Chapter 2: CUPID

*Introduction / Literature Review*

Measuring New and Expanded Opportunities

Cumulative advantage is a key driver for the development of scientific stars (Mali et al. 2012, 235), a term that refers to a specific network structure. Networks consist of actors (researchers) and the relationships among them (ties) (Mali et al. 2012, 216). Social Network Analysis (SNA) provides a framework for understanding these structures, focusing on the relationships among actors within a network (Borgatti et al. 2022, 2; Mali et al. 2012, 216). The modular structure of researcher networks operates across disciplinary, sectoral, and geographical boundaries (Mali et al. 2012, 219; Vacca et al. 2015). These actors, or nodes, can be characterized by various categorical attributes, such as department affiliation, or continuous, like years of geographical distances (Mali et al. 2012, 219). The relationship in this context, termed as ties or edges, connects researchers to each other and can be quantified in multiple ways, including the frequency of interactions over a given period (Borgatti et al. 2022, 2; Mali et al. 2012, 216).

For example, Newman (2001) undertakes a comprehensive study of social networks, specifically focusing on scientific collaborations. Newman leverages the co-authorship of scientific papers as an unbiased and scalable measure for mapping social connections within the scientific community. Gathered data from multiple scientific databases, such as MEDLINE and the Los Alamos e-Print Archive, Newman (2001) created the network that tied each researcher in the network to all other researchers with whom they co-authored a paper within a five-year window (1995-1999). These ties interlink through common nodes, forming paths and, ultimately, a network. Within this network, frequently interacting actors may form a distinct subgroup (Borgatti et al. 2022, 2). Newman (2001) found that researchers tend to collaborate with peers who have gained influence through numerous prior joint projects, following a pattern of preferential attachment.

Co-authorship is a common type of relationship used to study scientific collaboration. In their book chapter, Mali et al. (2012) explore the complexities of scientific collaboration using co-authorship networks for their example but highlight various other collaborative activities, such as shared editorship, joint supervision of research projects, collaborative research proposal writing, participation in formal research programs, and the organization of scientific conferences (Mali et al. 2012, p 213).

THE GRANT NETWORK methods

Historical grant proposal application data from 2016 and 2020 creates multiple networks, including five-year and yearly networks. Nodes are faculty who collaborate within the given time frame, and edges are formed when any two faculty co-propose. Another grant proposal within the bounds of the network links these faculty to other faculty, creating a co-occurrence network (Borgatti et al. 2022). Faculty who proposed alone are removed from the network because we are examining collaborative proposals. Faculty who did not propose within a single year are removed from the network during network modeling.

The grant proposal network does not reveal the true social relationship between the faculty. While some faculty take on the role of PI on grant proposals, the reasoning for the role varies. Because of this, all individuals who share a grant together are considered equal, with no particular direction that connects the nodes.

Describe ATTRIBUTES

Mali et al. highlight the foundational elements of modern social network analysis (SNA) as identified by Freeman (2004): a focus on structural analysis of actors within social relations, the use of systematic empirical data, extensive use of graphical imagery, and a foundation in formal, mathematical, and computational models (Mali et al. 2012, p 216). By leveraging SNA, I analyze the intricate web of grant proposal collaboration, indicating how relationships and network structures contribute to developing scientific work. Using network visualizations, node and network metrics, and exponential random graph models, I describe BSU’s grant proposal collaboration networks and how they evolve between 2016 and 2020.

Methods

For example, the local property of a node in the network is **degree centrality**, deﬁned as the number of ties a node has (Mali et al. 2012, 214; Borgatti et al. 2022, 171). A high degree centrality takes the shape of a star, where one node has many ties to other nodes compared to most other nodes in the network. Its interpretation can vary based on the nature of these ties (Borgatti et al. 2022, 172). A star structure in team science networks may indicate a significant inequality in collaborative offers, as few scientists or scholars receive disproportionate offers to collaborate (Moody 2004). The cumulative advantage in science posits that scientist already recognized for their contributions are more likely to gain further recognition and resources (Mali et al. 2012, 235). This concept, drawing parallels to the biblical passage in Matthew's Gospel and referred to as "The Matthew Effect," implies a disparity in the distribution of resources and opportunities within the scientific community, where established researchers gain disproportionately more funding and power while emerging scientists face challenges in achieving recognition and success (Mali et al. 2012, 235–36). This concept highlights how normal social behaviors can thwart the GCs’ investment goal to expand research opportunities across campus.

Networks formed through this *preferential attachment* suggest a scale-free structure characterized by a power-law degree distribution where burgeoning scientists tend to collaborate with established ‘scientific stars’, reflecting the principle of cumulative advantage in science (Mali et al. 2012, 215; Vacca et al. 2015). This scale-free structure could indicate a hierarchical network dominated by a few highly connected individuals or "hubs" (Mali et al. 2012, 236). I examine the **degree distribution** to identify this phenomenon.

**Betweenness centrality** measures a node’s frequency along the shortest paths between other node pairs (Borgatti et al. 2022, 182). It is interpreted as a node’s potential to control or regulate the flow through the network, playing a gatekeeper or broker role (Borgatti et al. 2022, 183). With their control over resources and opportunities, gatekeepers play a crucial role in shaping the network's topology (Mali et al. 2012, p 236). I examine the **betweenness distribution** to identify this phenomenon.

WORK HERE: GINI COEFFICIENT

The **connectedness** score illuminates the level of structural cohesion (Borgatti et al. 2022, 201–3). Comparing the connectedness across each year’s network depicts the change in structural cohesion of the grant proposal network over time. Analysis from degree distribution and connectedness could be used to intentionally connect researchers across diverse modules, such as spanning structural holes and counterbalancing preferential attachment, as Vacca et al. showcase. The potential of network interventions is to overcome inherent biases in collaboration patterns and to bridge gaps between disparate scientific communities. Vacca et al.’s (2015) approach offers a pragmatic pathway for fostering cross-disciplinary team science and enhancing the cohesion and diversity of scientific research networks.

Measuring Interdisciplinary

Increasing interdisciplinary and transdisciplinary collaborations is a core goal of the GCs investments. Scientific disciplines must work together to solve complex and large-scale societal challenges like resource sustainability and One Health. Collaborative research is often categorized into three distinct yet interconnected types: multi-, inter-, and transdisciplinary (e.g., Dalton, Wolff, and Bekker 2022; Sonnenwald 2007; Lieberknecht et al. 2023). Multidisciplinary research involves various disciplines working in parallel, each contributing their perspective without integrating their efforts (Dalton, Wolff, and Bekker 2021). In contrast, interdisciplinary research signifies a deeper level of collaboration where multiple disciplines actively merge their methodologies and viewpoints to tackle a common problem (Dalton, Wolff, and Bekker 2021). Transdisciplinary research transcends traditional academic boundaries, incorporating inputs from external entities such as industry, government, and community stakeholders, thus offering a holistic approach to complex societal issues (Dalton, Wolff, and Bekker 2021).

Bolger (2021) zeros in on the degree of interdisciplinary research by categorizing discipline distances. Through a study of three established research centers, the study surveys faculty members on their motivations for joining the centers, their perceptions of interdisciplinary research, and the nature of their collaborative activities. Bolger introduces a novel classification based on the 'distance' between collaborating disciplines: 'within-discipline' collaborations (e.g., between biologists with different specializations), 'short distance' within the same super-discipline (e.g., an engineer collaborating with a biologist), and 'long distance' across distinct super-disciplines (e.g., an ecologist working with a social scientist) (Bolger 2021). This final categorization distinguishes collaborations spanning 'hard' sciences (natural and applied sciences) and 'soft' sciences (social sciences and humanities), offering a more granular understanding of interdisciplinary research dynamics (Bolger 2021).

The analysis of subgroups and the overall network structure allows for the examination of shared attributes, offering insights into the collaborative dynamics in scientific communities (Borgatti et al. 2022, 2–3, p 214). Research specialties can be identified as a central cluster of collaborating scientists responsible for producing a significant number of innovative concepts and ideas (Moody 2004; Vacca et al. 2015). Collaboration within specific scientific disciplines often leads to the emergence of distinct clusters within research collaboration networks, indicative of a *small-world* network structure marked by high local clustering and minimal steps between clusters (Vacca et al. 2015; Mali et al. 2012, 215). I use **network visualizations** showing the researcher’s affiliated **college attribute** to illuminate possible within-discipline and short-distance disciplinary clustering. Networks with clusters that cross super-disciplines from “invisible colleges” drive the intellectual and creative output of the scientific community (Mali et al. 2012, p 236).

In a small-world network, local clustering is high, but the average number of steps between actors is minimal (Moody 2004). In contrast, a cohesive core is where a growing number of authors show a tendency toward collaboration across different specialties (Moody 2004)I focus on the network statistics **clustering coefficient** [`gtrans()` in the `sna` package] and **average path length** to analyze the 5-year grant network for small-world properties. [The data does not allow for short-distance examination because there are too many departments to analyze efficiently.]

ERGMs

The landscape of social network analysis (SNA) is has been profoundly transformed by the introduction of exponential random graph models (ERGMs) (Mali et al. 2012, 218). ERGMs are a specific category of statistical models that articulate the likelihood distribution of network graphs, premised on the assumption that network connections form patterns or configurations that recur more frequently than chance would predict (Harris 2014, 33). These configurations vary broadly, offering adaptability for various contexts, with a positive parameter value indicating a configuration’s propensity to occur more often within the network data (Caimo and Gollini 2020, 2).

Independence / dependence

ERGMs is a statistical network modeling method that addresses the issue of interdependence.

Standard statistical approaches assume independence of observations, but humans are intentional beings with multiple motivations for and expressions of social action (Lusher and Robins 2013). Many social processes occur simultaneously. Exponential Random Graph Models (ERGMs) incorporate dependency between network ties (Lusher and Robins 2013). These tie-based models permits an understanding of the complex combination of social processes by which network ties are formed (Lusher and Robins 2013).

Our networks are nonindependent because knowing one faculty member co-proposed with a second faculty member tells information about the second faculty member that depends on the first.

theory of dependence has the definition of local(Lusher and Robins 2013).

Certain network patterns are important based on more specific social science / social network theory: adopting a particular dependence hypothesis and definition of local (Lusher and Robins 2013). 4-cycle the presence of existing relationships creates the conditions whereby an old friend tie affects the chances of a new friendship. “social circuit dependence (Lusher and Robins 2013). Network ties that organize themselves into patterns because the presence of some ties encourages other to come into existence (Lusher and Robins 2013). Being popular may attract even more popularity. Popularity is defined through the diversity of in-degree distribution where few highly centralized nodes are popular. “preferential attachment” (Lusher and Robins 2013) Activity is defined by out-degree (Lusher and Robins 2013) Reciprocity is a form of dependency whereby the two possible directed ties within a dyad are dependent on each other. “dyadic dependence”(Lusher and Robins 2013) transitivity is closing the path, forming a third tie that produces a triangle. Also called network clustering in undirected. forming triads (Lusher and Robins 2013). from the social network theory that humans social propensity to operate in group like structures (Lusher and Robins 2013). a triangle of three is a simple archetypal expression of a small group. Many triangles together form clique-like structures forming a community, cohesive subgroups (Lusher and Robins 2013). Transitive triad where one node is receiving two ties and sending none. cyclic triad where the direction of all ties is consistent so that they for a 3-cycle (Lusher and Robins 2013). Transitivity or path closure from Markov dependence where ties are assumed dependent if they share a node (Lusher and Robins 2013).

ERMGS can also provide insight into other structural features of the network that are relevant to small-world nature. For example, if homophily of the college nodal attributes significantly predicts tie formation, this might provide context for the network’s small-world properties.

Interaction terms for nodal attributes account for the attributes of both members of a dyad (Harris 2014). Homophily is the most commonly used interaction term, two nodes sharing an attribute. Conversely, heterophily is where two nodes are different on an attribute (Harris 2014). Based on college affiliation, I look at the probability of a tie between grant proposers.

WORK HERE NVIVO metrics ERGMs CUPID Network Stats and up

In their study, Vacca et al. (2015) highlight the significance of recognizing and utilizing the modular structures within a university’s scientific collaboration networks to facilitate effective network interventions. They demonstrate that mapping co-authorship and grant proposal networks, identifying unconnected researcher groups, and employing strategic interventions can significantly enhance the network's overall structure, promoting interdisciplinary collaboration and innovation.

Mali et al. (2012) also acknowledges the complexities and challenges in fostering and measuring interdisciplinary research due to its demands for extensive networks, time, and mobility among researchers (Mali et al. 2012, p 222).

Vacca et al. not only mapped the university’s scientific collaboration network in their 2015 study but also surveyed the peripheral researchers receiving network treatments. The researchers discovered that familiarity with each other's work decreases as researchers are further apart in the network, implying that network data can effectively map a university's research activities (Vacca et al. 2015). Additionally, they observed that researchers are more skeptical about collaborating with more distant individuals in the network (Vacca et al. 2015). This skepticism is particularly pronounced in the context of grants and patents, potentially due to trust issues (Vacca et al. 2015). They recommend adding incentives for collaboration to motivate distant collaborations (Vacca et al. 2015).

The findings of Vacca et al. (2015) underscore the utility of employing methods like surveys to glean insights into the dynamics of scientific collaboration networks.