NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

ANALYSIS OF A SIMILARITY-BASED APPROACH TO POSITIVELY AFFECT MENTOR-MENTEE RELATIONSHIPS AMONG SERVICE MEMBERS

by

Thomas P. Reeder

June 2021

Thesis Advisor:	Ralucca Gera
Second Reader:	D'Marie E. Bartolf

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** June 2021	**3. REPORT TYPE AND DATES COVERED** Master's thesis	
4. TITLE AND SUBTITLE ANALYSIS OF A SIMILARITY-BASED APPROACH TO POSITIVELY AFFECT MENTOR-MENTEE RELATIONSHIPS AMONG SERVICE MEMBERS			**5. FUNDING NUMBERS**
6. AUTHOR(S) Thomas P. Reeder			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.			**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

Mentorship encourages growth between a mentor and a mentee and, as a result, positively affects relationships within an organization. Effective mentorship can influence professional development that can positively impact the Undersea Warfare (USW) community through increased performance, knowledge sharing, positive work-life culture, and retention. Currently, the United States Navy institutes a Command Sponsorship and Indoctrination Program to assist incoming service members as they settle into a new command. We focus on the scope of this program and recognize that a sponsor serves in a mentorship role to the incoming service member.

The existing process outlines general selection criteria for sponsors (mentors) before they are assigned to incoming service members (mentees). In our study, we complement those criteria with a network science perspective that uses personal attributes to connect mentors and mentees with the goal of supporting a stronger connection between them. We accomplish this through a modified assignment process, referred to as the Similarity-Based Node Pairing (SBNP) Model, that prioritizes similarity of the service members based on select attributes among both mentors and mentees. As a proof of concept, we test this process through a simulated network of people with randomly assigned attributes and deliver mentor-mentee pairings.

14. SUBJECT TERMS sponsorship, mentorship, network science, optimization, personalization			**15. NUMBER OF PAGES** 75
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

ANALYSIS OF A SIMILARITY-BASED APPROACH TO POSITIVELY AFFECT MENTOR-MENTEE RELATIONSHIPS AMONG SERVICE MEMBERS

Thomas P. Reeder
Ensign, United States Navy
BS, U.S. Naval Academy, 2020

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN APPLIED MATHEMATICS

from the

NAVAL POSTGRADUATE SCHOOL
June 2021

Approved by: Ralucca Gera
Advisor

D'Marie E. Bartolf
Second Reader

Wei Kang
Chair, Department of Applied Mathematics

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Mentorship encourages growth between a mentor and a mentee and, as a result, positively affects relationships within an organization. Effective mentorship can influence professional development that can positively impact the Undersea Warfare (USW) community through increased performance, knowledge sharing, positive work-life culture, and retention. Currently, the United States Navy institutes a Command Sponsorship and Indoctrination Program to assist incoming service members as they settle into a new command. We focus on the scope of this program and recognize that a sponsor serves in a mentorship role to the incoming service member.

The existing process outlines general selection criteria for sponsors (mentors) before they are assigned to incoming service members (mentees). In our study, we complement those criteria with a network science perspective that uses personal attributes to connect mentors and mentees with the goal of supporting a stronger connection between them. We accomplish this through a modified assignment process, referred to as the Similarity-Based Node Pairing (SBNP) Model, that prioritizes similarity of the service members based on select attributes among both mentors and mentees. As a proof of concept, we test this process through a simulated network of people with randomly assigned attributes and deliver mentor-mentee pairings.

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

THIS PAGE INTENTIONALLY LEFT BLANK

List of Figures

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

List of Acronyms and Abbreviations

CNO Chief of Naval Operations

CSIP Command Sponsor and Indoctrination Program

IP Integer Program

NAVPLAN Navigation Plan

OPNAVINST Office of the Chief of Naval Operations Instruction

PCS Permanent Change of Station

SBNP Similarity-Based Node Pairing

THIS PAGE INTENTIONALLY LEFT BLANK

Executive Summary

The United States Navy currently institutes the Command Sponsor and Indoctrination Program (CSIP) to assist incoming service members as they settle into a new command. In this work, we focus on the scope of this program and recognize that a sponsor serves in a mentorship role to the incoming service member; this program serves as a formal mentorship program due to the gap in the levels of experience at the new command between the sponsor and incoming service member. The existing process uses general selection criteria for sponsors (mentors) before they are assigned to incoming service members (mentees). Proper mentorship fosters growth between both a mentor and a mentee and, as a result, positively affects relationships within an organization. Effective mentorship can influence professional development which can positively impact the Undersea Warfare community through increased performance, knowledge sharing, positive work life culture, and retention. We propose to complement the current CSIP selection criteria with a network science perspective that uses personal attributes to connect mentors and mentees with the goal of supporting a stronger connection between them. We accomplish this proposal through a modified assignment process by prioritizing similarity of the service members based on select attributes among both mentors and mentees. As a proof of concept, we test this process through a simulated network of people with randomly assigned attributes.

We implement a Similarity-Based Node Pairing (SBNP) Model that will generate a network of people and attributes and assign attributes to people. We then calculate two similarity metrics, Pearson and Jaccard, for each potential mentor-mentee pairing. Using these coefficients, we pair mentors to mentees with three different methods: by randomizing their selection, by maximizing their Pearson coefficient sum, and by maximizing their Jaccard coefficient sum. We use integer programming to compute the two SBNP optimal pairings.

We perform a sensitivity analysis on each of the four input variables of the SBNP Model to observe any effects. While each input variable contributes to the SBNP solution in its own way, the SBNP optimal pairings consistently produced Pearson and Jaccard coefficient sums twice as large as the random pairing. We note that this trend held up in nearly every group of simulations. We also compared the two similarity metrics used in the SBNP Model and concluded that, while the Jaccard coefficient performed slightly better on average and is

the preferred metric, the Pearson solution offers value as well as a check against the Jaccard solution. Overall, the SBNP provides more personalized mentor-mentee pairings according to these two similarity metrics than with a random mentor-mentee pairing.

Acknowledgments

I would like to thank the Navy for affording me the opportunity to learn and research at the Naval Postgraduate School.

I would like to thank several people that assisted me along the way in completing this work. Ralucca Gera and D'Marie Bartolf both dedicated several hours and endless effort each week for at least six months prior to submission to ensure I was on the right track. I learned a lot about the academic writing process and applying critical thinking to a relevant military topic and could not have delivered a complete thesis paper or presentation without their guidance.

I would also like to thank Jen McClary who took time to provide feedback on my writing and offered her perspective on my paper.

I want to thank Matt Critchley and Andrew Dietz for their input and support throughout the entire thesis process, both inside and outside of our weekly thesis meetings.

I also want to thank students in the Undersea Warfare Department, Applied Mathematics Department, and various other departments throughout NPS for the relationships formed and the camaraderie developed throughout my time in Monterey.

I finally want to thank my family and any friends that listened to me share my frustrations throughout the thesis process and kept me moving in the right direction.

This work was not possible without all of you!

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 1:
Introduction

We all experienced entering an organization as the newest member of the team or community, or changing location for the first time, and inherently we adapt as quickly as possible. While some people thrive in a new environment, others may struggle to acclimate. The military is no exception; its dynamic nature demands service members to adjust to new situations under different circumstances, most commonly a service member's frequent Permanent Change of Station (PCS).

The Chief of Naval Operations (CNO), in his 2021 Navigation Plan (NAVPLAN), defines an objective to develop a seasoned team of naval warriors (Chief of Naval Operations 2021). He further emphasizes the "family" aspect of the Navy and the importance of taking care of Navy families while personnel are focused on their respective missions. As a result, he commits to strengthening programs that support families. Office of the Chief of Naval Operations Instruction (OPNAVINST) 1740.3D introduces and defines the Command Sponsor and Indoctrination Program (CSIP), one of the programs that supports the objective set out by the CNO. Per the Department of the Navy, the Navy currently implements the CSIP, last updated in 2017, with the goal of the program being to "facilitate the adaptation of service members and their families into new working and living environments (Department of the Navy 2017)." This program aims to minimize any administrative issues associated with a PCS in order to allow the service member to focus on their performance while ensuring their family has access to proper command family resources.

Upon arrival to a new Naval command, the new service member is assigned to a more senior service member (sponsor) stationed at the same command. Sponsors are members of the command who can assist in a multitude of areas to include, but not limited to, challenges faced during in-processing, connecting with command ombudsman and family readiness groups, and using their experiences to help provide a valuable foundation of knowledge for the new service member regarding the local area. We focus on a formal mentorship opportunity, as it pertains to the CSIP, that occurs when a sponsor is assigned an incoming service member. Throughout this work we refer to sponsors as mentors, and to the incoming service members as mentees. We make this change to standardize and simplify the repeated referral of both parties throughout the work, as well as to emphasize the purpose of the CSIP.

The instruction only indicates four baseline requirements to be met for qualification as an appropriate sponsor, a outlined by the CNO. Every command adopts their own policy that may or may not require additional criteria to be met. These four requirements to qualify as a sponsor are:

- is of equal or higher paygrade;
- is at least an E-5 or a senior E-4 on their second tour of duty with a must promote or early promote on their most recent evaluation;
- is unmarried for single arriving members or married with children for arriving married members who also have children; and
- should have at least 12 months remaining on board the present command and should not be the person the incoming Sailor is slated to relieve. (Department of the Navy 2017)

We note that of the attributes that the Navy currently considers, only one attribute (marital status) pertains to the personal aspect of the mentor and mentee while the others focus on professional attributes. We believe that incorporating personal attributes for consideration when pairing mentors and mentees is important. People rely on socialization when forming relationships. In today's world, especially with increasing online interactions, connecting with a like-minded individual has never been more accessible.

Social media websites implement recommender systems to users, whether it be recommended content or users to keep participants engaged with the platform (Guy et al. 2010). Various algorithms focus on user profiles, user activity on the platform, or a combination of both. Regardless of the application, today's technology provides extra avenues for socialization that bring people together (Terveen and McDonald 2005). As a result, we believe performing a personalized pairing in the CSIP based on attributes suited to both a mentor and a mentee creates a greater potential in supporting the objective outlined by the CNO.

Is there an effective method to pair incoming and current personnel to act as a resource specifically for the incoming service member? While the military has gradually implemented formal mentoring programs across its branches of service, these programs are still relatively new. In a number of different case studies conducted throughout the military, several service members polled about mentorship responded affirmatively about its impact on their careers (Johnson and Andersen 2010). The CSIP acts as a Navy-wide-instituted formal mentoring program in the sense that mentees are assigned to experienced mentors that they can seek for assistance. We emphasize this formal mentoring program specifically within the scope of the CSIP and its purpose. However, the current instruction does not regulate mentor-mentee interactions themselves, and so this relationship may not always foster the interaction that fits the purpose of facilitating adaptation to a new living and working environment. As a result, the mentor-mentee pairing may conclude as unfruitful and there exists room for improvement. We know, however, that individualized mentorships prove to be more effective than mentorship in a group setting (Lester et al. 2011). We aim to explore a method that provides a higher chance of meaningful mentor-mentee relationships created through the CSIP.

In this work, we propose personalized mentorship assignment that has the potential benefits of these new mentor-mentee relationships. We conduct the personalized mentorship assignment based on attributes assigned to each mentor and mentee. In a real-world application, these attributes would be selected by each person in the form of a questionnaire prior to creating personalized pairings; for our proof of concept, we will randomly generate and assign attributes to each person.

We aim to pair each mentee with a mentor that offers the highest potential for creating positive interaction in hopes of translating to a more positive command environment. Mentorship sparks an overwhelmingly positive advancement for individual members and the collective group, and formal mentorship programs, while more difficult to perfect, provide a vehicle to stimulate growth among even more members of the military (Johnson and Andersen 2015). This interaction does not have to cease to exist upon a mentee's adjustment to the new command; in fact, this mutually beneficial effect may support personal performance as mentors and mentees build more sustainable relationships and in turn affect the command climate.

1.1 Overview

In Chapter 1 we introduce the problem statement and the impact associated with the United States Navy's CSIP. In Chapter 2 we dive into published articles regarding mentorship definition, its various forms and existing solutions, its application in the military, and any existing concepts from the network science domain that are needed to support our work. In Chapter 3 we introduce our methodology by outlining our process for creating a synthetic network and applying different metrics for measuring the pairing of mentors to mentees to answer the problem statement. In Chapter 4 we provide our results and takeaways discovered through our methodology. In Chapter 5 we conclude our analysis and lay out further directions for this work.

CHAPTER 2:
Background

In this chapter, we provide information on the topic of mentorship programs and network science. We begin this section with the definition of mentorship and the establishment of mentorship programs across different fields and their effects, and we conclude with a set of definitions and concepts related to network science, optimization, and statistics that we implement to analyze the Navy's current CSIP. We explore past research and analysis in this process.

2.1 Mentorship

Mentorship encompasses a large background of ideas, studies, and implementations. In this section we first provide a definition of mentorship and its various designs. We then identify characteristics contributing to effective mentorship and explore its application in several professions.

2.1.1 Definition

Due to its diversity in application, people may misunderstand the meaning of mentorship in different circumstances. For this paper, we use the following definition provided by the Welsh National Board for Nursing to guide our methodology:

Definition 2.1.1 *Mentorship*

> *Mentorship is reserved for long term relationships between people, one of whom usually is significantly older and/or more experienced than the other...the nature of the relationship is implicit in the term protégé suggesting as it does a recognition of potential and a concern for the individual's well-being, advancement and general progress. (Ravey et al. 1992)*

In this regard, the key difference between a mentor and a mentee is the level of experience.

Within this definition, there are different types of mentorship, specifically formal and informal (Henry-Noel et al. 2019). Mentors and mentees self-select for informal mentoring, usually initiated by the mentee, and there is no standardized, training, goal, or outcome. Formal mentoring, on the other hand, requires a regulated selection and training process. Organizations set checkpoints for the mentorships to meet, demanding a serious time commitment from both the mentor and the mentee. Informal mentoring can occur over longer terms whereas formal mentoring programs are much shorter-term, typically a year or less (Inzer and Crawford 2005). These differences present different strengths and weaknesses to a mentorship and are worth noting when assessing effectiveness.

Informal mentors provide more benefits than formal mentors according to a study conducted by Lonnie D. Inzer (Inzer and Crawford 2005). Informal mentors typically stem from knowing someone or from taking interest in one's field, but these connections are not natural for everyone. They challenge their respective mentee more in respect to career development skills, but, as a result, they facilitate more positive interactions, such as counseling, social interactions, and friendship. Mentees expressed greater satisfaction with their informal mentor than mentees with formal mentors. Formal mentoring does, however, factor in scalability across an organization and allows for more opportunities of mentorship.

Each type of mentorship utilizes various processes such as coaching, role modeling, and collaboration (Henry-Noel et al. 2019). These processes allow for proper guidance, a stronger relationship, and better decision-making skills between the mentor and the mentee. While the mentee achieves personal growth and new goals, the mentor hones in on their professional knowledge, skills, and abilities and stimulates their passion for their profession while teaching the upcoming generation in that career (Henry-Noel et al. 2019).

2.1.2 Attributes Contributing to an Effective Formal Mentorship

Regardless of the type of mentorship or processes defined, effective mentorship plays a significant role in the careers of the mentor and mentee. Mentors influence the career guidance, personal growth, and productivity of their respective mentees (Kibbe et al. 2016). As a result, this relationship can increase career satisfaction and thus, retention rates, highlighting the value a proper mentorship provides.

While studies have demonstrated that formal mentors may not achieve the same success as informal mentors, they are easier to scale across entire organizations (Inzer and Crawford 2005). Formal mentoring programs, if structured properly, allow for an effective relationship developed with the goal of improving the overall career and personal path of both the mentor and the mentee. This structure includes a few key characteristics that influence its effectiveness, such as an attentive matching process, high mentor commitment, and quality training to ensure the participants understand the program's goals (Chao 2009). The critical first step is the matching process in order to foster a positive initial interaction.

The National Athletic Trainers' Association (NATA) performed a qualitative study on 14 mentorships, with half of the mentorships involving those who have participated in formal mentorships and half who have participated in informal mentorships. While there were no group differences between those who experienced prior formal mentoring versus those who experienced prior informal mentoring, NATA concluded that one of the most important characteristics that contributes to a positive mentoring relationship is the presence of similar interests (Barrett et al. 2017). These similar interests include, but are not limited to research agendas, personalities, or environments at university.

We can use the takeaways from the NATA study and apply them to the military; a member of the military inevitably shares interests with other members of the military. For example, every military member has to PCS at some point in their career, often times more than once. As a result, the CSIP inherently utilizes related circumstances of members when assigning mentors to mentees, so mentors and mentees usually share this commonality. A study conducted in the healthcare community across several organizations concluded that tailored mentor programs prove to be an effective strategy in recruiting and retaining personnel in an increasingly stressful work environment (Jones 2017). Therefore, we believe that the CSIP's current mentor-mentee assignment can be modified to include a more individualized assignment using personal attributes that would foster a higher chance of communication and thus positive interaction.

An undergraduate medical school in Munich, Germany implemented a number of mentoring programs over a span of five years and experienced high success with an attribute-based online matching system (Pinilla et al. 2015). Nine out of every 10 mentor-mentee matchings originated from an online matching. Mentees filled out a questionnaire regarding their personal and professional interests and an algorithm provided a list of 10 potential mentors that shared some or all attributes. Although this method did not clearly prove to be more effective, it overwhelmingly stood out as the primary means for obtaining a mentor. It is also worth noting that this method is scalable over entire organizations while also preserving individualized mentor-mentee matchings. The circumstances surrounding the CSIP require mentor-mentee pairings across entire Naval commands. Therefore, a similarity-based matchmaking algorithm, where similarity corresponds to shared attributes, provides an opportunity to deliver mentor-mentee pairings while conserving effective, individualized matches.

By understanding the basic definition of mentorship and its implementation and effect through various organizations, we now introduce a network science approach to the United States Navy's CSIP aimed at assigning mentors to mentees according to similar attributes. We generate a random network and, using statistical computation and optimization techniques, we provide a solution for mentor-mentee pairings in our network.

2.2 Network Science

Before examining any network, we first present necessary graph theory and network science terminology. It is worth noting that the terms network and graph are used interchangeably in this thesis. However, in general, the term "network" implies a graph with more complexity and information tags associated with the nodes and edges. We utilize a network to generate people and attributes and assign respective attributes to each person. From this network, we then determine mentor-mentee pairings based on shared attributes among different people. To understand a network, we first define a graph, G, using Bollobás' definition.

Definition 2.2.1 *Graph*

> *A graph is an ordered pair of finite disjoint sets (V, E) such that E is a subset of the set V^2 of unordered pairs of V. The set V is the set of vertices and E is the set of edges. If G is a graph, then $V = V(G)$ is the vertex set of G, and $E = E(G)$ is the edge set. An edge {x, y} is said to join the vertices x and y and is denoted by xy. Thus xy and yx means exactly the same edge; the vertices x and y are the end vertices of this edge. (Bollobás 2013)*

Because we deal with two distinct groups of nodes (people and attributes), we must define and understand the bi-modal and bi-partite graphs. McHugh defines a bi-partite graph as:

Definition 2.2.2 *Bi-partite/Bi-modal*

> *A graph G is called a bi-partite graph (or bi-graph) if its vertex set $V(G)$ is the disjoint union of sets V_1 and V_2, and every edge in $E(G)$ has the form (v_1, v_2), where $v_1 \in V_1$ and $v_2 \in V_2$. A complete bipartite graph is a bi-graph in which every vertex in V_1 is adjacent to every vertex in V_2. A complete bi-graph depends only on the cardinalities M and N of V_1 and V_2 respectively, and so is denoted by K(M,N). Generally, we say a graph G or G(V_1,...,V_k,E) is k-partite if the vertex set V is the union of k disjoint sets V_1,...,V_k, and every edge in E(G) is of the form (v_i, v_j), for vertices v_i î V_i and v_j î V_j, V_i and V_j distinct. A complete k-partite graph is defined similarly. A synonymous term for bipartite is bi-modal which indicates two distinct sets of nodes in the graph. (McHugh 1990)*

We use the characteristics of a bi-partite/bi-modal graph in our simulation of the CSIP. We also introduce the term "network" in relation to the term "graph." Although networks are more complex than graphs, they still focus the development of a model used for analysis. Newman continues this connection by defining a network as:

Definition 2.2.3 *Network*

> *A network is a simplified representation that reduces a system to an abstract structure capturing only the basics of connection patterns and little else. (Newman 2018)*

More specifically, we create random networks in our methodology to simulate realistic circumstances at various Naval commands. Barabasi defines a random network as:

Definition 2.2.4 *Random Network*

> *A random network consists of N nodes where each node pair is connected with probability p. (Loscalzo and Barabási 2016)*

When we create our synthetic random network, we note that it possesses the bi-partite structure; we further specify it as a bi-modal network as its node set is of two different types, namely people and attributes. Splitting the nodes up into separate groups allows us to compare the people nodes to each other based on their shared nodes in the attribute group. The people set of nodes is comprised of the union between mentors and mentees, and each person has edges connecting to nodes in the attribute set corresponding their associated attributes. As a result, this random network is not completely random, as two people will not have any edges connecting them. Rather, edges only exist between nodes of separate sets, i.e. a person and an attribute. We use this network's adjacency matrix to calculate similarity coefficients among all potential mentor-mentee pairings using the edges between each person and attribute. McHugh defines an adjacency matrix as:

Definition 2.2.5 *Adjacency Matrix*

> *The adjacency matrix A of a graph $G(V, E)$ is a matrix of size $|V| \times |V|$ where the following is true for each element, denoted as A_{ij}:*

$$A_{ij} = \begin{cases} 1 & \textit{if } (i, j) \in E(G), \\ 0 & \textit{if } (i, j) \notin E(G) \end{cases} \quad (McHugh\ 1990) \tag{2.1}$$

We analyze the newly created adjacency matrix to determine mentor-mentee pairings, and we depict this output, per reference Department of the Navy (2017), with a tree structure. McHugh defines a tree as:

Definition 2.2.6 *Tree*

> *A tree is a hierarchical graph where each edge (known as a child) has exactly one parent (node from which it originates). If there is a parent node from*

which the whole structure arises then it is known as the rooted tree. It is easy to prove that the number of nodes in a tree equals the number of edges plus one, that is, $N = E+1$. The deletion of any edge break a tree into disconnected components. (McHugh 1990)

We note that each mentor that is assigned to mentees creates a tree; that is, the mentor is the root node, while each mentee is connected to that mentor. We now define the similarity metrics we use to compare potential mentor-mentee pairings in order to create our tree outputs.

2.3 Network Statistics

There are two similarity metrics we calculate across the network to use in selecting mentor-mentee pairs.

The first metric we use is the Pearson Correlation Coefficient. This coefficient calculates the correlation based on the degrees of two nodes and their neighbors. We utilize the adjacency matrix generated based on the edges between each person and their associated attributes.

Definition 2.3.1 *Pearson Correlation Coefficient*

The Pearson Correlation Coefficient is a metric that compares the number of common neighbors between two nodes to the expected number of common neighbors if the edges from both nodes were randomly selected (Gera 2020). We calculate the Pearson Correlation Coefficient, denoted r_{ij}, as:

$$r_{ij} = \frac{\sum_k [a_{ik} - < a_i >][a_{jk} - < a_j >]}{\sqrt{\sum_k [a_{ik} - < a_i >]^2}\sqrt{\sum_k [a_{jk} - < a_j >]^2}} \tag{2.2}$$

In this equation a_{ik} refers to the value in the i-th row and k-th column of the network's adjacency matrix. This value represents the number of edges between node i and node k. The term $< a_i >$ is the average value in the i-th row of the adjacency matrix and is $\frac{\deg(i)}{n}$, where n is the total number of nodes in the network. We sum over all k nodes in the network to compute the Pearson Correlation Coefficient between nodes i and j, which ranges from -1 to 1. Negative values represent two nodes who have fewer common neighbors than expected, and positive values represent two nodes who have more common neighbors than expected. A coefficient of zero indicates that two nodes have exactly the number of common neighbors expected. Thus, pairs of nodes with positive Pearson Correlation Coefficients are considered more similar than pairs with negative Pearson Correlation Coefficients.

The second metric we implement is the Jaccard Coefficient. This coefficient calculates the normalized number of common neighbors between two nodes.

Definition 2.3.2 *Jaccard Coefficient*

> *The Jaccard Coefficient is a metric that compares the number of common neighbors between two nodes to the total number of neighbors between two nodes (Leicht et al. 2006). We calculate the Jaccard Coefficient, denoted J_{ij}, as:*

$$J_{ij} = \frac{|N(i) \cap N(j)|}{|N(i) \cup N(j)|} \tag{2.3}$$

In this equation $N(i)$ refers to the set of nodes that neighbor node i, and $N(j)$ refers to the set of nodes that neighbor node j. For each node pairing i and j, we calculate the cardinality of both the intersection and the union of their respective neighbor sets. This value ranges from 0 to 1 and represents the proportion of shared neighbors to total neighbors. A value of 0 indicates no shared neighbors between nodes i and j, while a value of 1 indicates complete sharing of neighbors between nodes i and j. Therefore, pairs of nodes with Jaccard Coefficients closer to 1 are considered more similar than pairs of nodes with Jaccard Coefficients closer to 0.

We now define the computational framework for determining optimal mentor-mentee pairings specific to our methodology.

2.4 Optimization Modeling

We utilize optimization modeling to provide the optimal mentor-mentee pairings based on the similarity metrics defined. By optimizing the values of the similarity metrics among mentor-mentee pairs, we provide more opportunities for positive interaction among mentors and mentees by pairing the most similar mentors and mentees within the constraints of our optimization models. We input the similarity metrics defined above into their own respective integer programs with the intent of maximizing the respective sum of similarity coefficients for selected mentor-mentee pairs. To understand the purpose of integer programs, we first use Ronald Rardin's definition to introduce the broader category of mathematical programs:

Definition 2.4.1 *Mathematical Program*

> *The general form of a mathematical program or (single objective) optimization model is:*
>
> $$\text{min or max } f(x_1, \ldots, x_n)$$
> $$\text{subject to}$$
> $$g_i(x_1, \ldots, x_n) \leq b_i \quad i = 1, \ldots, m$$
>
> *where $f, g_1, \ldots, g_m$ are given functions of decision variables $x_1, \ldots, x_n$, and $b_1, \ldots, b_m$ are specified constant parameters. (Rardin 1998)*

Specifically, we implement two integer programs, one for each similarity metric. Rardin goes on to define integer programs and their unique characteristics:

Definition 2.4.2 *Integer Program*

> *An optimization model is an integer program (IP) if any one of its decision variables is discrete. If all variables are discrete, the model is a pure integer program; otherwise, it is a mixed-integer program (Rardin 1998).*

Because all of our decision variables are discrete, we technically employ two pure Integer Program (IP)s, but for the sake of understanding and simplicity, we label them as IPs. We use these IPs to select mentor-mentee pairings based on maximizing the respective similarity metrics of each pairing. We further define our IPs in Section 3.1.3.

We have introduced the definition of mentorship and discussed characteristics that impact its effectiveness. We then used this foundation to incorporate network science, statistics, and optimization concepts into mentorship. We create a mathematical model that addresses formal mentorship through the Navy's current CSIP and present this model in Chapter 3.

CHAPTER 3: Methodology

The literature in Chapter 2 demonstrates that, while mentorship programs have been established and studied across many professions, there are several ways to approach such programs for possible improved outcomes of the mentorship process. We propose exploring the benefits of a modified approach to conducting the CSIP based on attributes associated with prospective mentors and mentees in a given Naval command. To perform this modification, we introduce the Similarity-Based Node Pairing (SBNP) Model, a model that uses networks and attributes to provide an alternative process driven by attribute matching.

3.1 The Similarity-Based Node Pairing Model

The SBNP Model utilizes network structure and attribute assignment to ultimately guide mentor-mentee pairings. We describe the basic framework of the model in Section 3.1.1. We introduce our two similarity metrics in Section 3.1.2 and their respective IPs in Sections 3.1.4 and 3.1.5. These similarity metrics are used to determine mentor-mentee pairings based on levels of similarity between mentors and mentees.

The SBNP Model provides recommended mentor-mentee pairings based on attributes assigned to each person. We specifically center on the attributes of each person in the SBNP model because it enables us to personalize pairings based on said attributes. We define and calculate similarity between the possible people to be paired, based on the common attributes between a mentor and a mentee. Similar mentors and mentees share a majority of attributes, while dissimilar mentors and mentees share little to no attributes. As a result, we do not focus on any specific attributes in question, such as skills, hobbies, or personality, rather the network's structural relationship between a mentor's attributes and mentee's attributes. With the SBNP Model, we aim to maximize similarity across the network by pairing mentors and mentees that exhibit the highest similarity. By utilizing network structure containing people and attributes as nodes, and edges representing their relationships, we compare mentors and mentees based on their assigned attributes. We compute "similarity metrics" that we employ to select mentor-mentee pairings based on shared attributes among each other.

3.1.1 The SBNP Model Network

The first step of the SBNP Model is to generate a synthetic, static, bi-modal network whose nodes are divided into two partite sets: people (mentors/mentees) and attributes. The edges are undirected and exist between a person and an attribute if a person is assigned said attribute. In other words, edges exist only between a person and an attribute, not between nodes of the same partition in the bi-modal network. The SBNP Model includes variables to generate networks of different sizes for our simulations. Because no two Naval commands are the same, the input variables allow for the flexibility to build a network that can be fit to any command size.

The variables we use in this first step to create the synthetic network are:

- number_of_people,
- percentage_of_mentors, with 0 < percentage_of_mentors ≤ 0.5,
- number_of_attribute_sets, and
- number_of_choices_per_attribute_set,
  with number_of_choices_per_attribute_set ≥ 2.

We now provide an intuition behind the variables. The number_of_people variable specifies the number of people in the network, namely the mentor and mentees together. The percentage_of_mentors variable specifies a value between 0 and 0.5 that is multiplied to the number_of_people to obtain the number of mentors in the network; the difference between the number_of_people and the number of mentors is the number of mentees. We restrict the percentage of mentors to below 0.5 to ensure we don't have more mentors than mentees. In our literature review, we determined that a typical mentorship program has one mentor for every one to three mentees (Pinilla et al. 2015). Additionally, we have the number_of_attribute_sets variable that specifies the number of unique attribute sets to create in the network – these capture the different questions users can answer to describe themselves based on their association or preferences.

We utilize multiple attribute sets to capture a more realistic network, where people use more than one attribute to describe themselves. For example, when asked to describe themselves, unless instructed to do so in one word, people will use multiple skills, hobbies, or personality traits that interest them. Multiple attribute sets to select from also provide a more realistic solution, as the SBNP Model does not determine pairings solely off of a single attribute. The number_of_choices_per_attribute_set variable specifies how many unique attributes to create in each attribute set. We restrict the number of choices per attribute set to any integer greater than or equal to 2. We provide more than one attribute to choose per attribute set in order to realistically simulate different inputs from various mentors and mentees, otherwise every person would have the exact same attributes and similarity computation would be redundant. Additionally, we recognize that certain attribute sets inherently contain a broader scope of answers than others. Therefore, allowing for flexibility in the number of choices per attribute set based on the attribute set simulates a more realistic scenario in a given Naval command. For the purposes of this work, however, the number_of_choices_per_attribute_set input value applies to every attribute set.

We are now ready to generate the synthetic network. For our work, only one attribute per attribute set created is assigned to each person. This constraint allows us to capture each attribute set's selected attribute and compute similarity metrics among different people based on the count of shared attributes and the categories they come from. We generate a random distribution that each attribute set follows and randomly assigns an attribute for each person. By randomly assigning these attributes, our work is simulating the diverse choices of a command environment, where people have different attributes that describe them. The newly created synthetic network captures these assignments of attributes to people through the edges that connect nodes from the two partite sets, namely if a person associates themselves with an attribute.

3.1.2 The SBNP Model Adjacency Matrix

The next step of the SBNP Model is to create the adjacency matrix of the synthetic network we just built, as defined in Equation 2.2.5. The ordering of the elements of the adjacency matrix begins with mentors, then mentees, and finally attributes. Any edge in our network between a person and attribute corresponds to a 1 in the matrix between the person and the attribute, otherwise it is 0. Using the adjacency matrix, we calculate both the Pearson coefficient and Jaccard coefficient of each potential mentor-mentee pairing. These coefficients serve as the two similarity metrics we use to pair mentors and mentees. We are calculating two different coefficients to observe any potential unique effects each coefficient may have on selecting mentor-mentee pairs.

We input each similarity metric into its own respective IP to maximize network similarity according to that metric. At this point, to allow for comparison of results, we proceed with two different methods: a random mentor-mentee assignment and a SBNP optimal mentor-mentee assignment. We note that the SBNP optimal assignment contains two separate solutions: optimal assignment by Pearson coefficient and optimal assignment by Jaccard coefficient.

We create an IP formatted for each similarity metric with their respective objectives aimed at maximizing the similarity across the entire synthetic network. This objective enables us to focus on selecting mentor-mentee pairs across the network that share more similarity than a random pairing. By aiming to maximize network similarity, we are selecting mentor-mentee pairs with the greatest similarity while ensuring each mentee is considered. Each IP's respective objective function sums up all of the coefficients of the mentor-mentee pairs selected.

The Pearson IP considers potential mentor-mentee pairings based on their Pearson coefficients, with the objective function adding up all of the Pearson coefficients of selected mentor-mentee pairs. Likewise, the Jaccard IP considers potential mentor-mentee pairings based on their Jaccard coefficients, with the objective function adding up all of the Jaccard coefficients of selected mentor-mentee pairs. We compute the sums of the Pearson coefficients and Jaccard coefficients for the random pairing and compare them to their respective SBNP optimal objective values. We expect to notice a higher sum of Pearson and Jaccard coefficients for the SBNP optimal solutions than the random pairing. This difference in values indicates that the mentor-mentee pairs are more similar in the SBNP optimal solutions than in the random pairing, and as a result, the overall network possesses greater similarity.

3.1.3 Notation for the Integer Programs

Before we present the IPs implemented across each similarity metric, we introduce the notation that is common to both IPs.

Sets

I := the set of all mentors.
J := the set of all mentees.

Decision Variables

x_{ij} := 1 if mentor i is assigned to mentee j, 0 otherwise, for all $i \in I, j \in J$.
y_i := 1 if mentor i is selected, 0 otherwise, for all $i \in I$.
z_j := 1 if mentee j is selected, 0 otherwise, for all $j \in J$.

3.1.4 Pearson Integer Program

In the Pearson IP, the objective is to maximize the sum of the Pearson coefficients for mentor/mentee pairs that are selected. There are four (classes of) constraints in the IP. Constraint (1) requires that both the mentor and mentee are selected before they can be paired together. Constraint (2) enforces that exactly one mentor is assigned to a mentee. Constraint (3) ensures that every mentee is assigned to a mentor. Constraint (4) ensures that no mentor is assigned more than the ceiling of the quotient of number of mentees divided by number of mentors.

Parameters

r_{ij} := the Pearson coefficient of mentor i and mentee j, for all $i \in I, j \in J$.

Model

$$\text{Maximize } Z_{pearson} := \sum_{i \in I, j \in J} r_{ij} x_{ij}$$

subject to

$$\begin{aligned}
x_{ij} &\le y_i z_j \ , \ \forall\, i \in I, \ j \in J && (1)\\
\sum_{i \in I} x_{ij} &= 1 \ , \ \forall\, j \in J && (2)\\
\sum_{i \in I, j \in J} x_{ij} &= |J| && (3)\\
\sum_{j \in J} x_{ij} &\le \left\lceil \frac{|J|}{|I|} \right\rceil, \ \forall\, i \in I && (4)
\end{aligned}$$

3.1.5 Jaccard Integer Program

In the Jaccard IP, the objective is to maximize the sum of the Jaccard coefficients for mentor/mentee pairs that are selected. The same four (classes of) constraints from the Pearson IP apply.

PARAMETERS

c_{ij} := the Jaccard coefficient of mentor i and mentee j, for all $i \in I, j \in J$.

MODEL

$$\text{Maximize } Z_{jaccard} := \sum_{i \in I, j \in J} c_{ij} x_{ij}$$

subject to

$$\begin{aligned}
x_{ij} &\le y_i z_j \ , \ \forall\, i \in I, \ j \in J && (1)\\
\sum_{i \in I} x_{ij} &= 1 \ , \ \forall\, j \in J && (2)\\
\sum_{i \in I, j \in J} x_{ij} &= |J| && (3)\\
\sum_{j \in J} x_{ij} &\le \left\lceil \frac{|J|}{|I|} \right\rceil, \ \forall\, i \in I && (4)
\end{aligned}$$

3.1.6 Pseudo Code for Methodology

We summarize the SBNP Model described in Sections 3.1.1 - 3.1.5 by introducing pseudo code in Algorithm 1. This pseudo code provides a step-by-step outline of the SBNP process that can be referenced or recreated in further pieces.

Algorithm 1 SBNP Model

```
Input number_of_people, percentage_of_mentors, number_of_attribute_sets, number_of_choices_per_attribute_set
Create synthetic network
Create nodes for every person and attribute and add to network
Randomize a distribution applied to each attribute set
for Every person do
    for Every attribute set do
        Randomly assign one attribute from each attribute set to each person
        Create edge from person to attribute and add to network
    end for
end for
Create adjacency matrix for network (Equation 2.2.5)
for Every mentor do
    for Every mentee do
        Calculate Pearson coefficient (Equation 2.3.1)
        Calculate Jaccard coefficient (Equation 2.3.2)
    end for
end for
Randomly select mentor-mentee pairings
for Each similarity metric do
    Run similarity metric IP to determine SBNP optimal pairing and SBNP optimal objective value
    Sum up coefficients of random pairing
    Compare SBNP optimal pairings to random pairings and their respective objective values (sums)
end for
```

3.2 Example SBNP Model

To properly depict and validate the SBNP Model structure, for this section we introduce a small visual example. We input the following values for the variables:

- number_of_people = 10
- percentage_of_mentors = 0.3
- number_of_attribute_sets = 5
- number_of_choices_per_attribute_set = 4.

We note that we input number_of_choices_per_attribute_set = 4. This creates the same number of unique attributes per attribute set. As a result, attribute assignment acts similar to a multiple-choice questionnaire that each person fills out, with their answers corresponding to the attributes in each attribute set with which they most closely associate. We decide to constrict the number of attributes per attribute set to one value to standardize the computation of similarity metrics and clearly illustrate the SBNP model's process. In reality, each attribute set is likely to have its own unique number of attribute choices.

With the values we specify above, we create a synthetic network composed of 10 people and 20 attributes. Of the 10 people, three are mentors and the remaining seven are mentees. Of the 20 attributes, there are five attribute sets with four attribute choices in each set, as shown in Table 3.1. We create nodes in our network for each person and for each attribute, totalling 30 nodes.

Table 3.1. Attribute sets in the example SBNP Model.

Attribute Set 0	Attribute Set 1	Attribute Set 2	Attribute Set 3	Attribute Set 4
att0	att4	att8	att12	att16
att1	att5	att9	att13	att17
att2	att6	att10	att14	att18
att3	att7	att11	att15	att19

Focusing on the attributes, we randomize a distribution of attribute choices per attribute set. This distribution prescribes a weighted assignment of each attribute choice in a given attribute set. For our example, the attribute distribution is [2,1,2,5]. This distribution tells us that, on average, for every 10 attribute assignments within a single attribute set, the fourth attribute choice in an attribute set is selected five times. Theoretically, this fourth attribute in each attribute set should be assigned to five people, but with a small sample size this characteristic may not always hold true.

Focusing on the people now, we iterate through each person and assign an attribute from each attribute set based on the randomized distribution, leaving each person with five attributes assigned to them. To finish building our network, we add edges between each person and their attributes, totaling 50 edges. The resulting bi-modal network is pictured in Figure 3.1, showing two types of nodes, namely people and the choices of attributes they can attach to, while the edges capture the attributes assigned to each person.

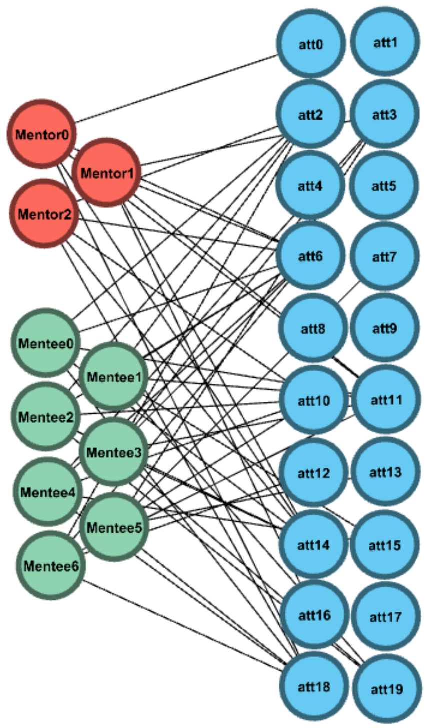

Figure 3.1. Bi-modal Network: Mentors and mentees are randomly assigned attributes and edges only connect mentor/attribute or mentee/attribute.

We then create the 30×30 adjacency matrix for the network introduced in Figure 3.1. The ordering of the elements of the adjacency matrix is as follows: Mentor0, Mentor1, Mentor2, Mentee0, Mentee1, . . . , Mentee6, att0, att1, . . . , att19. The example we introduce in Matrix 3.1 captures the adjacency of Mentor0 in the first row, of Mentor1 in the second row, showing the same pattern down to att19 in the 30^{th} row.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ \vdots & & \ddots & & & & & & & & & & & & & & & & & & & & & & & & & & \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{3.1}$$

From this adjacency matrix, we calculate the Pearson and Jaccard coefficients for every potential mentor-mentee pairing. For the example in Figure 3.1, with three mentors and seven mentees, we have 21 potential mentor-mentee pairings. We obtain the following values for each pairing and display them in Table 3.2.

Table 3.2. Potential mentor-mentee pairing coefficient values in the example SBNP Model.

Mentor	Mentee	Pearson	Jaccard
0	0	0.747	0.667
0	1	0.242	0.250
0	2	0.242	0.250
0	3	0.495	0.429
0	4	0.242	0.250
0	5	-0.011	0.111
0	6	0.495	0.429
1	0	0.495	0.429
1	1	0.242	0.250
1	2	0.242	0.250
1	3	0.747	0.667
1	4	0.242	0.250
1	5	-0.011	0.111
1	6	0.242	0.250
2	0	0.747	0.667
2	1	0.242	0.250
2	2	0.747	0.667
2	3	0.242	0.250
2	4	0.495	0.429
2	5	0.242	0.250
2	6	0.495	0.429

Following coefficient calculations, we now introduce the three different ways we compute the overall network similarity, followed by the comparison of each method's results in Figure 3.2.

Random mentor-mentee pairing

We first randomly select mentor-mentee pairs to use as a base comparison. This process follows the same constraints applied in both SBNP IPs, but there is no objective function. Once the mentor-mentee pairings are selected, we sum the respective Pearson coefficients and Jaccard coefficients of the randomly selected pairings to compare to the objective values of each IP.

Similarity-Based Node Pairing using Pearson coefficients

Using the Pearson coefficients calculated, we input these coefficients as a parameter into an IP. The objective function of the IP is to maximize the summation of the Pearson coefficients of selected mentor-mentee pairings. This SBNP Pearson optimal solution corresponds to a network containing the greatest global similarity according to the Pearson similarity metric.

Similarity-Based Node Pairing using Jaccard coefficients

Following a similar process to the Pearson IP, we input the Jaccard coefficients as a parameter into its own IP. This IP maximizes the summation of the Jaccard coefficients for selected mentor-mentee pairings. This SBNP Jaccard optimal solution corresponds to a network containing the greatest global similarity according to the Jaccard similarity metric.

We note that in Figure 3.2, for each similarity metric, the SBNP optimal objective value is greater than the respective random pairing objective value. This indicates that the IPs produce more pairings across the network with greater similarity values than the random pairing.

We also note that in this example, the Pearson IP and the Jaccard IP both select the same mentor-mentee pairings. We revisit this observation in our analysis in Chapter 4 to observe whether these metrics consistently select the same pairings.

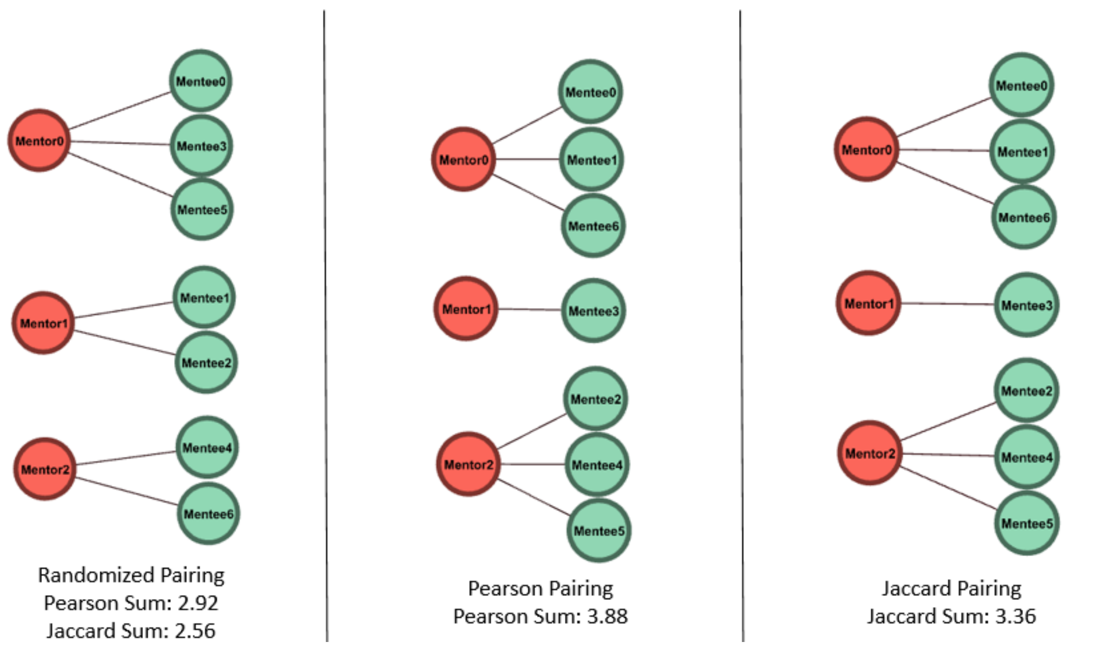

Figure 3.2. Mentor-Mentee Pairings: Mentors and Mentees are paired randomly, SBNP Pearson optimally, and SBNP Jaccard optimally. Their objective values are reported underneath each pairing.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 4:
Results and Analysis

We established the workflow for the SBNP Model and presented a visual example in Chapter 3. In Chapter 4, we conduct results and analysis for the SBNP Model when executed with various input values for the variables. We use Python programming language, particularly the NetworkX packages, to build a network and perform the SBNP Model that provides mentor-mentee pairings. We then define a base case model and modify the input variable values to both analyze the impact of each variable on the effectiveness of the SBNP Model, and to extrapolate this impact to implementation in Naval commands of different sizes.

4.1 Base Case SBNP Model

In order to perform an analysis on each input variable, we define a base case with set input variable values. We note that we assign placeholder letters for each variable for ease of constant reference throughout Chapter 4. For each input variable in the base case, we create a range of values surrounding these base case variable values and produce results for each range of values. We analyze the sensitivity of each variable in regard to its impact on the SBNP Model results.

The base case SBNP Model uses the following input variable values:

- n = number_of_people = 100,
- p = percentage_of_mentors = 0.33,
- s = number_of_attribute_sets = 10, and
- u = number_of_choices_per_attribute_set = 20.

We select these values as a base case to provide a median reference value for each variable's range of values. The number_of_people value provides a standard example with enough choices to provide realistic results. The percentage_of_mentors value considers the study stating that the typical mentor has one to three mentees in a mentorship program and represents the 1:2 median ratio (Pinilla et al. 2015). Both the number_of_attribute_sets and number_of_choices_per_attribute_set values simulate a realistic attribute questionnaire with a limited number of questions that each have several answer choices. This setup allows for a scalable questionnaire that can be developed and applied to any Naval command. The length of the questionnaire remains concise with the number of questions yet contains enough answer choices per question to appeal to a variety of preferences.

We input these values into the SBNP Model and run 10 simulations to eliminate the effect of any extreme results. The average of these results are listed in Table 4.1, where "Pearson Random" captures the sum of the Pearson coefficients for the random pairings and "SBNP Pearson Optimal" captures the sum of the Pearson coefficients for the SBNP Pearson optimal pairings. Likewise, we compare the sum of the Jaccard coefficients for the random pairings, "Jaccard Random," to the sum of the Jaccard coefficients in the "SBNP Jaccard Optimal" pairings:

Table 4.1. SBNP Model base case results.

Method	Coefficient Sum
Pearson Random	20.128
SBNP Pearson Optimal	38.541
Jaccard Random	14.276
SBNP Jaccard Optimal	29.324

Over the 10 simulations, the SBNP values of the optimal pairings are nearly twice as large as the values for the random pairing in relation to the Pearson coefficients, and slightly above twice as large as the values of the Jaccard coefficients. While this doesn't necessarily mean that the matching is twice as effective, it provides superior matching of mentor-mentees, providing greater opportunity for the positive interaction between a mentor and a mentee.

Additionally, we note that each simulation fully executed in seconds, with the total time for 10 simulations being 39.72 seconds. We also note that, on average, the SBNP Pearson optimal pairings and SBNP Jaccard optimal pairings shared 74.18% of the same mentor-mentee pairs. We now use this base case as a reference to observe the effect on our solution when we perform sensitivity analysis for each variable. Since the Pearson coefficient measures similarity of two nodes based on the common number of friends relative to a similar situation in a random graph, the Pearson optimal pairings identify two nodes that are expected to be related based on their associated attributes in different networks of related structure. On the other hand, the Jaccard coefficient measures the similarity of two nodes based on the proportion of common friends out of potential common friends, and thus the Jaccard optimal pairings capture a ratio indicating how similar two nodes are in the specific network.

4.2 Input Variable Sensitivity Analysis

We now isolate each input variable and observe the effects on the solution. We use the base case defined in Section 4.1 as a reference in our sensitivity analysis. We select one variable at a time and conduct simulations with a range of values while the other three variables keep their base case input value. To match the base case analysis, we perform 10 simulations at each different value to mitigate extreme results' influence on our analysis. We record the average coefficient sums of each similarity metric in both the random pairing and optimal pairings, the average shared percentage of pairings between SBNP Pearson optimal and SBNP Jaccard optimal solutions, and the total solve time at each variable value. We note that, in each results table, the base case results are shown in bold for clarification and reference. We analyze the results of each variable's impact below.

4.2.1 Number of People

To isolate the number_of_people variable, we define the following variables for each simulation:

- p = percentage_of_mentors = 0.33,
- s = number_of_attribute_sets = 10, and
- u = number_of_choices_per_attribute_set = 20.

We test five separate input values for our number_of_people variable, ranging from 10 people to 500 people. Similar to the base case, we record each method's respective coefficient sums, as well as the average shared pairings between both SBNP optimal pairings and the average total time for 10 simulations. We present the results in Table 4.2 for each number_of_people input value saved as a different column of the table, while the rows depict the values of the Pearson and Jaccard coefficients for a random pairing versus their respective SBNP optimal pairing.

Table 4.2. SBNP Model results for each number_of_people value.

	n = 10	n = 50	**n = 100**	n = 250	n = 500
Pearson Random	0.194	5.548	**20.128**	64.153	153.555
SBNP Pearson Optimal	0.259	13.250	**38.541**	121.609	267.628
Jaccard Random	0.259	3.856	**14.276**	46.246	115.905
SBNP Jaccard Optimal	0.541	9.205	**29.324**	100.040	233.010
Common Pairing (%)	87.14	77.05	**74.18**	68.452	56.42
Solve Time (sec)	1.88	8.43	**39.72**	253.351	1687.43

With every number_of_people value, the SBNP optimal solutions overwhelmingly outperform the random pairing, showing that the SBNP model provides superior solutions.

We also compare these coefficient sum values visually in Figure 4.1. The x-axis depicts each n-value we implement, while the y-axis shows the average coefficient sum for each respective output.

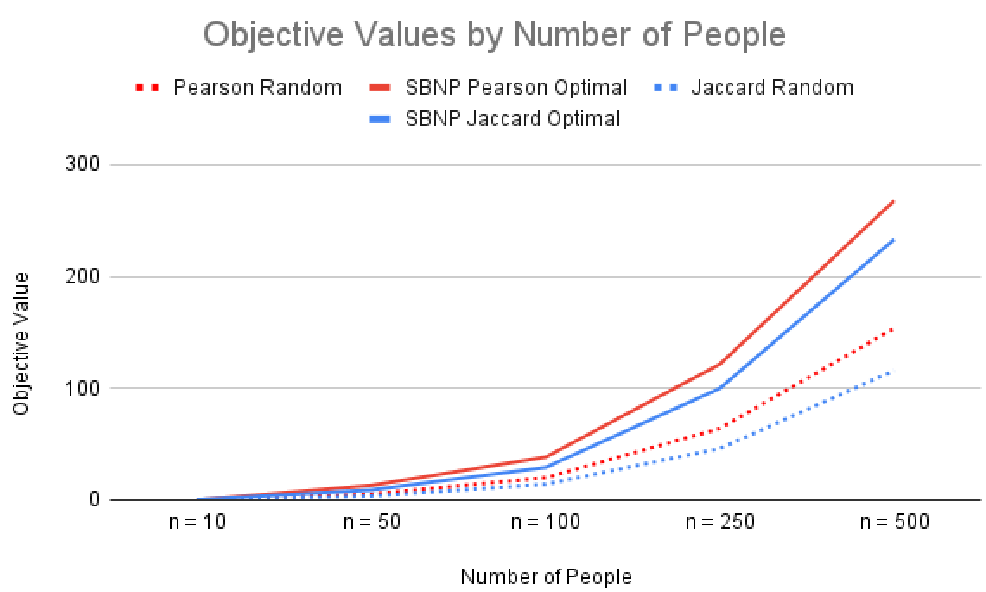

Figure 4.1. Number of People Comparison: The relationship between each method's objective value is compared across each number_of_people value.

We note that, for each input value for n, each SBNP optimal pairing produces objective values that are either close to double their respective random pairing coefficient sum or better. The closest results between SBNP optimal pairings and the random pairing is when $n = 10$; this result understandably provides closer results since there are 20 attribute choices per attribute set assigned to 10 people, so the likelihood that no two people have the same attribute per attribute set is higher than the base case.

Due to the different numbers of people and thus mentor-mentee pairings, we cannot directly compare objective values between networks of different size. We calculate, however, the average similarity coefficient between each mentor-mentee pairing and display the results in Table 4.3.

Table 4.3. Average Similarity Comparison: Average similarity coefficients per mentor-mentee pairing compared by number_of_people value.

	$n = 10$	$n = 50$	$\boldsymbol{n = 100}$	$n = 250$	$n = 500$
SBNP Pearson Optimal	0.043	0.402	**0.575**	0.728	0.799
SBNP Jaccard Optimal	0.090	0.279	**0.438**	0.599	0.696

We observe, that, when $n = 10$, the average similarity coefficients for the SBNP optimal pairings are 0.043 for Pearson and 0.090 for Jaccard. When $n = 500$, however, the average similarity coefficients for the SBNP optimal pairings are 0.799 for Pearson and 0.696 for Jaccard.

We identify an increasing average similarity per mentor-mentee pairing as the number of people in the network increases. This relationship relies heavily on the person-to-attribute ratio; that is, when the number of people begins to exceed the number of total attributes in the network, people will inevitably be constrained by the different attributes they can choose. With more people and the same number of attributes, the chances of people selecting the same attributes increases with each person added. While we do indicate the relationship and impact between the number of people and number of attributes in the network, we do not recommend a specific person-to-attribute ratio to obey.

The SBNP Pearson optimal pairing shares less mentor-mentee pairs with the SBNP Jaccard optimal pairings due to the increase in the number of potential pairings. The network contains more choices between mentors and mentees than the base case, and a larger number of people increases the chances of two people sharing more attributes with each other. As a result, while the SBNP Pearson IP might indicate a specific optimal mentor-mentee pairing, the SBNP Jaccard IP might suggest a different optimal mentor-mentee pairing for said mentor and mentee. Since the Jaccard coefficient measures the count of common friends normalized by the size of their combined neighborhood, it captures the potential friendships between its neighbors independent of the network's size or structure. On the other hand, the Pearson correlation coefficient compares if two nodes have more or less neighbors when compared to similar degree nodes in a random network, thus capturing the network's behavior locally versus what is expected that chance would create. These observations indicate that the Jaccard IP aims to provide a locally optimal solution to the immediate network while the Pearson IP aims to provide a globally optimal solution to different networks of similar structure.

We also note the relationship between the number of nodes and the total solve time. With 10-person network generation, the average solve time for a single iteration of the SBNP Model was 0.188 seconds. On the other hand, with 500-person network generation, the average solve time for a single iteration of the SBNP Model was 168.743 seconds. If we relate these solve times to amount of time required per person, we obtain a 0.019 second-per-person solve time versus a 0.337 second-per-person solve time. This relationship makes sense since, for every additional person, we also add additional constraints to each IP that accounts for said person. Therefore, it is important to note that the SBNP Model requires exponentially more time for every person added.

4.2.2 Percentage of Mentors

We isolate the percentage_of_mentors variable by defining the following variables for this simulation:

- n = number_of_people = 100,
- s = number_of_attribute_sets = 10, and
- u = number_of_choices_per_attribute_set = 20.

We simulate with five different values for percentage_of_mentors, ranging from 0.1 to

0.5. We multiply this value to the number_of_people input value to obtain the number of mentors in our network; the difference between the number_of_people input and the number of mentors gives us the number of mentees. Following the same process we used in Section 4.2.1, we present the results for our sensitivity analysis on percentage_of_mentors in Table 4.4 for different percentages of mentors captured in the columns:

Table 4.4. SBNP Model results for each percentage_of_mentors value.

	$p = 0.1$	$p = 0.2$	$\boldsymbol{p = 0.33}$	$p = 0.4$	$p = 0.5$
Pearson Random	26.821	29.680	**20.128**	12.959	11.321
SBNP Pearson Optimal	45.207	49.710	**38.541**	32.400	28.690
Jaccard Random	19.018	21.750	**14.276**	8.832	8.007
SBNP Jaccard Optimal	33.411	39.130	**29.324**	23.808	21.940
Common Pairing (%)	76.78	68.63	**74.18**	73.67	67.4
Solve Time (sec)	17.67	25.81	**39.72**	36.359	36.218

Once again, our SBNP IPs outperform the random pairing. We note that, as we increase the percentage of mentors in the network, the SBNP optimal pairings produce objective values much larger relative to their random pairing counterpart. The SBNP optimal pairings produce coefficients sums more than twice as large as the random pairing coefficient sums for $p = 0.33$ and greater.

We present Figure 4.2, where the x-axis shows the different values of p we test with their corresponding average objective values of the random and each SBNP optimal pairings on the y-axis. This plot reinforces the idea that a higher percentage of mentors corresponds to a larger selection of mentors per mentee. As a result, there exists a greater chance of a mentee matching with a similar mentor.

We note that there is no clear relationship between the percentage of mentors and the percentage of shared mentor-mentee pairs between each SBNP optimal IP. We cannot determine a logical interpretation of the results based on these two factors. While a larger pool of mentors allows for a larger selection for each mentee, there is no guarantee that a mentee matched to a mentor in one of the SBNP IPs will fail to be matched in the other SBNP IP. Regardless, the percentage of shared mentor-mentee pairs remains at least two-thirds of

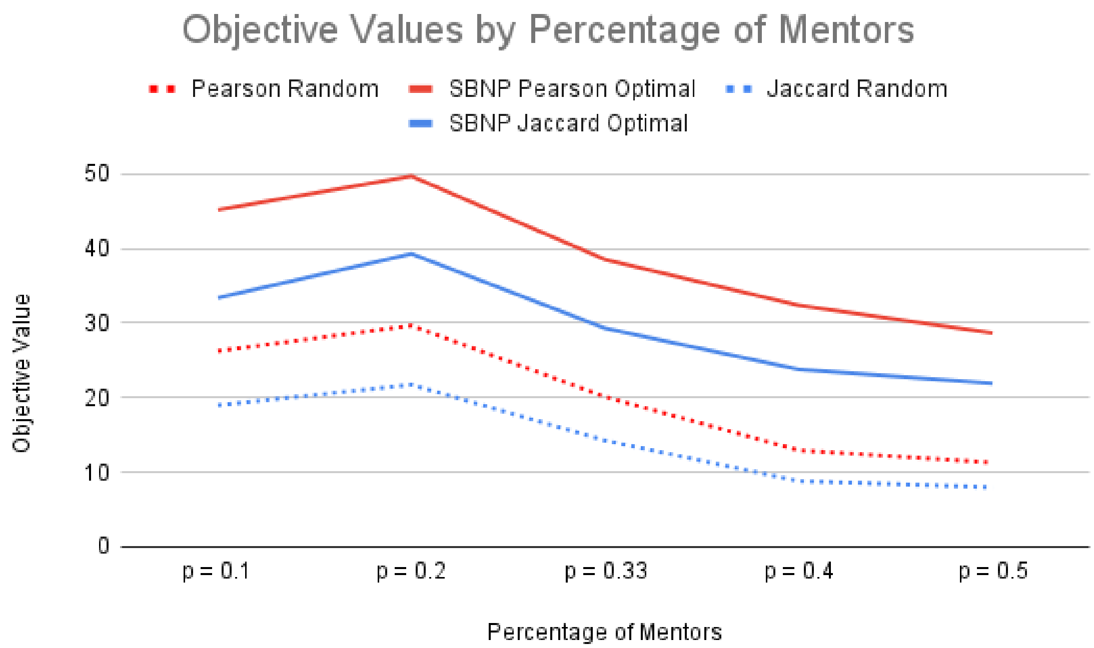

Figure 4.2. Percentage of Mentors Comparison: The relationship between each method's objective value is compared across each percentage_of_mentors value.

all pairings, upwards of three-fourths of all pairings.

While the objective value decreases as the percentage of mentors increases, we also note that the number of mentor-mentee pairings also decreases. This relationship shows us that, similar to Section 4.2.1, we cannot directly compare objective values for each percentage_of_mentors value. We assess the average mentor-mentee pairing similarity coefficients at each percentage_of_mentors input and record the results in Table 4.5. We identify $p = 0.2$ to produce the largest average similarity coefficients per mentor-mentee pairing, however, this average is not significantly larger than averages of other percentage_of_mentors values. Therefore, we cannot identify a trend corresponding to the percentage_of_mentors and their respective average similarity coefficients per pairing.

The percentage of mentors did not significantly affect the solve time of the simulations. This variable affects the size of the set of mentors, $|I|$, and the set of mentees, $|J|$. While these sets directly determine the number of constraints in each IP, their variations in size do not

Table 4.5. Average Similarity Comparison: Average similarity coefficients per mentor-mentee pairing compared by percentage_of_mentors value.

	$p = 0.1$	$p = 0.2$	$\boldsymbol{p = 0.33}$	$p = 0.4$	$p = 0.5$
SBNP Pearson Optimal	0.502	0.621	**0.575**	0.540	0.574
SBNP Jaccard Optimal	0.371	0.489	**0.438**	0.397	0.439

affect the number of constraints as drastically as the number_of_people variable.

We determine the number of constraints in constraint class (1), defined in Section 3.1.4, by multiplying $|I| \cdot |J|$. For example, if $p = 0.1$, the network contains 10 mentors and 90 mentees; the total constraints in class (1) will be $10 \cdot 90 = 900$. Now suppose $p = 0.5$—the network contains 50 mentors and 50 mentees. The total constraints in class (1) would now be $50 \cdot 50 = 2500$.

We observe the difference in the number of constraints, however, this increase in constraints is also balanced with the number of mentor-mentee pairings. Using the same example, with $p = 0.1$, the network requires 90 pairings, whereas when $p = 0.5$, the network only requires 50 pairings. Modifications in the percentage_of_mentors do not notably impact the solve time given the current input variables, but it is worth considering as the number_of_people in the network increases.

4.2.3 Number of Attribute Sets

We now move to isolating the number_of_attribute_sets variable. In order to do so, we define the following variables:

- n = number_of_people = 100,
- p = percentage_of_mentors = 0.33, and
- s = number_of_choices_per_attribute_set = 20.

Using five different values for number_of_attribute_sets, ranging from 1 to 20, we present the results for our sensitivity analysis on number_of_attribute_sets captured as columns in Table 4.6:

Table 4.6. SBNP Model results for each number_of_attribute_sets value.

	s = 1	s = 5	**s = 10**	s = 15	s = 20
Pearson Random	20.877	20.549	**20.128**	21.385	12.707
SBNP Pearson Optimal	66.605	44.477	**38.541**	36.988	26.896
Jaccard Random	21.2	14.615	**14.276**	14.892	8.686
SBNP Jaccard Optimal	55.7	36.009	**29.324**	27.535	28.568
Common Pairing (%)	19.10	57.16	**74.18**	74.02	84.18
Solve Time (sec)	25.31	28.13	**39.72**	34.89	40.28

We observe that the SBNP optimal pairings produce coefficient sums almost triple the value of their respective random pairing coefficient sums when $s = 1$. As s increases, this multiple decreases to just about twice as large.

We plot these results in Figure 4.3. where the x-axis displays the different values of s we test with their corresponding average objective values of the random pairing and each SBNP optimal pairings on the y-axis.

We notice that, as s increases, the SBNP optimal objective values decrease. Because we introduce a greater selection of attributes to each person with each attribute set, there exists a greater combination of attributes that each person may possess with each additional attribute set. As a result, two people are not as likely to have as large of a similarity coefficient, whether it be Pearson or Jaccard, as they did with less attribute sets. We confirm this observation by comparing the objective value when $s = 1$ (66.605) versus when $s = 20$ (26.896). Mentor-mentee pairings reveal a much larger coefficient when pairing based on one attribute set as opposed to pairing based on 20 attribute sets.

The percentage of shared mentor-mentee pairings increases as the number_of_attribute_sets increases. We find this relationship interesting; as we increase the number of attribute sets, the most similar mentor-mentee pairings are more apparent. That is, the variability in the similarity metrics between each mentor-mentee pairing is more precise as more attribute sets are added. While the average similarity coefficient per mentor-mentee pairing may

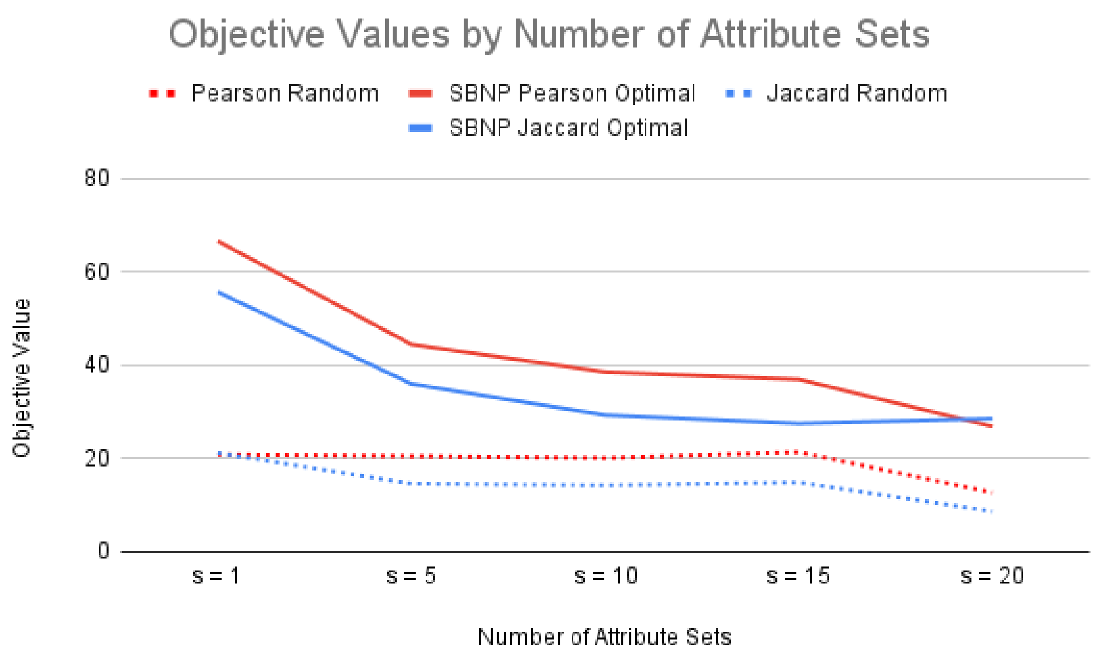

Figure 4.3. Number of Attribute Sets Comparison: The relationship between each method's objective value is compared across each number_of_attribute_sets value.

decrease, the most similar mentor-mentee pairings become clear due to the higher number of attribute combinations a person could possess.

The number_of_attribute_sets did not affect the solve time drastically. We note that, as we add attribute sets to the network, the SBNP Model requires more time to assign attributes to each person and calculate the similarity coefficients. The SBNP IPs are not affected by this variable.

4.2.4 Number of Attribute Choices Per Set

Finally, we isolate the number_of_choices_per_attribute_set by defining the other three input variables as:

- n = number_of_people = 100,
- p = percentage_of_mentors = 0.33, and

- s = number_of_attribute_sets = 10.

We choose five different number_of_choices_per_attribute_set values to test on our SBNP Model, ranging from 5 to 35 choices per attribute set. We present the results of each group of simulations in Table 4.7, where each column displays choices of number_of_choices_per_attribute_set for comparison purposes:

Table 4.7. SBNP Model results for each number_of_choices_per_attribute_set value.

	$u = 5$	$u = 12$	$\boldsymbol{u = 20}$	$u = 28$	$u = 35$
Pearson Random	30.557	26.631	**20.128**	14.375	14.822
SBNP Pearson Optimal	50.168	45.650	**38.541**	32.050	31.880
Jaccard Random	25.091	19.929	**14.276**	9.603	9.766
SBNP Jaccard Optimal	43.755	37.342	**29.324**	22.766	22.423
Common Pairing (%)	67.61	68.36	**74.18**	71.493	75.373
Solve Time (sec)	26.689	28.79	**39.72**	34.24	36.467

This table illustrates that, as u increases, the SBNP optimal pairings producer larger coefficient sums than the random pairings. The smallest multiple we observe is when $u = 12$, where each SBNP optimal pairing produce coefficient sums just under twice the values of its random pairing counterpart.

We display the objective values for each value of u in Figure 4.4. The x-axis shows the different values of u we test while the y-value shows the coefficient sums for each respective pairing at each u value.

As the number_of_choices_per_attribute_set increases, the SBNP optimal objective values for each similarity metric decreases. Understandably, when we provide more attribute choices per attribute set, there will overall be less common attributes between mentors and mentees.

Aside from this relationship, we do not identify any other significant relationships produced by modifying number_of_choices_per_attribute_set. We observe that the percentage of shared pairings between SBNP Pearson optimal and SBNP Jaccard optimal pairings, as

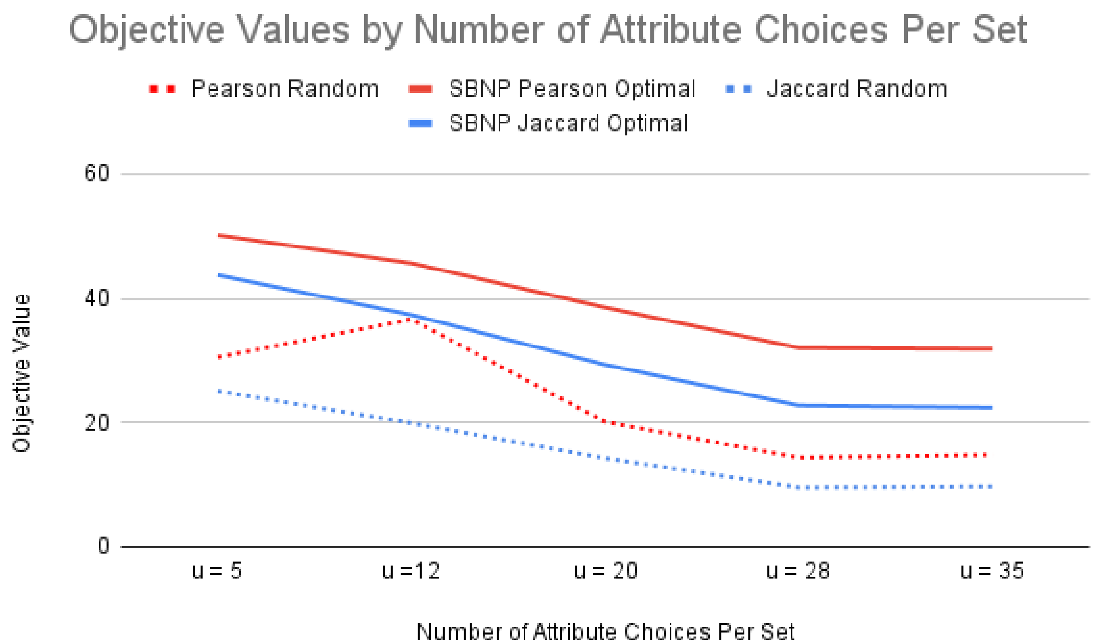

Figure 4.4. Number of Choices Per Attribute Set Comparison: The relationship between each method's objective value is compared across each number_of_choices_per_attribute_set value.

well as the total solve time of each simulation group, slightly increase as we increase u. These increases are not notable enough to infer any impact on the results in the SBNP Model.

4.3 Similarity Metric Comparison

Following completion of sensitivity analysis for each input variable, we examine the differences between each similarity metric and their respective performances in the SBNP Model. Because the Pearson coefficient and Jaccard coefficient each implement different formulas and contain different range values, we cannot directly compare them to each other. Instead, we will compare the performance of each similarity metric's SBNP IP optimal objective value to its respective coefficient sum for the random pairing by calculating the percent change. We apply this calculation to each input variable's simulation results displayed throughout Section 4.2.

It is important to note that, relative to its respective random pairing coefficient sum, we expect the SBNP Pearson optimal objective value to provide a higher improvement percentage than the SBNP Jaccard optimal objective value. We can expect this relationship due to the fact that Pearson coefficients can be negative values, all the way down to -1. The Jaccard coefficient, on the other hand, has a minimum possible value of 0. Because both coefficients have a maximum potential value of 1, we believe the Pearson coefficient sum has more room for improvement since pairings with negative coefficient values could be selected in the random pairing, lowering the sum of coefficients for selected pairings.

We construct bar graphs for each input variable and place the SBNP Pearson optimal improvement next to the SBNP Jaccard optimal improvement. We present results for each input variable, beginning with the number_of_people input variable in Figure 4.5.

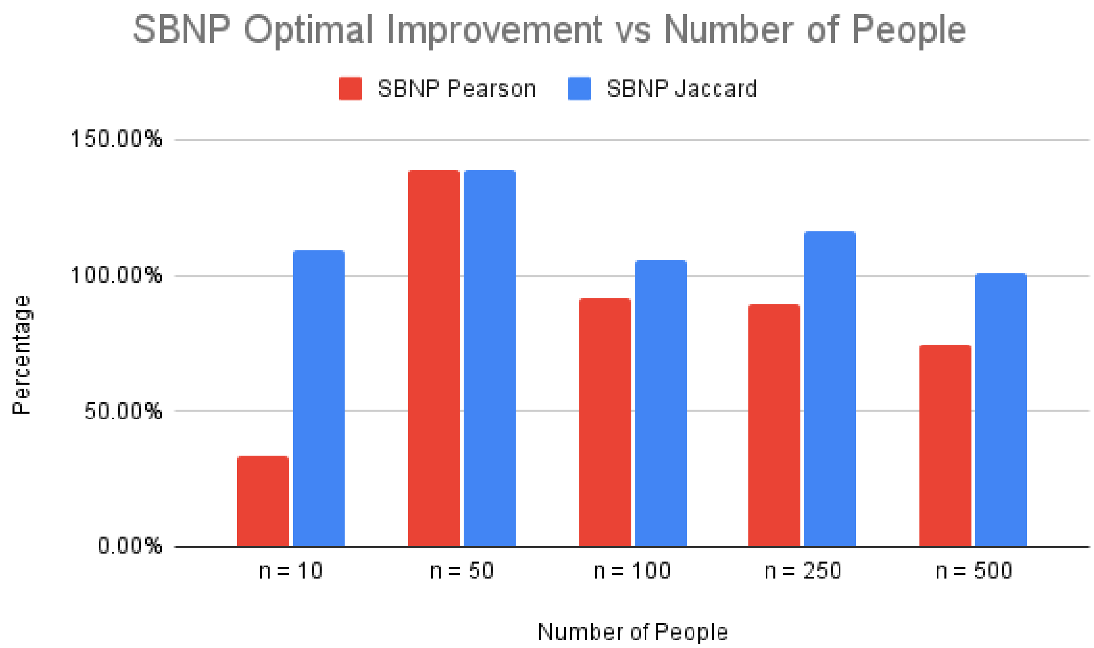

Figure 4.5. Number of People Similarity Metric Analysis: The percentage of improvement for each similarity metric's performance is compared across different number_of_people values.

For the number_of_people variable, we observe that the SBNP Jaccard optimal pairing provides a larger percentage of improvement than the SBNP Pearson coefficient. Both IPs improve the random pairing coefficient sums the most when $n = 50$, and besides the results from $n = 5$, the comparison between both similarity metrics remains relatively stable. As a result, we observe that this input variable does not affect either similarity metric consistently, and the SBNP Jaccard IP outperforms the SBNP Pearson IP in each simulation group.

We now compare the similarity metrics with respect to the change in percentage_of_mentors and present the results in Figure 4.6.

Once again, the SBNP Jaccard IP outperforms the SBNP Pearson IP in terms of improvement with each percentage_of_mentors value. We note that as p increases, the percentage of improvement for both SBNP IPs also increases. As a result, the percentage_of_mentors value is proportionally related to the percentage of improvement in each SBNP. While the SBNP Jaccard IP consistently performs better than the SBNP Pearson IP, both provide

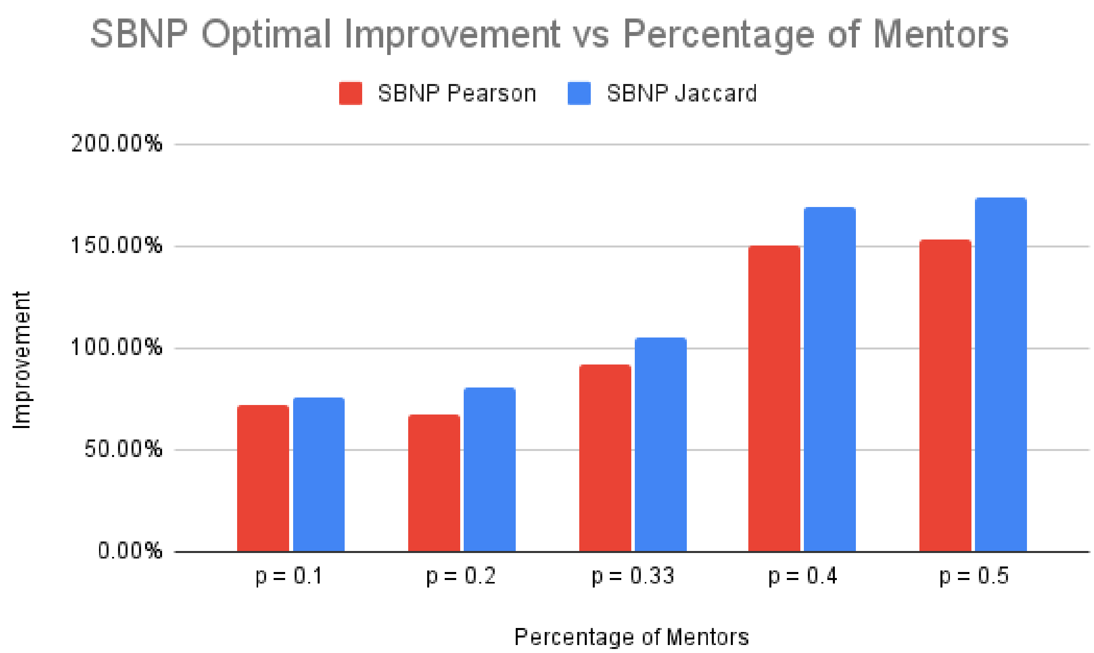

Figure 4.6. Percentage of Mentors Similarity Metric Analysis: The percentage of improvement for each similarity metric's performance is compared across different percentage_of_mentors values.

significant improvement with respect to the percentage_of_mentors variable.

We now move to our most interesting input variable regarding similarity metric comparison: number_of_attribute_sets. Depicted in Figure 4.7, we observe a unique relationship between each similarity metric.

The smaller the number_of_attribute_sets value, the better the SBNP Pearson IP performs relative to the SBNP Jaccard IP. As number_of_attribute_sets increases, however, the SBNP Jaccard IP provides a higher improvement percentage. We note that, for the minimum and maximum number_of_attribute_sets values, there is an extreme improvement percentage from the SBNP Pearson IP and the SBNP Jaccard IP, respectively. The SBNP Pearson optimal results start high, and as number_of_attribute_sets increases, this percentage of improvement decreases, increasing in the highest number_of_attribute_sets value. The SBNP Jaccard optimal results, in a similar fashion, decrease as number_of_attribute_sets increases, but then achieves its highest value in the largest number_of_attribute_sets value. Because

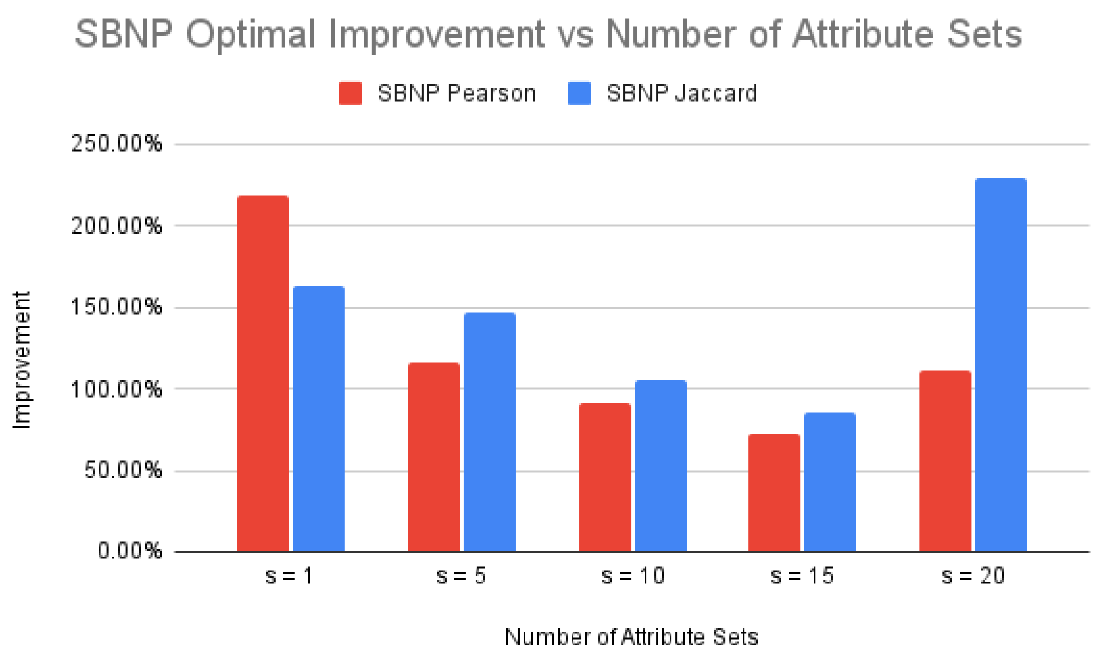

Figure 4.7. Number of Attribute Sets Similarity Metric Analysis: The percentage of improvement for each similarity metric's performance is compared across different number_of_attribute_sets values.

we cannot necessarily identify a trend for either SBNP IP, we note modifying this variable cannot guarantee better results for one similarity metric versus the other.

We finally analyze our last input variable: number_of_choices_per_attribute_set. We chart these results in Figure 4.8.

Similar to percentage_of_mentors, each SBNP IP consistently increases as our number_of_choices_per_attribute_set value increases. The only exception is when $u = 12$, where the SBNP Pearson optimal objective value does not uphold the trend mentioned. Aside from this exception, we note that an increase in number_of_choices_per_attribute_set corresponds to an increase in each SBNP IP percentage of improvement in coefficient sums. We add that, similar to the rest of the input variables, the SBNP Jaccard IP outperforms the SBNP Pearson IP in terms of percentage of improvement.

Based on the similarity metric analysis above, the SBNP Jaccard IP provides consistently

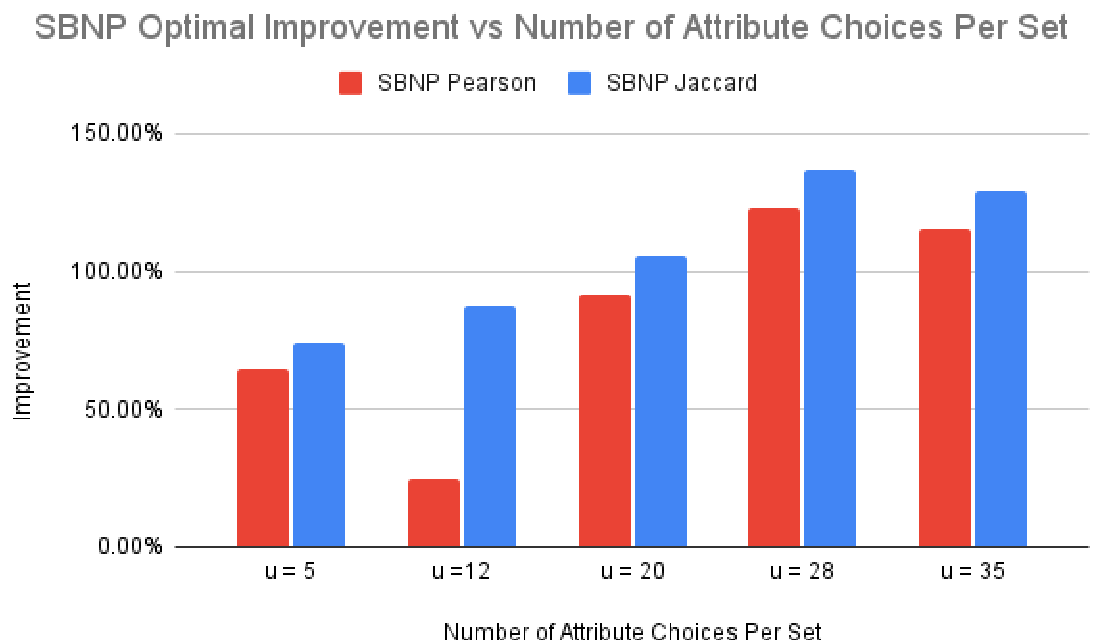

Figure 4.8. Number of Choices Per Attribute Set Similarity Metric Analysis: The percentage of improvement for each similarity metric's performance is compared across different number_of_choices_per_attribute_set values.

higher percentages of improvement from its random pairing coefficient sum counterpart than the SBNP Pearson IP. We do not, however, regard the Pearson coefficient useless as a similarity metric. It does not quite match the Jaccard coefficient performance, but it also outputs similar performance in many simulations. The Pearson coefficient also provides an alternative result which can act as both a comparison and a check to the Jaccard results. We also note that, while the Jaccard coefficient may perform better out of the two similarity metrics, several of the simulation results included a majority of the same SBNP optimal mentor-mentee pairings in each SBNP IP.

In conclusion, we observed the effect each input variable played on the SBNP Model results. While each input variable impacts the solution in its own way, we recognize the significant improvement the SBNP optimal mentor-mentee pairings provide when compared to the random mentor-mentee pairings. When comparing the two similarity metrics used in the SBNP Model, we concluded that the Jaccard coefficient consistently provided a larger percentage of improvement than the Pearson coefficient when compared to their respective coefficient sums of the random pairings. We still deem each similarity metric useful as they capture different characteristics of the network, and they act as a check against each other.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 5: Conclusion and Future Direction

In Chapter 5 we summarize what we have learned in this research and propose further extensions of our work.

5.1 Conclusion

The United States Navy's CSIP implements a mentor-mentee assignment process that considers a number of professional attributes, of which only one is a personal attribute for each of the mentors and mentees. Mentorship in an organization leaves an impact across more than individual relationships, ultimately, both directly and indirectly, affecting the entire organization. While the intention of the CSIP is to provide each new member of the command a resource as they settle in, we believe we can enrich the mentor-mentee assignment method to provide further opportunity for positive interactions based on similar personal interests that the mentors and mentees have. In this work, we propose modifications to the CSIP through the introduction and application of the SBNP Model on a synthetic network in order to simulate delivering personalized mentor-mentee assignments at a given Naval command.

Our SBNP Model accounts for different Naval commands in terms of size and variability by allowing for specification of number of people and attribute affiliation for each person. The number_of_people and percentage_of_mentors input variables determine both the total number of people as well as the split of this total between mentors and mentees. The number_of_attribute_sets and number_of_attribute_choices_per_set input variables specify how many different attribute groups to consider, as well the number of attribute choices within each attribute group. These two pairs of input variables allow the SBNP Model to be tailored to fit varying sizes of people and attributes that realistically represent different Naval commands.

While we introduce a general model, for this work we input specific variables into Python programming language in order to define the size and structure of the synthetic network for the SBNP Model. The SBNP Model then uses this synthetic network to assign mentor-mentee pairings both randomly for control, and through SBNP optimally for our improved method. We then compare our method to the random assignment using two different similarity metrics. With each combination of input variable values, we run 10 simulations and, for each similarity metric, we average the coefficient of sums for the random and SBNP optimal pairings. We compare the values of random mentor-mentee pairings to the values of SBNP optimal mentor-mentee pairing, using both the Jaccard and the Pearson coefficients.

We conclude with a sensitivity analysis on each of the four input variables and observe their effect on the SBNP Model results. While each variable impacts the SBNP Model results, we observe that both the SBNP Pearson and Jaccard optimal pairings consistently provide coefficient sums that are at least twice the value compared to their respective coefficient sums of the random mentor-mentee pairing. As a result, we conclude that the SBNP Model provides higher personalization in the mentor-mentee pairings, therefore having a higher chance of fostering positive interaction than a random mentor-mentee pairing. We note that the Jaccard coefficient provides a consistently higher improvement in coefficient sum value compared to the random pairing's respective coefficient sum than the Pearson coefficient. Each similarity metric, however, produces similar results in terms of percentage of improvement from their respective coefficient sum in the random pairing, so these metrics provide a check for each other.

5.2 Future Directions

While the SBNP Model proposes a modified approach to satisfying the CSIP requirements by delivering a higher personalization of the mentor-mentee pairings in a Naval command, it fails to capture many aspects that contribute to real-world applications. We outline a number of potential directions that can improve our work.

An immediate step to improve this model is to select actual attributes to use in a realistic questionnaire. While the specific attributes themselves do not affect the structure of the model, the SBNP Model only remains a concept unless it can be applied to real world data. We deferred the selection of attributes due to our focus of the paper being on network science and the structure of the model, not necessarily the content. We think the best way to determine said attributes is to create a questionnaire and test on a small sample size. The answers from every person would then be input into the SBNP Model and it completes the same process, eliminating what we created as the synthetic network generation based on the input variables. The results would display a random pairing and SBNP optimal pairings based on the input data from the questionnaire.

There may be cases where a person associates with either more than one attribute per attribute group or does not associate with any attributes in an attribute group. We decided not to capture this variation in our work due to the lack of standardization across different people. We argue that, if each person had to choose *one* attribute that they associate the most with from each attribute group, we can perform the SBNP optimal pairings from there. If we focus on how many different attributes a person can select from each attribute group, the structure of the network shifts significantly, and we cannot standardize similarity metrics across everyone. On one hand, if no one answers any questions, we cannot perform a similarity-based analysis to pair similar mentors and mentees. We believe that, realistically, there may be a small percentage of questions left unanswered, but ideally every person selects at least one answer to every question. We note that, should a person leave a question unanswered, this implies a lack of breadth of attribute choices for that questions. Therefore, allowing user input for attributes to be considered in each question could prevent any one person leaving a question blank. On the other hand, if every person selects multiple answers to each question, more people would likely be more similar to each other. While multiple answer selection could still result in disjoint attribute selection between two people depending on the size of the attribute pool, the chances of two people selecting the same attribute increases as each additional attribute is selected.

We recognize another potential layer for pairing the most similar mentors and mentees. In the questionnaire that contains each attribute group with its respective attribute choices, one can introduce the weighting of the categories of questions based on the person's association with that attribute. For example, if you we provide a five-question questionnaire, people could answer each question and rank the questions from one to five based on how much they want to be matched by each question (five being the highest preference). This weight would apply to the edge associated with the person and that attribute and be properly reflected in the network's adjacency matrix. So, an attribute ranked a five would matter five times as much as an attribute ranked one. This difference in magnitude would affect mentor-mentee pairings, as shared attributes themselves would not be the only factor. The goal to would to be match mentors and mentees with shared *preferred* attributes.

Additionally, the SBNP Model can be improved by considering the optimization programs. We could modify the objective functions of the current IPs we implement to provide SBNP optimal pairings. In this work, we maximize the sum of the coefficients since we believe it captures the maximum similarity across the network, but we could implement both a greedy approach and a maximum-minimum approach. The greedy approach walks through each mentee and checks the coefficient to each mentor being paired with that mentee. It then selects the mentor with the highest similarity coefficient and repeats the process for each mentee. The maximum-minimum approach focuses on the least similar mentor-mentee pairing, based on the lowest similarity coefficient, selected in the optimization model and looks to find pairings that maximize that number. Each of these approaches may provide different results and would be interesting to compare to the current IPs.

Assuming all of these concerns are solved and the SBNP Model is ready for use, the big question is how to implement it in an operational system or app. This questionnaire may be part of a transfer package when a service member begins the PCS process; it could be part of an online profile that can be updated at any time. No matter what, there needs to be a system that tracks each person's answers to allow for updated mentor-mentee pairing. This model can be implemented however each Naval command handles sponsor assignment, whether by person or by system.

List of References

Barrett JL, Mazerolle SM, Nottingham SL (2017) Attributes of effective mentoring relationships for novice faculty members: Perspectives of mentors and mentees. *Athletic Training Education Journal* 12(2):152–162.

Bollobás B (2013) *Modern graph theory*, volume 184 (Springer Science & Business Media).

Chao GT (2009) Formal mentoring: Lessons learned from past practice. *Professional Psychology: Research and Practice* 40(3):314.

Chief of Naval Operations (2021) *Chief of Naval Operations Navigation Plan 2021*. CNO NAVPLAN 2021, Washington, DC, https://media.defense.gov/2021/Jan/11/2002562551/-1/-1/0/CNO%20NAVPLAN%202021%20-%20FINAL.PDF.

Department of the Navy (2017) *Command Sponsorship and Indoctrination Program*. OPNAV 1740.3D, Washington, DC, https://www.secnav.navy.mil/doni/Directives/01000%20Military%20Personnel%20Support/01-700%20Morale,%20Community%20and%20Religious%20Services/1740.3D.pdf.

Gera R (2020) Ma4404 complex networks: Node similarity. http://faculty.nps.edu/rgera/MA4404/15-nodeSimilarity.pdf, accessed: 2021–05-07.

Guy I, Zwerdling N, Ronen I, Carmel D, Uziel E (2010) Social media recommendation based on people and tags. *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, 194–201.

Henry-Noel N, Bishop M, Gwede CK, Petkova E, Szumacher E (2019) Mentorship in medicine and other health professions. *Journal of Cancer Education* 34(4):629–637.

Inzer LD, Crawford C (2005) A review of formal and informal mentoring: Processes, problems, and design. *Journal of Leadership Education* 4(1):31–50.

Johnson WB, Andersen GR (2010) Formal mentoring in the U.S. military: Research evidence, lingering questions, and recommendations. *Naval War College Review* 63(2):113–126.

Johnson WB, Andersen GR (2015) Mentoring in the U.S. navy: Experiences and attitudes of senior navy personnel. *Naval War College Review* 68(3):76–90.

Jones SJ (2017) Establishing a nurse mentor program to improve nurse satisfaction and intent to stay. *Journal for nurses in professional development* 33(2):76–78.

Kibbe MR, Pellegrini CA, Townsend CM, Helenowski IB, Patti MG (2016) Characterization of mentorship programs in departments of surgery in the United States. *JAMA Surgery* 151(10):900–906.

Leicht EA, Holme P, Newman ME (2006) Vertex similarity in networks. *Physical Review E* 73(2):026120.

Lester PB, Hannah ST, Harms PD, Vogelgesang GR, Avolio BJ (2011) Mentoring impact on leader efficacy development: A field experiment. *Academy of Management Learning & Education* 10(3):409–429.

Loscalzo J, Barabási A (2016) *Network science* (Cambridge: Cambridge University Press), ISBN 1107076269.

McHugh JA (1990) *Algorithmic graph theory*, volume 68056 (Citeseer).

Newman MEJ (2018) *Networks: An Introduction,* 2nd ed. (Oxford University Press, New York).

Pinilla S, Pander T, von der Borch P, Fischer MR, Dimitriadis K (2015) 5 years of experience with a large-scale mentoring program for medical students. *GMS Zeitschrift für Medizinische Ausbildung* 32(1).

Rardin RL (1998) *Optimization in operations research* (Upper Saddle River, N.J: Prentice Hall), ISBN 0023984155.

Ravey DA, Welsh National Board for Nursing M, Health Visiting CUK (1992) *Mentors, Preceptors and Supervisors: Their Place in Nursing, Midwifery and Health Visitor Education* (Welsh National Board for Nursing, Midwifery and Health Visiting).

Terveen L, McDonald DW (2005) Social matching: A framework and research agenda. *ACM transactions on computer-human interaction (TOCHI)* 12(3):401–434.

Initial Distribution List

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California