Department of Built Environment

Synthesizing Social Networks for Travel Behavior Models

by

Zhi Yan Chew

7Z45M0 USRE Graduation Project

Graduation	
Program:	Architecture, Building, and Planning
Supervisor:	Full Professor Soora Rasouli
	Assistant Professor Valaria Caiati
	Assistant Professor Gamze Dane
Date of defense:	August 29, 2024
Student ID:	1777556
Study load (ECTS):	45
Track:	Urban Systems and Real Estate

Abstract

This thesis explores the application of Bayesian Belief Networks (BBNs) for synthesizing social networks at both the ego and ego-alter levels for travel behavior modeling. BBNs are employed to capture and model the complex interdependencies between socio-demographic characteristics, social network size and composition, types of social relationships, and the frequency of joint activities. Two distinct models were developed: Model I, which examines the influence of socio-demographic factors on the composition and size of an individual's social network and the aggregated frequency intervals of joint activities, and Model II, which provides a disaggregated analysis of individual social relationships with social network member, focusing on how specific socio-demographic differences, geographic distance between them, and relationship duration impact joint activity frequencies.

The findings reveal that individuals with larger networks of friends tend to engage more frequently in activities such as dining out, while those with a greater number of siblings in their network are more likely to participate in regular family vacations. Additionally, the analysis shows a higher likelihood of individuals visiting or hosting friends rather than colleagues and a stronger tendency to dine out with close family members, such as parents or children. These complementary insights from both models underscore the significant role of social networks in shaping travel behavior, emphasizing the importance of both social network size and the nature of specific social relationships.

This research demonstrates the potential of BBNs to enhance travel behavior models by incorporating synthetic social networks, offering a means to address data limitations and privacy concerns. While the study highlights the effectiveness of BBNs, it also identifies limitations related to data accuracy and the evolving nature of social interactions, particularly in the post-COVID-19 era. The results provide a foundation for further research, particularly in integrating synthetic social networks with travel behavior models to improve the accuracy of travel behavior predictions. This work contributes to both the academic understanding and practical application of social network synthesis in transportation planning and policy.

Preface

As I reflect upon the journey that culminated in the completion of this thesis, I am filled with a profound sense of growth and accomplishment. This project has been more than an academic requirement; it has been a transformative experience that has challenged my perspectives, deepened my understanding, and sharpened my skills. To be honest, this research project turned out to be completely different from what I had initially imagined. It led me into a new area that I had not explored before, and there were times when I struggled with the challenges it presented.

The most challenging part of this project was the initial phase of confirming the research direction. It took a considerable amount of time to review existing literature and identify a meaningful research gap. Compounding this difficulty was the need to ensure that the research direction aligned with the data we had, even though, at that time, I had little understanding of what the data would actually look like. Additionally, the pressure to prepare, clean, model, and analyze the data quickly as deadlines approached added to the challenge. Despite these obstacles, I persevered and ultimately completed the graduation project. Although the outcome may not be perfect, it is a product of my sweat and tears, and I take great pride in it. This is also the nature of research, which is always open to critique and improvement, and I hope future work will build upon it. Throughout this process, I have learned new modeling techniques, enhanced my data analysis skills, improved my critical thinking, and refined my academic writing. However, my greatest achievement has been the personal growth I experienced along the way. I used to handle everything independently and rarely sought help, but this project pushed me to step out of my comfort zone, speak up, and seek assistance when needed. Admitting that I am not perfect and being honest with myself about my limitations have been key factors in my development. As the saying goes, what doesn't kill you makes you stronger.

This journey, however, was not undertaken alone. I owe a great deal of gratitude to those who have guided, supported, and inspired me along the way. First and foremost, I would like to extend my heartfelt thanks to my supervisors, Prof. Soora and Valeria. Their unwavering support, insightful guidance, and thoughtful critiques have been invaluable throughout this entire process. From the very beginning, they have supported me, offering encouragement when needed and challenging me to think more critically and deeply about my work. Prof. Soora, responsible and insightful, consistently pointed out key aspects that helped clarify doubts and moved the project forward. Her ability to identify and address the most crucial points in my research was instrumental in shaping the direction and quality of this thesis. Valeria, always kind and approachable, was equally essential to this journey. She paid meticulous attention to detail, ensuring that every aspect of my work was thoroughly examined and refined. Her keen eye for detail, coupled with her genuine care and understanding, provided not only academic guidance but also emotional support during the most stressful moments. Valeria's thoughtful critiques and careful attention to the finer points of my research consistently pushed me to improve and polish my work to a higher standard.

I would also like to express my sincere appreciation to Gamze, whose contributions in reviewing my thesis and providing feedback helped refine my arguments and ensure the clarity of my writing.

In closing, this thesis represents not just my own efforts but also the collective support and wisdom of those who have accompanied me on this academic journey. For that, I am deeply grateful.

Contents

Abstract i

Preface iii

Contents v

List of Figures vii

List of Table ix

Chapter 1 Introduction 1

Chapter 2 Literature Review 5

Chapter 3 Methodology 11

Chapter 4 Data 15

Chapter 5 Results 27

Chapter 6 Conclusion 36

Bibliography 39

List of Figures

Figure 1: Conceptual Framework. 12

Figure 2: Boxplot of the total number of social network members from all social categories by respondents' sociodemographic characteristics. (a) by gender; (b) by age group; (c) by household size; (d) by marital status; (e) by education level; and (f) by income level. 23

Figure 3: Data distributions according to the social categories. (a) gender differences; (b) age differences; (c) education differences; (d) relationship distance; (e) relationship duration; and (f) type of interaction. 26

Figure 4: DAG of Model I (log(p)=-333864.4). 28

Figure 5: DAG of Model II (log(p)=-419302). 28

Figure 6: Posterior probability distribution of child node "Frequency intervals of dining out" by instantiating parent node "The number of friends"..32

Figure 7: Posterior probability distribution of child node "Frequency intervals of dining out" by instantiating parent node "The number of colleagues". 33

Figure 8: Posterior probability distribution of child node "Frequency intervals of going on vacation" by instantiating parent node "The number of siblings". 33

Figure 9: Posterior probability distribution of child node "Frequency intervals of going on vacation" by instantiating parent node "The number of other family members". 33

Figure 10: Posterior probability distribution of child nodes "Frequency of visiting", "Frequency of hosting" and "Frequency of dining out" by instantiating parent node "Type of social relationship". 35

List of Table

Table 1: Summary of variables collected through the questionnaire 16

Table 2: Descriptive statistics of respondents' socio-demographics and their social network characteristics (N=1278) 20

Table 3: Descriptive statistics of aggregated joint activity frequency (N=1278) 22

Table 4: Descriptive statistics of differences in socio-demographic characteristics and joint activity frequency (N=16682) 24

Table 5: The accuracy, recall, precision, and F1-score of predicted variables in Model I (N=1278). 30

Table 6: The accuracy, recall, precision, and F1-score of predicted variables in Model II (N=16682). 30

Chapter 1

Introduction

Understanding travel behavior is essential for transportation planning, infrastructure development, policy-making, and disaster risk management (Wang et al., 2024). These topics are vital for creating sustainable and resilient urban systems. Generally, travel behavior studies explore when and where people engage in different activities and how they travel to these locations. It examines individual's choices, including transport modes, departure times, travel durations, activity choices, and destination locations (Sharma et al., 2021). Factors such as socioeconomic attributes, accessibility to transport modes, and built environmental characteristics are often considered. Urban planners and policymakers utilize travel behavior models to incorporate these factors to make rational, evidence-based decisions. By analyzing the spatial and temporal balance between travel demand and supply, mobility gaps can be identified, transportation inequalities can be addressed, and social equity can be promoted, particularly for vulnerable groups (Barajas, 2021; Li & Landis, 2019). As such, improving the accuracy and depth of travel behavior models remains a priority.

Recently, the social context of travel behavior has gained increasing attention, particularly focusing on with whom individuals engage in activities (Calastri et al., 2017). Joint activities involving social network members represent a significant portion of urban trips and differ from work-related trips in terms of regularity and purpose, contributing to traffic congestion. A deeper understanding of joint activities and their impact on travel patterns allows for more accurate behavior predictions. Studies have shown that social networks significantly influence an individual's activity-travel patterns and mode choices (Carrasco & Miller, 2009; Van Den Berg et al., 2009, 2012a), constraining travel-related decisions based on their spatial distribution (Arentze et al., 2012; Frei & Axhausen, 2007). Moreover, social networks act as vital sources of information and decision support are crucial not only for daily activities (Ryley & Zanni, 2013) but also for shaping attitudes towards new mobility concepts like Electric Vehicles (EVs), Autonomous Vehicles (AVs), shared mobility, and Mobility as a Service (MaaS). For example, studies have demonstrated how social networks affect decisions on EV purchases (Kim et al., 2014; Rasouli & Timmermans, 2013, 2016) and MaaS subscriptions (Caiati et al., 2020). In short, social networks profoundly influence various aspects of travel behavior, from daily activities to long-term transportation choices.

Modeling travel behavior traditionally focuses on individual characteristics, such as accessibility, and trip attributes, like cost (Sharma et al., 2021). However, incorporating social network variables offers a more nuanced understanding of travel decisions (Calastri et al., 2017; Dugundji & Walker, 2005), especially regarding joint activities and trips. Discrete choice modeling has been used to measure aggregated social influences (Walker et al., 2011), while agent-based models simulate individual behaviors (Joubert & De Waal, 2020). As examples for choice modeling, Pike (2014) found that students with larger social networks were more likely to choose bicycles for commuting, and Maness and Cirillo (2016) showed how social influences within communities drive bicycle ownership. Despite these advancements, many models overlook the disaggregated effects of social networks, such as variations in size, structure, and composition.

This research project addresses the gap regarding the linkage between joint activities and travel behavior. Studying how frequent individuals perform joint activities with their social network members can provide valuable insights into travel behavior, which in turn shapes activity-travel patterns. Furthermore, this research project aims to empirically examine the relationship between individuals and their social network members and the frequency of conducting joint activities at ego (individual) and ego-alter (individual and certain social network member) levels. Eight joint activities are considered such as visiting, hosting, dining out, going to a bar or disco, going to the cinema or theatre, doing recreational activities, going on vacation, and shopping. By employing Bayesian Belief Networks (BBNs), this study leverages their ability to reason probabilistically and simultaneously model numerous variables (Sun & Erath, 2015). BBNs are ideal for capturing the complex interactions between individuals, their social networks, and joint activity frequencies.

In this research project, two models were developed, Model I at the ego level and Model II at the ego-alter level. Developing models at these two different levels is essential to fully investigate how the frequency of joint activities is influenced by individuals and by both individuals and their social network members. Model I examines individuals' socio-demographic characteristics, the size of various social network member groups (e.g., friends, colleagues, neighbors, grandparents, parents, siblings, children, and other family members), and the frequency intervals of various joint activities, such as visiting and hosting social network members, and so on. Model II explores the differences between individuals and their social network members, the type of social relationship, the distance between ego and alter, the relationship duration, and the frequency of different joint activities.

The data used was collected through a web-based survey in 2014 in the Netherlands. The collected dataset contains information about the size and composition of respondents' social networks and the frequency and nature of their interactions with their social network members for various joint activities. The data preparation procedure is described, and descriptive statistics are presented. This research project contributes to social network synthesis by integrating joint activity frequency at both ego and ego-alter levels. The findings offer insights into the relationship between social networks and joint activities, with implications for enhancing travel behavior simulations and informing urban mobility strategies.

The following section outlines the problems and research questions addressed in this research project. Next, the societal and academic relevance are discussed. The remainder of this final thesis is structured as follows: Chapter 2 provides a thorough literature review of social influence on travel behavior, core concepts, and methods of social network synthesis. Chapter 3 explains the methodology for synthesizing social networks in this research project. Chapter 4 describes the data used in synthesizing social networks. Chapter 5 discusses the results and findings. Chapter 6 concludes and reflects on the research project.

1.1 Problem Analysis and Research Questions

In recent years, scholars have expanded the traditional framework for explaining travel behavior by incorporating social influence (Calastri et al., 2017), recognizing that social networks are complex and dynamic. Social relationships evolve and influence travel behavior, as individuals form new connections and existing relationships change, often in response to relocation and other life events (Axhausen, 2008). This dynamic nature of social networks is associated with changing patterns of joint activity and travel, presenting significant challenges for modeling. Two key challenges must be addressed to effectively model the complex composition of social networks and joint activities with social network members.

Firstly, limited sources of detailed social network data. Travel behavior modeling traditionally relies on travel surveys that record activity and travel diaries, including travel modes, time, duration, and socio-demographic characteristics of respondents (Sharma et al., 2021). However, these surveys typically do not include detailed social network information. Collecting comprehensive data on individuals' social connections and the attributes of these relationships is often impractical and resource-intensive (Axhausen, 2008; Koelle et al., 2006). Detailed social network data such as types of social relationships, strength of social relationships, frequency of interaction, and types of joint activity are required for travel behavior models.

Secondly, privacy concerns on social network data. The use of detailed and disaggregated social network data raises significant privacy issues. Individuals can be easily identified through their demographic and social network characteristics, making such data highly sensitive (Sun & Erath, 2015). These privacy concerns restrict the study of social influence at a disaggregated level, limiting the integration of social networks in travel behavior modeling. While some studies have incorporated social influence at an aggregated level using stated choice and discrete choice modeling methods to examine the impact on individuals' travel decisions (Cherchi, 2017; Kim et al., 2014, 2017; Pan et al., 2019; Rasouli & Timmermans, 2013, 2016), there remains a lack of exploration into social networks at a disaggregated level and integration into current travel behavior models.

Due to these data limitations and privacy concerns, researchers have turned to incorporating synthetic social networks instead of actual observed social networks to model the social influences on travel behavior at a disaggregated level (Arentze et al., 2012a; Arentze et al., 2012b; T. Arentze & Timmermans, 2008; Axhausen & Hackney, 2006). Despite the advantages of using synthetic social networks, there is a significant gap in how social networks can be synthesized for travel behavior modeling purposes. To obtain valuable and referable insights from travel behavior simulations and models, it is essential to synthesize realistic social networks that reflect real-world social interactions, showing the interactions between individuals and their social network members.

Considering this, the main research question addressed in this project is: **How can realistic social networks for travel behavior modeling be effectively synthesized?** Building upon this, the project will delve into the following research sub-questions:

1. **How do synthetic social networks capture various types and sizes of social relationships?** This sub-question explores to what extent synthetic social networks capture various types and sizes of social relationships, such as parents, siblings, friends, colleagues, and neighbors.
2. **How do synthetic social networks capture different types and frequencies of joint activities?** This sub-question explores to what extent synthetic social networks capture different types and frequencies of joint activities with social network members.
3. **How to assess the realism of synthetic social networks?** This sub-question assesses the realism of synthetic social networks by comparing them to observed social networks. Several accuracy indicators can be measured to discuss the accuracy of the models.

1.2 Academic and Societal Relevance

This research project holds significant academic and societal relevance in several key aspects. Academically, it advances the understanding of the interdependencies between social networks and joint activities, shedding light on how different types and strengths of social relationships influence the frequency of these activities. By synthesizing realistic social networks, this research addresses critical limitations in current travel behavior modeling approaches, including data scarcity, privacy concerns, and ethical considerations. The integration of synthetic social networks enables the development of more accurate and predictive travel behavior models, enhancing their reliability and applicability. Furthermore, the study provides a unique contribution by analyzing social networks at two levels of granularity: ego level and ego-alter level. This multi-level analysis offers a nuanced perspective on the role of social relationships in shaping travel behavior, enriching behavioral explanations, and supporting evidence-based decision-making in urban and transportation planning.

From a societal perspective, this research has profound implications for sustainable urban and transportation planning. Travel behavior simulations, enriched through synthetic social networks, enable planners to more accurately predict how individuals and communities will respond to changes in transportation infrastructure, policies, and emerging technologies. These insights are essential for informing decisions on congestion pricing, public transit expansions, and the implementation of low-emission zones. Furthermore, the findings are particularly relevant for promoting sustainable mobility solutions such as electric vehicles, autonomous vehicles, shared mobility services, and Mobility as a Service (MaaS) platforms. Social networks play a critical role in shaping attitudes and behaviors toward these innovations, and understanding their influence allows for the creation of more targeted and effective interventions. In sum, this research bridges significant gaps in academic understanding while delivering actionable insights that can support the development of sustainable, socially inclusive transportation systems. Overall, this research bridges important gaps in academic knowledge while providing actionable insights that support the development of sustainable, socially responsive transportation systems.

.

Chapter 2

Literature Review

The literature review explores the impacts of social influence on travel behavior. This includes social network variables (e.g. size, composition, and strength of relationship) that affect different aspects of individuals' travel behavior for example transport mode choices, activities scheduling, destination selection, and the adoption of new mobility services or products. Additionally, the review discusses the core concepts and methods of social network synthesis relevant to travel behavior models. The research gap identified, and the potential contributions of this research project will also be discussed at the end of Chapter 2.

2.1 Social influence on travel behavior

The existing body of literature on travel behavior modeling underscores the crucial role of social networks in shaping individuals' travel patterns. Various studies have demonstrated that social networks significantly influence multiple aspects of travel behavior, such as activity scheduling, mode choice, and destination selection (Arentze & Timmermans, 2008; Axhausen & Hackney, 2006; Kim et al., 2018).

For instance, Van Den Berg et al. (2012a) investigated the relationship between social networks and travel behavior, finding that stronger social connections correlate with more frequent social activities and increased travel. The size of social networks has also been shown to impact travel behavior, as individuals with larger networks tend to be more socially active, leading to heightened engagement in social activities and, consequently, more travel (Van Den Berg et al., 2012b). Furthermore, the relationship between social activities and travel has been explored, revealing that social activities, typically involving multiple participants, are often associated with longer travel distances and durations (Van Den Berg et al., 2009).

Moore et al. (2013) employed a structural equation approach to study the influence of individual and interactional attributes on the duration, distance, and number of people involved in daily activities. They concluded that socio-demographic characteristics such as race, income, educational level, and accessibility to transport modes are not sufficient to explain the spatiotemporal dimension of daily activity patterns and including social network factors such as the size of social networks, characteristics of social activities, and strength of the relationships as additional determinants to travel behavior modeling increases the accuracy of activity scheduling prediction (Moore et al., 2013).

Beyond social network size and relationship strength, the composition of social networks also plays a crucial role in travel behavior. The diversity of interpersonal relationships influences travel-related decisions, with the level of influence varying based on the nature of the relationship. Kim et al. (2017) investigated the social influence in car-sharing decisions and found that individuals are more likely to mimic the behaviors of close social contacts who have aligned interests and preferences, such as friends of similar age or opposite gender. Similarly, Rasouli and Timmermans (2013a) discovered that the influence of relatives on individuals' electric car purchase decisions is weaker than that of friends or colleagues, highlighting the varying degrees of influence depending on social relationship type. However, Caiati et al. (2020) found that relatives had the most significant influence on the decision to subscribe to Mobility as a Service (MaaS), contrasting with Rasouli & Timmermans (2013a)' findings. This indicates that the impact of social influence is context-dependent and shaped by factors beyond network size, relationship strength, and composition.

In turn, the influence of travel behavior on social networks has also been explored. Van Den Berg et al. (2009) found that individuals with a higher income level or vehicle ownership are more likely to have larger social networks extending over greater distances. This often involves family members or relatives, suggesting that people are willing to travel further to maintain stronger social connections. Therefore, travel behavior not only reflects the influence of social networks but also contributes to their expansion and reshaping.

The advent of Information and Communication Technology (ICT) has further influenced the dynamics between social networks and travel behavior. Several studies have investigated how ICT serves as either a substitute or complement to in-person social interactions. Van Den Berg et al. (2009) estimated an ordinal regression model that implied a higher use of ICT leads to increased face-to-face contact. Carrasco and Miller (2009) discovered that the frequency of phone contact is associated with a rise in social activity, while instant messaging serves as a substitute for social activities, along with emails for distant connections. When examining the strength of the relationship between individuals and their social contacts, Van Den Berg et al. (2012a) found that people tend to reach out more often to those with whom they have close relationships using different forms of communication. More recently, Parady et al. (2021) highlighted that social network size significantly influences contact frequency through ICT, enabling more frequent interactions at a low cost. Collectively, these studies suggest that ICT complements rather than replaces traditional travel behavior and social network interactions.

In the context of long-term mobility decisions, Axhausen (2008) argued that travel is essential for fostering and sustaining social capital within networks, influencing decisions such as residential moves, workplace changes, holiday destinations, and regular meeting spots. These long-term mobility decisions can, in turn, lead to changes in the size and composition of an individual's social network. For example, relocating to a new neighborhood or workplace may result in the maintenance or diminishment of old relationships and the formation of new connections (Axhausen, 2008). Therefore, changes in social networks can be better understood by examining people's long-term mobility decisions.

In summary, extensive research has established the significant role social networks play in shaping individuals' travel behaviors, highlighting the influence of network size, composition, and relationship strength on travel decisions. Moreover, the bidirectional relationship between travel behavior and social networks has been emphasized, showing that social networks influence travel behavior, while travel behavior can also expand or reshape social networks. However, much of the literature has focused on aggregated social influence, often overlooking the nuanced impacts of disaggregated social networks on individual travel behaviors. This gap underscores the need for further research that explores the effects of individual social network members on travel and activity decisions. To address this, capturing and analyzing social interactions between individuals and their social network members at a granular level, and integrating these insights of social networks into travel behavior models, is crucial for enhancing predictive accuracy.

2.2 Core concepts and methods of social networks synthesis

To effectively integrate social networks into travel behavior models, a comprehensive understanding of core social network concepts in travel and activity behavior is essential. These concepts can be categorized into four primary areas: (1) social cooperation, (2) social influence, (3) social capital, and (4) social learning (Maness, 2020). Social influence involves the alteration of an individual's travel and activity decision-making due to the actions, behaviors, attitudes, and beliefs of others. Social cooperation refers to the active coordination of travel and activities between individuals. Social learning occurs when individuals adopt new behaviors and beliefs through observations and social experiences. Finally, social capital enables individuals to achieve objectives through collective efforts that would otherwise be difficult or impossible to achieve alone. All four factors are critical in understanding travel and activity decision-making (Maness, 2020). Despite their relevance, the integration of these social network concepts into travel behavior models remains underexplored due to the limited availability of realistic and representative social networks.

Some researchers have suggested that a detailed representation of social networks is necessary to provide a robust foundation for their integration into travel behavior models (Carrasco & Miller, 2009; Kim et al., 2018; Moore et al., 2013). However, collecting comprehensive social network data, which includes every social network member across various categories, requires substantial resources, time, and ethical considerations (Axhausen, 2008; Koelle et al., 2006). As a result, there is growing interest in synthetic social networks as an alternative approach. Synthetic social networks not only address the challenges of data collection but also offer solutions to privacy concerns, serving as a form of data anonymization (Narayanan & Shmatikov, 2009). Consequently, synthetic social networks present a viable option for travel behavior modeling without compromising data privacy, and they can be used as foundational input for travel behavior analyses and simulations to model real-world social interactions.

One common method for generating synthetic social networks is the Exponential Random Graph Model (ERGM), which captures the underlying mechanisms of social network formation (Robins et al., 2007). This probabilistic approach predicts social ties among a specific group of individuals by representing the graph structure based on an exponential function of potential links that form specific patterns within the graph (Robins et al., 2007). These patterns may involve reciprocal and nonreciprocal relationships between pairs or triples of nodes (Robin et al., 2007). Additionally, the attributes of nodes can also be considered in the model, allowing for the incorporation of effects like homophily. The likelihood of observing certain structures is determined by parameters estimated from actual social network data. Estimating these parameters is challenging. Methods used are generally based on the principle of maximum likelihood estimation. Standard likelihood estimations can become difficult and could be biased for larger networks (Robins et al., 2007). Markov chain Monte Carlo methods are gaining attention as a solution in this case for large-scale network generation. They sample from the posterior distribution to estimate parameters for exponential random graph models (Robins et al., 2007).

Another prominent approach for generating synthetic social networks is the stochastic actor-based model (Snijders et al., 2010). Unlike ERGMs, it simulates the evolution of a social network over time by incorporating individual-level attributes and behavior (Snijders et al., 2010). The stochastic actor-based model is flexible enough to represent several different explanations of change by objective functions formulated as a linear combination of effects that bear resemblance to homophily, transitivity, and reciprocity (Arentze et al., 2012a; Arentze et al., 2012b). In this approach, parameters of the objective function are estimated based on longitudinal social network data which are crucial to examine how social networks evolve and change over time. Both ERGMs and stochastics actor-based models are capable of reproducing network characteristics such as homophily, reciprocity, and transitivity. However, stochastics actor-based models are behaviorally richer than ERGMs, as they are designed to model the actual choice behavior of individuals.

Some first approaches using synthetic social networks representing real-world social networks in travel behavior modeling have been proposed by Arentze et al. (2012b). They developed a stochastic actor-based model to generate a whole social network for a given population focusing on friendship relationships. The friendship formation model predicts the probability of the existence of friendship between two individuals based on homophily and geographic distance. For each pair of nodes (agents), the friendship formation model is applied, and the decision is made using Monte Carlo simulation. This procedure is repeated until all pairs of agents have been processed and eventually generate a synthetic social network. They proved that the method could generate a synthetic social network that matches the relevant characteristics of real networks. However, the standard deviation of the number of generated social ties per individual was underestimated due to the consideration of transitivity in friendship tie formation. The friendship formation model was then applied to the Swiss context for large-scale micro-simulation of the travel demand (Arentze et al., 2012a). In the latter application of the generation of synthetic social networks, they included transitivity in friendship formation by extending the process of simulating friendship ties. Each pair of agents has been processed twice in two rounds: in the first round, the same step as in the previous work was taken to determine whether a tie existed based on homophily and geographic distance, and in the second round, transitivity was considered, whether there was any common friend. In their work, only friendship relationships are considered in the model, ignoring other types of social ties such as family, colleagues, etc. (Arentze et al., 2012a; Arentze et al., 2012b).

Social networks can be synthesized based on small world theory, which argues that everyone is connected within a certain boundary (Jiang et al., 2022). Jiang et al. (2022) generated three types of synthetic social networks based on the same household, same workplace, and same educational institute through a synthetic reconstruction approach. They also demonstrated the application of synthetic social networks in travel behavior modeling to simulate traffic dynamics, disease outbreaks, and community response in a high-density urban setting. These demonstrations have proved the re-usability and generalizability of synthetic social networks which can be used in exploring a variety of urban problems. However, the synthetic social networks used in their research assumed that every node within the network is connected, it remains arguable regarding its realism in representing real-world social interaction between individuals.

Recently, some studies used Bayesian Belief Networks (BBNs) to synthesize social networks. The results showed that BBNs are a promising data-driven framework for studying complex relationships and generating large samples of synthetic data with well-fitting distributions (Ilahi & Axhausen, 2019). Koelle et al. (2006) synthesized a social network via BBNs to study the probability of an individual within the network being an organizational leader. They claimed that the advantages of using BBNs include augmenting social network algorithms with uncertainty, searching the network for nodes, and inferring new links in the network (Koelle et al., 2006). Zhang et al. (2019) utilized BBNs and the exponential random graph model to synthesize social networks for transportation simulation. The key benefit of synthesizing social networks through the BBNs approach is that it could be released for wider use while the original data remains confidential (Young et al., 2009). Furthermore, BBNs have been extensively applied for probabilistic inference and reasoning problems, especially in analyzing and interpreting survey data (Sun & Erath, 2015). Despite the advantages and possibilities of BBNs for generating synthetic data, their application in social network synthesis for travel behavior models is underexplored. We think the BBNs approach is an exciting approach to synthesizing social networks that suit our research objectives and therefore can be applied in this research project.

2.3 Research Gap and Contribution

Existing literature has extensively discussed the relationship between social networks and travel behavior, with a focus on the size, composition, and strength of social ties from an aggregated perspective. Several studies have also examined the role of ICT in reshaping the way people socialize and their activity-travel patterns. However, there remains a notable research gap in studying the impacts of social networks on travel behavior at a disaggregated level. The current focus has primarily been on aggregated social influence, overlooking the nuanced effects that each social network member may have on an individual's travel decisions and activity choices. The review emphasizes the importance of incorporating social network into travel behavior models to enhance their predictive accuracy.

One significant barrier to advancing this area of research is the difficulty in collecting detailed and comprehensive social network data. This challenge is compounded by privacy concerns, which further restrict the application of social network data in travel behavior modeling. While the concept of synthetic social networks has been introduced as a solution to these challenges, there has been limited progress in synthesizing realistic social networks that can be effectively integrated into travel behavior simulations and models. Current approaches to synthesizing social networks such as the ERGMs and the stochastics actor-based models can reproduce certain network characteristics, but the existing applications do not include considerations of travel and activity aspects. Moreover, the BBNs approach has a strong ability to perform probabilistic inference and reasoning, particularly in analyzing and interpreting survey data. We identified that BBNs offer the potential to synthesize social networks while ensuring data privacy, making them a promising tool for travel behavior modeling.

To contribute to this field, this research will explore and extend the application of BBNs for synthesizing social networks in travel behavior models. By leveraging BBNs, we aim to synthesize realistic social networks that preserve real-world social interactions while addressing privacy concerns. This approach will provide a robust foundation for integrating social networks into travel behavior simulations, enabling more accurate predictions of how social interactions influence individual and collective travel patterns.

In summary, this research addresses the gap in understanding the disaggregated effects of social networks on frequency of joint activities and provides a foundation for integrating social networks into travel behavior models. By doing so, it contributes to the advancement of travel behavior modeling, offering new insights into the role of social networks in shaping travel and activity decisions and supporting the development of more effective transportation policies and interventions.

Chapter 3

Methodology

To address the research questions stated in Chapter 1.1, this research project utilizes Bayesian Belief Networks (BBNs) to synthesize social networks' size, composition and frequency of joint activities. This includes synthesizing social network compositions and sizes at the ego level, as well as detailed social relationships at the ego-alter level, with different considerations of frequency intervals and frequency of joint activities. The data used in this research project and data preparation process will be discussed in next chapter, Chapter 4. The methodology is detailed below.

Firstly, a brief introduction to BBNs. BBNs are graphical models representing probabilistic relationships among a set of variables, consisting of a directed acyclic graph (DAG) and a set of conditional probability distributions (Joubert & De Waal, 2020). They provide a framework for probabilistically encoding domain knowledge and uncertainties about the relationships (Joubert & De Waal, 2020). Given their efficiency and advantages in graphical representation, BBNs can identify critical relationships among different attributes and determine the core structure of a model with a limited number of samples. The characteristics and abilities of BBNs suit our research objectives of examining the relationship between social networks and the frequency of joint activities.

In this research project, two models are developed. Model I examines the overall composition and size of social networks and the frequency of various joint activities with all members of their network (regardless of the social categories) at the ego (individual) level. It includes individual socio-demographic characteristics, the size of each social category in the social network, and the frequency of each type of activity with all social network members. This model focuses on how socio-demographic characteristics contribute to the size of each social category (e.g., friends, colleagues) in an individual's social network and to what extent the social demographic of an ego as well as the size of various social categories impacts the frequency intervals of different joint activities.

Model II delves deeper into the relationships between egos and alters. This model aims to understand how the differences (or similarities) of egos and alters influence the frequency of their joint activities. This BBN includes variables such as differences in gender, age, and educational level between ego and alter, geographical distance between them, duration of the relationship, and type of social relationship. Specifically, Model II links the social relationships between egos and alters to the types and frequencies of joint activities, offering insights into how different social relationships influence the frequency of joint activities.

Figure 1 illustrates the conceptual framework of Model I and Model II described above. By utilizing these models, this research aims to synthesize realistic social networks that effectively capture types and sizes of social relationships and joint activity frequencies at ego and ego-alter levels to address sub-questions one and two.

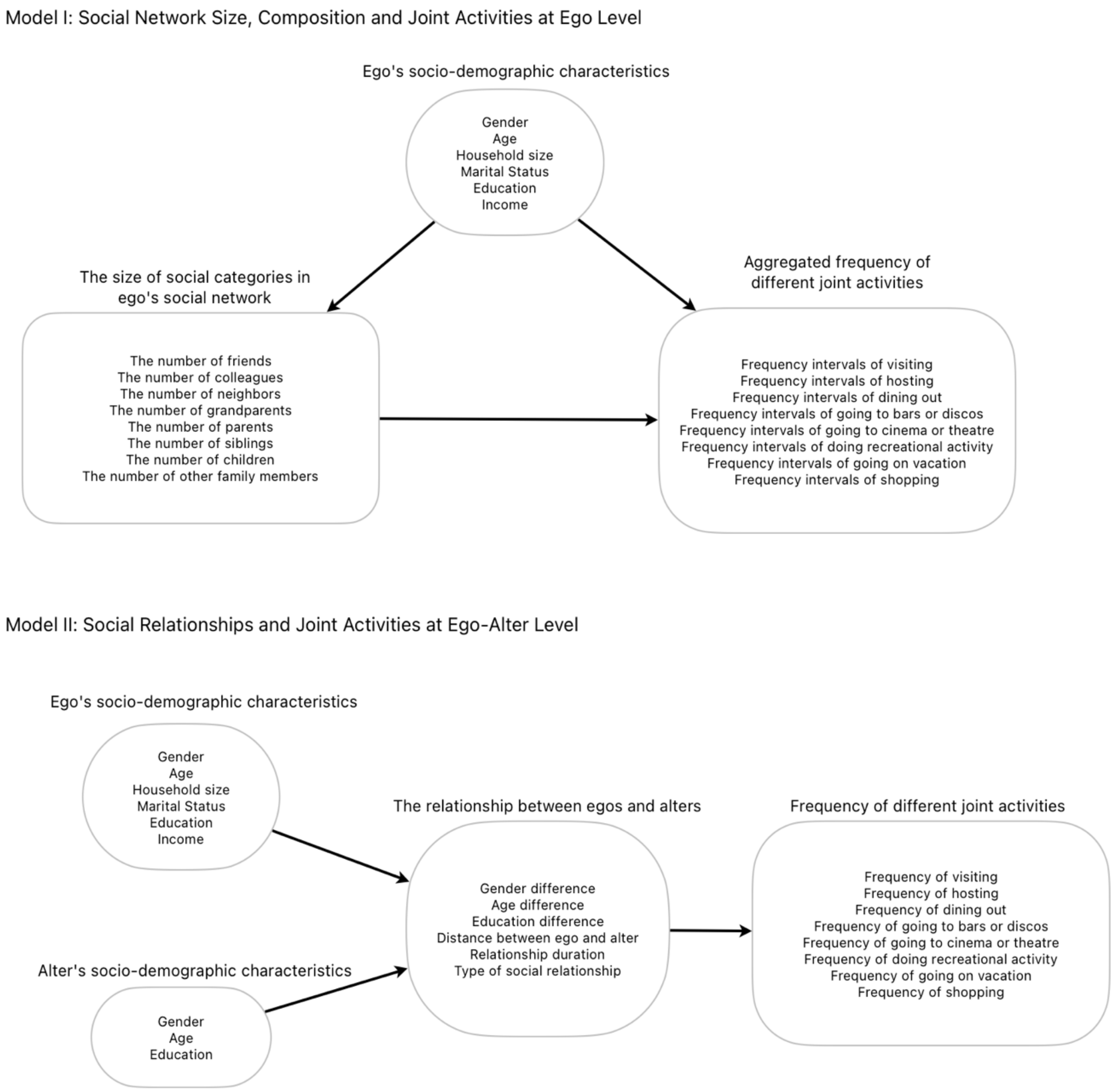

Figure 1: Conceptual Framework.

The following parts explain the theory of BBNs.

3.1 Construction of BBN

Constructing BBN involves defining the nodes X and edges $P(X)$ that represent variables and their relationships within the network (Sallard & Balać, 2023). Nodes can represent different attributes of interest such as demographic characteristics (gender, age, income, etc.), relationship type (family, friends, colleagues, etc.), duration of the relationship, joint activity types and frequency, etc. Edges between nodes capture probabilistic dependencies between variables, showing the influence of one variable on another.

$$X = \{X_1, \dots \dots, X_n\}$$

denotes the set of variables in BBN.

$$\Theta = \{ P(X_1 \mid \pi_1), \dots \dots, P(X_n \mid \pi_n)\}$$

represents local probability distributions for each node X_i, which is conditional on its parents π_i. These conditional probability distributions demonstrate the probability of X_i by giving each combination of its parent variables as evidence.

The network structure *G* of the BBN can be informed by domain knowledge and expert input, or by a data-driven approach using collected data (Sallard & Balać, 2023). By integrating model structure *G* and model parameters Θ, the joint distribution $P(X)$ for X can be decomposed into a factorized form with smaller, local probability distributions, each involving one node and its parents only:

$$P(X) = \prod_{i=1}^{n} P(X_i \mid \pi_i)$$

Building BBNs involves two primary steps: structure learning and parameter estimation (Zhang et al., 2019). Structure learning defines the network structure *G* that captures the conditional independence of random variables. This can be approached using two types of algorithms: constraint-based, which relies on domain knowledge to set dependencies and interpret the resulting models as causal, or score-based, which selects the network structure *G* with the highest score based on a selected criterion (Zhang et al., 2019). Concurrently, probabilistic parameters, such as conditional probabilities, are estimated from observed data through parameter learning (Sun & Erath, 2015).

3.2 Model selection and optimization

The purpose of structure learning is to determine the best structure of the model that captures the causal dependencies between the variables in the dataset.

Given the possible combinations of DAG, the scoring method is used to find the best fit of the data. For example, the Bayesian information criterion (BIC) is commonly used for model selection due to its balance of likelihood and penalty function, reducing overfitting. The Akaike information criterion (AIC) is another option, differing in the strength of the penalty function (Sun & Erath, 2015). After selecting a score function, the goal of the optimization stage is to identify the hypothetical structure with the highest score, which represents the best fit of the data.

There are several search algorithms to optimize the search space of all possible DAGs such as exhaustive search, hill climb search, and naïve bayes. However, exhaustive enumeration is impractical due to the exponential growth of potential structures with many nodes (Robinson, 1973; Sun & Erath, 2015). Therefore, heuristic search techniques, including hill-climbing, hill-climbing with restarts, tabu search, best-first search, K2, and Markov Chain Monte Carlo (MCMC) methods, are preferable (Heckerman, 1998; Sun & Erath, 2015).

Hill climbing search implements a greedy local search from a disconnected DAG and proceeds by iteratively performing single-edge manipulations that maximally increase the score. The search will terminate when a local maximum is found. It is suitable for structure learning of DAG with many nodes due to its computational feasibility. Furthermore, tree-augmented naïve bayes is a tree-based approach that can be used to model huge datasets involving lots of uncertainties among its various interdependent feature sets. It is more suitable for performing classification (Friedman et al., 1997). Naïve bayes only determines the feature variables to the dependent variable which is not suitable for the problem in this research project (Friedman et al., 1997).

Parameter learning involves estimating the values of the conditional probability distributions. It can be done by maximum likelihood estimation or Bayesian parameter estimation to find the parameter values that make the observed data most probable. However, the maximum likelihood estimation can lead to overfitting if the observed data does not represent the underlying distribution. In this case, the Bayesian parameter estimation is introduced. The Bayesian parameter estimation begins with existing prior conditional probability distributions, that express our beliefs about the variables before inputting observed data. Commonly, we assume uniform priors that deem all states equiprobable. Through parameter learning, conditional probability distributions quantitatively describe the statistical relationship between each node and its parents, thereby allowing for the precise quantification of relationships within the network.

This research project uses GeNIe software to analyze BBNs. There are several structure learning algorithms such as “Bayesian Search”, “Greedy Thick Thinning”, “PC algorithm”, “Naïve Bayes”, “Augmented Naïve Bayes”, and “Tree Augmented Naïve Bayes” found in GeNIe. As described above, “Naïve Bayes”, “Augmented Naïve Bayes”, and “Tree Augmented Naïve Bayes” are not suitable for our research objectives. The “PC algorithm” is used to learn the structure with continuous data. “Bayesian Search” algorithm undergoes a hill-climbing procedure, guided by a scoring function of log-likelihood. We implemented “Bayesian Search” for structure learning of BBNs in this research project.

3.3 Validation of BBN

To answer sub-question three, the constructed BBN can be validated to assess their accuracy and reliability. To show the performance, “Accuracy”, “Recall”, “Precision” and “F1-score” were calculated (Joubert & De Waal, 2020; Yang et al., 2023).

The “Accuracy” shows the proportion of correct predictions. The “Recall” is measured as the proportion of correctly predicted positive cases relative to the number of all true positives while the “Precision” represents the proportion of correctly predicted positive occurrences relative to all predicted positive. The “F1-score” is the weighted harmonic mean of the “Recall” and “Precision”.

$$Accuracy = \frac{TP + TN}{\sum test\ dataset}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F1 - score = \frac{2 * (Recall * Precision)}{Recall + Precision}$$

True positive (TP) is the number of cases predicted and observed in the data. False positive (FP) represents the number of cases in which the simulation predicts an event occurs but is not observed in the data. True negative (TN) represents the number of cases for which a certain event neither happened in the sample nor appeared in the simulation results. The False negative (FN) represents cases for which the event occurred in the data but did not appear in the simulation results.

Chapter 4

Data

4.1 Data collection

The data was collected in 2014 in the Netherlands via a web-based survey distributed to 1870 respondents through a market research company. The aim was to collect data on the size and composition of respondents' social networks, and the extent to which respondents have interactions with their social network members for various joint activities.

At the beginning of the questionnaire, the socio-demographic characteristics of respondents are acquired. Then, respondents were asked to specify the categories within their social networks, such as friends, colleagues, neighbors, grandparents, parents, siblings, children, and other family members. Each respondent was requested to provide a detailed list of individuals within each social category, revealing the size and composition of their social network. Four series of questions are asked in the following part of the questionnaire for each social category of contacts. Firstly, the sociodemographic characteristics of the identified social contacts. Respondents were prompted to furnish information on each contact, including their name, type of relationship (e.g., grandparents, parents, siblings, child, friends, colleagues, or neighbors), gender, age, nationality, education level, geographical distance from the respondent, and the duration of the relationship.

Secondly, the contact methods and the frequency of interaction per method are asked. The contact methods were divided into ten categories: telephone, facetime, email, SMS, Skype, WhatsApp, Facebook, LinkedIn, face-to-face, and others. Respondents could specify multiple contact methods for each social contact and the frequency of each method based on their actual interaction. In terms of face-to-face interaction, respondents were asked to provide detailed information on joint activities and the frequency of interaction for each activity type. Joint activities were categorized into eight types: visiting social contacts at their place, hosting social contacts at the respondent's place, dining out at restaurants, going out to bars or discos, attending theatres or cinemas, shopping together, traveling together, and participating in recreational activities.

In summary, the questionnaire aimed to collect detailed social network data, including the size, composition of social relationships, and joint activity patterns within respondents' social networks. Table 1 provides an overview of the collected data.

Table 1: Summary of variables collected through the questionnaire

Variable classes	Variables	Description	Categorical Levels
Ego level (Respondents' sociodemographic and social network characteristics)	id	Unique id of each respondent	-
	Gender	Gender of the respondent	1: Male 2: Female
	Age	Age group of the respondent	1: 18 to 25 2: 26 to 35 3: 36 to 50 4: 51 to 65 5: 66 to 75 6: Over 75
	Household size	The number of persons in respondents' household	
	Marital Status	The marital status of respondents	1: Single 2: Couples without children 3: Couples with children younger than 12 4: Couples with children older than 12
	Education	The highest education level of respondents	1: Low and middle 2: High
	Income	The income level of respondents	1: < 625euros 2: 625–1250euros 3: 1251-1875euros 4: 1876-2500euros 5: > 2500euros
	Friends	The total number of friends of each respondent	
	Colleagues	The total number of colleagues of each respondent	
	Neighbors	The total number of neighbors of each respondent	
	Grandparent	The total number of grandparents of each respondent	
	Parents	The total number of parents of each respondent, including mother and father-in-law or stepfather and stepmother	

	Siblings	The total number of siblings of each respondent, including brother and sister-in-law or stepbrother and stepsister	
	Children	The total number of children of each respondent, including grandchildren	
	Other family members	The total number of other family members of each respondent, including uncle, aunt, niece and nephew	
	Social Network	The size of respondent's social network	
Alter level (Social contacts' sociodemographic characteristics)	id_contact	Unique id of each social contact of each respondent	
	Gender	The gender of social contact	1: Male 2: Female
	Age	The age group of social contact	1: 18 to 25 2: 26 to 35 3: 36 to 50 4: 51 to 65 5: 66 to 75 6: Over 75
	Nationality	The nationality of social contact	
	Education	The education level of social contact	1: Low and middle 2: High
Ego-alter level (Social relationship characteristics)	Relationship duration	The duration of relationship between respondent and social contact	1: 1 year 2: 1-2 years 3: 2-5 years 4: 5-15 years 5: >15 years
	Distance	The geographical distance between respondent and social contact	1: < 500m 2: 500m-1km 3: 1-2km 4: 2-5km 5: 5-10km 6: 10-20km 7: 20-50km

			8: >50km 9: Abroad
	Type of social relationship	The category of social network which social network members belong	1: Friends 2: Colleagues 3: Neighbors 4: Grandparents 5: Parents 6: Siblings 7: Children 8: Other
Physical interaction (Joint activities)	Activity type	The type of face-to-face activities	1: Visiting 2: Hosting 3: Dining out 4: Bar 5: Cinema 6: Recreation 7: Vacation 8: Shopping
	Activity frequency	The frequency of certain type of face-to-face activity	1: Everyday 2: 2-3x per week 3:1x per week 4: 2-3x per month 5: 1x per month 6: 2-3x per year 7: 1x per year 8: < 1x per year

4.2 Data Preparation

This section provides a detailed explanation of the data preparation process for the datasets used in Models I and II.

The dataset for Model I focuses on variables at the ego level and the physical interaction level. Key variables include socio-demographic characteristics such as gender, age, education, household size, marital status, and income. Additionally, social network characteristics are captured, including the number of friends, colleagues, neighbors, grandparents, parents, siblings, children, and other family members. The dataset also includes aggregated joint activity frequencies for activities such as visiting, hosting, dining out, going to bars or discos, going to the cinema or theatre, engaging in recreational activities, shopping, and going on vacation. The aggregated joint activity frequencies are derived from the collected data at the ego (individual) level, representing an individual's total annual frequency intervals for conducting joint activities. Initially, these frequencies are recorded as categorical variables, indicating how often the individual engages in a specific joint activity with social network members on a weekly, monthly, or yearly basis. **To better capture the heterogeneity among individuals with different social network sizes, these categorical variables are transformed into continuous values representing annual frequencies. For instance, if a category indicates a frequency of two to three times per week, the lower bound is used to calculate an annual frequency, resulting in 96 occurrences per year (2 times per week * 4 weeks per month * 12 months per year). Similarly, if the frequency is once per month, the annual frequency is calculated as 12 times (1 time per month * 12 months per year).**

To aggregate the frequency of joint activities at the ego level, the individual frequencies for different social network members are summed, resulting in the total annual frequency for each joint activity. These aggregated annual frequencies are then transformed back into categorical variables, as Bayesian Belief Networks (BBNs) require discrete inputs. The transformed categorical variables represent frequency intervals for conducting joint activities, showing how often an individual engages in specific joint activities for a year. The frequency intervals vary depending on the nature of the joint activity. For instance, activities like visiting and hosting social network members may occur daily (with the highest frequency interval exceeding 365 times per year), while activities such as going on vacation are less frequent (with the highest frequency interval exceeding 12 times per year). The specific categories for different joint activities' frequency intervals can be found in Table 3.

For Model II, the dataset includes variables at multiple levels: ego, alter, ego-alter, and physical interaction levels. Socio-demographic characteristics at the ego level, such as gender, age, education, household size, marital status, and income, are considered. Additionally, three new variables are computed to capture differences in gender, age, and education between the ego and alter, representing homophily within social networks. Model II includes variables such as gender differences, age differences, education differences, the distance between ego and alter, relationship duration, and relationship types, all of which define the social relationships between ego and alter. The dataset also includes the frequency of various joint activities, such as visiting social network members, hosting social network members, dining out together, going to bars or discos together, going to the cinema or theatre together, engaging in recreational activities together, shopping together, and going on vacation together. The frequency categories for joint activities in Model II differ from those in Model I because Model II is disaggregated, focusing on the interactions between the ego and specific alters. This could provide additional insights from a disaggregated perspective, emphasizing the unique interpersonal connection between the individual and his or her certain social network member. As a result, the original frequency categories presented in the questionnaire are used in Model II. The specific frequency categories used in the BBN for Model II are detailed in Table 4.

The initial dataset comprised responses from 1870 individuals. However, due to missing data, a cleaning process was conducted to remove rows with incomplete information. After this process, the dataset for Model I consisted of 1278 respondents with complete socio-demographic and social network information. The dataset for Model II included 1272 respondents and 16682 unique ego-alter relationships, where respondents provided complete information regarding their social network members and at least one joint activity frequency.

In the following section, descriptive statistics for the datasets of both models will be presented.

4.3 Descriptive Statistic

Table 2 shows the descriptive statistics of respondents’ demographic and social network characteristics. In the dataset for Model I, 50.4% of the respondents are male and 49.4% are female. Regarding age group, 29.4% of the respondents are between 51 to 65 years old, 29.1% are 36 to 50, and 19.6% are between 26 to 56 years old. 41.6% of the respondents have a 2-person household, and 35.7% are household size in between 3 to 5 persons. 78.4% of the respondents are couples, and among them, 37.7% are without children and 40.7% have kids. 21.6% of the respondents are single. Most of the respondents have a low and middle-education level and 38.3% have high education. In terms of income level, 26.9% of the respondents earn between 1876 euros to 2500 euros per month, and 25.8% earn between 1251 euros to 1875 euros. As described in the previous section, respondents’ social network members are categorized into friends, colleagues, neighbors, grandparents, parents, siblings, children, and other family members. Friends and other family members make up a higher proportion of respondents' social networks, followed by colleagues and neighbors. On average, each respondent has 6.2 other family members (correspond to relatives such as uncle, aunt, niece, nephew, granduncle, grandaunt, grand-niece, and grand-nephew), 5.1 friends, 3.8 colleagues, and 3.4 neighbors.

The overall average social network size of the respondents is 25.1, which consists of all mentioned social categories. This figure aligns with the average network size reported in other studies (Arentze et al., 2012; Kim et al., 2020; Lin & Wang, 2014; Van Den Berg et al., 2009).

Table 2: Descriptive statistics of respondents' socio-demographics and their social network characteristics (N=1278)

Variables	Levels	Descriptions	Proportion (%)
Gender	1	Male	50.62
	2	Female	49.37
Age	1	18-25 years old	10.56
	2	26-35 years old	19.64
	3	36-50 years old	29.11
	4	51-65 years old	29.42
	5	66-75 years old	10.09
	6	>75 years old	1.17
Household size	1	1 person	21.52
	2	2 persons	41.55
	3	>2 persons	36.93
Marital status	1	Single	21.60
	2	Couples without children	37.72
	3	Couples with children below 12 years old	18.47
	4	Couples with children above 12 years old	22.22
Education	1	Low and middle	61.74
	2	High	38.26
Income	1	<625 euros	11.58
	2	625-1250 euros	22.46
	3	1251-1875 euros	25.82
	4	1876-2500 euros	26.92
	5	>2500 euros	13.22
The number of friends	Mean (SD)		5.1(5.4)
The number of colleagues	Mean (SD)		3.8(5.3)
The number of neighbors	Mean (SD)		3.4(4.2)
The number of grandparents	Mean (SD)		0.6(1.1)
The number of parents	Mean (SD)		1.6(2.1)
The number of siblings	Mean (SD)		2.9(3.5)
The number of children	Mean (SD)		1.5(3.0)
The number of other family members	Mean (SD)		6.2(11.5)
Social Network Size	Mean (SD)		25.1(20.9)

Table 3 presents the descriptive statistics of joint activity frequencies aggregated at the ego (individual) level. The aggregated joint activity frequencies show the total frequency of different joint activities taken by individuals with all their social network members in a year, and the levels of variables represent different annual frequency intervals.

In general, joint activities such as visiting and hosting social network members have a higher frequency than other joint activities. There are 35.4% of the respondents have a frequency of more than 365 times of visit social network members per year meaning that they at least visit one of their social network members every day. 43.7% of the respondents visit their social network members in a frequency between 48 to 365 times per year, meaning they visit one of their social network members at least once per week, while 9.7% visit their social network member 12 to 48 times a year meaning that at least once per month. Furthermore, 6.8% of the respondents visit their social network members 1 to 12 times a year. For the joint activity of hosting social network members, there are 28.3% of the respondents have a frequency of more than 365 times a year, 39.3% are 48 to 365 times, 13.8% are 12 to 48 times, and 7.8% are 1 to 12 times.

Joint activities such as dining out, doing recreational activities, and shopping are not daily activities, and they could be considered weekly activities. 16.0% of the respondents dine out with their social network members more than 48 times a year, meaning at least once a week. 10.7% of the respondents are in the frequency range of 12 to 48 times which means they dine out with their social network members at least once a month. There are 14.6% of the respondents dine out with their social network members 4 to 12 times a year and 32.2% are 1 to 4 times a year. More information on descriptive statistics for doing recreational activities and shopping can refer to Table 3, the levels of variables can be explained similarly to the joint activity of dining out.

Moreover, joint activities such as going to a bar or disco, cinema or theatre, and going on vacation with social network members are not as frequent as other joint activities mentioned above, the highest level of the variables is defined as at least once a month which is more than 12 times a year. However, there are only 16.8% of the respondents go to a bar or disco, 14.0% going to a cinema or theatre, and 6.2% going on vacation at this frequency interval. In the frequency interval of 1 to 4 times a year, 15.0% of the respondents for joint activity going to a bar or disco, 20.8% for a cinema or theatre, and 31.8% for vacation. Also, around 50 to 60% of the respondents are doing these joint activities with their social network members less than once a year.

To simply look at the differences in social network size according to different socio-demographic characteristics, we boxplot social network size by socio-demographic characteristics such as gender, age, household size, marital status, education, and income levels. Figure 2 presents the boxplot graphs that show the distribution of social network size by respondents' socio-demographic characteristics. There is no apparent difference between male and female respondents in their social network size although females are shown to have slightly larger networks. In terms of age, the total number of social network members gradually decreases with increasing age, indicating that younger respondents have larger social networks compared to older respondents(Lin & Wang, 2014; Van Den Berg et al., 2009). Furthermore, larger household sizes are associated with larger social networks. Respondents living with spouses and children have larger social networks than those who are single or couples without children. Respondents with higher education and income levels generally have a larger social network.

Table 3: Descriptive statistics of aggregated joint activity frequency (N=1278)

Variables	Levels	Description (times/year)	Proportion (%)
Frequency intervals of visiting	1	> 365	35.4
	2	48 - 365	43.7
	3	12 - 48	9.7
	4	1 - 12	6.8
	5	< 1	4.5
Frequency intervals of hosting	1	> 365	28.3
	2	48 - 365	39.3
	3	12 - 48	13.8
	4	1 - 12	7.8
	5	< 1	10.8
Frequency intervals of dining out	1	> 48	16.0
	2	12 - 48	10.7
	3	4 - 12	14.6
	4	1 - 4	32.2
	5	< 1	26.4
Frequency intervals of going to bar or disco	1	> 12	16.8
	2	4 - 12	5.0
	3	1 - 4	15.0
	4	< 1	63.1
Frequency intervals going to cinema or theatre	1	> 12	14.0
	2	4 - 12	7.7
	3	1 - 4	20.8
	4	< 1	57.4
Frequency intervals of doing recreational activity	1	> 48	16.0
	2	12 - 48	5.5
	3	4 - 12	6.7
	4	1 - 4	20.3
	5	< 1	51.5
Frequency intervals of going on vacation	1	> 12	6.2
	2	4 - 12	8.6
	3	1 - 4	31.8
	4	< 1	53.4
Frequency intervals of shopping	1	> 48	15.1
	2	12 - 48	8.2
	3	4 - 12	10.2
	4	1 - 4	22.1
	5	< 1	44.4

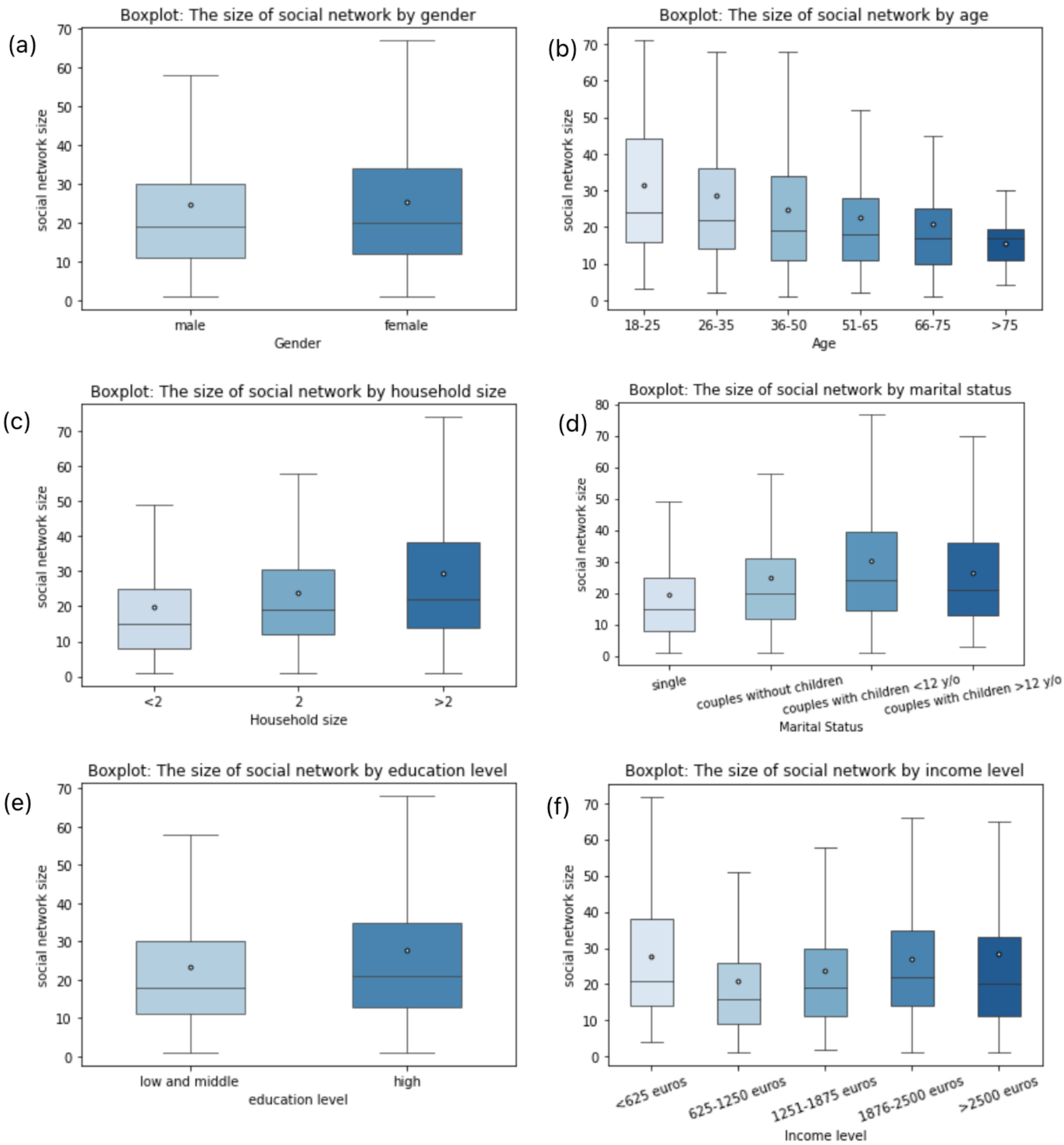

Figure 2: Boxplot of the total number of social network members from all social categories by respondents' sociodemographic characteristics. (a) by gender; (b) by age group; (c) by household size; (d) by marital status; (e) by education level; and (f) by income level.

In the dataset for Model II, 1272 respondents with 16682 ego-alter relationships were observed. Each ego has an average number of 13.2 social network members with whom they have physical interaction in their social network. The average number of social network members differs from the average social network size mentioned earlier because it only considers members with completed relationship information such as relationship duration and distance between ego and alter. To examine the effect of the relationship between respondents and their social network members on joint activity frequency, differences in gender, age, and education are discussed. Detailed descriptive statistics of the social relationship between ego and alter and joint activity frequency are presented in Table 4.

Figure 3 shows the distribution of differences in socio-demographic characteristics across different types of social relationships in social networks. Friends and colleagues are more likely to be the same gender and other social relationships are evenly distributed for the same and different gender according to Figure 3(a). Regarding age differences in Figure 3(b), most friends have no age gap, while respondents are at least 20 years younger than their parents, and respondents are at least 40 years younger than their grandparents. Regarding the relationship duration shown in Figure 3(d), siblings, and other family members are the groups with the largest number of members who respondents have known for more than 15 years. Social category colleagues are more likely to have a relationship duration of less than 5 years. As for the distribution of distance between egos and their social network members, as expected, neighbors are the ones with the largest share of members located very near to the respondents. The other social categories are distributed in the ranges of geographical distances from 1km to more than 50km. There is also a small proportion of social network members who have been indicated to be located abroad as shown in Figure 3(e). From Figure 3(f), we can conclude that respondents tend to interact face-to-face more with neighbors followed by family members. Unexpectedly digital interactions with friends have been higher than physical interaction, especially because the data was collected before COVID-19.

Table 4: Descriptive statistics of differences in socio-demographic characteristics and joint activity frequency (N=16682)

Variables	**Levels**	**Descriptions**	**Proportion (%)**
Gender difference	0	Same gender	59.8
	1	Different gender	40.2
Age difference	-3	At least 40 years old younger	11.5
	-2	At least 20 years old younger	11.2
	-1	At least 10 years old younger	16.2
	0	No age gap	39.1
	1	At least 10 years old older	13.9
	2	At least 20 years old older	6.0
	3	At least 40 years old older	2.2
Education difference	-1	1 level lower	8.4
	0	Same level	52.2
	1	1 level higher	39.5
Relationship duration	1	<5 years	25.4
	2	5-15 years	26.7
	3	>15 years	47.9
Distance between ego and alter	1	<1km	27.0
	2	1-10km	30.2
	3	10-50km	22.9
	4	>50km	16.3
	5	Abroad	3.6

Type of social relationship	1	Friends	26.7
	2	Colleagues	14.6
	3	Neighbors	13.9
	4	Grandparents	4.8
	5	Parents	7.7
	6	Siblings	12.7
	7	Children	6.2
	8	Other family members	13.4
Frequency of visiting	1	Everyday	3.4
	2	At least once a week	13.4
	3	At least once a month	22.7
	4	At least once a year	29.9
	5	Less than once a year	9.6
	6	Never	20.9
Frequency of hosting	1	Everyday	3.0
	2	At least once a week	10.4
	3	At least once a month	17.1
	4	At least once a year	27.5
	5	Less than once a year	8.9
	6	Never	33.1
Frequency of dining out	1	Everyday	0.8
	2	At least once a week	2.4
	3	At least once a month	5.4
	4	At least once a year	21.1
	5	Less than once a year	8.7
	6	Never	61.6
Frequency of going to bar or disco	1	Everyday	0.7
	2	At least once a week	1.5
	3	At least once a month	2.7
	4	At least once a year	4.9
	5	Less than once a year	9.6
	6	Never	80.6
Frequency of going to cinema or theatre	1	Everyday	0.7
	2	At least once a week	0.9
	3	At least once a month	1.9
	4	At least once a year	6.0
	5	Less than once a year	9.2
	6	Never	81.3
Frequency of doing recreational activity	1	Everyday	0.9
	2	At least once a week	2.4
	3	At least once a month	3.6
	4	At least once a year	7.8
	5	Less than once a year	7.5
	6	Never	77.8

Frequency going on vacation	1	Everyday	0.7
	2	At least once a week	0.9
	3	At least once a month	1.2
	4	At least once a year	7.2
	5	Less than once a year	10.2
	6	Never	79.8
Frequency of shopping	1	Everyday	0.8
	2	At least once a week	1.9
	3	At least once a month	3.6
	4	At least once a year	6.5
	5	Less than once a year	8.0
	6	Never	79.2

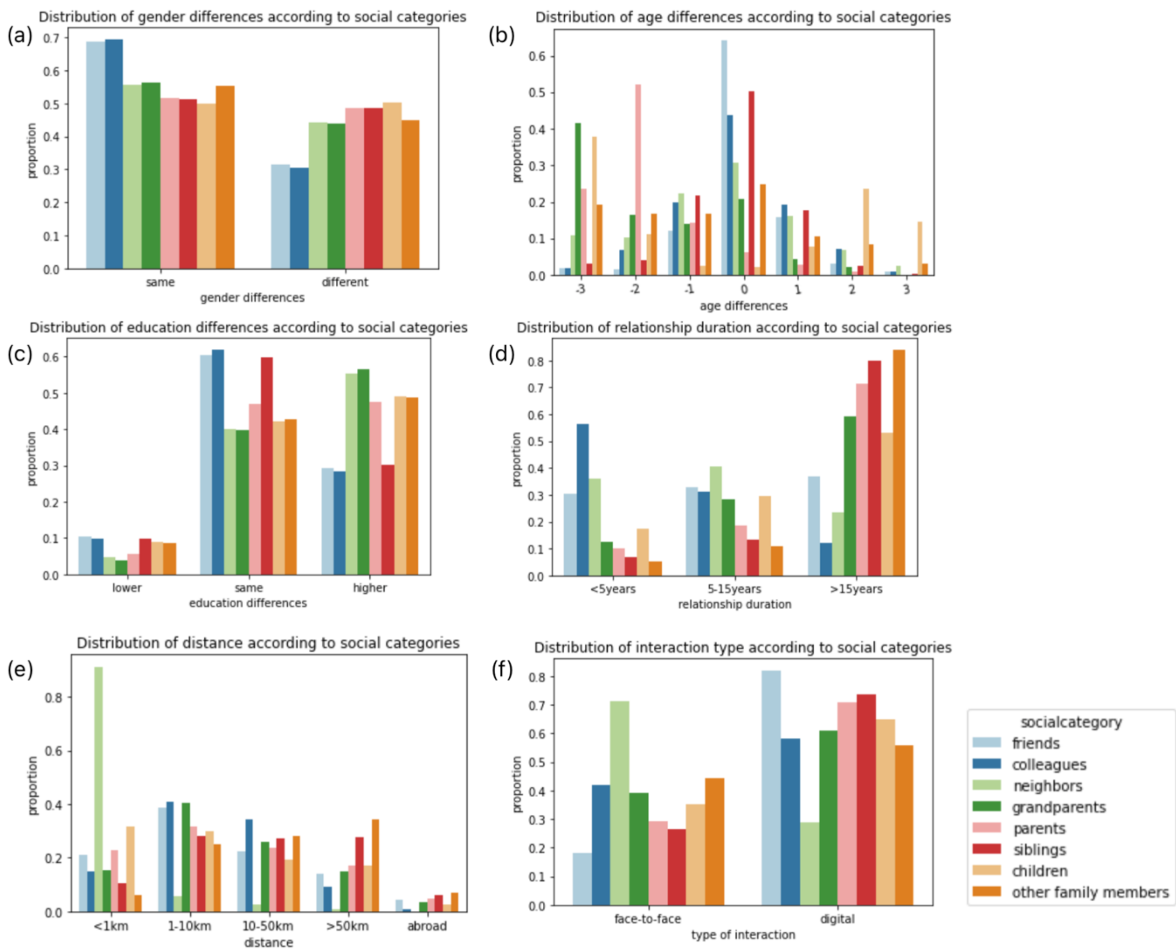

Figure 3: Data distributions according to the social categories. (a) gender differences; (b) age differences; (c) education differences; (d) relationship distance; (e) relationship duration; and (f) type of interaction.

Chapter 5

Results

This research project employs BBNs to synthesize social networks that capture various types of social relationships and frequencies of joint activities. Two models are developed. First, we model social network size, composition, and joint activity patterns at the ego level in Model I. Second, at the ego-alter level, the type of social relationship between the ego and alter and the frequency of joint activities are modeled in Model II. The BBN is constructed using the software GeNIe. The “Bayesian Search” algorithm is chosen for structure learning. Then, parameter learning is carried out to estimate the conditional probability distribution of variables. By making inferences through BBNs, we can explore how social networks affect the frequency of joint activities.

The results of both models explained in Chapter 3 will be presented and discussed in the following section. The BBN structure of both models will be first discussed, and then we present the validation of BBNs. In the end, instantiations for both models are shown.

5.1 Model I: Social Network Size, Composition, and Joint Activities at Ego Level

Model I consists of three sets of variables: the ego’s demographic characteristics, the ego’s social network size, and the frequency of different joint activities. It aims to model individuals’ social network size and total frequency of joint activities by social network members regardless of the type of social relationship. The frequency is represented by the number of times an activity took place in a year. This variable of joint activity frequencies is categorical and the levels of variable show different intervals of annual frequencies, referring to Table 3. In addition, the model examines how the size and composition of an individual's social network vary based on their socio-demographic characteristics.

Figure 4 shows the network structure of Model I, displaying the interdependencies between variables, providing an overview of social networks and the frequency of conducting different types of activities (times/year) with social network members at the individual level. The age of an individual influences the number of colleagues, parents, children, grandparents, and other family members in his or her social network. The marital status of an individual also influences the number of parents, and it is understandable as parents-in-law are included in the social category of parents. The age of an individual affects the frequency of joint activity of going to bars or discos directly. The frequency of joint activity hosting social network members is influenced by the number of siblings and household size of an individual. Furthermore, the number of other family members and siblings are the parent nodes of the frequency of various joint activities such as doing recreational activity, shopping, and going on vacations. The number of grandparents and other family members also affects the frequency of joint activity of going to the cinema or theatre. The frequency of joint activity dining out with social network members is affected by the number of friends and colleagues in an individual’s social network. Through Figure 4, we can observe the interdependencies between socio-demographic characteristics, social network size, and frequency of conducting joint activities.

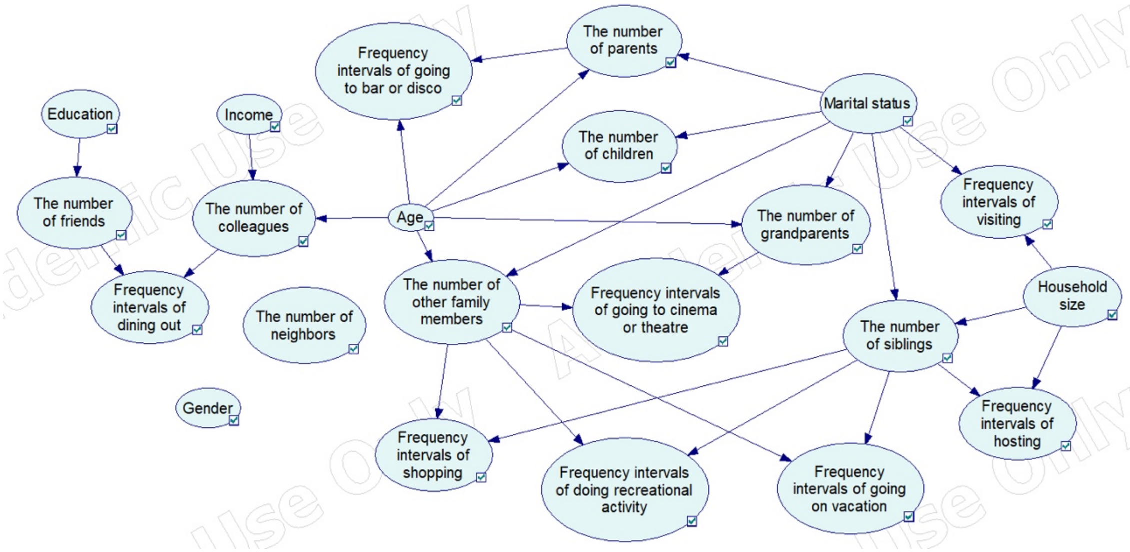

Figure 4: DAG of Model I (log(p)=-333864.4).

5.2 Model II: Social Relationship and Joint Activities at Ego-Alter Level

Model II consists of variables of the ego's socio-demographic characteristics, the social relationship between ego and alter, and the frequency of different joint activities. It aims to model the social relationship between ego and alter and the frequency of joint activities at a disaggregated level of individuals and each of their social network members. The links between ego's socio-demographic and social relationship between ego and alter will also be examined. Specifically, Model II explores how social relationships with certain social network members influence the frequency of various joint activities.

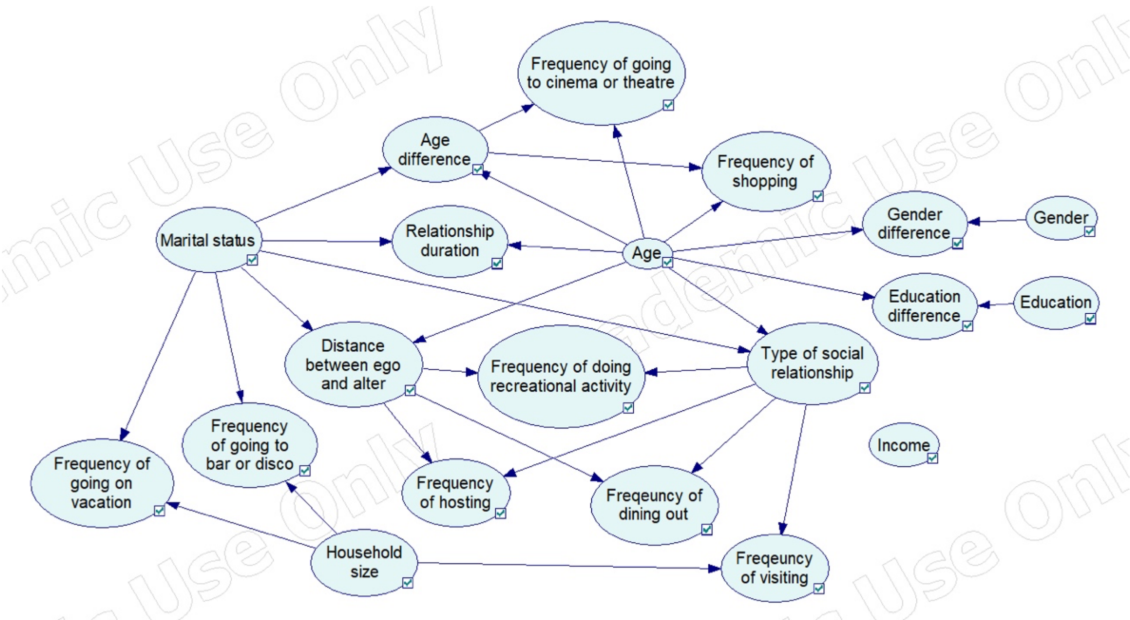

Figure 5: DAG of Model II (log(p)=-419302).

Figure 5 shows the DAG representing the network structure of Model II, displaying interdependencies between variables of socio-demographic characteristics, the social relationship between ego and alter, and the frequency of joint activities. This provides an overview of synthesizing social networks and the frequency of joint activities conducted at the ego-alter level. The age and marital status of individuals are key determinants of social relationship types with their social network members. The age of an individual influences the gender difference, age difference, and education difference between individuals and their social network members. The gender difference is affected by the gender of an individual while the education difference is influenced by the education level of an individual. The age of an individual together with marital status influences the relationship duration of social network members, the distance between ego and alter, and the type of social relationship. In addition, age and age differences affect the frequency of shopping and going to the cinema or theatre with social network members. The distance between ego and alter and the type of social relationship affect the frequency of joint activities such as dining out, hosting social network members, and doing recreational activity. Through Figure 5, we can define the interdependencies between the socio-demographic characteristics of individuals, the social relationship between ego and alter, and the frequency of joint activities.

5.3 Validation of the BBNs

As described in Chapter 3, four indicators, "Accuracy", "Recall", "Precision", and "F1-score" are calculated to assess the validity of BBNs constructed (Yang et al., 2023). The calculation and explanation of these indicators check the equations in Chapter 3.3.

Table 5 records the accuracy, recall, precision, and F1-score of predicted variables in Model I. As seen, the highest accuracy of prediction related to the number of social network members is associated with the number of other family members followed by the number of siblings and parents. The lowest accuracy is related to the prediction of the number of children and then grandparents. In terms of accuracy of activity frequency jointly conducted with all members of the social network, Table 5 suggests the highest accuracy is achieved for going on vacation followed by going to a cinema or theatre while the lowest accuracy is associated with going to a bar or disco with members of social network.

Table 6 records the accuracy, recall, precision, and F1-score of predicted variables in Model II. As for the accuracy of prediction related to the difference in socio-demographic characteristics, the best prediction was made for the age difference followed by the difference in education. The lowest accuracy belongs to the prediction of gender difference. The model performs relatively well in predicting the type of social relationship and the distance between ego and alter. When it comes to the prediction accuracy of the frequency of joint activities, the highest accuracy is achieved for going to cinema or theatre followed by shopping. However, the type of activity for which the lowest accuracy is attained is different from Model I and is for the visiting social network members (in their houses) in Model II.

The results indicate that while both models perform reasonably well, Model II demonstrates a better predictive accuracy compared to Model I (computing average accuracy for both models results in 62.69% for Model I and 80.12% for Model II). This higher accuracy can be attributed to Model II's more granular approach, which considers the disaggregated level of social relationships between individuals and their social network members and their specific characteristics. By capturing detailed social relationships between ego and alter, Model II can provide more precise predictions.

Overall, the accuracy figures reflect the models' ability to capture the complexities of social networks and their influence on joint activities, albeit with some limitations. The higher accuracy of Model II suggests that incorporating disaggregated social relationship between ego and alter enhances the model's predictive power, offering a better understanding of social network impacts on joint activity frequencies.

Table 5: The accuracy, recall, precision, and F1-score of predicted variables in Model I (N=1278).

Variables	**Accuracy (%)**	**Recall(%)**	**Precision(%)**	**F1-score(%)**
The number of friends	60.02	26.24	8.25	12.36
The number of colleagues	57.22	35.96	35.89	35.23
The number of neighbors	49.14	33.33	7.90	12.78
The number of grandparents	33.12	36.88	36.48	32.89
The number of parents	63.15	27.30	24.82	25.30
The number of siblings	63.58	24.40	22.79	19.66
The number of children	28.48	28.79	27.01	25.79
The number of other family members	63.65	25.07	24.54	21.90
Frequency intervals of visiting	66.79	18.33	11.93	11.65
Frequency intervals of hosting	68.08	23.52	8.84	11.62
Frequency intervals of dining out	68.04	21.05	12.03	15.14
Frequency intervals of going to bar or disco	59.47	25.68	32.01	17.66
Frequency intervals of going to cinema or theatre	68.78	24.17	18.07	19.32
Frequency intervals of doing recreational activity	65.01	21.75	7.57	10.49
Frequency intervals of going on vacation	73.55	27.61	20.78	21.78
Frequency intervals of shopping	67.36	20.00	3.68	6.21

Table 6: The accuracy, recall, precision, and F1-score of predicted variables in Model II (N=16682).

Variables	**Accuracy (%)**	**Recall(%)**	**Precision(%)**	**F1-score(%)**
Gender difference	59.85	50.00	29.92	37.44
Age difference	82.70	15.69	24.47	11.21
Education difference	71.31	40.55	37.95	38.97
Relationship duration	65.73	36.16	30.81	28.84
Distance between ego and alter	72.64	22.36	18.87	18.81
Type of social relationship	81.69	12.73	7.92	5.77
Frequency of visiting	76.64	16.66	6.18	7.69
Frequency of hosting	77.69	16.67	5.51	8.29
Frequency of dining out	87.21	16.67	10.27	12.71
Frequency of going to bar or disco	83.75	23.27	13.80	12.02
Frequency of going to cinema or theatre	93.76	16.67	13.55	14.95
Frequency of doing recreational activity	92.60	16.67	12.97	14.59
Frequency of going on vacation	83.03	22.33	13.69	12.00
Frequency of shopping	93.08	16.67	13.21	14.74

5.4 Instantiation of Model I

To better understand the impact of social network size on frequency intervals of joint activity, we instantiated certain levels of parent nodes and inferred the posterior distribution of child nodes. Figure 6 shows the posterior probability distribution of joint activity dining out with social network members by instantiating the number of friends in an individual's social network. Figure 6(a) presents the probability distribution before instantiation, while Figures 6(b) to 6(e) show the probability distribution when instantiating a certain level of friends' numbers.

Individuals having more than eight friends have the highest probability of conducting the joint activity of dining out with their social network members more than 48 times a year (which is the most frequent category of frequency intervals), indicating a frequency of at least once a week compared to the baseline scenario. This group is followed by those who have less than three friends as these people have the second highest probability of dining out more than 48 times a year. They may suggest that having very strong connections with the few friends they have leads to dining out together at least once a week. According to Figure 6(b), 32% of joint activity dining is predicted to occur between 4 to 11 times a year, indicating a frequency interval of less than once a month but more than four times a year when fewer than three friends are instantiated. Conversely, 27% of joint activity dining out with social network members is predicted to occur between 12 to 48 times a year when more than eight friends are instantiated, indicating that individuals dine out with their social network members at least once a month.

Figure 7 shows the posterior probability distribution of frequency intervals of dining out with social network members by instantiating the number of colleagues in an individual's social network. There are three states for the number of colleagues: less than two; between two to four; and more than four, and they are instantiated accordingly. Figure 7(a) presents the probability distribution before instantiation, while Figures 7(b) to 7(d) show the probability distribution when instantiating a certain state of colleagues' numbers.

The frequency interval of more than 48 times a year for joint activity dining out with social network members has a higher probability of being predicted for individuals with a larger number of colleagues (22% for two to four colleagues versus 17% for less than two colleagues). A trend is identified where a larger size of colleagues leads to a higher frequency of joint dining activity within a certain range. The probability of dining out between 12 to 48 times a year increases gradually with increasing numbers of colleagues (8% in Figure 7(b), 23% in Figure 7(c), 37% in Figure 7(d)). Similarly, the probability of dining out between 1 to 3 times a year decreases gradually with increasing colleagues' numbers (32% in Figure 7(b), 20% in Figure 7(c), 13% in Figure 7(d)). It is important to note that there is only 9% of joint activity dining out at the frequency interval of more than 48 when the category of more than four colleagues is instantiated. We assume that colleagues represent a social relationship that is not as strong as other social categories (e.g. friends), making it less likely to dine out regularly at least once a week. However, there is a higher chance that individuals dine out with their colleagues at least once a month occasionally due to company activities.

Based on the network structure of Model I represented in Figure 4, "Frequency intervals of going on vacation" has parent nodes of "The number of siblings" and "The number of other family members". Thus, the number of siblings and other family members in individuals' social networks influences the frequency intervals of going on vacation. Figure 8 shows the posterior probability distribution of "Frequency intervals of going on vacation" with social network members by instantiating "The number of siblings" in an individual's social network. Figure 8(a) presents the probability distribution before instantiation, while Figures 8(b) to 8(e) show the probability distribution when instantiating a certain level of sibling' numbers. Higher frequency intervals of joint activity vacations are more likely to be predicted with a larger number of siblings. There is a significant increase in the probability of frequency intervals more than 12 times a year for going on vacation with social network members when the number of siblings increases from less than 3 siblings (21% in Figure 8(b), to between 3 to 5 siblings (30% in Figure 8(c), and further to between 6 to 8 siblings (37% in Figure 8(d)). This indicates that an individual has at least one social network member who goes on vacation together at least once a month. In Figure 8(e), the probability of never going on vacation with social network members or less than once a year is 33% when the category of more than eight siblings is instantiated. We assume that it could be difficult to maintain close relationships when the number of siblings is so large, and this causes fewer interactions among them. Also, the number of siblings could be large since in this variable, sisters, brothers, stepsisters, stepbrothers, sister-in-law, and brother-in-law are included.

Figure 9 shows the posterior probability distribution of the frequency intervals of going on vacation with social network members by instantiating the number of other family members in an individual's social network. Specifically, Figure 9(a) presents the probability distribution before instantiation, while Figures 9(b) to 9(e) show the probability distribution when instantiating a certain level on the number of other family members. The impact of the number of other family members shows the same trend as the number of siblings on the frequency intervals of going on vacation with social network members. Therefore, the interpretation of the dependencies among variables "The number of other family members" and "Frequency intervals of going on vacation" is identical to Figure 8. However, there is a 30% probability of going on vacation at the frequency interval of 1 to 3 times a year when the category of more than eight other family members is instantiated. This indicates that higher probability of an individual going on vacation once a year with a larger number of other family members (30% in Figure 9(e)) than with a larger size of siblings (23% in Figure 8(e)). This is backed up by the finding from Lin and Wang (2014) stated that social companionship by other family members as a source of instrumental support leads to more joint activities.

In conclusion, the findings from Model I illustrate the intricate relationships between individuals' socio-demographic characteristics, the size and composition of their social networks, and the frequency intervals of various joint activities. By analyzing the posterior probability distributions, we have elucidated how the number of different social categories within an individual's social network, such as friends, colleagues, siblings, and other family members, impacts their likelihood of engaging in specific activities like dining out and going on vacations. The data reveals distinct trends, such as individuals with fewer or more friends dining out more frequently and those with larger numbers of siblings or other family members having higher probabilities of frequent vacations. A limited discussion can be made referencing the existing literature since there are limited studies on social categories of individual social networks and the type and frequency of joint activities. Our findings of Model I contribute to current knowledge and literature by adding empirical investigation of how the number of different social categories in an individual's social network influence the frequency intervals of certain joint activity.

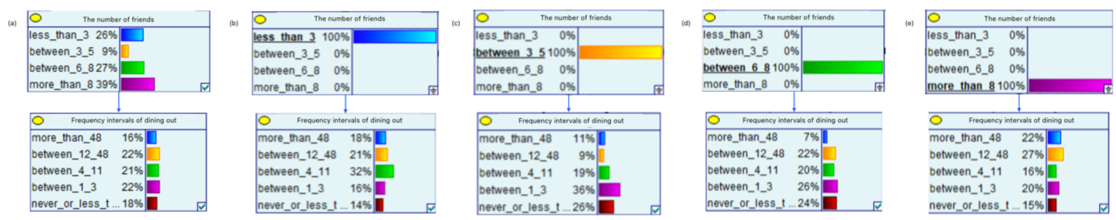

Figure 6: Posterior probability distribution of child node "Frequency intervals of dining out" by instantiating parent node "The number of friends".

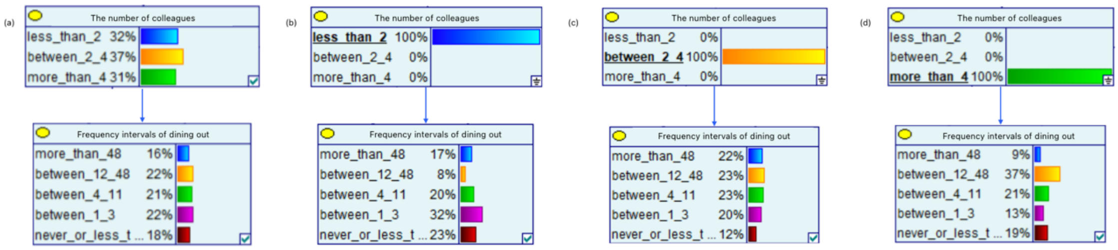

Figure 7: Posterior probability distribution of child node "Frequency intervals of dining out" by instantiating parent node "The number of colleagues".

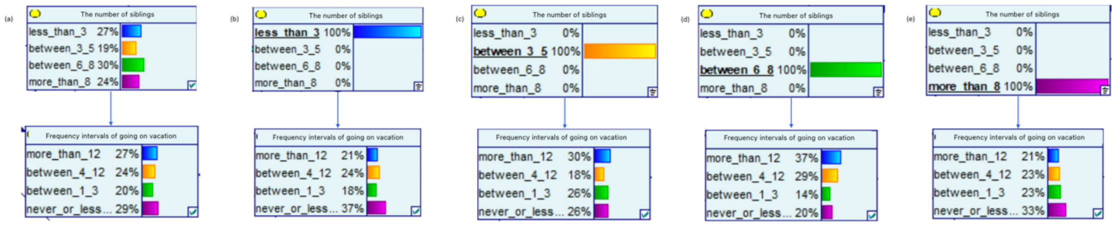

Figure 8: Posterior probability distribution of child node "Frequency intervals of going on vacation" by instantiating parent node "The number of siblings".

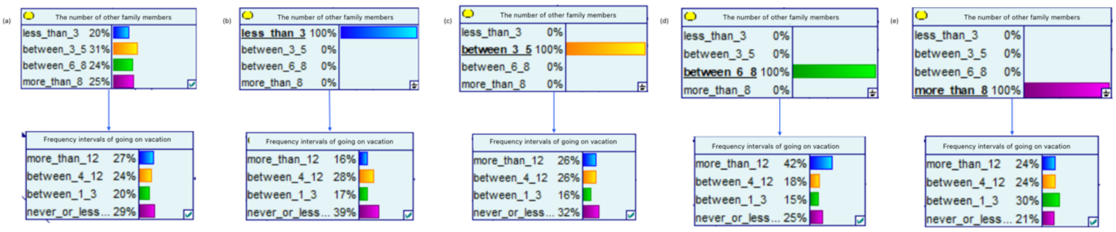

Figure 9: Posterior probability distribution of child node "Frequency intervals of going on vacation" by instantiating parent node "The number of other family members".

5.5 Instantiation of Model II

Like in Model I, we instantiated certain levels of parent nodes and inferred the posterior distribution of child nodes to better understand the impacts of social relationship types on joint activity frequencies. Figure 10 shows the probability distribution of the frequency of three joint activities by instantiating the type of social relationship between individuals and their social network members. Figure 10(a) presents the probability distribution before instantiation, while Figures 10(b) to 10(i) show the probability distribution when instantiating a certain level of type of social relationship.

Some trends can be identified in Figure 10 between the type of social relationship and the frequency of joint activities. When the social relationship between an individual and social network member as friends is instantiated, the probability of visiting a social network member is 4% at a frequency of every day, 15% at the frequency of at least once a week, 28% at least once a month, 31% at least once a year, 6% less than once a year and 16% never. While the probability of hosting a social network member is 4% at the frequency of every day, 12% at least once a week, 21% at least once a month, 29% at least once a year, 7% less than once a year, and 27% never. However, we examined a significant drop in the probability of these regular frequencies and an increase in never when the social category of colleagues is instantiated. There is only a probability of 2% of visiting a social network member every day, 6% at least once a week, 10% at least once a month, 16% at least once a year, 16% less than once a year, and 49% never. For hosting social network members, 2% at the frequency of every day, 8 % at least once a week, 8% at least once a month, 12% at least once a year, 15% less than once a year, and 55% never. We can conclude that individuals are more likely to visit or host their friends than colleagues. This is supported by the finding of Lin and Wang (2014) stated that people tend to perform more social activities with those who they receive emotional and instrumental support from.

Furthermore, there is a higher probability of visiting or hosting social network members every day when types of social relationships such as parents and children are instantiated. It is 8% of visiting their parents and 5% of hosting their parents every day for an individual. On the other hand, 15% of visiting their children, and 11% of hosting their children every day. The distribution of other frequency levels can be referred to in Figures 10(f) and 10(h), a large proportion occurs either at least once a week or once a month. Moreover, it is significant that individuals visit or host their other family members at least once a year based on Figure 10(i). There is a probability of 47% for individuals visiting their other family members and 30% of hosting them at least once a year. For joint activity dining out together, most probably happens at a frequency of at least once a year. Parents (37%) and children (31%) are the social categories in an individual's social network that are found to be more likely to dine out with. However, social categories such as friends (27%), colleagues (21%), and siblings (24%) also pose a relatively high possibility that individuals will dine out together.

In conclusion, Model II provides a comprehensive understanding of how socio-demographic characteristics and social relationships between ego and alter influence the frequency of joint activities at the ego-alter level. By examining the disaggregated level of social relationships between individuals and their social network members, the model highlights the complex ways in which different types of social relationships impact the frequency of various joint activities. Similarly, to my best knowledge, there is a limited comparison that can be made between our findings and current studies due to the lack of exploration about the frequency of joint activities. However, our findings add to the existing body of literature by examining the dependencies among social relationships between ego and alter and the frequency of various joint activities.

Figure 10: Posterior probability distribution of child nodes “Frequency of visiting”, “Frequency of hosting” and “Frequency of dining out” by instantiating parent node “Type of social relationship”.

Chapter 6

Conclusion

6.1 Main Findings and Contribution

This research project has demonstrated the effectiveness and potential of Bayesian Belief Networks (BBNs) in synthesizing social networks at both the ego and ego-alter levels for travel behavior modeling. BBNs have proven to be particularly adept at capturing and modeling the complex interdependencies between various factors, such as socio-demographic characteristics, social network size, composition, social relationships, and the frequency of joint activities. We have successfully constructed BBNs that elucidate how these socio-demographic characteristics, as well as the types and strengths of social relationships, interact with different types and frequencies of joint activities. Two models were developed: Model I and Model II. Both models aim to clarify the impact of socio-demographic characteristics and social networks on the frequency of joint activities, though they approach the problem from different perspectives and levels of granularity. Overall, this research provides a foundation for the further integration of social networks into travel behavior modeling. By incorporating synthetic social networks, this work could contribute to addressing challenges such as data limitations and privacy concerns in the development of more accurate and predictive travel behavior models.

Model I demonstrates how an individual's socio-demographic characteristics influence the number of different social categories within their social network and to what extent social network size impacts the frequency of aggregated joint activities. For instance, we found that individuals with a larger number of friends in their social network tend to have higher frequency intervals of dining out. Additionally, individuals with six to eight siblings or other family members are more likely to go on vacation with their social network members at least once a month (with a frequency interval of more than 12 times a year). Model II delves into the disaggregated levels of social relationships between individuals and their social network members. It specifically examines the impact of ego socio-demographic characteristics on differences between the ego and alter, as well as how these differences along with factors such as geographic distance and the duration of the relationship affect the frequency of joint activities. The findings indicate a higher probability of individuals visiting or hosting friends rather than colleagues. Additionally, individuals are more likely to dine out with family members, such as their parents or children.

Model I and Model II offer complementary perspectives on the impact of social networks on joint activity frequencies. Model I provides an aggregated view that is useful for identifying broad patterns in the frequency intervals of joint activities. In contrast, Model II offers a detailed, disaggregated analysis that is essential for understanding the nuances of individual social relationships with their social network members. Together, these models provide a comprehensive understanding of how social networks at ego and ego-alter levels influence the frequency of joint activities, highlighting the importance of both network size and the nature of specific social relationships in shaping the frequency of various joint activities. By synthesizing social networks, researchers and policymakers can simulate and study social network and travel behaviors without compromising privacy or requiring extensive data collection.

6.2 Limitations and Future Research

The successful application of BBNs in this research opens several avenues for future work. The methodology developed here can be extended to other domains where social networks play a critical role, such as public health, marketing, and organizational behavior. The ability to model and synthesize social networks can significantly aid in understanding and predicting the spread of information, behaviors, and diseases. Moreover, this research highlights the importance of considering different levels of granularity and incorporating additional variables to enhance the interpretability and accuracy of the models.

A particularly promising area for future research is the integration of synthetic social networks into travel behavior models. This integration could provide more comprehensive and realistic simulations of travel behavior by accounting for the influence of social interactions on decision-making processes. Future research could explore how synthetic social networks could be dynamically linked with travel demand models to simulate the impact of social relationships on travel choices more accurately. This could lead to better-informed transportation policies and more personalized travel recommendations.

However, several limitations of this research project must be addressed in future studies. First, data collection for joint activity frequencies was done categorically, providing only approximate frequencies within certain ranges (e.g., per week, per month, per year). Most of the data collected accumulated in one category, with many missing data points for other joint activities except for visiting and hosting social network members. This introduced biases into our analysis. Although efforts were made to recompute frequencies to examine individual heterogeneity, the results and accuracy of the models were not ideal. Notably, a significant gap was found between the accuracy of Models I and II. Future researchers should prioritize more precise data collection methods regarding joint activity frequencies. Additionally, the data used in this research was collected in 2014, nearly a decade ago. Social behaviors and interactions have likely evolved since then, particularly in light of the digital interactions that became more prevalent during and after the COVID-19 pandemic. Thus, our findings may not fully represent the current situation, and future research should consider these changes in social dynamics.

In conclusion, this research project showcases the robust capability of Bayesian Belief Networks in modeling and understanding social networks and the frequency of joint activities. The experimental approach not only validates the theoretical potential of BBNs but also provides practical tools for synthesizing social networks with the frequency of joint activities. This work offers significant contributions to both academic research and practical applications in travel behavior field, although there is still progress to be made.

Bibliography

[1] Arentze, T. A., Kowald, M., & Axhausen, K. W. (2012). *A method to model population-wide social networks for large scale activity-travel micro-simulations* [Application/pdf]. 18 p. https://doi.org/10.3929/ETHZ-B-000038374

[2] Arentze, T., & Timmermans, H. (2008). Social Networks, Social Interactions, and Activity-Travel Behavior: A Framework for Microsimulation. *Environment and Planning B: Planning and Design*, *35*(6), 1012–1027. https://doi.org/10.1068/b3319t

[3] Arentze, T., Van Den Berg, P., & Timmermans, H. (2012). Modeling Social Networks in Geographic Space: Approach and Empirical Application. *Environment and Planning A: Economy and Space*, *44*(5), 1101–1120. https://doi.org/10.1068/a4438

[4] Axhausen, K. W. (2008). Social Networks, Mobility Biographies, and Travel: Survey Challenges. *Environment and Planning B: Planning and Design*, *35*(6), 981–996. https://doi.org/10.1068/b3316t

[5] Axhausen, K. W., & Hackney, J. K. (2006). *An agent model of social network and travel behavior interdependence* [Application/pdf]. 47 p. https://doi.org/10.3929/ETHZ-A-005228628

[6] Barajas, J. M. (2021). The roots of racialized travel behavior. In *Advances in Transport Policy and Planning* (Vol. 8, pp. 1–31). Elsevier. https://doi.org/10.1016/bs.atpp.2021.06.007

[7] Caiati, V., Rasouli, S., & Timmermans, H. (2020). Bundling, pricing schemes and extra features preferences for mobility as a service: Sequential portfolio choice experiment. *Transportation Research Part A: Policy and Practice*, *131*, 123–148. https://doi.org/10.1016/j.tra.2019.09.029

[8] Calastri, C., Hess, S., Daly, A., & Carrasco, J. A. (2017). Does the social context help with understanding and predicting the choice of activity type and duration? An application of the Multiple Discrete-Continuous Nested Extreme Value model to activity diary data. *Transportation Research Part A: Policy and Practice*, *104*, 1–20. https://doi.org/10.1016/j.tra.2017.07.003

[9] Carrasco, J.-A., & Miller, E. J. (2009). The social dimension in action: A multilevel, personal networks model of social activity frequency between individuals. *Transportation Research Part A: Policy and Practice*, *43*(1), 90–104. https://doi.org/10.1016/j.tra.2008.06.006

[10] Cherchi, E. (2017). A stated choice experiment to measure the effect of informational and normative conformity in the preference for electric vehicles. *Transportation Research Part A: Policy and Practice*, *100*, 88–104. https://doi.org/10.1016/j.tra.2017.04.009

[11] Dugundji, E. R., & Walker, J. L. (2005). Discrete Choice with Social and Spatial Network Interdependencies: An Empirical Example Using Mixed Generalized Extreme Value Models with Field and Panel Effects. *Transportation Research Record: Journal of the Transportation Research Board*, *1921*(1), 70–78. https://doi.org/10.1177/0361198105192100109

[12] Frei, A., & Axhausen, K. W. (2007). *Size and structure of social network geographies* (p. 37 p.) [Application/pdf]. ETH Zurich. https://doi.org/10.3929/ETHZ-A-005562753

[13] Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian Network Classifiers. *Machine Learning*, *29*(2), 131–163. https://doi.org/10.1023/A:1007465528199

[14] Ilahi, A., & Axhausen, K. W. (2019). Integrating Bayesian network and generalized raking for population synthesis in Greater Jakarta. *Regional Studies, Regional Science*, *6*(1), 623–636. https://doi.org/10.1080/21681376.2019.1687011

[15] Jiang, N., Crooks, A. T., Kavak, H., Burger, A., & Kennedy, W. G. (2022). A method to create a synthetic population with social networks for geographically-explicit agent-based models. *Computational Urban Science*, *2*(1), 7. https://doi.org/10.1007/s43762-022-00034-1

[16] Joubert, J. W., & De Waal, A. (2020). Activity-based travel demand generation using Bayesian networks. *Transportation Research Part C: Emerging Technologies*, *120*, 102804. https://doi.org/10.1016/j.trc.2020.102804

[17] Kim, J., Bae, Y. K., & Chung, J.-H. (2020). Modeling Social Distance and Activity-Travel Decision Similarity to Identify Influential Agents in Social Networks and Geographic Space and Its Application to Travel Mode Choice Analysis. *Transportation Research Record: Journal of the*

Transportation Research Board, *2674*(6), 466–479. https://doi.org/10.1177/0361198120919412

[18] Kim, J., Rasouli, S., & Timmermans, H. (2014). Expanding scope of hybrid choice models allowing for mixture of social influences and latent attitudes: Application to intended purchase of electric cars. *Transportation Research Part A: Policy and Practice*, *69*, 71–85. https://doi.org/10.1016/j.tra.2014.08.016

[19] Kim, J., Rasouli, S., & Timmermans, H. J. P. (2017). Investigating heterogeneity in social influence by social distance in car-sharing decisions under uncertainty: A regret-minimizing hybrid choice model framework based on sequential stated adaptation experiments. *Transportation Research Part C: Emerging Technologies*, *85*, 47–63. https://doi.org/10.1016/j.trc.2017.09.001

[20] Kim, J., Rasouli, S., & Timmermans, H. J. P. (2018). Social networks, social influence and activity-travel behaviour: A review of models and empirical evidence. *Transport Reviews*, *38*(4), 499–523. https://doi.org/10.1080/01441647.2017.1351500

[21] Koelle, D., Pfautz, J., Farry, M., Cox, Z., Catto, G., & Campolongo, J. (n.d.). *Applications of Bayesian Belief Networks in Social Network Analysis*.

[22] Li, S., & Landis, J. D. (n.d.). *Exploring Racial and Poverty-based Disparities in Transport Equity Among Large U.S. Metropolitan Areas*.

[23] Lin, T., & Wang, D. (2014). Social networks and joint/solo activity–travel behavior. *Transportation Research Part A: Policy and Practice*, *68*, 18–31. https://doi.org/10.1016/j.tra.2014.04.011

[24] Maness, M. (2020). Choice modeling perspectives on the use of interpersonal social networks and social interactions in activity and travel behavior. In *Mapping the Travel Behavior Genome* (pp. 399–411). Elsevier. https://doi.org/10.1016/B978-0-12-817340-4.00020-6

[25] Moore, J., Carrasco, J.-A., & Tudela, A. (2013). Exploring the links between personal networks, time use, and the spatial distribution of social contacts. *Transportation*, *40*(4), 773–788. https://doi.org/10.1007/s11116-013-9467-4

[26] Narayanan, A., & Shmatikov, V. (2009). De-anonymizing Social Networks. *2009 30th IEEE Symposium on Security and Privacy*, 173–187. https://doi.org/10.1109/SP.2009.22

[27] Pan, X., Rasouli, S., & Timmermans, H. (2019). Modeling social influence using sequential stated adaptation experiments: A study of city trip itinerary choice. *Transportation Research Part A: Policy and Practice*, *130*, 652–672. https://doi.org/10.1016/j.tra.2019.10.007

[28] Rasouli, S., & Timmermans, H. (2013). Incorporating Mechanisms of Social Adoption in Design and Analysis of Stated-Choice Experiments: Illustration and Application to Choice of Electric Cars. *Transportation Research Record: Journal of the Transportation Research Board*, *2344*(1), 10–19. https://doi.org/10.3141/2344-02

[29] Rasouli, S., & Timmermans, H. (2016). Influence of Social Networks on Latent Choice of Electric Cars: A Mixed Logit Specification Using Experimental Design Data. *Networks and Spatial Economics*, *16*(1), 99–130. https://doi.org/10.1007/s11067-013-9194-6

[30] Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (*p**) models for social networks. *Social Networks*, *29*(2), 173–191. https://doi.org/10.1016/j.socnet.2006.08.002

[31] Ryley, T. J., & Zanni, A. M. (2013). An examination of the relationship between social interactions and travel uncertainty. *Journal of Transport Geography*, *31*, 249–257. https://doi.org/10.1016/j.jtrangeo.2013.05.014

[32] Sallard, A., & Balać, M. (2023). Travel demand generation using Bayesian Networks: An application to Switzerland. *Procedia Computer Science*, *220*, 267–274. https://doi.org/10.1016/j.procs.2023.03.035

[33] Sharma, A., Gani, A., Asirvatham, D., Ariyaluran Habeeb, R. A., Hamzah, M., & Asli, M. (2021). Travel Behavior Modeling: Taxonomy, Challenges, and Opportunities. *International Journal of Advanced Computer Science and Applications*, *12*, 2021. https://doi.org/10.14569/IJACSA.2021.0120590

[34] Snijders, T. A. B., Van De Bunt, G. G., & Steglich, C. E. G. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks*, *32*(1), 44–60. https://doi.org/10.1016/j.socnet.2009.02.004

[35] Sun, L., & Erath, A. (2015). A Bayesian network approach for population synthesis. *Transportation Research Part C: Emerging Technologies*, *61*, 49–62. https://doi.org/10.1016/j.trc.2015.10.010

[36] Van Den Berg, P., Arentze, T. A., & Timmermans, H. J. P. (2009). Size and Composition of Ego-Centered Social Networks and Their Effect on Geographic Distance and Contact Frequency. *Transportation Research Record: Journal of the Transportation Research Board*, *2135*(1), 1–9. https://doi.org/10.3141/2135-01

[37] Van Den Berg, P., Arentze, T., & Timmermans, H. (2012a). A Multilevel Path Analysis of Contact Frequency Between Social Network Members. *Journal of Geographical Systems*, *14*(2), 125–141. https://doi.org/10.1007/s10109-010-0138-0

[38] Van Den Berg, P., Arentze, T., & Timmermans, H. (2012b). Involvement in clubs or voluntary associations, social networks and activity generation: A path analysis. *Transportation*, *39*(4), 843–856. https://doi.org/10.1007/s11116-012-9403-z

[39] Wang, R., Wang, Y., & Li, N. (2024). Revealing disaster dynamics and disparities in urban facility accessibility using an improved utilization-based metric. *Cities*, *144*, 104636. https://doi.org/10.1016/j.cities.2023.104636

[40] Yang, Y., Rasouli, S., & Liao, F. (2023). Effects of life events and attitudes on vehicle transactions: A dynamic Bayesian network approach. *Transportation Research Part C: Emerging Technologies*, *147*, 103988. https://doi.org/10.1016/j.trc.2022.103988

[41] Young, J., Graham, P., & Penny, R. (n.d.). Using Bayesian Networks to Create Synthetic Data. *Journal of Official Statistics*.

[42] Zhang, D., Cao, J., Feygin, S., Tang, D., Shen, Z.-J., & Pozdnoukhov, A. (2019). Connected population synthesis for transportation simulation. *Transportation Research Part C: Emerging Technologies*, *103*, 1–16. https://doi.org/10.1016/j.trc.2018.12.014