations, notably low response rate and low interrater reliability, valid methods exist to impute and clean the data. An exploratory data analysis coupled with ERGM, such as I have described, can provide both breadth and depth of information to Pederson and her colleagues. These methods are scientifically sound, accepted by researchers, and cited often in the literature.

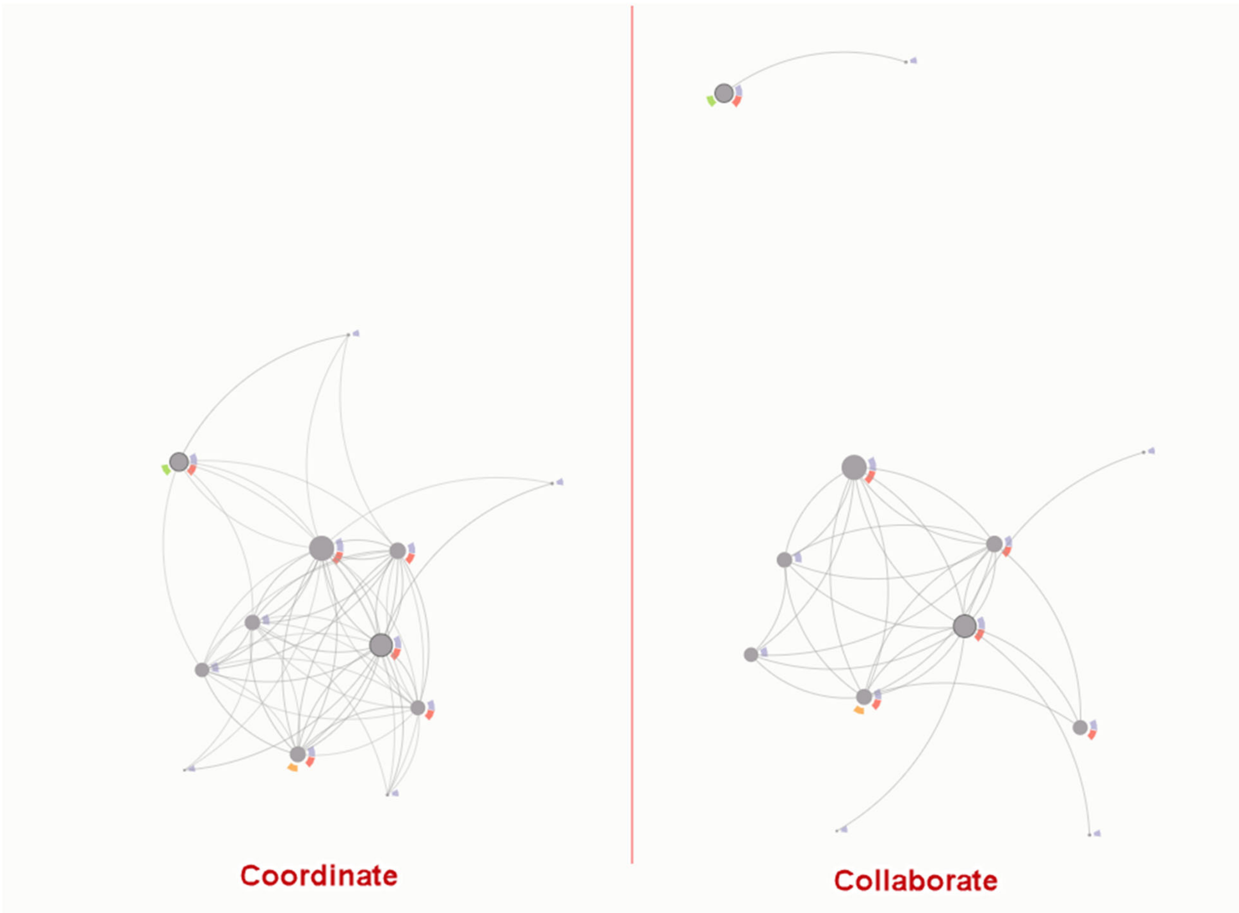
Chapter 4: Findings/Results

The Visualization

The first objective of this project was to create a visualization of the network. An interactive, publicly available visualization allows the clients and network members to explore their relationships. Through filtering by role, group affiliation, location, and even individual network member, the clients can explore the relationships within and across the networks. For example, one area of interest is food systems educators. Figure 3 shows the difference between the relationships among food systems educators in the Coordinate and Collaborate networks. Notably, the sub-network of food systems educators breaks into two components when moving from Coordinate to Collaborate, highlighting that two of the food systems educators lack collaborative relationships with the main cohort of food systems educators.

Figure 3

Visualization Comparison of Food Systems Educators



Many findings are discoverable via the exploration of the interactive visualization, but the interactive visualization is based on available data only, which limits its overall usefulness. Therefore, while I encourage the clients and the reader to explore, the remainder of the findings in this chapter focus on the more complete, imputed data set.

Five Number Summary and Comparison to Simulated Networks

Table 1 contains the Five Number Summary results for each of the networks and fulfills the second objective of the project. The “All Possible Connections” network is included for comparison’s sake. It is a simulated network including all possible edges, resulting in a density of one with a single component, no isolates, a diameter of six, and a transitivity of one. Each of the other networks can be interpreted in relation to that complete network.

Table 1

Five Number Summary

Network	Edges	Vertices	Density	Number of Components	Size of Largest Component	Number of Isolates	Diameter	Transitivity
All Possible Connections	4950	100	1.00	1	100	0	6	1.00
Coexist	4402	100	0.89	1	100	0	6	0.93
Communicate	2036	100	0.41	1	100	0	10	0.56
Coordinate	692	100	0.14	1	100	0	17	0.37
Collaborate	376	100	0.08	12	89	11	30	0.38
Collaborate Complete	275	73	0.10	7	67	6	30	0.42

The research shows that as relationships move to higher levels of the Spectrum of Collaboration, density and transitivity decrease while diameter increases. This is not a surprising or unexpected result. A coexist relationship requires little effort, while a collaboration relationship requires considerable effort. As one research participant stated, “just because I know them, doesn't mean I do anything with them” (personal communication, anonymous participant, May 17, 2020). Given this, we expected to see far fewer collaborative relationships than coexisting relationships.

Density

Notable, though, is the high density in the Coexist network. It is rare in real-world networks to see a density over .5 (Harris, 2014), so seeing a density of .89 is unexpected. Pederson was concerned that it appeared that many people in the network did not know each other, but the research results do not support that concern. Most people are at least aware of one another. In fact, there are only 548 coexist relationships that do not exist out of 4950 possible relationships, suggesting a very tightly connected network. The density for the Communicate network is also quite high, at .41. This suggests that just under half of the possible communication relationships exist.

Density drops off precipitously for the Coordinate and Collaborate network. Each of these networks has a relatively low density. While not unexpected, each of these networks could benefit from an increase in density. Higher density increases the efficiency of information flow in a network (Yamaguchi, 2014) and rapid and accurate information flow increases network success.

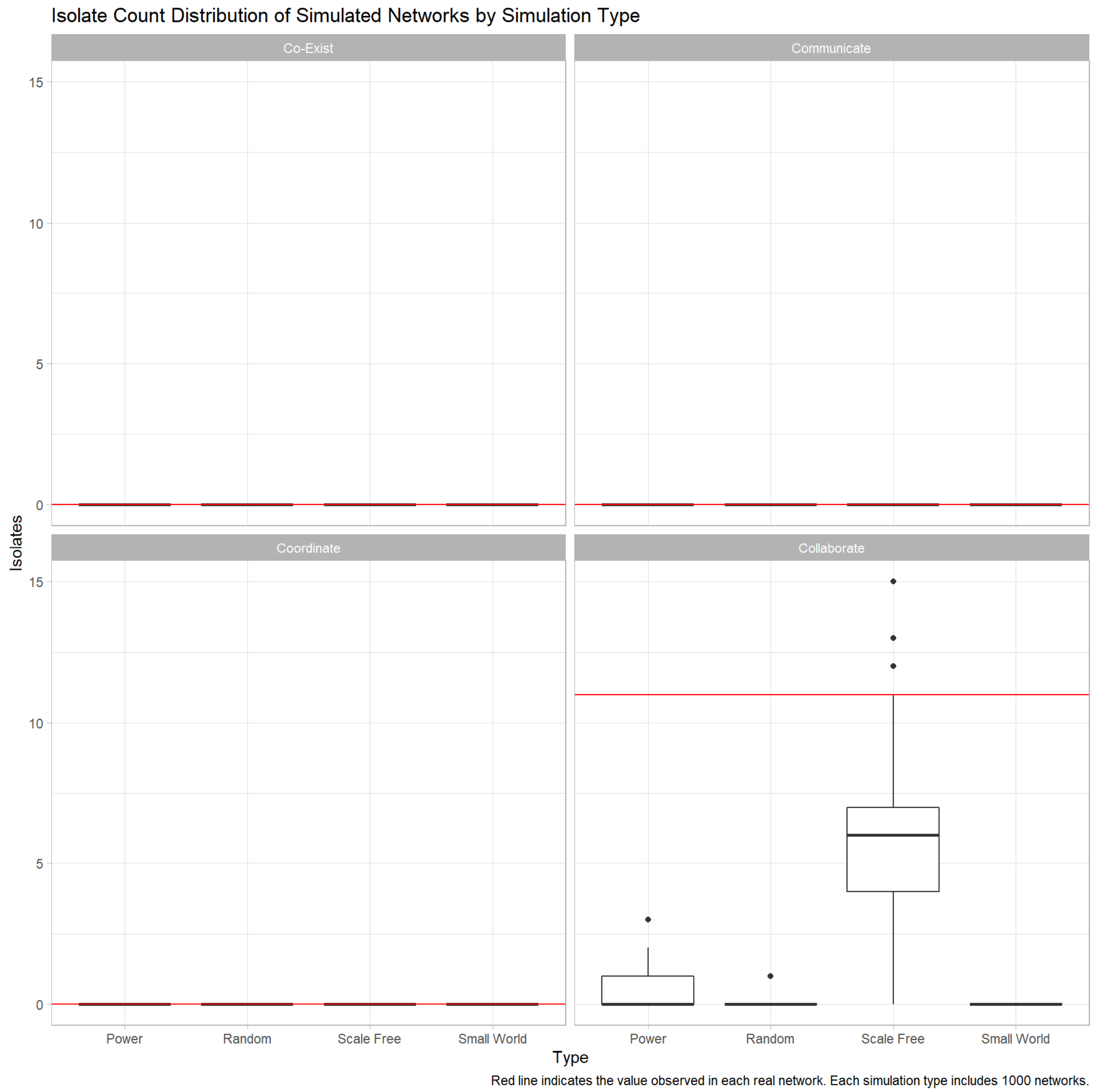
Isolates/Component Counts/Component Size

Isolates, component counts, and component sizes are all tightly connected statistics. The research shows that in our observed network, everyone is either in the main component or an isolate. The simulations for component counts and size of largest component echoes that finding. The number of components in the simulated networks is consistently equal to one "giant" component plus the total number of isolates. The size of the largest component consistently equates to all nodes except isolates. For all but the Collaborate network, our observed network and all simulated networks have a single component with size 100 (indicating no isolates). This single component means that it is possible to reach all people within each network by starting with any single person in the network, which is beneficial for knowledge transfer, technical assistance, and emotional support of network members. In the Collaborate network, those that are connected enjoy that same information flow, while those that are not part of the main component are without access to any of the benefits of collaborative relationships.

There are 11 people in the Collaborate network that lack a collaborative relationship. While having zero isolates for the first three networks is not an unusual finding and matches the findings found in simulated networks, a network with 11 isolates is an unusual finding and would be considered an outlier even for scale-free networks (see Figure 3). These isolates are potentially at risk of leaving the network unless new relationships are fostered. One caveat to that concern is that these isolates do all have relationships of other types within the broader network structures. Each of the isolates has at least one coordinate relationship, at least 7 communicate relationships, and at least 19 coexist relationships. A list of isolates and their alternate relationship counts is included in Appendix C.

Figure 4

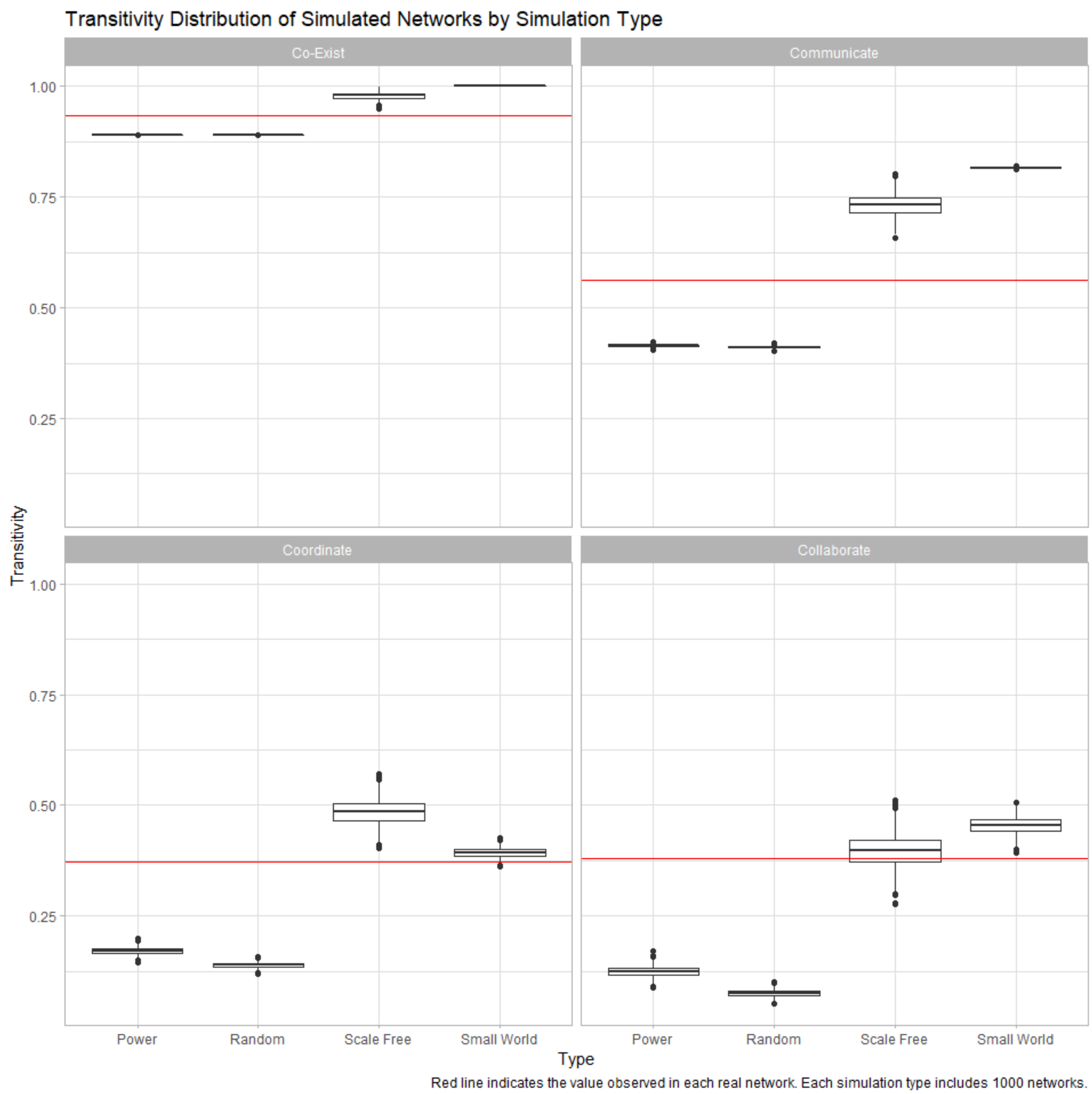
Comparison of Isolate Counts Across Simulated Networks



Note. The red line indicates the observed isolate count. The Collaborate network has far more observed isolates than expected.

Figure 5

Comparison of Transitivity Across Simulated Networks



Note. The red line indicates observed transitivity in each network.

Transitivity

Figure 4 shows the transitivity distribution found in the simulated networks, with the red line indicating the transitivity found in each of our observed networks. A transitivity of .38, as found in the Collaborate network, means that 38% of all triangles that could exist in the network do exist. Interestingly, transitivity is slightly higher in the Collaborate network than in the Coordinate network. This suggests that it is more likely for two individuals that collaborate with the same person to also collaborate with each other than it is for two individuals who coordinate with the same person to also coordinate with each other.

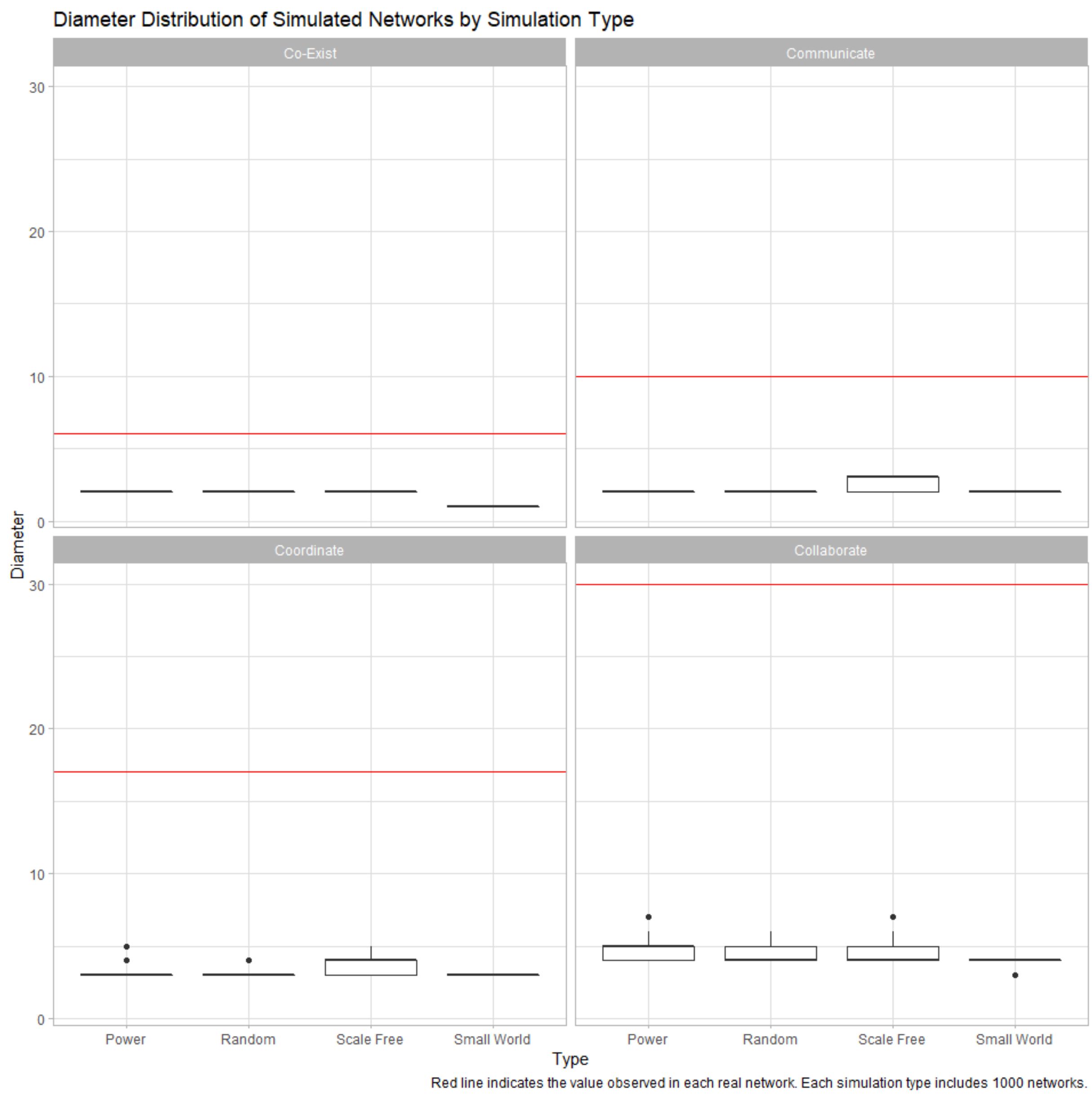
Transitivity for all four networks is consistently higher than the transitivity typically seen in a power or random network of similar size and lower than that typically seen in a scale-free or small-world network of similar size. The former is expected. Transitivity increases quickly in most real-world networks, typically much faster than in random networks. The latter may be an artifact of the missing data imputation, as I would have expected the transitivity to be quite like the small-world or scale-free simulations for both our Coordinate and Collaborate networks.

Diameter

Our most surprising finding from the Five Number Summary is diameter. Figure 5 shows the diameter distribution for our simulated networks, with the observed diameter for each network type in red. Across all four networks, observed diameter exceeds simulated diameter. The diameter in the observed Collaborate network is approximately six times larger than what we would have expected to see, based on simulations. A network with a high diameter indicates that members of the network are more difficult to reach (Stoltenberg et al., 2019), and high diameter has been shown to have a negative impact on a company's ability to innovate (Schilling & Phelps, 2007).

Figure 6

Comparison of Diameter Across Simulated Networks



Note. Red line indicates diameter seen in observed networks, which is higher than expected across all simulation types

Network Centrality

Table 2 shows the network centrality statistics for each network and type of centralization.

Table 2

Network Centrality Statistics for Each Network

Network	Degree Centralization	Betweenness Centralization	Closeness Centralization
All Possible Connections	0.00	0.00	0.00
Coexist	0.10	0.00	0.16
Communicate	0.51	0.04	0.57
Coordinate	0.53	0.19	0.51
Collaborate	0.44	0.28	0.02
Collaborate Complete	0.45	0.28	0.04

Remember that centrality refers to the tendency for actors within the network to hold positions of power, and it is measured as a decimal between zero and one. A network with all possible connections will always have centrality measurements of zero, because no individual holds a position with more power than any other (all are equal). A very dense network, such as we see in the Coexist network, will also often have low centrality.

We begin to see centrality as an important factor when we reach the Communicate network. This network has large centralization statistics for both degree and closeness centrality, though it has very low centrality for betweenness. The low betweenness suggests that there are few structural holes – places where bridging is required to connect different parts of the network. The high closeness centrality indicates that many people in the Communicate network can reach other people via few intermediaries. This network has many “information spreaders” – fitting for a network of communicators.

The Coordinate network has similar measures of degree and closeness centrality as the Communicate network. This network, however, has a higher betweenness metric, suggesting that there are parts of the network that are more loosely coupled than the lower-level networks.

The Collaborate network has extremely low closeness, as expected given the diameter of 30 we saw earlier. This network lacks people that hold key information spreading positions. Also notable about this network is that degree centralization is .45 – slightly lower than the Coordinate or Communicate networks. You could think of this network as slightly more egalitarian, with fewer people holding positions of power. The one exception is betweenness. The Collaborate network has the highest betweenness centrality, suggesting that there are key people that either play brokering or blocking roles between different parts of the network, and that the network likely has some structural holes.

Important Actors

The third objective of the project was to determine who the important actors are in the networks. I accomplished this through identifying centrality of actors and cutpoint actors.

Centrality Importance

Who are these people that hold positions of power? Principal component analysis revealed that degree centralization is the type of centralization that provides the most information in the Coexist and Coordinate networks, closeness is the one that provides the most information for the Communicate network and betweenness is the one that provides the most information for the Collaborate network. Figure 6 identifies the most important actors in each network.

Figure 7

Important Actors for Each Network

Note. Individuals of highest importance are indicated in red with a named node.

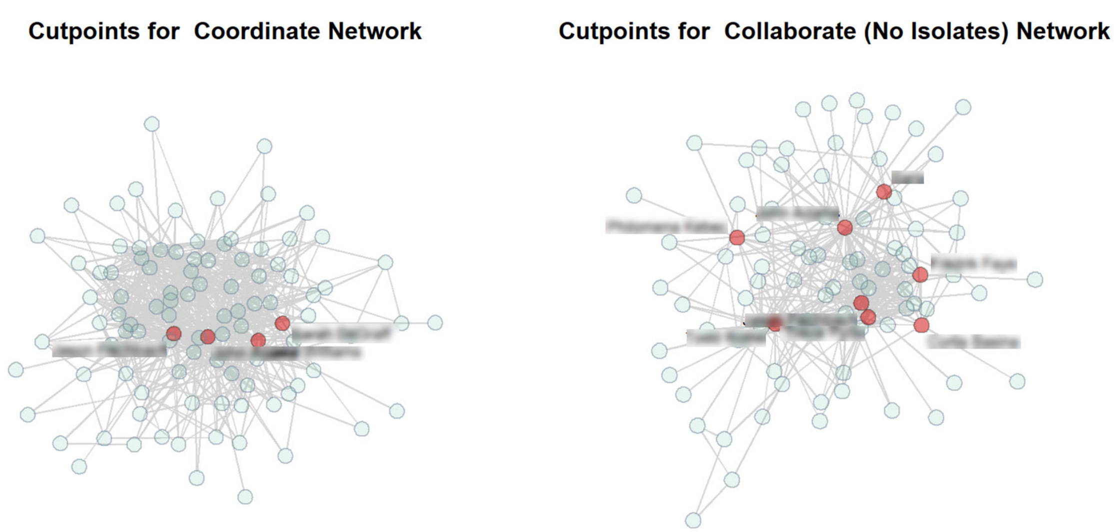
The Coexist network has multiple individuals that are all deemed important. This is common in a network with low overall centrality, so this is not an unexpected result. When centrality is shared by multiple people, the network itself is not highly centralized. The other networks each have only a single individual that dominates the centrality metric that provides the most information. This is also in line with the higher overall network centrality scores for those networks. Notable is that one of the Extension educators, holds the highest position for closeness in the Communicate network. He could be a key information spreader or key sensor in the network due to having short paths to reach many of the other network members. A colloquial way of expressing it would be that he probably has “his finger on the pulse of the network” – an important role for Extension to play.

Cutpoint Importance

As a reminder, cutpoints are individuals who, if removed from the network, would disconnect parts of the network, causing more components than were previously in the network (Luke, 2015). There are no cutpoints in either the Coexist or Communicate networks. This is not surprising given the density of those two networks. Both the Coordinate and Collaborate networks have multiple cutpoints (shown in Figure 7). Notable is that both networks have Extension educators as cutpoints. There is a high number of cutpoints in the collaborate network, suggesting that the network may have some fragility.

Figure 8

Cutpoints for the Coordinate and Collaborate Networks



Note. Cutpoints for each network are nodes colored red and identified by named label.

Community Detection

The fourth objective for the project was to determine if sub-communities could be identified, and if so, around what factors they formed. After running many different manual and automated community detection algorithms (available in Appendix D), only two approaches provided much in the way of findings – cliques and k-cores.

Cliques

Cliques are groups of individuals that are completely connected to one another (Luke, 2015). In the Collaborate network, there are three cliques that each contain 10 people. The three cliques encompass only 14 unique individuals, though, meaning that there is a lot of overlap within these three cliques. Further analysis of the individuals in these cliques could reveal information about what draws people into collaborations. A complete list of people in the three cliques is available in the rendered code in Appendix D.

K-cores

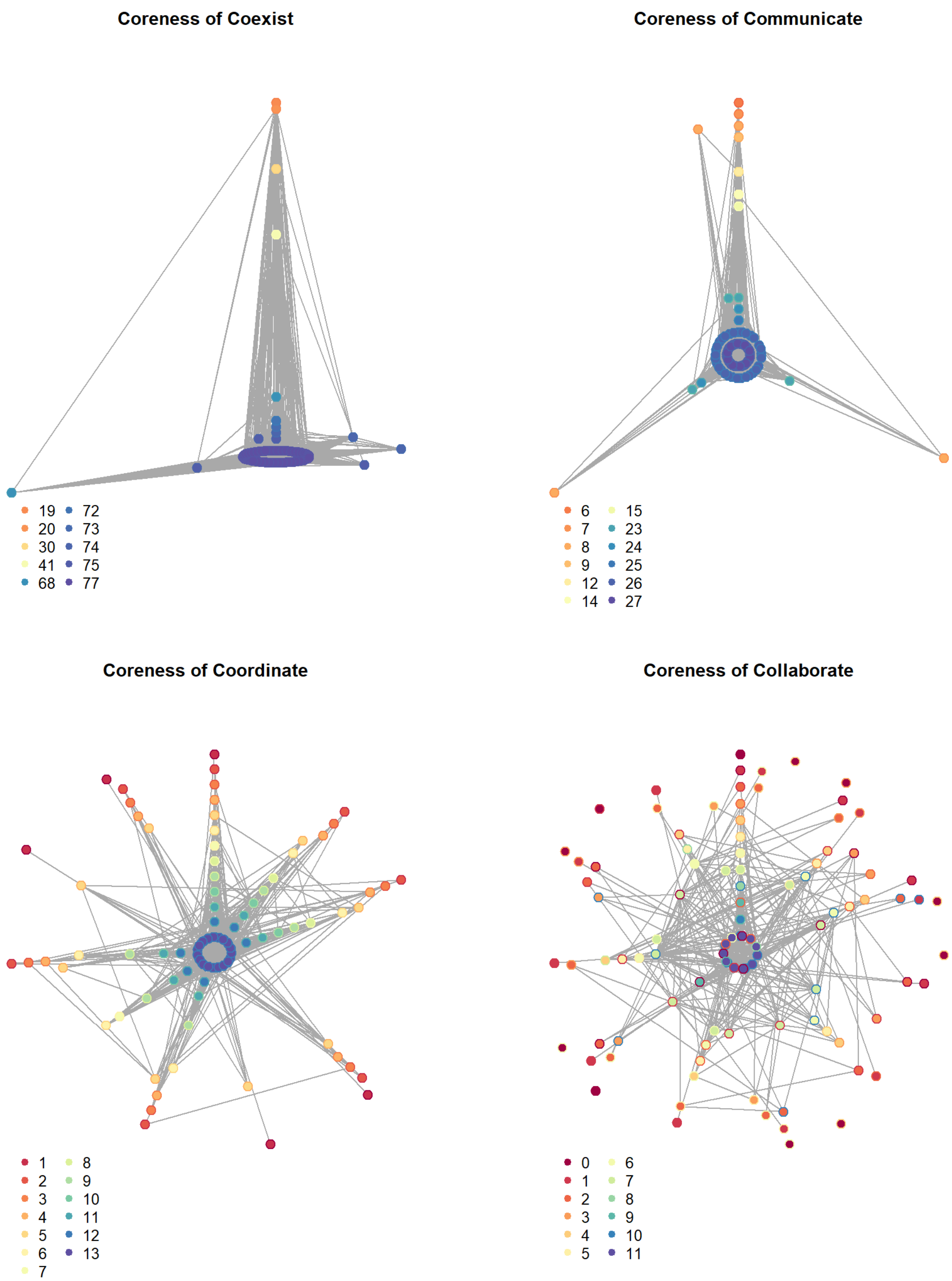
K-cores identify subgraphs based on degree, identifying the sets of actors who all have at least “k” degrees (Giuliani & Pietrobelli, 2011). K-core visualizations show nodes colored by their k-core community (each node that shares the same number of connections is shown in the same color). Additionally, networks that have more individuals in high-degree relationships will have a target-like appearance, with rings of k-cores. Networks with more “points” – lines connecting nodes far from the central ring – have more low-degree individuals.

The visualizations of the network’s k-cores (shown in Figure 8) strengthen the findings that we have already seen in our network centralization statistics – the Coexist network has many actors with high degree, therefore little overall network centralization, while higher-order networks have more centralization, due to fewer actors with high-degree. The visualizations also make obvious that only a few people have a low number of coexist relationships, as evidenced by only a few “points” in that network graph. Each network past Coexist has more “points” to the network and less of a centralized ring of high-degree individuals. By the time we reach the Collaborate network, there are 14 people that only have a single collaborative relationship and 15 people that have only two collaborative relationships, and the highest degree found in the Collaborate network is 11 connections (15 people form this k-core), compared to 85 connections in the Coexist network (shared by 77 people).

Perhaps more important than what I did find is what I did not find. There are no node attributes by which we can identify a sub-community. This indicates that none of the data we collected about individuals in the network represents something around which people are naturally choosing to form ties. For instance, one of Pederson’s hypotheses was that people who are part of the same organizations or have the same roles might be more apt to form ties with one another. Community detection based on group membership did not identify this to be the case, though. The highest modularity score – a metric used to identify how cohesive a sub-group is (Luke, 2015) – was .07 for affiliation with Bayfield Foods in the Collaborate Complete network. This is very low modularity. The lack of sub-community detection by node attribute may be due in part to the fact that many people in the network belong to more than one group or hold more than one role. Or it may be because we did not ask the right questions to surface the attributes around which individuals naturally cluster.

Figure 9

K-Cores of Each Network



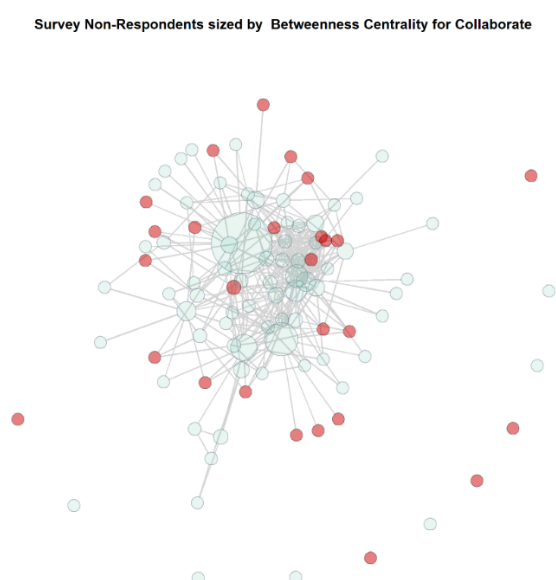
Note. *Network k-cores range from 10 k-cores (Coexist) to 13 k-cores (Coordinate). Nodes are colored by the number of degrees (the k-core). The lowest-degree k-cores are the isolates, with 0 connections. The highest degree k-core are seen in the Coexist network, where 85 people have 77 connections each.*

Exponential Random Graph Modeling

The fifth and final objective for the project was to identify factors that were predictive of strong network bonding. Exponential Random Graph Modeling (ERGM) was utilized to meet this objective. Before diving into ERGM findings, it is important to remind the reader that when conducting ERGM, I removed the 27 non-respondents from the collaborate network. Recall that betweenness was the centrality measure that provided the most information about the important actors in the network. Figure 9 identifies the 27 non-respondents that were removed from the network prior to ERGM analysis, sized by betweenness. While it is clear from the visualization that none of these network members held particularly “important” positions in the network (none of the red nodes are very large or central), it is still wise to remember that the model may be biased due to this exclusion of non-responders.

Figure 10

Non-Respondents Sized by Betweenness in the Collaborate Network

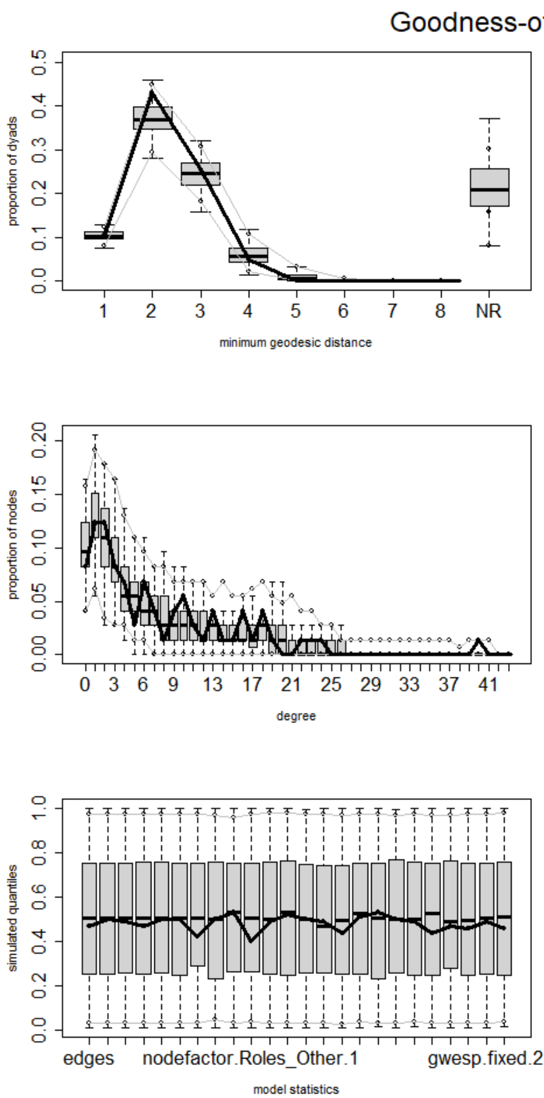


Note. *Non-respondents that were removed prior to ERGM are identified in red.*

Goodness of Fit

Figure 11

Collaborate Network ERGM Model Goodness of Fit



Note. Goodness of Fit visualizations for our final model. Note that the black line should pass roughly through the center of the grey boxes, which represent the middle quantile of the distributions found in model simulations. These visualizations signify good fit for geodesic distance, model statistics, and triad census, and acceptable fit for degree and edgewise shared partners.

Before interpreting any model, it is important to verify the goodness of fit. Our final model is within acceptable parameters for all goodness of fit measures. Figure 11 shows a visual representation of the goodness of fit of the model. In a perfectly fit model, the black line in each graph would pass roughly through the center of each of the grey boxes that denote the distributions found when simulating networks based on the model. However, an acceptable fit is one in which the black line stays within the confidence bounds (the light grey/dotted lines on either side of the boxes). We have very good fit for all the variables that were included in the model. (This is the sub-graph labeled “model statistics”.) We do not, however, have perfect fit for edgewise-shared partners. Despite including two edgewise shared partner terms in the model itself, our simulations can only roughly approximate the observed network. It is possible that this is an artifact of dropping the non-respondents, which resulted in a strange pattern of edgewise shared partners in our observed network. Or it is possible that there is some other factor influencing edgewise shared partners that we have yet to uncover because we did not ask the right questions. Either way, based on the simple goodness of fit visualizations and the complete diagnostics available in Appendix D, this model is acceptable to use, and we can go on to interpret the terms.

Model Terms

Table 3 identifies the name, coefficients, odds-ratios, and odds-ratio confidence intervals for each of the terms included in the final model. Some explanation of ERGM terms is required to understand the model. Common to each term is that they each have a statistically significant effect on the odds of a collaborative tie forming in the network, and each can only be interpreted individually when all other terms are held constant. Also common is that positive coefficients (odds-ratio more than one) have a positive effect on the likelihood to form collaborative ties, while negative coefficients (odds-ratios less than one) have a negative effect on the likelihood to form ties. Beyond that, each type of term is interpreted in a slightly different way.

Table 3

Terms in the Final ERGM Model

term	coefficient	conf.low	odds	conf.high
mix.Roles_FoodSystemsEducator.1.1	1.51	2.08	4.55	9.92
mix.GroupAffiliations_FarmToSchool.1.1	1.03	1.57	2.81	5.02
nodefactor.VendorMarkets_DuluthSuperior.1	1.00	1.88	2.73	3.96
mix.GroupAffiliations_BadRiverFoodSovereignty.1.1	0.92	1.44	2.50	4.34
nodefactor.SocialMediaFlag.Unknown	0.90	1.26	2.46	4.81
nodematch.GroupAffiliations_BayfieldFoods.1	0.72	1.50	2.05	2.79
esp11	0.71	1.30	2.04	3.21
nodefactor.GroupAffiliations_NorthlandCollegeprograms.1	0.59	1.44	1.80	2.26
nodefactor.GroupAffiliations_FarmBureau.1	0.50	1.13	1.65	2.41
nodefactor.VendorMarkets_Ashland.1	0.45	1.21	1.57	2.05
nodefactor.SocialMediaFlag.Yes	0.45	1.11	1.56	2.20
mix.Roles_FoodSystemsEducator.0.0	0.42	1.05	1.53	2.22
nodefactor.Roles_Other.1	0.35	1.13	1.42	1.79
nodefactor.CEA_StronglyAgree.1	0.28	1.09	1.33	1.61
gwesp.fixed.2	0.22	1.18	1.25	1.33
edgecov.dMatrix	-0.03	0.96	0.97	0.99
mix.GroupAffiliations_BadRiverFoodSovereignty.0.1	-0.43	0.48	0.65	0.90
nodefactor.PSD_Agree.1	-0.53	0.41	0.59	0.86
nodefactor.PSD_Disagree.1	-0.62	0.37	0.54	0.79
nodefactor.PSD_StronglyAgree.1	-0.63	0.35	0.53	0.80
nodefactor.Roles_NoneofThese.1	-0.69	0.27	0.50	0.94
nodefactor.PSD_Neutral.1	-0.80	0.30	0.45	0.67
nodefactor.Roles_MarketManager.1	-0.91	0.20	0.40	0.84
edges	-3.91	0.01	0.02	0.06

Node Factor Terms. Because all categorical variables are dummy-encoded, node factor terms can be interpreted as “if the individual has this attribute” or “if the individual answers this way.” For instance, the term “nodefactor.VendorMarkets_DuluthSuperior.1” means that if the individual vends at the Duluth/Superior farmer’s market, they are 2.73 times more likely to form a collaborative tie than someone that does not vend at the Duluth/Superior farmer’s market, all other terms held constant. All the node.factor group affiliations, roles, and market affiliation terms can be interpreted in this way.

The PSD and CEA node factor terms are slightly more complicated. PSD refers to the Personal Self Dependence question, “In my work, I rely on myself most of the time; I rarely rely on others.” CEA refers to the Collaborations are an Economic Advantage question, “Generally speaking, I believe that businesses that collaborate have an economic advantage.” These questions had a multi-level scale from Strongly Disagree to Strongly Agree. The “nodefactor.PSD_Agree.1” term means that if an individual answers “agree” to the Personal Self Dependence question the odds that they’ll form a collaborate tie are .59 times as large as the odds of someone that does not answer “agree,” all other terms held constant. In other words, people that answer yes to this question are 41% less likely to form collaborative ties than people that do not answer yes. Keep in mind, however, that “disagree,” “strongly agree,” and “neutral” all had statistically significant results. Having all these terms in the model, all with similar negative statistically significant effect is confusing. However, removing any of them resulted in a model that did not converge. The takeaway from this set of terms is, perhaps, that there is little consistency in how people answer it. The term “nodefactor.CEA_StronglyAgree.1” was the only level of that question that had statistical significance. People that answer “Strongly

Agree" are 1.33 times more likely to collaborate than people that did not answer "Strongly Agree" for that question.

Node Match Terms. Node match terms assess the likelihood that two people will collaborate if they match on some characteristic. The important subtlety of node match terms is that they are interpreted in relation to matching a different way, not in relation to not matching. For instance, the only node match term in the model is "GroupAffiliations_BayfieldFoods.1." This term is interpreted as two individuals who are both affiliated with Bayfield Foods are 2.05 times more likely to collaborate than two individuals who are not both affiliated with Bayfield Foods. This kind of matching is referred to as homophily. This term does not tell us anything about a pairing in which one person is affiliated with Bayfield Foods and one person is not (heterophily).

Node Mix Terms. Node mix terms take the next step to look at pairs of individuals both in terms of matching (homophily) and not matching (heterophily). The term with the greatest effect in our model is the term "mix.Roles_FoodSystemsEducator.1.1." This term can be interpreted as two individuals who are both food systems educators are 4.55 times more likely to collaborate with each other than a pair that includes one food systems educator and one person that is not a food systems educator. This term is interpreted in conjunction with the "mix.Roles_FoodSystemsEducator.0.0" term, which suggests that two individuals who are not food systems educators are 1.53 times more likely to collaborate with each other than a pair that includes one food systems educator and one person that is not a food systems educator. From a practical standpoint, these two terms taken in combination suggest that the food systems educators in the regional food network are not regularly forming collaborative relationships with non-food systems educators.

Edge Covariate Terms. Our model contains one edge covariate term, “edgecov.dMatrix.” This term refers to the geographic Euclidian distance (“as the crow flies”) between two individuals. This term suggests that for every additional mile apart two people are, the chance of collaborating is .97 times smaller. In other words, for every mile further apart two people are, they are about 3% less likely to collaborate. While this is a small percentage, it is important to remember that it is cumulative for each mile further apart. People one mile apart are 3% less likely to collaborate, while people 10 miles apart are 30% less likely to collaborate. The two people that are the farthest apart in our network are 96 miles apart, therefore 288% less likely to collaborate. (These two individuals do not know each other.)

Structural Terms. The model contains three structural terms, “edges, “gwesp.fixed.2” and “esp11.” “Edges” is a default term that simple starts the model with the same density seen in our observed network. Both other terms attempt to explain aspects of edge formation in relation to an edge that will complete a local clique or triangle. The “gwesp.fixed.2” term is a geometrically weighted edgewise shared partners term. An in-depth explanation of these terms can be found in at Stanford’s Short Course on Social Network Analysis (Jones et al., n.d.). Briefly, the term suggests ties that would close triangles are more likely than ties that would not. The “esp11” term suggests that ties that would create cliques of 11 edgewise partners are more likely than ties that would create cliques of other sizes. From a practical standpoint, these terms both suggest that if two people are already collaborating with the same person, it is likely they would also collaborate with each other.

Chapter 5: Recommendations/Next Steps

Pederson is particularly interested in how social network analysis can help her improve the Chequamegon Bay Regional Food Network. The analysis highlights several immediately actionable items, along with some next steps related to assessing change. Actions that could be taken immediately include engaging the periphery, reinforcing structural holes, and re-thinking the role of food systems educators. Future actions include setting benchmarks for change and establishing a re-analysis timeline.

Engaging the Periphery

In the Coordinate and Collaborate networks the analysis shows that density is low, and diameter is high. A solution to both these issues is to engage the periphery. Starting with the isolates in Appendix C, Pederson and her colleagues should facilitate collaborative relationship-building. After isolates, the next priority should be the individuals with a low number of higher-order relationships. These individuals are also listed, along with the total count of their relationships, in Appendix C. Engaging the periphery will de-centralize the network and decrease diameter, creating less reliance on the few central network members and creating more resilience within the network.

Reinforcing Structural Holes

Structural holes are parts of the network that are less well-connected than other parts of the network. Spanning structural holes has a positive impact on actors in the network because spanning a hole is related to developing more social resources and building non-redundant access to information (Lubell et al., 2014). Both individuals with high betweenness centrality and the presence of cutpoints are indicative of structural holes. Pederson and her colleagues can work to reinforce these structural holes by explicitly fostering redundant connections at cutpoints and by examining the individuals along the shortest paths of individuals with high betweenness scores.

Cutpoints

Examining each of the cutpoints, I found that in all cases removing a cutpoint created an additional isolate. While these are people that were already listed in Appendix C as individuals that would benefit from relationship redundancy, they are also listed separately in Appendix C with the cutpoint that would impact them, were the cutpoint to leave the network. If Pederson and her colleagues are concerned about any of these cutpoints leaving the network soon, then the individuals who would become isolates should take priority for fostering relationship-building.

Shortest Path Examination

People with a very high betweenness score are central to the network because they sit between other people that have weaker connections to the network. By examining the shortest paths that the people with high betweenness scores sit on, Pederson and her colleagues can identify individuals to target for cultivating additional relationships which bypass the person with high betweenness. This will close a structural hole. While this decreases the power and prestige of the person with high betweenness, it benefits the network by increasing resiliency. Appendix C contains the shortest paths for some key network actors with high betweenness scores. The person with the highest betweenness score in the Collaborate network sits on 52 shortest paths. For brevity, those network paths are included in the linked documents in Appendix D.

Re-Thinking the Role of Extension Educators

Currently the analysis shows that an Extension educator is very central to both the Coordinate and Collaborate networks. There are advantages to having an Extension educator playing a key role in these networks. It indicates that the community trusts and values Extension's contributions. Note, though, that Lubell et al. (2014) recommend that Extension educators transition out of these central roles. Specifically, they encourage Extension educators to instead focus on boundary-spanning partnerships and utilizing train-the-trainer approaches to facilitate knowledge transfer in such a way that Extension educators can encourage the growth of network actors that can become the new influencers and knowledge-spreaders. Leveraging the educator's current connections to identify new emerging leaders and offering training to those individuals could be one potential avenue for network growth that decreases the dependence on Extension.

Relatedly, the ERGM model showed that food systems educators tend to collaborate with each other more often than they collaborate with non-food systems educators. Extension educators are one type of food system educator. Extension educators could focus on forming collaborative ties with non-educators, directly engaging with growers, market managers, and value-added producers.

Setting Benchmarks for Change

One limitation of this analysis is that it is exploratory in nature. To make future analysis more meaningful, Pederson and her colleagues should identify targets for what an improved network would look like. Possible improvements could be a decrease in diameter or an increase in density for the Coordinate and Collaborate networks. Specific benchmark metrics should be set, and a plan made to reach those benchmarks. For instance, based on the previous recommendations, Pederson could identify five additional collaborative relationships she hopes to nurture, or she could identify certain individuals that are currently in tenuous, peripheral parts of the network and work to strengthen their network ties.

When appropriate benchmarks have been set, Pederson and her colleagues should identify reasonable timelines for meeting those benchmarks. Timelines in hand, we can plan for future analysis.

Conclusion

Pederson's objective going into this project was to gain a better understanding of the current state of the Chequamegon Bay Regional Food Network. She was curious about how well-connected people were, and at what level of the Spectrum of Collaboration they were operating. Pederson's specific objectives included visualizing the network, understanding basic network statistics, discovering important actors, unearthing sub-communities within the networks, and identifying factors that lead to strong collaborative relationships. Through a literature review, social network analysis combined with exponential random graph modeling were identified as viable research methodologies to answer her research questions. After collecting and cleaning survey data from the network, I was able to meet all the proposed objectives of this project. I visualized the network in a myriad of ways. I produced, explained, and put into context core statistics. I identified important actors over multiple categories. I searched for the presence of sub-communities (though none materialized). Finally, I identified those factors that contributed to collaborative ties.

In-depth social network analysis is a time-consuming but rewarding approach to understanding social networks. In this instance, social network analysis illuminated key aspects of the networks, such as the finding that the higher order networks are more sparsely populated and have greater diameter. Social network analysis also identified areas for future growth, such as repositioning the food systems educators and engaging the periphery. Overall, SNA proved a valuable tool for developing greater understanding of this regional food system.

References

Bastos, M., Piccardi, C., Levy, M., McRoberts, N., & Lubell, M. (2018). Core-periphery or decentralized? Topological shifts of specialized information on Twitter. *Social Networks*, *52*, 282-293. https://doi.org/10.1016/j.socnet.2017.09.006

Bell, G. G., & Zaheer, A. (2007). Geography, networks, and knowledge flow. *Organization Science*, *18*(6), 955-972. https://doi.org/10.1287/orsc.1070.0308

Borgatti, S. P., & Cross, R. (2003). A relational view of information seeking and learning in social networks. *Management Science*, *49*(4), 432-445. https://doi.org/10.1287/mnsc.49.4.432.14428

Borgatti, S. P., & Everett, M. G. (1997). Network analysis of 2-mode data. *Social Networks*, *19*(3), 243-269. https://doi.org/10.1016/s0378-8733(96)00301-2

Brinkley, C. (2018). The small world of the alternative food network. *Sustainability*, *19*(8), 2921. https://doi.org/10.3390/su10082921

Carboni, J. L., Siddiki, S., Koski, C., & Sadiq, A. (2017). Using network analysis to identify key actors in collaborative governance processes. *Nonprofit Policy Forum*, *8*(2), 133-145. https://doi.org/10.1515/npf-2017-0012

Carolan, B. V. (2013). *Social network analysis and education: Theory, methods & applications*. SAGE Publications.

Carr, A., & Wilkinson, R. (2005). Beyond participation: Boundary organizations as a new space for farmers and scientists to interact. *Society & Natural Resources*, *18*(3), 255-265. https://doi.org/10.1080/08941920590908123

Chiffoleau, Y. (2009). From politics to co-operation: The dynamics of embeddedness in alternative food supply chains. *Sociologia Ruralis*, *49*(3), 218-235. https://doi.org/10.1111/j.1467-9523.2009.00491.x

Chiffoleau, Y., & Touzard, J. (2013). Understanding local agri-food systems through advice network analysis. *Agriculture and Human Values*, *31*(1), 19-32. https://doi.org/10.1007/s10460-013-9446-6

Christensen, L., & O'Sullivan, R. (2015). Using social networking analysis to measure changes in regional food systems collaboration: A methodological framework. *Journal of Agriculture, Food Systems, and Community Development*, 113-129. https://doi.org/10.5304/jafscd.2015.053.013

Clark, F. E. (2010). Space to choose: Network analysis of social preferences in a captive chimpanzee community, and implications for management. *American Journal of Primatology*, *73*(8), 748-757. https://doi.org/10.1002/ajp.20903

Correa, C. D. (2017). History and evolution of social network visualization. *Encyclopedia of Social Network Analysis and Mining*, 1-17. https://doi.org/10.1007/978-1-4614-7163-9_102-1

Crespo, J., Réquier-Desjardins, D., & Vicente, J. (2014). Why can collective action fail in local agri-food systems? A social network analysis of cheese producers in Aculco, Mexico. *Food Policy*, *46*, 165-177. https://doi.org/10.1016/j.foodpol.2014.03.011

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695. http://igraph.org

Dianati, N. (2016). Unwinding the hairball graph: Pruning algorithms for weighted complex networks. *Physical Review E*, *93*(1). https://doi.org/10.1103/physreve.93.012304

Doreian, P., & Woodard, K. L. (1992). Fixed list versus snowball selection of social networks. *Social Science Research*, *21*(2), 216-233. https://doi.org/10.1016/0049-089x(92)90016-a

Falck, O., & Heblich, S. (2007). Dynamic Clusters. *Working Papers 16, Bavarian Graduate Program in Economics*. http://www.bgpe.de/texte/DP/016_falck_heblich.pdf

Festinger, L. (1949). The analysis of sociograms using matrix algebra. *Human Relations*, *2*(2), 153-158. https://doi.org/10.1177/001872674900200205

Freedman, D. A., & Bess, K. D. (2011). Food systems change and the environment: Local and global connections. *American Journal of Community Psychology*, *47*(3-4), 397-409. https://doi.org/10.1007/s10464-010-9392-z

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215-239. https://doi.org/10.1016/0378-8733(78)90021-7

Fruchterman, T. M., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, *21*(11), 1129-1164. https://doi.org/10.1002/spe.4380211102

Furht, B. (2010). *Handbook of social network technologies and applications*. Springer Science & Business Media.

Geng, J., Bhattacharya, A., & Pati, D. (2018). Probabilistic community detection with unknown number of communities. *Journal of the American Statistical Association*, *114*(526), 893-905. https://doi.org/10.1080/01621459.2018.1458618

Giuliani, E., & Pietrobelli, C. (2011). Social network analysis methodologies for the evaluation of cluster development programs. *Technical Notes #IDB-TN-317, Inter-American Development Bank, Capital Markets and Financial Institutions Division*.

Goodreau, S. M. (2007). Advances in exponential random graph (P*) models applied to a large social network. *Social Networks*, *29*(2), 231-248. https://doi.org/10.1016/j.socnet.2006.08.001

Handcock, M., Hunter, D., Butts, C., Goodreau, S., Ktivitsky, P., & Morris, M. (2019). ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks. *The Statnet Project*. https://CRAN.R-project.org/package=ergm

Harris, J. K. (2013). *An introduction to exponential random graph modeling*. SAGE Publications.

Huang, F., Zhang, M., & Li, Y. (2019). A comparison study of tie non-response treatments in social networks analysis. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02766

Huisman, M. (2009). Imputation of missing network data: some simple procedures. *Journal of Social Structure*, *10*(1), 1-29. https://www.exeley.com/journal_of_social_structure/doi/10.21307/joss-2019-050

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). Ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, *24*(3). https://doi.org/10.18637/jss.v024.i03

Jarosz, L. (2000). Understanding agri-food networks as social relations. *Agriculture and Human Values*, *17*, 279-283. https://doi-org.ezproxy.library.wisc.edu/10.1023/A:1007692303118

Jones, J. H., Ready, E., & Hazel, A. (n.d.). *Eehh-Stanford/SNA-workshop*. GitHub. https://github.com/eehh-stanford/SNA-workshop

Kemkes, R., Hofstedt, B., & Tochterman, A. (2018). *Ashland and Bayfield Counties Household Food Survey*. Center for Rural Communities, Northland College.

https://www.northland.edu/wp-content/uploads/2018/01/Ashland-Bayfield-Counties-Food-Suvey_NC_CRC_Jan2018.pdf

Killworth, P., & Bernard, H. (1976). Informant accuracy in social network data. *Human Organization*, *35*(3), 269-286. https://doi.org/10.17730/humo.35.3.10215j2m359266n2

Kossinets, G. (2006). Effects of missing data in social networks. *Social Networks*, *28*(3), 247-268. https://doi.org/10.1016/j.socnet.2005.07.002

Larson, L. E. (2005). *Chequamegon Bay and its communities*.

Lionberger, H. F. (1959). Community prestige and the choice of sources of farm information. *Public Opinion Quarterly*, *23*(1), 110. https://doi.org/10.1086/266851

Lubell, M., Niles, M., & Hoffman, M. (2014). Extension 3.0: Managing agricultural knowledge systems in the network age. *Society & Natural Resources*, *27*(10), 1089-1103. https://doi.org/10.1080/08941920.2014.933496

Luke, D. A. (2015). *A user's guide to network analysis in R*. Springer.

McLeod, M., & Vaughan, R. (2014). *Knowledge networks and tourism*. Routledge.

Morgan, K. (2004). The exaggerated death of geography: Learning, proximity and territorial innovation systems. *Journal of Economic Geography*, *4*(1), 3-21. https://doi.org/10.1093/jeg/4.1.3

Naderifar, M., Goli, H., & Ghaljaie, F. (2017). Snowball sampling: A purposeful method of sampling in qualitative research. *Strides in Development of Medical Education*, *14*(3). https://doi.org/10.5812/sdme.67670

Nicolosi, A., Laganà, V. R., Laven, D., Marcianò, C., & Skoglund, W. (2019). Consumer habits of local food: Perspectives from northern Sweden. *Sustainability*, *11*(23), 6715. https://doi.org/10.3390/su11236715

Okamoto, K., Chen, W., & Li, X. (n.d.). Ranking of closeness centrality for large-scale social networks. *Frontiers in Algorithmics*, 186-195. https://doi.org/10.1007/978-3-540-69311-6_21

Porter, M. E. (2008). *On competition*. Harvard Business Press.

Porter, M. E. (2014, October 29). *Michael Porter: "reshaping regional economic development: clusters and regional strategy"* [Video]. YouTube. https://www.youtube.com/watch?v=mF2CsUcwFrw

Radke, L., & Deller, S. (2020). *A simple overview of the Ashland county economy*. Ashland Area Development Corporation. https://static1.squarespace.com/static/56eac8c9044262c09442c67d/t/5e581bcf22d7e90201d3781f/1582832595762/Ashland+County+Economic+Profile.pdf

Reid, N., Smith, B. W., & Carroll, M. C. (2008). Cluster regions: a social network perspective. *Economic Development Quarterly*, *22*(4), 345-352. https://doi.org/10.1177/0891242408322719

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks*, *29*(2), 173-1919. https://doi.org/10.1016/j.socnet.2006.08.002

Rogers, E. M. (2010). *Diffusion of innovations* (4th ed.). Simon & Schuster.

Scheinert, S., Zia, A., Koliba, C., & Merrill, S. (2017). Growing collaborations: Forecasting changes in partnership networks using a bottom-up approach. *Journal on Policy and Complex Systems*. https://doi.org/10.18278/jpcs.3.1.1

Schilling, M. A., & Phelps, C. C. (2007). Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. *Management Science*, *53*(7), 1113-1126. https://doi.org/10.1287/mnsc.1060.0624

Scott, J. (2017). *Social network analysis*. SAGE.

Scott, J., & Carrington, P. J. (2011). *The SAGE handbook of social network analysis*. SAGE.

Stoltenberg, D., Maier, D., & Waldherr, A. (2019). Community detection in civil society online networks: Theoretical guide and empirical assessment. *Social Networks*, *59*, 120-133. https://doi.org/10.1016/j.socnet.2019.07.001

Stork, D., & Richards, W. D. (1992). Nonrespondents in communication network studies. *Group & Organization Management*, *17*(2), 193-209. https://doi.org/10.1177/1059601192172006

Sánchez Herrera, J., & Dimitri, C. (2019). The role of clustering in the adoption of organic dairy: A longitudinal networks analysis between 2002 and 2015. *Sustainability*, *11*(6), 1514. https://doi.org/10.3390/su11061514

Tamarack Institute. (2017). *The collaboration spectrum*. https://www.tamarackcommunity.ca/hubfs/Resources/Tools/Collaboration%20Spectrum%20Tool%20July%202017.pdf

Ting Helena Chiu, Y. (2008). How network competence and network location influence innovation performance. *Journal of Business & Industrial Marketing*, *24*(1), 46-55. https://doi.org/10.1108/08858620910923694

Triandis, H. C., & Gelfand, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology*, *74*(1), 118-128. https://doi.org/10.1037/0022-3514.74.1.118

Trivette, S. A. (2018). The importance of food retailers: Applying network analysis techniques to the study of local food systems. *Agriculture and Human Values*, *36*(1), 77-90. https://doi.org/10.1007/s10460-018-9885-1

Tropp, D. (2019). From anecdote to formal evaluation: Reflections from more than two decades on the local food research trail at USDA. *Journal of Agriculture, Food Systems, and Community Development*, 1-18. https://doi.org/10.5304/jafscd.2019.091.044

U.S. Census Bureau QuickFacts: Ashland County, Wisconsin. (n.d.). Census Bureau QuickFacts. https://www.census.gov/quickfacts/fact/table/ashlandcountywisconsin/PST045219

U.S. Census Bureau QuickFacts: Bayfield County, Wisconsin. (n.d.). Census Bureau QuickFacts. https://www.census.gov/quickfacts/fact/table/bayfieldcountywisconsin,US/PST045219

United States Department of Agriculture. (n.d.). *Extension*. National Institute of Food and Agriculture. https://nifa.usda.gov/extension

University of Wisconsin-Madison, Division of Extension. (n.d.). *Extension staff directory*. Extension Staff Directory. https://people.extension.wisc.edu/

Wang, M., Wang, C., Yu, J. X., & Zhang, J. (2015). Community detection in social networks. *Proceedings of the VLDB Endowment*, *8*(10), 998-1009. https://doi.org/10.14778/2794367.2794370

Yamaguchi, K. (1994). The flow of information through social networks: Diagonal-free measures of inefficiency and the structural determinants of inefficiency. *Social Networks*, *16*(1), 57-86. https://doi.org/10.1016/0378-8733(94)90011-6

Zou, Q., Li, J., Hong, Q., Lin, Z., Wu, Y., Shi, H., & Ju, Y. (2015). Prediction of microrna-disease associations based on social network analysis methods. *BioMed Research International*, *2015*, 1-9. https://doi.org/10.1155/2015/810514

Žnidaršič, A., Ferligoj, A., & Doreian, P. (2017). Actor non-response in valued social networks: The impact of different non-response treatments on the stability of blockmodels. *Social Networks*, *48*, 46-56. https://doi.org/10.1016/j.socnet.2016.06.001

Appendix A: Initial Survey Questions

The following fields/questions were in the stage one survey delivered to 108 network actors on April 22, 2020.

1. First and last name
2. Email address
3. Street Address (used for geolocation)
4. What best describes your role in the Food System in the Chequamegon Bay Region?
    a. Value-Added Producer
    b. Grower
    c. Distributor/Wholesaler
    d. Buyer
    e. Restauranteur/Cook/Chef
    f. Food Systems Educator
    g. Market Manager
    h. None of These
    i. Other (fill in option)
5. Are you self employed?
    a. Yes
    b. No
6. What is the name of your business?

7. Beyond working for your own business, do you have a job (or second job) in the food-related economy (restaurant, value-added production, distribution, wholesale, food or nutrition education, farming, etc.)
    a. Yes
    b. No
8. What is the name of the business or institution you work for?
9. What percentage of your adjusted gross income comes from food/farming business?
10. Which farmers' markets have you attended in the past year as a vendor?
    a. Ashland
    b. Washburn
    c. Bayfield
    d. Madeline Island
    e. Cornucopia
    f. Cable
    g. Duluth/Superior
    h. None
11. Which farmers' markets have you attended in the last year as a consumer?
    a. Ashland
    b. Washburn
    c. Bayfield
    d. Madeline Island
    e. Cornucopia

f. Cable

g. Duluth/Superior

h. None

12. Which groups have you been affiliated with in the past 5 years?

a. A.E.R.C.

b. Alliance for Sustainability

c. Bad River Food Sovereignty

d. Bayfield Foods

e. Beaser Community Garden

f. Chequamegon CSA

g. Chequamegon Food Coop Board of Directors

h. F.E.A.S.T.

i. Farm Bureau

j. Farm to School

k. Farmer's Market Board

l. In Her Boots

m. National Farmers Organization (NFO)

n. Northland College programs (Growing Connections, Hulings Rice Food Center)

o. Re Cliff Community Farm

p. Some other multi-producer CSA

q. UW Extension Farmily or Bootlakers Guild

r. Wisconsin Farmer's Union

s. None

t. Other (fill-in option)

13. Have you used social media in the past year?

    a. Yes

    b. No

14. What social media platforms have you used in the last year?

    a. Facebook

    b. Instagram

    c. LinkedIn

    d. Snapchat

    e. Twitter

    f. Tumblr

    g. None

    h. Other (fill in option)

15. About how many hours per day have you spent on social media platforms in the past year?

    a. Less than 1

    b. 1

    c. 2

    d. 3

    e. 4

f. 5

g. More than 5

16. Does your food/farm business have a website?

a. Yes

b. No

17. What is the URL of your food/farm business website?

18. Have you actively marketed your business online in the past year? (e.g. made a Facebook or Instagram post, updated content, run web promotions, etc.)

a. Yes

b. No

19. In the past year, when you have had a question about your business, where did you look for answers?

a. Search Engine (Google/Bing)

b. Social media

c. Extension agent

d. Business or trade group

e. Podcasts

f. How-to videos

g. Books/magazines

h. Library/librarian

i. Call neighbor, friend, or an acquaintance

j. Networking event or conference

   k. None
   l. Other (fill-in option)

20. In the past year, how much as your social network (friends, family) and your business network (customers, employees, colleagues, coworkers) overlapped?
    a. They were the same
    b. They overlapped a lot
    c. There was some overlap
    d. There was no overlap
    e. I did not have both of these networks
21. Do you consider yourself primarily and introvert or primarily an extrovert?
22. Please rank yourself on a scale of strongly disagree to strongly agree for the following statements:
    a. In my work, I'd rather depend on myself than others
    b. In my work, I rely on myself most of the time; I rarely rely on others
    c. Generally speaking, I believe that businesses that collaborate have an economic advantage
    d. Generally speaking, I think that collaborating with other businesses and business owners is not worth the hassle
23. Excluding yourself and your immediate family, who were the most important people in YOUR food/farm network in the past 5 years? (primary customers or buyers, production partners, suppliers, people whom you've gone to for help/advice, etc.). Please list first and last names, not business names.

a. Person 1

b. Person 2

c. Person 3

Appendix B: Survey Response Calculations

A network is made up of dyads. For every 2 people in the network, there is one relationship. A person's relationship with themselves is not of interest. Therefore, in a 100 x 100 complete network matrix, we would have (100 * 99)/2 total possible relationships, or 4950 total possible relationships. We had 54 of the 100 people respond to the network survey. However, one of the individuals was added late and was not in the matrix to which other people responded. Therefore, we have 53 individuals who rated everyone in the network. That results in (53 * 52)/2 completely described relationships (1378). Those 53 people also rated the 46 non-respondents, which results in 53 * 46 or 2438 partially described relationships via incoming ties. The person added late rated the other 99 individuals in the network, for 99 partial outgoing ties. Finally, we have (46 * 45)/2 completely missing relationships (1035).

Therefore, to determine overall missingness of data, we divide the described relationships by the possible relationships.

- Completely described: 1378/4950 = 28%
- Partially described: (2438 + 99)/4950 = 51%
- Completely or partially described: (2438 + 99 + 1378)/4950 = 79%
- Completely missing: 1035/4950 = 21%

Appendix C: Individuals to Target for Facilitated Relationship Building

Isolates

The following individuals were identified as isolates in the collaborative network. While everyone had relationships of other types, some of the individuals had very limited higher-order (coordinate) relationships. These individuals would be ideal candidates to target for relationship-building facilitation.

Table 4

Collaborative Network Isolates

Redacted for privacy

Weakly Connected Individuals

The following individuals have less than three relationships in either the Coordinate or Collaborate network. These are secondary individuals to target for relationship-building facilitation. (Note that some isolates are repeated in this table.)

Table 5

Low-Degree Members of the Coordinate & Collaborate Network

Redacted for privacy

Individuals at Risk Via Cutpoints

The following individuals are the cutpoints and the people they would turn into isolates if they left the Coordinate network.

Table 6

Individuals in Danger of Becoming Isolates in the Coordinate Network

Redacted for privacy

The following individuals are the cutpoints and the people they would turn into isolates if they left the Coordinate network.

Table 7

Individuals in Danger of Becoming Isolates in the Collaborate Network

Redacted for Privacy

Individuals on Shortest Paths

The following individuals are on the shortest paths of individuals with high-ranking betweenness centrality. These individuals could be directly connected (bypassing the person with high betweenness) to increase network density and redundancy. (Additional betweenness paths are available in the generated results file noted in Appendix D.)

Table 8

Coordinate Network Betweenness Paths for Jason Fischbach

Redacted for Privacy

Table 9

Collaborate Network Betweenness Paths for Jason Fischbach

Redacted for Privacy

Appendix D: Code and Rendered Code

This project was completed using a combination of Jupyter Notebooks and R Markdown. The data is confidential. The code (minus the data sources) is available in a Github repository: https://github.com/DeannaDS/portfolio/tree/master/Capstone