This dissertation examines the effects of online social recommendation systems (RS) on consumer preference similarity under different social network structural properties. There is a debate on recommendation system’s effects on consumer preference similarity. One view suggests that RS heterogenizes consumer preferences (i.e., making consumers less similar); while another view proposes an opposite effect (i.e., RS makes consumers to be more similar). This study first resolves the debate by revealing RS’s effects depends on: (1) the level of the analysis; and (2) the type of recommendation used.

Secondly, based on a large archival data set from an online music recommendation provider, we examine the effects of social recommendation systems (SRS) on consumer preference diversity and preference similarity at three different levels: individual level, ego-network level, and cluster level. Our findings show that the homogenizing and heterogenizing effects of a SRS depend on consumer’s social network structural properties. At the individual level, SRS tends to diversify consumer preferences as consumer’s centrality increases. At the ego-network level, a central consumer in a network (e.g., a consumer with many connections) is
more likely to have preferences that are different to her connections’. Moreover, SRS’s effect on preference similarity is non-linear: it becomes weaker as a consumer has more friends. At the cluster level, SRS homogenizes consumer preferences if consumers in the cluster connect to each other densely. This dissertation provides important insights to both academics and practitioners.

INDEX WORDS: social recommendation system; consumer preferences; social network structure; social influence; online shopping
YOUR FRIENDS LIKE IT. DO YOU? EXAMINING SOCIAL RECOMMENDATION
SYSTEM’S EFFECTS ON CONSUMER PREFERENCE SIMILARITY

by

SIYUAN LI

B.B.A., Hong Kong University of Science and Technology, P.R. China, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2013
YOUR FRIENDS LIKE IT. DO YOU? EXAMINING SOCIAL RECOMMENDATION SYSTEM’S EFFECTS ON CONSUMER PREFERENCE SIMILARITY

by

SIYUAN LI

Major Professor: Elena Karahanna
Committee: Richard T. Watson
Dale L. Goodhue
Dawn T. Robinson

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2013
DEDICATION

To my parents, Hao Li and Hui Hou,

for their love, concern, support, and strength.

And to my wife, Yingri Yu, for her patience

and understanding over the years.
ACKNOWLEDGEMENTS

Foremost, I would like to gratefully and sincerely thank my dissertation chair, Prof. Elena Karahanna, for her guidance, support, patience, and understanding during my graduate study and research at the University of Georgia. From the very beginning of my doctoral studies, she continuously stimulated me and provided me opportunity to explore. During the dissertation-writing process, her mentorship was paramount. Every meeting with her was fruitful and I always felt enlightened. Her scientific attitude and perpetual enthusiasm for research will motivate me over the whole course of my career.

I would also like to express my grateful attitude to my other dissertation committee members, Prof. Rick Watson, Prof. Dale Goodhue, and Prof. Dawn Robinson. In particular, I thank Rick for his insightful comments and guidance at different stages of my research. His creative thinking always inspired me. I am very grateful to Dale for his suggestions and advice on my research. He sets an excellent example of researcher with his rigor and passion for research. I thank Dawn for her helpful advice concerning social theories and methodologies in my research.

My thanks go to many other faculty members and students at the University of Georgia who supported and encouraged me in the last five years. I would like to especially thank Maric Boudreau, Hugh Watson, and Amrit Tiwana for their kind advice, support, and guidance in both my research and teaching at UGA.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGEMENTS</th>
<th>LIST OF TABLES</th>
<th>LIST OF FIGURES</th>
<th>CHAPTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>vi</td>
<td>viii</td>
<td>ix</td>
<td></td>
</tr>
</tbody>
</table>

CHAPTER

1 INTRODUCTION

1.1 Research Phenomenon .......................................................... 1
1.2 Research Gap ................................................................. 2
1.3 Research Objectives and Questions ........................................ 4
1.4 Structure of the Dissertation .................................................. 6

2 LITERATURE REVIEW AND THEORETICAL BACKGROUND ...................... 7

2.1 Types of Recommendation Systems ......................................... 8
2.2 Recommendation System’s Effects on Consumer Preference Similarity .... 13
2.3 Theoretical Underpinnings of the Effects of Social Recommendation System .............................................................. 20

3 RESEARCH MODEL AND HYPOTHESES .......................................... 30

3.1 Definitions of Key Constructs and Research Model ...................... 30
3.2 Hypotheses ........................................................................... 33

4 METHODOLOGY ........................................................................ 48

4.1 Data Description ................................................................... 48
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Recommendation Systems’ Effects by Levels of Study</td>
<td>16</td>
</tr>
<tr>
<td>2.2</td>
<td>Social Influence Types, Processes, and Goals (Burnkrant and Cousineau 1975)</td>
<td>22</td>
</tr>
<tr>
<td>2.3</td>
<td>Structural Characteristics of Social Network</td>
<td>27</td>
</tr>
<tr>
<td>3.1</td>
<td>Definitions of Key Constructs</td>
<td>30</td>
</tr>
<tr>
<td>4.1</td>
<td>Summary of the Key Variables and Their Measures</td>
<td>56</td>
</tr>
<tr>
<td>5.1</td>
<td>Descriptive Statistics at the Individual Level</td>
<td>59</td>
</tr>
<tr>
<td>5.2</td>
<td>Descriptive Statistics at the Ego-network Level</td>
<td>60</td>
</tr>
<tr>
<td>5.3</td>
<td>Descriptive Statistics at the Cluster Level</td>
<td>60</td>
</tr>
<tr>
<td>5.4</td>
<td>Variable Correlations at the Individual Level</td>
<td>62</td>
</tr>
<tr>
<td>5.5</td>
<td>Variable Correlations at the Ego-network Level</td>
<td>62</td>
</tr>
<tr>
<td>5.6</td>
<td>Variable Correlations at the Cluster Level</td>
<td>64</td>
</tr>
<tr>
<td>5.7</td>
<td>Model Results at the Individual Level</td>
<td>66</td>
</tr>
<tr>
<td>5.8</td>
<td>Model Results at the Ego-network Level</td>
<td>70</td>
</tr>
<tr>
<td>5.9</td>
<td>Model Results at the Cluster Level</td>
<td>74</td>
</tr>
<tr>
<td>5.10</td>
<td>Summary of Findings</td>
<td>77</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Example of Social Recommendation System</td>
<td>2</td>
</tr>
<tr>
<td>2.1</td>
<td>Content-based Recommendation System</td>
<td>9</td>
</tr>
<tr>
<td>2.2</td>
<td>Collaborative Recommendation System</td>
<td>11</td>
</tr>
<tr>
<td>2.3</td>
<td>Social Recommendation System</td>
<td>12</td>
</tr>
<tr>
<td>3.1</td>
<td>Research Model at Different Analysis Levels</td>
<td>32</td>
</tr>
<tr>
<td>3.2</td>
<td>Individual Level Research Model</td>
<td>34</td>
</tr>
<tr>
<td>3.3</td>
<td>Actor with high betweenness centrality</td>
<td>36</td>
</tr>
<tr>
<td>3.4</td>
<td>Actor with high closeness centrality</td>
<td>36</td>
</tr>
<tr>
<td>3.5</td>
<td>Ego-network Level Research Model</td>
<td>38</td>
</tr>
<tr>
<td>3.6</td>
<td>A High Density Ego-network</td>
<td>42</td>
</tr>
<tr>
<td>3.7</td>
<td>Three Clusters</td>
<td>43</td>
</tr>
<tr>
<td>3.8</td>
<td>Cluster Level Research Model</td>
<td>44</td>
</tr>
<tr>
<td>4.1</td>
<td>Screenshots of Last.fm Social recommendations</td>
<td>49</td>
</tr>
<tr>
<td>4.2</td>
<td>Percentage of Total Listen Counts by Rank</td>
<td>50</td>
</tr>
<tr>
<td>4.3</td>
<td>Lorenz Curve</td>
<td>53</td>
</tr>
<tr>
<td>4.4</td>
<td>Gini Coefficient (high vs. low) at the Ego-network Level</td>
<td>54</td>
</tr>
<tr>
<td>4.5</td>
<td>Gini Coefficient (high vs. low) at the Individual Level</td>
<td>55</td>
</tr>
<tr>
<td>4.6</td>
<td>Two Examples of Ego Networks</td>
<td>57</td>
</tr>
<tr>
<td>4.7</td>
<td>Two Examples of Clusters</td>
<td>57</td>
</tr>
</tbody>
</table>
Figure 5.1: Examples of Lorenz curve when degree centrality is low or high............................72

Figure 6.1: The Indirect Influence ..........................................................................................79
CHAPTER 1
INTRODUCTION

1.1 Research Phenomenon

In the online business to user (B2C) context, recommendation systems are applications that recommend products and services that are tailored to individual users based on their preferences and behaviors (Adomavicius and Tuzhilin 2005). Implementing recommendation systems is a way for e-vendors to enhance users’ online purchase experience and promote sales. Studies find that recommendation systems can significantly increase e-vendors’ revenue, customer satisfaction, and customer loyalty (AberdeenGroup 2011; Ansari and Mela 2003; Komiak and Benbasat 2006; Peppers and Rogers 1997). There are two widely adopted recommendation systems exist in online market nowadays: one is content-based and the other one is collaborative filtering (Adomavicius and Tuzhilin 2005). With the widespread diffusion of social network platforms (e.g., Facebook, Twitter, Last.FM, etc.), e-vendors are now leveraging social network information to provide personalized offerings through a new type of recommendation system: the social recommendation system. This provides a completely new method for understanding consumers (i.e., elicit consumers’ preferences) and promoting sales. Different from the two traditional types of recommendation systems\(^1\), social recommendation system generates recommendations based on items purchased or liked by focal consumer’s friends in the same social network. Amazon is among the first major e-vendors which implemented social recommendation systems. As shown in Figure 1.1, Amazon first identifies

---

\(^1\) Content-based approach recommends items similar to the ones a consumer purchased or liked (i.e., rated highly) in the past. Collaborative filtering recommends items purchased or liked by other people with similar preferences (e.g., people who purchased the same item).
the connections of the focal user based on information from Facebook (in the dashed box). Then it recommends items purchased by these connections to the focal user.

![Popular Among Your Friends on Facebook](image)

**Figure 1.1. Example of Social Recommendation System**

Though this approach is still at a nascent stage in the B2C field, it has already shown its ability to influence user’s preferences. A credit card company tripled its ad click rate from 0.9% to 2.7% by tailoring its offers based on connections’ responses. Yahoo! also observed similar patterns. It found that after someone clicks on an ad, if the same ad was offered to her connections, her connections were three to four times more likely to click on the ad (Baker 2009). Using social network information to personalize recommendations to consumers is considered as the next trend in the online market (Rogers and Sexton 2012).

1.2 Research Gaps

In recent years, however, there has been debate on negative effects of recommendation systems in general. A focus of this debate has been on whether these systems homogenize or heterogenize user preferences. One view is that personalized recommendations by an e-vendor
limit users’ ability to explore a variety of products and services. Users are constrained by an invisible filter that only passes contents that fit\textsuperscript{2} with users’ preferences (Pariser 2011). As a result, users’ preferences are more and more homogenized and narrowly focused because only similar products are recommended to them. Fleder et al. (2010) provide empirical evidence for this homogenizing effect. On the other hand, an alternative view is that recommendation systems, in fact, diversify users’ preferences. The argument supporting this view suggests that recommendations systems increase the exposure of niche products, i.e., of the unpopular products that may have been previously ignored by users. By introducing a variety of products (both popular and unpopular) to users, recommendation systems heterogenize users’ preferences (Brynjolfsson et al. 2011; Fleder and Hosanagar 2009). In this case, e-vendors may not be able to achieve economies of scale because users might become very dissimilar to each other and thus each product might only fit the preferences of a small group of users. Depending on whether recommendation systems have a homogenizing or heterogenizing effect, e-vendors need to use complementary methods to overcome negative consequences (i.e., highly homogenized or highly diversified users).

Prior empirical studies examining homogenizing or heterogenizing effects, based on content-based and collaborative filtering recommendation systems, yielded inconsistent findings. Further, there have been no studies (of which we are aware) on the effects of social recommendation systems on user preference similarity. In the next chapter, we first resolve the inconsistent findings on recommendation systems’ effects on user preference similarity by showing that recommendation systems have different effects on user preference depending on level of analysis (i.e., the individual user vs. the aggregate cluster of users) and type of

\textsuperscript{2} In addition, the fit is not matching personalized recommendations to a user’s real-time preferences but to what the recommendation system thinks the user likes.
recommendation system (content-based vs. collaborative filtering). We, then, build upon and extend this line of research to examine the effects of social recommendation systems on user preference similarity at three different levels of analysis: the individual level, the ego-network level\(^3\), and the cluster level. Furthermore, given that users have different positions in their social network, from a structural point of view, we examine how individual’s and cluster’s social network characteristics influence a social recommendation system’s effects on users’ preference similarity.

1.3 Research Objectives and Questions

As social recommendation system is built upon user’s social network, the recommendations generated by the system will vary depending on the user’s social network characteristics. Therefore, the objective of this dissertation is to examine the effects of social recommendation system on user preference diversity and similarity; and how the effects will change as users’ network structural characteristics (i.e., degree centrality, betweenness centrality, closeness centrality, and density) change. We examine the effects at three levels of analysis: (a) the individual level, where we focus on examining the effects of social recommendation systems on a user’s preference diversity; (b) the ego-network level, where we examine social recommendation system’s effects on the preference similarity between a focal user (the ego) and her connections; and (c) the cluster level, where we focus on examining the effects of social recommendation systems on user preference homogeneity within the cluster. Specifically, we focus on the following three research questions:

\(^3\) Ego-network is composed by a focal user (ego) and her direct connections (alters).
(a) At the individual level, how do a user’s network structural characteristics (i.e., centrality\(^4\)) influence a social recommendation system’s effect on user’s preference diversity?

(b) At the ego-network level, how do a user’s network structural characteristics (i.e., centrality and density) influence a social recommendation system’s effects on the user’s preference similarity to her connections?

(c) At the cluster level, how do a cluster’s structures (i.e., density and centrality) influence a social recommendation system’s effects on user preference similarity in the cluster?

We draw from structural view of social network (Burt 1992; Wellman 1997), social influence theories (Festinger 1954; Friedkin 1998, 2011), and human information processing theories (Cowan 1988; Payne 1982) to develop specific hypotheses of the effects and we present empirical evidence on whether social recommendation system homogenizes or diversifies user preferences.

The motivation of this paper is two-fold. First, academics can strengthen their understanding of the social recommendation system and its impact on user preference diversity and similarity. Since social recommendation system relies on information (i.e., social network information) that is completely different from traditional recommendation systems, findings on traditional recommendation systems may not be applicable to this new type of recommendation system. Indeed, our literature review has shown that the effects of recommendation systems on user preferences depend on the type of the recommendation system used. Identifying the impacts of social recommendation system then contributes to the theoretical debate of the homogenizing or diversifying effect of recommendation systems on user preferences. In addition, it identifies

\(^4\) Centrality reflects an ego’s network position. Prior literature has different measures to proxy this concept, including degree centrality, betweenness centrality, and closeness centrality. These (as well as the cluster characteristics) are defined later in the paper.
structural characteristics of the social networks that influence such effects (e.g., the diversifying effect in the ego-network is dependent on the user’s degree centrality).

Second, from a practitioner’s perspective, this study provides several marketing implications. Based on the findings of this study, e-vendors are able to tell whether implementing social recommendation systems aligns with their business goals (i.e., to have more homogenized user preferences or to create diversified user preferences). In the case of contradictory between social recommendation systems' effects and e-vendors’ business goals, e-vendors may want to increase the variety of contents exposed to users if social recommendation systems homogenize user preferences; or e-vendors may need to adjust their market campaigns to target on different preference segments if social recommendation systems diversify user preferences.

1.4 Structure of the Dissertation

This dissertation proceeds as follows. In chapter 2, we review prior studies on recommendation systems’ impacts on user preference similarity and take a closer look at the homogenizing/diversifying debate in the literature mentioned earlier. Then we introduce the theoretical foundations of social recommendation system’s impacts. In Chapter 3, we present the conceptual model and ensuing hypotheses. Chapter 4 presents the methodology section that explains how we are going to conduct the empirical analysis. Results of the empirical study are presented in Chapter 5. Chapter 6 then discusses the research findings, implications for research and practice, possible future directions, and then concludes the dissertation.
CHAPTER 2
LITERATURE REVIEW AND THEORETICAL BACKGROUND

Recommendation systems serve key roles in online market nowadays. E-vendors implement various types of recommendation systems to promote their sales and achieve their economic goals. In this chapter, we first review three different types of recommendation systems: content-based, collaborative filtering, and social recommendation system. Their unique characteristics and how they are implemented are discussed.

Though recommendation systems vary in types, the prior literature tends to treat them equally. This is the reason why there are inconsistent findings on recommendation systems’ effects, especially on consumer preference similarity. Therefore, we try to resolve the inconsistent findings by distinguishing the types of recommendation systems (content-based and collaborative filtering) and bringing in an often neglected factor in the prior literature – the levels of analysis. From the review, we can see that content-based recommendation systems and collaborative recommendation systems do have different effects on consumer preference similarity. Even for the same type of recommendation systems, their effects can be different at different analysis levels (e.g., for collaborative recommendation system, the effects at the individual level are opposite to the effects at the market level).

Different types of recommendations have distinct effects on consumer preferences. This is because they are based on different information sources. Social recommendation system, which is based on friends’ preferences, should have its unique effects on consumer preferences.
Our next step is to review how and why a social recommendation system influences consumer preferences. Three underpinning theories for social recommendation systems are then reviewed. They are social influence theory, structural view of social networks, and human cognitive capacity theories. These theories are the foundations for our hypothesis development in the next chapter.

2.1 Types of Recommendation Systems

To generate accurate recommendations, e-vendors must first get information from consumers about their preferences and behaviors. For this purpose, most e-vendors rely on information about the consumers past purchases, ratings, and browsing behaviors. Based on the source of this information, Adomavicius and Tuzhilin (2005) identified two types of recommendation systems that were used in online markets at that time – content-based recommendation systems and collaborative recommendation systems. Collaborative recommendation systems are also called collaborative filtering systems in some studies. In recent years, with the widespread diffusion of social network platforms (e.g., Facebook, Twitter, Last.FM, etc.), e-vendors are increasingly leveraging social network information to provide recommendations through a new type of recommendation system: the social recommendation system. In this section, we briefly introduce all three types of recommendation systems and focus on the unique features of the social recommendation system.

2.1.1. Content-based Recommendation Systems

*Content-based recommendation systems* recommend services or products similar to the ones the consumer preferred in the past (Adomavicius and Tuzhilin, 2005). The information sources can be historical ratings, purchases, webpage viewings, and other personal information
stored either in the e-vendor’s database or as browser cookies. Such information indicates the consumer’s preferences and therefore can be utilized as the foundation to generate personalized recommendations. By analyzing the similarities of items viewed, purchased, and rated, a content-based recommendation system will build up its knowledge about consumer preferences (Bakabanovic and Shoham, 1997). This type of recommendation system is the most widely adopted approach today. Amazon.com, Youtube.com, eBay.com, and most other major e-vendor sites already use content-based recommendation systems. Figure 2.1 shows an example from Amazon.com where it recommends books that are similar in topics with a consumer’s past purchases (i.e., a series of social network analysis books).

In most cases, content-based recommendation system can provide recommendations that fit with consumers’ preferences quite well. However, as pointed out by Bakabanovic and Shoham (1997), it has two major shortcomings: shallow analysis and over-specification. *Shallow analysis* means the analysis done by the recommendation technique is mostly based on limited characteristics of a product. The limited characteristics sometimes may not be the key features for consumers to make a purchase decision. For example, a consumer just purchased a laptop computer. Based on the basic content-based recommendation technique, the system recommends products with similar characteristics as the past purchase -- more laptop computers are then
recommended to the consumer. Obviously, the consumer will ignore these recommendations because she has no incentives at all at this point to purchase another computer. The problem is due to the fact that the content-base recommendation system does not consider the product’s lifetime when generating recommendations. The analysis could be too shallow to get to the consumer’s real preferences.

The second shortcoming of content-based recommendation system is over-specification. This means that the system recommends items that are similar to what a customer has already rated or viewed. The recommendations are limited to a specific range of items. In addition, recommendations will not change if consumers do not make more purchases, ratings, or viewings (Bakabanovic and Shoham 1997). Products that do not have similar characteristics to consumers’ past purchases are not likely to be recommended by the system. Thus, the system creates an invisible filter that only allows through contents that fit with consumers’ preferences (Pariser 2011).

2.1.2. Collaborative Recommendation Systems

**Collaborative, or peer-based, recommendation systems** are also known as collaborative filtering. This type of recommendation system recommends items to the consumer based on the preferences of other people who have similar tastes as the focal consumer (Adomavicius and Tuzhilin, 2005). The system supporting “consumer who bought this item also bought…” section on Amazon.com is a typical example of collaborative personalization. As shown in Figure 2.2, when a consumer browses an item on Amazon.com, the system will first identify the consumers that have purchased this item before (i.e., the “consumers who bought this item”). Then, the system recommends other items purchased by those consumers to the focal consumer.
From Figure 2.2, we can clearly see that not all recommendations are similar to the item that the consumer was browsing. This is the biggest difference between content-based recommendations and collaborative recommendations. It overcomes the over-specification limitation of content-based recommendation systems. A collaborative recommendation system is an excellent way to introduce new products that are outside a consumer’s current preference list (Balabanovic and Shoham 1997).

However, it has its own shortcomings – consumers with unique tastes do not receive high quality collaborative recommendations. This is because the system does not have a big enough pool to execute analysis and generate recommendations (i.e., it cannot find other consumers who bought the same item). Even if consumers with the same purchase can be found, the match on this one item could be largely by chance. Other products in the consumers’ preference baskets are still widely different. In this case, recommending products liked by those consumers can be ineffective.
2.1.3. Social Recommendation Systems

Since content-based and collaborative recommendation systems have each their own unique limitations, e-vendors seek to use complementary methods to overcome these shortcomings. In recent years, with the explosive growth in the use of social networking platforms, a new type of recommendation system, the social recommendation system, caught e-vendors’ attention. With the support of various social networking platforms (e.g., Facebook and Twitter), social recommendation systems are able to identify people in the same social network as the focal consumer (e.g., the focal consumer’s friends) and then recommends items purchased, used, or liked by these people to the focal consumer (Arazy et al. 2010). Amazon is among the first major e-vendors which implemented social recommendation systems. As shown in Figure 2.3, Amazon first identifies the friends of the focal consumer based on information from Facebook (in the dashed box). Then it recommends items purchased by these friends to the focal consumer.

![Social Recommendation System](image)

Figure 2.3. Social Recommendation System

Compared to content-based recommendation systems, social recommendation systems are not limited to a small range of products that are similar to focal consumers’ past purchases and thus do not suffer from the over-specification problem. Compared to collaborative
recommendation systems, social recommendation systems generate recommendations based on people within the same social network as the focal consumer. Therefore, the likelihood of having some common preferences among these people is higher than the likelihood of finding common preferences among the strangers who are the information source for collaborative recommendation systems – a limitation of collaborative recommendation systems. In fact, research has found that social recommendation systems does have a higher or at least equivalent accuracy in providing recommendations that match the consumer’s preferences than collaborative recommendation system (Li and Karahanna 2012).

Though this approach is still at a nascent stage in the B2C field, it has already shown its effects on consumers and product sales. A credit card company tripled its advertisement click rate from 0.9% to 2.7% by tailoring its offers based on friends’ responses. Yahoo! also observed similar patterns. It found that after someone clicks on an online advertisement, if the same advertisement is offered to her friends, her friends are three to four times more likely to click on the ad (Baker 2009).

While recommendation systems can have a variety of effects, this dissertation focuses on the effects of on consumer preference similarity. Therefore, in the following section, we review studies on the effects of collaborative and content-based recommendation systems on consumer preference similarity. There have been no studies thus far on the effects of social recommendation systems - which is the objective of the current study.

2.2. Recommendation System’s Effects on Consumer Preference Similarity

Prior studies on recommendation systems have examined their effects on consumer attitudes and perceptions toward the system (Greer and Murtaza 2003; Komiak and Benbasat
2006), consumer purchasing behavior (Chen and Hitt 2002; Zahay and Griffin 2004), and consumer’s decision accuracy (Tam and Ho 2006). However, one aspect that has received relatively little research attention is the effects on consumer preferences. In recent years, both policy makers and e-vendors started paying attention to such effects. From a policy perspective, policy makers are concerned about consumer well-being as it is possible that recommendation systems constrain consumers’ exposure to the variety of products online. If a recommendation system shows consumers only what they previously “liked”, the system can weaken the “social architecture” that offers “both shared experiences and unanticipated exposures” (Sunstein 2001). From the vendor perspective, e-vendors want to know the recommendation system’s effects on consumer preferences, including whether consumers’ preferences become more or less homogeneous. By knowing this, e-vendors can decide on business strategies (e.g., achieve economies of scale if consumer preferences become more homogeneous) or utilize complementary techniques to counteract any negative effects (e.g., if consumer preference heterogeneity is important to the e-vendor’s strategy and social recommendation systems increase sales but homogenize preferences, then another technique that diversifies preferences may need to be used).

There are two perspectives on whether recommendation systems homogenize or diversify consumer preferences. One view is that personalized recommendations by an e-vendor limit consumers’ ability to explore a variety of products and services. Consumers are constrained by an invisible filter that only passes contents that fit5 with consumers’ preferences (Pariser 2011). As a result, consumers’ preferences are more and more homogenized and narrowly focused

---

5 In addition, the fit is not matching personalized recommendations to a consumer’s real-time preferences but to what the recommendation system thinks the consumer likes.
because only similar products are recommended to them. Fleder et al. (2010) provide empirical evidence for this homogenizing effect.

On the other hand, an alternative view is that recommendation systems diversify consumers’ preferences. The argument supporting this view suggests that recommendations systems increase the exposure of niche products, i.e., of the unpopular products that may have been previously ignored by consumers. By introducing a variety of products (both popular and unpopular) to consumers, recommendation systems diversify consumers’ preferences (Brynjolfsson et al. 2011; Fleder and Hosanagar 2009). In this case, e-vendors may not be able to achieve economies of scale because consumers might become very dissimilar to each other and thus each product might only fit the preferences of a small group of consumers.

Both arguments have merit and, in fact, empirical evidence on the effects of recommendation systems on consumer’s preference is inconsistent. While some studies have found support for homogenizing of consumer preferences (Fleder et al. 2010), others found support for a diversifying effect (Brynjolfsson et al. 2011). However, a close review of the extant empirical studies shows distinct differences that may account for the inconsistencies. Specifically, we have identified two differences that, if taken into account when examining recommendation systems’ effects on consumer preference similarity, reveal clear patterns and resolve inconsistent and conflicting findings. These two factors are the level of analysis (cluster or individual) and the type of recommendation system examined (content-based or collaborative filtering). Table 2.1 presents the 2x2 categorization and summarizes prior findings. Using Table 2.1 as the organizing framework, we next review this literature and examine the empirical evidence and supporting theoretical arguments by level of analysis.
2.2.1. Recommendation Systems Effects at the Market Level

Brynjolfsson et al. (2011; 2006) argue that recommendation systems, together with the search engines on the Internet, reduce consumer’s search costs and thus significantly increase the sales of niche products on the market, which is called the “long tail” phenomenon. These studies finding diversifying effects at the market level all focused on content-based recommendation systems, which provide recommendations on products and services that are similar to a consumer’s past purchases (Adomavicius and Tuzhilin 2005). For example, one study states that “the website always recommends five other products...these five products are picked by the company’s experts based on their similarity and relevance to the focal product” (Brynjolfsson et al. 2011); another study states: “recommend[ing] additional products and services which are similar to those purchased or considered in the past” (Van Alstyne and Brynjolfsson 2005). As the foundation for generating recommendations in content-based system, similar products as products previously purchased by the focal consumer are identified. The system only compares the characteristics of two products to calculate similarity. Whether or not the product has been

<table>
<thead>
<tr>
<th>Table 2.1. Recommendation Systems’ Effects by Levels of Study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Market Level</strong></td>
</tr>
<tr>
<td><strong>Content-based Recommendation System</strong></td>
</tr>
<tr>
<td>Recommendation systems <em>diversify</em> consumer preferences</td>
</tr>
<tr>
<td><strong>Collaborative Recommendation System</strong></td>
</tr>
<tr>
<td>Recommendation systems <em>homogenize</em> consumer preferences</td>
</tr>
<tr>
<td><strong>Empirical Support:</strong> Fleder and Hosanagar (2009).</td>
</tr>
</tbody>
</table>
previously purchased by others is not considered when calculating similarity. Therefore, it is possible for a niche product (i.e., a product that has only been purchased by a few people or not at all — in other words, a non-popular product) to be identified as a similar item and be recommended to the focal consumer. Once a consumer purchases a niche product, contributing to the long tail phenomenon, the likelihood of recommending another similar niche product will increase. Therefore, the long tail can lead consumers further down the long tail, “allowing them to cultivate deeper tastes for these niche products” (Brynjolfsson et al. 2006). Therefore, a niche product’s probability of being purchased is increased by content-based recommendation systems.

On the other hand, studies that find homogenizing effects all focus on collaborative recommendation systems, which recommend items liked by consumers who have similar preferences as the focal consumer (Adomavicius and Tuzhilin 2005). For example, the recommendation system used in one study uses information about “segment shares (market shares within a segment of similar users)” (Fleder and Hosanagar 2009). Recommendation systems concentrate sales on more popular product and thus homogenize consumer preferences. They find that common recommendation systems are biased toward popularity. More weight is given to popular products when generating recommendations for consumers (Fleder and Hosanagar 2009) since these are purchased by many more consumers and, therefore, have a higher chance of being recommended in collaborative recommendation systems that take into account other consumers’ preferences. Collaborative recommendation systems that compensate for a product’s popularity may increase its sales to a greater extent. This leads to a rich-get-richer effect, which concentrates the sales on products with higher popularity level. In other words, for a specific product, collaborative recommendation systems increase its probability of being liked by consumers if it is a popular or modest popular product; whereas the probability may decrease
if the product is a niche product. In their conclusions, Fleder and Hosanagar (2009) state that even though an individual consumer may “be aware of more items and purchase more unique items,” the aggregate preference diversity may still decrease as the collaborative recommendation systems “push consumers toward the same products.”

Similar homogenizing effects on consumer preferences are also found in the study by Fleder et al. (2010). Specifically, Fleder et al. (2010) compare consumer preference similarities to others before and after they adopt the recommendation system. Two groups were identified in this study – one group in which consumers were exposed to the recommendation system; while consumers in the other group (which served as the control group) received no recommendations. Results show that consumer preferences become more similar after they are exposed to collaborative filtering recommendations. In other words, consumer preferences are more concentrated after implementing a collaborative recommendation system. The study provides two reasons why consumer preferences become more similar after recommendations. One is that the collaborative recommendation system shifts consumer preferences toward similar products, which is called the “taste effect”; the other reason is that collaborative recommendation system increases the overall purchases in the cluster and thus increases the likelihood of similar purchases in the cluster, which is called the “volume effect”.

They do not find a “far ones become farther” effects in this study. That is, after implementing the system, the overall preferences of consumers in the treatment group are not more dissimilar to those of the consumers in the control group than they were prior to implementing the system. It is possible, however, that sub-clusters of consumers who are all exposed to the collaborative recommendation system have more homogenous preferences within their own sub-clusters while
they exhibit higher heterogeneity across these clusters (i.e., the far ones becoming farther). The study, however, did not examine effects at this more granular level.

2.2.2 Recommendation Systems’ Effects at the Individual Level

Collaborative recommendation system’s effect is reversed at the individual level as compared to that at the cluster/market level. At the individual level, collaborative recommendation systems diversify a consumer’s preferences over time. Since this type of recommendations is not limited to products that are highly similar to the consumer’s past preferences, they are designed to introduce products outside the focal consumer’s preference list (i.e., recommending other consumers’ preferred items) (Balabanovic and Shoham 1997). Fleder and Hosanagar (2009) find empirical evidence that collaborative recommendations increase consumers’ awareness of more items. Similar diversifying effects of collaborative recommendation are also found in data of over 250,000 books sold on Amazon.com (Oestreicher-Singer and Sundararajan 2012). However, even though individual consumer’s preferences become more diversified, studies find no support that preferences across the two consumers become less similar (i.e., no support for “far ones becoming farther”). In fact, the average preference similarity across consumers remained unchanged (Fleder et al. 2010).

On the other hand, because content-based recommendation systems only recommend products similar to what a consumer has purchased in the past, they homogenize an individual’s preferences over time and have the possibility to create tunnel view of products (Balabanovic and Shoham 1997). In fact, Mooney and Roy (2000) found that their content-based intelligent book recommender reinforced individual consumer’s preferences over time cultivating a deeper preference for a consumer’s existing preferences.
Given that social recommendation systems are an emerging type, little is known about their effects. Specifically, we are currently not aware of any studies that examine the effect of social recommendation systems on consumer preference similarity. Further, our understanding of the recommendation technique itself is limited. Therefore, in the next section we first review the information foundation of social recommendation systems that is used to produce recommendations – i.e., the social network and its characteristics. Then we review the underpinning theories that explain how and why social recommendation systems influence consumer preferences.

2.3. Theoretical Underpinnings of the Effects of Social Recommendation Systems

2.3.1. Social Influence Theory

Social influence is an important process in our study since we are investigating how an information system (i.e., the social recommendation system) facilitates the social influence process among people in a social network. More specifically, our study examines how the social network structure that underlies the social recommendation system, and via which social influence is exerted, influences consumer preference similarity. A fundamental theory for the above question is social influence theory (Festinger 1954; Friedkin 1998, 2011). This theory argues that connected members within a network may influence each other to produce similarity of beliefs and behavior (Friedkin and Johnsen 1990, 1999). The influence can take two forms, informational influence and normative influence. Informational influence occurs when a person conforms to an influence because she perceives the source of the influence is valid and real. Normative influence happens when a person conforms to the influence in order to meet the positive expectations of others (Bearden et al. 1986; Deutsch and Gerard 1955).
In the current context, social influence is exerted via the social recommendation system. When a consumer browses recommendations that are based on her friends’ preferences, she is influenced by her friends. Since recommendations are generated by the recommendation system and not sent directly by consumers, consumers who use social recommendation systems are all influence receivers. Kelman (1958) proposed three processes via which influence affects another person’s beliefs, attitudes, preferences, and behaviors: compliance, identification, and internalization.

*Compliance* happens when a person accepts influence (i.e., conforms to similar attitudes and/or behaviors exhibited by other people in the network) because she wants rewards or positive reactions from other people in the network. For example, in a music community, many members may purchase a newly released album by an artist favored by the community. Even though a member might not like the artist very much, she purchases the album because she does not want to be considered as an outlier in the community (i.e., avoiding punishment) or because she wants to be able to engage in the community’s discussions related to the new album (i.e., a form of rewarding). The key point for compliance is the reaction of others. Therefore, Kelman (1958) found that a person tends to comply only under conditions of surveillance and control by the influencing agent.

*Identification* occurs when a person espouses attitudes and behaviors because of their role in a relationship with another person or group. In this case, the individual conforms to attitudes and behaviors that are matched with the role. The exact content of the attitude or behavior could be irrelevant in the process. The critical part is the real action of conformity (Kelman 1958). A typical example of this type of conforming process is receiving influence from a leader (e.g., president) in a network. The “followers” believe that *having* (i.e., the action) the same attitude
and behavior as the leader is correct while the contents of the attitude or behavior are irrelevant in the influencing process.

*Internalization* places more focus on the exact content of the influence than the previous two types. This type of conformity happens when a person believes the content of the influence is true, correct, and valid and the influence fits with the person’s internal value system (Kelman 1958). Learning new things from teachers, in most cases, is a process of internalization because students think teachers are credible. Therefore, if the influence comes from expert sources, conformity tends to take the form of internalization (Kelman 1958).

As shown in Table 2.2, each of the three processes described by Kelman relates to one of the social influence types described by Friedkin and Johnsen (1990, 1999). The first two processes underlie normative influence while the latter process underlies informational influence. In addition, each type of influence process achieves a unique goal. These goals are the motivations for a person to accept an influence (Burnkrant and Cousineau 1975).

<table>
<thead>
<tr>
<th>Influence Type</th>
<th>Influence Process</th>
<th>Goal Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normative</td>
<td>Compliance</td>
<td>External Reward/ Avoid</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Punishment</td>
</tr>
<tr>
<td></td>
<td>Identification</td>
<td>Self Maintenance or</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Enrichment</td>
</tr>
<tr>
<td>Informational</td>
<td>Internalization</td>
<td>Knowledge</td>
</tr>
</tbody>
</table>

Normative social influence occurs through either the process of compliance or of identification. When a consumer wants to realize a reward to avoid a punishment from others in the same social network, the consumer is likely to conform to the influence. However, this compliance will only occur when the consumer believes her conformity is visible to others (Kelman 1958). In the Amazon’s example (Figure 2.3), the social recommendation system displays the number of friends that have purchased the item. When the focal consumer realizes
that most of her friends prefer the same item, she is more likely to accept the recommendation because she might not want to be an “outlier” among her friends (avoiding negative consequences). In this case, the conforming process takes the form of compliance because the content of the influence is not relevant to the process and the visibility of her behavior is a critical reason of accepting the influence (i.e., her friends can find out if she likes the item by browsing her preferences or through communications in other channels).

A consumer that is motivated to strengthen her identity is expected to accept influence from others by associating herself with positive referents or dissociating herself from negative referents (Burnkrant and Cousineau 1975). Therefore, a consumer adopts her friends’ preferred items because she perceives such behavior as a supporting or enhancing action to their friendship. In other words, she wants to associate herself with her friends (by having common preferences) in order to strengthen their friendship. Whether or not the recommended item really fits with her preferences or whether her adoption behavior is visible to others is not relevant in the conforming process. The primary goal in this case is to maintain and strengthen her role in the satisfying self-defining relationship (i.e., the friendship) (Kelman 1958).

According to Deutsch and Gerard (1955), an informational social influence is accepted when it is perceived as a solution of some problems or the information is come from a credible source. This is tightly related to the process of internalization in Kelman’s study (1958). Therefore, when social recommendations fit with a consumer’s preferences, she is likely to be influenced and accept the recommendations as a solution to her product selection. In addition, if the recommendations are based on the preferred products of a friend who knows a lot about the product (i.e., an expert), there is also a high possibility for the consumer to accept the recommendations and be influenced by the recommendations.
In sum, social recommendation systems can influence consumer preferences through various social influence processes and influence types. However, social influence occurs only when people in social networks are connected. Ties among people are considered as conduits along which information or influence flows (Borgatti and Foster 2003; Harrison and Carroll 2002). Therefore, how people connect to others (i.e., the structural characteristics of the network) affects the process of social influence (Katona et al. 2011; Krackhardt 1998). We will discuss this next.

2.3.2. Social Network Characteristics

To study social network effects, the first step is to understand the relationships among people and structures of the social network (Wasserman and Faust 1994). According to Wellman’s summary of structural analysis characteristics (1997), one most important feature of the network is that researchers who study social network structure assume “the patterned relationships among multiple alters jointly affect network members’ behavior” while they do not assume people “engage in only multiple duets with separate alters”. Instead of treating people as isolated individuals, structural analysis considers people as members of networks and is thus more powerful in explaining social outcomes (Wellman 1997). Based on the structural view of social networks, a set of network attributes (e.g., network density, centralities, and other network properties) have been proposed and shown to have effects on network members’ behavior. For example, network structure is found to have significant impact on the attitude formation process. How people are connected with others (i.e., their network structural characteristics) predicts attitude similarity in the network (Erickson 1988). For online social networks, structural effects

---

6 Alter is a terminology used in social network analysis meaning a person who is directly connected with the focal person.
could be more salient because people in online networks are widely connected to others, which may lead to more variance and unrevealed effects of the social network structure.

Centrality is the most fundamental and popular concept in describing social network structure. It reveals an actor’s prominence and visibility within a complete network (Knoke and Yang 2008). Centrality has been used to examine power (Burt 1982; Brass 1984), advantage in exchange and competition (Blau 1963), and adoption of innovation (Coleman et al. 1996). Freeman (1979) summarizes three general measures of centrality, which are degree, betweenness, and closeness. Each of the three centrality measure has its own conceptual foundation and its applicable contexts.

Degree centrality is an index of communication activity (Freeman 1979). It shows the number of alters (in our study, connections) an actor (in our study, the focal consumer) has. Degree centrality tells the visibility of an actor in the network, which is similar to “prominence” in Knoke and Burt’s study (1983). In traditional offline social network contexts, prior studies find that degree centrality is closely related to cohesion and influence (e.g., Borgatti 2005; Kempe et al. 2003). An actor with a large degree centrality is adjacent to many other actors and is considered as a major channel of information (Wasserman and Faust 1994). Her likelihood of influencing others and being influenced by others is greater than an actor with small degree centrality. Degree centrality is more appropriate in measuring the immediate effects, especially when the content flowing in the network can be duplicated\(^7\) (Borgatti 2005).

“Actor in the middle” is often used to describe an actor with high betweenness centrality. Betweenness centrality refers to the extent to which other actors lie on the geodesic path (shortest distance) between pairs of actors in the network (Knoke and Yang 2008). It is an

---

\(^7\) Content duplication means after sending the piece of information to an alter, ego will still keep a copy of it, e.g., email broadcast and attitude influencing. It is different from content transfer, such as package delivery or money exchange, the content will be removed from the ego after it is transferred to an alter.
important indicator of control over resource or information flow. This is because an actor with high betweenness serves as a “gatekeeper” in the traditional social network. She can choose which information or resources can flow through her position. Therefore, “actor in the middle” has more interpersonal influence on the others (Freeman 1979; Friedkin 1991).

_Closeness centrality_ measures how close an actor is to all the other actors in the social network. Actors with high closeness centrality have very short communication paths to others. Therefore, they can quickly interact with all others and communicate with others more frequently (Beauchamp 1965; Wasserman and Faust 1994). For two persons that are far away in the social network to communicate with each other, the most efficient way is to go through people with high closeness centrality because they own the shortest communication paths to all people in the network. Therefore, it is an index of efficiency in the network (Freeman 1979). However, for closeness centrality to be considered as a reflection of efficiency relies on one implicit assumption – the contents flowing in the network will always take the shortest path between two connected nodes (such as package delivery). When this assumption is true, closeness centrality is a good measure of efficiency or the time-till-arrival (Borgatti 2005; Friedkin 1991); on the other hand, if this assumption is not met in some contexts, closeness centrality is more appropriate to be interpreted as a measure of indirect effects and a node’s reachability (Valente and Foreman 1998).

Other than actor level attributes, there is one critical network-level attribute: density. Density is the extent to which actors of a network are connected among themselves (Knoke and Yang 2008). It is a common measure of network cohesion (Blau 1977) and “knittedness” (Barnes 1969; Bott 1957). Density is also very important for block models in social network (Wasserman and Faust 1994). Higher density means actors in a social network have more ties with each other.
Thus, together with social influence theory, density can predict firm performance, consumer preference similarity, and other social network effects (e.g., (Athanassiou and Nigh 1999; Borgatti 2005; Fleder et al. 2010).

According to prior literature (Burt 1982, 1992; Granovetter 1983; Putnam 2000; Woolock 2001), the above social network structural characteristics can be broadly categorized into two types: bonds and bridges. Bonds refer to social ties among actors within tightly connected networks and emphasize homogenizing effects while bridges refer to connections that link one network to another network and emphasize heterogenizing effects (Burt 1992; Granovetter 1973). Individual level degree centrality (i.e., an ego’s direct connections to other actors in the network), closeness centrality, and cluster density are three characteristics that describe bonding connections among users in a network. Therefore, they represent the bonding status in the network/cluster. Betweenness centrality, on the other hand, describes the bridging status of an actor. This is because betweenness represents the extent to which a focal actor lies between two clusters/networks. Table 2.2 summarizes the concepts in the structural view of social network.

<table>
<thead>
<tr>
<th>Type</th>
<th>Network Characteristic</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonding</td>
<td>Degree Centrality</td>
<td>Degree centrality refers to the extent to which an actor connects to all other actors in a social network (Freeman 1979).</td>
</tr>
<tr>
<td></td>
<td>Density</td>
<td>Density refers to the proportion of possible connections that are actually present in the cluster (Wasserman and Faust 1994)</td>
</tr>
<tr>
<td></td>
<td>Closeness Centrality</td>
<td>Closeness centrality refers to the extent to which an actor is closely connected to all the other actors in the social network (Freeman 1979)</td>
</tr>
<tr>
<td>Bridging</td>
<td>Betweenness Centrality</td>
<td>Betweenness centrality refers to the extent to which other actors lie on the geodesic path between pairs of actors in the network (Freeman 1979; Knoke and Yang 2008)</td>
</tr>
</tbody>
</table>
2.3.3. Human Cognitive Constraint

As the number of friends increases in online social networks, Hill and Dunbar (2003) and Mok et al. (2007) find that the level of intimacy and level of interaction among people in the network decreases. This finding is consistent with the fact that online social networks now have many more weak ties than strong ties (Gross and Acquisti 2005). When the number of connections increases, the average strength (e.g., the amount of information transactions) of these connections is likely to be weaker. Katona et al. (2011) empirically demonstrated that people with many friends have a lower average influence than those with fewer friends. Since both the influence and tie strength are closely related to the communication and information exchange over the connection, all the above findings reveal one fact: that as the number of friends increases, the average amount of communication and information exchange occurring through each friendship tie is reduced. This finding is exactly what human cognitive constraint theories imply (Cowan 1988; Payne 1982).

Human cognitive constraint theories argue that each person has her own maximum level of cognitive capacity. One can spend only a limited amount of cognitive effort on performing one or several tasks (Payne 1982). In the context of social network, actors have limited cognitive capability and time to interact with others. In order to keep a minimum level of tie strength (to distinguish from unrelated couples), there appears to be an upper bound to the absolute number of friends an actor can have and interact with (Dunbar 2008; Milardo et al. 1983; Roberts et al. 2009). After this upper bound is reached, an actor is not able to process more friendship connections. She will either establish new connections by dropping older ones or establish new connections but the average tie strength of connections will decrease. This is because, given
limited resources of time and cognitive effort, the average effort the actor can devote to each connection will decrease.

These three theoretical perspectives set the foundation for our hypotheses. Our overarching thesis is that consumer preferences are influenced by social recommendation systems through a social influence process between the consumer and her friends. Our theorizing suggests that social recommendation systems will largely diversify consumer preferences. The amount of influence, and the extent to which this influence will diversify preferences, will depend on the consumer’s position in the social network. Given limited cognitive resources, we propose that this effect will exhibit a non-linear relationship to the number of friends a consumer has in her social network. We present our hypothesis development next.
CHAPTER 3

RESEARCH MODEL AND HYPOTHESES

Based on the theories and prior research discussed in the previous chapter, this chapter presents the research model and hypotheses that will be empirically tested in the dissertation. Given that recommendation systems can have different effects on consumer preference homogeneity at different levels of analysis, we will investigate social recommendation systems’ effects at the individual level, the ego-network level, and the cluster level of analysis. Definitions of key constructs used in this dissertation are presented before discussing the hypotheses.

3.1 Definitions of Key Constructs and Research Model

Definitions of the key constructs used in the study are presented in Table 3.1 and the research model is presented in Figure 3.1.

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preference Similarity</td>
<td>Preference similarity at the ego-network level refers to the extent to which two actors (in this dissertation, consumers) in the social network have similar product preferences.</td>
</tr>
<tr>
<td></td>
<td>At the cluster level, preference similarity refers to the extent to which all actors in a specific cluster have similar product preferences.</td>
</tr>
<tr>
<td>Disparity in Preference Similarity</td>
<td>Disparity in preference similarity is a measure of variability in the levels of preference similarity between an actor and his or her connections in the social network. In other words, it captures the extent to which the level of preference similarity between an actor and his or her connections exhibits variability such that the actor has a higher level of similarity with some connections than with others. It is the variance of the above preference similarity measure.</td>
</tr>
</tbody>
</table>
At the individual level, preference diversity refers to the extent to which an actor has diverse preferences (i.e., shows preference for a wide (rather than narrow) set of products).

Centrality refers to an actor’s prominence and importance in a complete social network (Knoke and Yang 2008).

Density refers to the proportion of possible connections that are actually present in the cluster (Wasserman and Faust 1994).

The main purpose of this study is to examine how social recommendation systems influence consumer preference similarity at different analysis levels. The source of recommendations for a social recommendation system is a consumer’s connections in the social network. Thus, structural differences in a consumer’s position and connections in the social network can influence the nature, quality, and quantity of recommendations provided by the social recommendation system.

As discussed in Chapter 2, centrality construct has three different measures (i.e., degree centrality, betweenness centrality, and closeness centrality). Each of the measure has its unique theoretical foundation and implications. Given that the structural characteristics of each consumer’s social network are likely different (e.g., different centralities, density, etc) we hypothesize that the effect of a social recommendation system on consumer preference similarity is a function of the social network’s structural characteristics. Further, since recommendation systems can have different effects at different levels of analysis (Fleder and Hosanagar 2009), we posit separate hypotheses for effects at three different analysis levels: (a) the individual level; (b) the ego-network level; and (c) the cluster level.

At the individual level, we examine how a consumer’s social network characteristics influence her preference diversity (i.e., the extent to which a consumer has a wide set of preferences rather than a narrow set of preferences). At the ego-network level, we examine how the focal consumer’s social network structure influences the preference similarity and the
disparity in preference similarity between the consumer and all her connected connections. (an ego-network is a composition of a focal consumer (called the ego) and all her connected connections (called the alters)). At the cluster level, we examine how social network characteristics of the cluster influence preference similarity within the cluster. A cluster is an aggregation of connected consumers. It is a term used to describe a block or clique. In this dissertation, it simply means a partition of a whole social network. Figure 3.1 shows the research models of this study.

---

* SRS stands for social recommendation system

---

8 For our hypotheses development and analysis, the term “social network characteristics” refers to the characteristics of the social network based on which social recommendations are generated.
### 3.2 Hypotheses

A social recommendation system presents products that are preferred by a focal consumer’s connections (i.e., connections). When browsing these recommendations (e.g., Amazon’s example in Chapter 2, Figure 2.3), the focal consumer is able to see how many of her connections “like” the same item and who they are. The focal consumer is able to browse all such recommendations. The consumer can also click on a friend’s name or profile picture to browse the full list of products preferred by this friend. In short, social recommendation systems in this study allow consumers to freely browse all preferred items of any and all of their connections. By revealing connections’ preferences in this manner, the social recommendation systems can influence consumer preferences and, thus, can potentially diversify the consumer’s preferences as well as influence the level of preference similarity\(^9\) between a consumer and her connections. With this in mind, we discuss our hypotheses next.

#### 3.2.1. Individual Level Hypotheses

At the individual level, we examine how different social network structural characteristics (i.e., degree centrality, betweenness centrality, and closeness centrality) influence the impact of a social recommendation system on an individual consumer’s preferences and specifically the extent to which a consumer’s preferences are wide and diverse versus narrowly focused on a small set of products. The research model at this level is presented in Figure 3.2.

---

\(^9\) Though tie strength (e.g., how close a focal consumer is to a friend) is very important in social networks, current social recommendation systems do not use this information in generating recommendations. Therefore, we do not take tie strength into account in this study.
3.2.1.1. Actor Degree Centrality

Degree centrality refers to the extent to which an actor connects to all other actors in a social network (Freeman 1979). In the current context, degree centrality means the number of connections a focal consumer has in the specific social network. If a consumer has high degree centrality, she is connected to many other consumers in the network and will receive influence from them via the social recommendation system. One assumption we make in this study is that no two consumers in the social network are likely to have an identical set of preferences (e.g., they differ in at least some of the products that they have previously purchased). Even though two consumers may have common preferences, they also have unique preferences. Therefore, when a consumer browses social recommendations based on her connections’ preferences, she will be exposed to a mix of overlapping and unique connections’ preferences. As such, she receives heterogeneous (diverse) influences from her connections. The more connections a consumer has, the more likely she is to be exposed to a larger set of unique recommendations. These heterogeneous influences will diversify a consumer’s preferences.

However, a consumer’s cognitive capability is limited (Payne 1982). There is a maximum amount of time, effort, and information processing ability a consumer can devote to browse social recommendations. As the number of connections increases, the average effort that a consumer can spend on browsing recommendations from each friend decreases (i.e., average effort on each friend= cognitive capability / number of connections). In other words, after a
certain point, adding more connections to the social network will have less of an impact on diversifying consumer preferences. As such, we expect a non-linear effect of degree centrality on consumer preference diversity such that the marginal effect of additional connections on preference diversity declines as degree centrality increases. Therefore, we have the following hypothesis.

**H1:** As degree centrality increases, a social recommendation system has a non-linear and decreasing diversifying effect on a consumer’s preferences.

### 3.2.1.2. Actor Betweenness Centrality

High betweenness centrality means an actor lies on the shortest distance between pairs of other consumers in the network (Knoke and Yang 2008). In traditional networks, an actor with high betweenness controls and facilitates the information or resource flows (Freeman 1979). However, online social recommendation systems remove the role of control from an actor with high betweenness. This is because the consumer’s connections can view all the products purchased or used by the focal consumer via the social recommendation system. The focal consumer does not control what products are recommended to her connections. However, even though the actor with high betweenness centrality does not control the information flow, she is still at a very important position in the social network. Actors with high betweenness centrality often represent structural holes in a social network (Burt 1992) or have weak ties that connect to other groups of consumers (Granovetter 1983). Both structural holes and weak ties indicate that the connected consumers are likely to have different preferences (e.g., in Figure 3.3, the group of consumers at the right side of A are likely to have different preferences as compared to the group of actors at the left side of A). Since the connected consumers have different preferences, the
influences a focal consumer with high betweenness centrality receives are more heterogeneous than someone with low betweenness centrality.

![Figure 3.3. Actor with high betweenness centrality](image)

As suggested in the discussion of degree centrality, a consumer’s cognitive capability is limited. As betweenness centrality increases, indicating that the focal consumer connects to more groups of consumers with different preferences, the average capability she can devote to processing influences from each of these groups decreases. Therefore, the marginal diversifying effect of these additional connections becomes weaker. Thus, as in the case of degree centrality, we expect that betweenness centrality, via the social recommendation system, has a non-linear effect on consumer’s preference diversity.

**H2: As betweenness centrality increases, a social recommendation system has a non-linear and decreasing diversifying effect on a consumer’s preferences.**

### 3.2.1.3. Actor Closeness Centrality

A consumer with high closeness centrality has shorter distances to all other consumers in the social network (Freeman 1979). In the context of social recommendation systems, a consumer with high closeness centrality will be influenced by others in a more direct (rather than indirect) way.

![Figure 3.4. Actor with high closeness centrality](image)
As in Figure 3.4, Consumer A has the highest closeness centrality while Consumer D has the lowest closeness centrality. For Consumer B to influence consumer A, it is very easy because she is a direct friend of A (with distance of one unit). Thus, B can influence A directly as A can browse B’s full list of preferences via the social recommendation system. However, it is much harder for B to influence D (with distance of three units) because social recommendation systems will only present B’s preferences to A. Only when A has accepted some recommendations (i.e., influenced by B), will the social recommendation system recommend them to C, and in the same manner, if C accepts the recommendations, to D. Whenever the influence is transmitted through a consumer (a node on the graph), it becomes weaker since the consumer is not likely to accept all recommendations by the system. Therefore, the longer the distance the influence travels, the weaker effect it will have. In other words, the consumer with the highest closeness centrality (i.e., the shortest distance to all others) will receive the strongest influence from all other consumers. Given that consumers’ preferences are not identical, this implies that this consumer will be more heterogeneously influenced. Thus, she will possess a more diverse set of preferences. Different than degree and betweenness centrality, here we do not expect a non-linear effect of closeness centrality on a consumer’s preference diversity. This is because closeness centrality depicts the ease with which one can be influenced by all others. That is, it captures the total distance one consumer has to all other consumers in the social network, and, thus how directly (vs. indirectly) she is influenced by others in the social network. This is not limited by the consumer’s cognitive capability. Therefore, we expect that closeness centrality, via social recommendation system, has a linear effect on a consumer’s preference diversity.

**H3:** As closeness centrality increases, a social recommendation system linearly increases a consumer’s preference diversity.
3.2.2. Ego-Network Level Hypotheses

An ego-network is composed by a focal consumer and all her direct connections (Freeman 1982). At this level, we examine how social network structural characteristics (i.e., degree centrality, betweenness centrality, closeness centrality, and density) influence the average preference similarity between the focal consumer and her direct connections in the ego network, as well as the variance of preference similarities, between a focal consumer and her direct connections in the ego network. The research model at this level is presented in Figure 3.5.

![Ego-network Level Research Model](image)

**Figure 3.5. Ego-network Level Research Model**

### 3.2.2.1. Actor Degree Centrality

Degree centrality at the ego-network level has the same meaning as it has at the individual level – the extent to which an actor (the ego) connects to all other actors in a social network (Freeman 1979). According to social influence theory (Friedkin 1995), when two actors are connected, they will influence each other. In the context of social recommendation systems, such influence is represented by a consumer adopting the recommendations of her connections – that is, by her preferences being influenced by her connections’ preferences. Since no two consumers in the ego-network have the exact same set of preferences, the focal consumer (the ego) will receive heterogeneous influences from all her connections. Because of the heterogeneous influences, the average level of similarity between the focal consumer and all her connections will decrease.
Take an example of a small ego-network with only three consumers (i.e., one focal consumer (A) with two connections (B and C)). Before A is connected to C, B is the only connection that A has. Due to social influence, A and B can have a high level of preference similarity. Now, Consumer C, whose preferences are not identical to B’s, becomes a friend of A. A receives influences from C and thus will like items that are not in B’s preferences. Therefore, the preference similarity between A and B decreases. Since A’s preferences still have items from B’s preferences, the preference similarity between A and C is not likely to be very high. As a result, the average preference similarity between A and her connections decreases as A adds more connections.

As we suggested at the individual level, a consumer has limited cognitive capability. She is not able to browse all her connections’ preferences when degree centrality is large (i.e., when she has a large number of connections). Therefore, we also expect the heterogeneous influences received by the focal consumer to become weaker as degree centrality increases. This means that as the focal consumer’s degree centrality increases, the average preference similarity between the focal consumer and her connections’ decreases at a decreasing rate.

$H4a$: As degree centrality increases, a social recommendation system has a negative non-linear effect on a consumer’s average level of preference similarity with her connections.

When a consumer has a large number of connections and she is not able to browse all recommendations based on all her connections’ preferences, she tends to exhibit information selectivity (Walsh 1988; Miller et al. 1956) and, thus, be selective in what she browses. Information selectivity can be displayed in two possible ways: (a) the consumer will selectively choose a few connections and only browse recommendations from them; and/or (b) she will still
browse recommendations from all connections but selectively browse only a few recommendations from each. In the first case, the consumer will be similar in preferences with only the few connections whose recommendations she browses; while in the latter case, the consumer receives equal but weak influences from all her connections and thus she will share a uniform (but lower) level of similarity with all her connections. In other words, in the latter case, the level of preference similarity one shares with her connections will likely be more uniform across all connections; while in the former case, the consumer will share high preference similarity with some connections and low with others.

We hypothesize that information selectivity is displayed by browsing the recommendations of a selected group of connections. According to information processing theories, when a consumer has information that exceeds her processing capability, she is likely to first screen the information based on a few key criteria (Bettman et al. 1998; Payne et al. 1988). In the context of social recommendations, two likely selection criteria are the strength of connection and the credibility of recommendations (Arazy et al. 2010; Komiak and Benbasat 2006). In either case, it is more likely that the consumer will tend to browse recommendations from only a few connections who are either her close connections or who have expertise in the specific products. Thus, the focal consumer will be more similar in preferences to these few connections whose recommendations she browses and less similar to the rest. In other words, when degree centrality is high, there is a higher disparity in the levels of preference similarity a consumer shares with her connections.

**H4b:** As the degree centrality increases, the focal consumer is likely to be influenced by only a few of her connections, which increases the disparity in preference similarity between the consumer and her connections in the ego-network.
3.2.2.2. Actor Betweenness Centrality

As mentioned at the individual level, high betweenness centrality also indicates structural holes of the ego-network. A consumer with high betweenness centrality has a bridging role – she connects different groups of consumers in the social network. Consumers in different groups tend to have different sets of preferences because they have no direct influence to each other. Therefore, the focal consumer, as a connector, receives high level of heterogeneous influences from different groups. The more heterogeneous influences a consumer receives, the less similar she will be to any other consumers in the ego-network. Thus, as betweenness centrality increases, we expect social recommendation systems to decrease the average preference similarity between the focal consumer and her connections. In addition, since a consumer has limited cognitive capability, she is not able to process all the heterogeneous influences she receives from different groups. Therefore, the decreasing effect on preference similarity becomes weaker as betweenness centrality increases.

H5: As betweenness centrality increases, a social recommendation system has a negative non-linear effect on a consumer’s average level of preference similarity with her connections.

3.2.2.3. Actor Closeness Centrality

Similar to closeness centrality at the individual level, closeness centrality at the ego-network level describes the ease with which a focal consumer can be influenced by all others in the social network. Given that consumers in a social network are unlikely to possess identical preferences, a consumer with high closeness centrality will then receive heterogeneous influences from others easier and faster (i.e., heterogeneous influences can reach the focal consumer by travelling shorter distances and thus with less information loss). Therefore, we
expect that focal consumers with higher closeness centralities will have preferences which are less similar to those of their connections.

*H6: As closeness centrality increases, a social recommendation system decreases a consumer’s average level of preference similarity with her connections.*

### 3.2.2.4. Ego-network Density

Density is an aggregate level indicator, which measures the extent to which members in a group are connected among themselves (Knoke and Yang 2008). Given a certain number of actors in an ego-network, the more the connections that exist among them, the higher the ego-network density. In the online social network, the maximum density implies that, in an ego-network, a focal consumer’s connections are also connected with each other. This will lead to high information redundancy in the group (Burt 1992). That is, an actor will receive the same information from multiple connections.

![Figure 3.6. A High Density Ego-network](image)

Figure 3.6 depicts an ego-network with high density (either A or B can be the ego, as all other consumers are the direct connections with him/her). Let’s assume it is A’s ego network. As shown in the figure, A is connected to C and D while B is also connected to C and D. Therefore,
A and B both receive recommendations based on C and D’s likings. In other words, the influences received by A and B are mostly redundant (at least 2/3 of the influences). As a result, A and B’s preferences are likely to be similar because they receive similar influences. The same logic applies to other consumers in the ego-network. Therefore, the overall consumer preference similarity in this ego-network is high. Thus, we expect that, when an ego-network’s density is high, as the influences transmitted in the ego-network would be highly redundant, the average preference similarity between the focal user and all her connections will be high.

H7: As ego-network density increases, a social recommendation system increases a consumer’s average level of preference similarity with her connections.

3.2.2. Cluster Level Hypotheses

Cluster is a group of consumers who are tightly connected with each other, for instance, the “Groups” in Facebook and “Circles” in Google+. Figure 3.7 shows an example of three connected clusters. Each node represents a consumer and the dashed circles represent the boundaries of clusters. The next set of hypotheses examines the effects of social recommendation systems on consumer preference heterogeneity within a cluster. The rationale
underlying the three centrality hypotheses is the following: (a) preferences within clusters are homogeneous while preferences across clusters are heterogeneous; (b) the higher the cluster’s centrality the more influences the cluster will receive from other clusters; (c) since influences from other clusters are heterogeneous (i.e., they each include (at least partly) a unique set of preferences), the more heterogeneous influences the cluster receives, the more diverse its preferences will become - i.e., high centrality decreases the preference similarity across all consumers within the cluster. We discuss the hypotheses at the cluster level next. Figure 3.8 presents the research model at this level.

![Cluster Level Research Model](image)

Figure 3.8. Cluster Level Research Model

3.2.2.1. Cluster Degree Centrality

Cluster degree centrality refers to the number of other clusters that are connected with the focal cluster. Since consumers within the same cluster are tightly connected, there is a large volume of information exchange and mutual influence among members of the same cluster. This leads to high similarity in attitudes and behaviors among consumers within the cluster (Burt 1992). Compared to consumers outside the cluster, consumers within the clusters then are more similar to each other – that is, there is more homogeneity of preferences within a cluster and more heterogeneity of preferences across clusters. Therefore, the more connections a cluster has to other clusters (i.e., higher degree centrality), the more heterogeneous influences it receives as
it is connected to a larger number of clusters each of which having (at least partly) its own unique set of preferences. Thus, as degree centrality of the cluster increases, social recommendation systems will heterogenize consumer preferences in the cluster.

Unlike our individual level hypothesis involving degree centrality, at the cluster level we posit a linear relationship between degree centrality and preference similarity. At the individual level, the effects of social recommendation systems are constrained by a consumer’s maximum cognitive capacity and the marginal effect of the social recommendation system on preference homogeneity decreases as a consumer’s number of connections increases. However, a cluster as a whole does not have limits in cognitive capacity. That is, there is no cognitive restriction on the number of influential connections a cluster can hold with other clusters. Therefore, we expect social recommendation systems to decrease the average preference similarity in a linear manner.

H8: As degree centrality of a cluster increases, a social recommendation system decreases the average preference similarity among all consumers in the cluster.

3.2.2.2. Cluster Betweenness Centrality

Similar to the definition of betweenness centrality at the individual level, a cluster with high betweenness centrality lies on the shortest paths between pairs of other clusters. We have already described that preferences tend to be homogeneous within a cluster but heterogeneous across clusters as each cluster has at least some unique preferences. A cluster with high betweenness centrality will then be heterogeneously influenced by many other clusters in the network. Therefore, we expect that social recommendation systems will decrease the average preference similarity within the cluster when the cluster’s betweenness centrality increases.
As betweenness centrality of a cluster increases, a social recommendation system decreases the average preference similarity among all consumers in the cluster.

3.2.2.3. Cluster Closeness Centrality

Cluster closeness centrality refers to the distance to all other clusters in the network. The shorter the distance, the higher is the cluster’s closeness centrality. Since the cluster with high closeness centrality is closer to other clusters, it will be influenced by other clusters in a more direct way. Similar to the rationale we provided at the individual level, we expect that a cluster with high closeness centrality will receive more heterogeneous influences from other clusters. As a result, consumer preferences within the cluster are more heterogeneous as compared to those of a cluster with low closeness centrality.

H10: As closeness centrality of a cluster increases, a social recommendation system decreases the average preference similarity among all consumers in the cluster.

3.2.2.4. Cluster Density

Density at the cluster level has similar effects as it does at the ego-network level. High density at the cluster also indicates a high level of redundant influences among the group of consumers. For a social recommendation system, this means a consumer in the cluster is exposed to the same, or a very similar, set of preferences from multiple connections. This strengthens the influential power of that set of products on the focal consumer, and thus increases the likelihood for the consumer to adopt a few or all products in that set. As a result, the overall preference similarity across consumers in the cluster is likely to be very high. Thus, we expect that
consumers in densely connected clusters will have more homogeneous preferences than consumers in sparsely connected clusters.

*H11: As the density of a cluster increases, a social recommendation system increases the average preference similarity among all consumers in the cluster.*
CHAPTER 4
METHODOLOGY

4.1 Data Description

Our empirical work is based on a music data set provided by Last.FM, which is an online music recommendation service (Last.FM 2011). We chose Last.FM as our analysis focus because it is a popular social recommendation provider and has a large user base. For each user, Last.FM shows songs and artists to whom the user’s “friends” listen. “Friends” are added to Last.FM through a “friend invitation” sent by the user. To be listed as a friend, the invitation has to be accepted by the other user. As shown in Figure 4.1(1), Last.FM displays all the user’s friends. By clicking on a friend, the user can see the songs her friend has been listening to. In addition, Last.FM users can mark a red heart on songs they like. Therefore, by browsing friends’ listen history, the focal user will know the songs that her friends like (as shown in Figure 4.1(2)).

\[10\] The word “friends” refers to a user’s direct connections in the Last.FM social recommendation network. Since this is the terminology used by Last.FM and by most other social recommendation networks, we will use the term “friends” in this and the next chapter (that discusses our results) to refer to connections in the social network.
The data set contains social networking and music artist listening information from a set of 1,892 users from the Last.fm online music system. In total, these users listened to 17,632 distinct artists and established 12,717 user-friend relations. The system records the artists to whom a user listens as well as the number of times listened. The top 50 artists [not songs] to

\[11\] Pictures, names, and other personal information have been removed from the screenshot to preserve privacy.
which each user listens most frequently are written into the data set, which is a reasonable set to show a user’s music preferences.

From a social networking perspective, the friendships in this data set are not directed such that if A is a friend of B, B is also a friend of A. This is because to establish a friendship connection B needs to “accept” A’s friendship request. On average, users in this social network tend to have a relatively small number of friends (86.67% of the users have fewer than 30 friends).

On average, each user has 6.72 friends and listens to 49.067 artists. However, the latter is an artifact of the data. The fact that each user listens to 49.067 artists, on average, is to be expected since data are only maintained on each user’s top 50 artists. Figure 4.2 shows the percentage of total listen counts for users’ top fifty artists. Since each user has his/her own top 50 listened artists, the percentage of total listen counts here are by rank but not specific artists. For example, the percentage for the Top 1 artist across all users is the average of the “listen counts of all users’ most frequently listened artist” divided by the “total listen counts” (i.e., listen counts of all 1,892 users). What the graph shows is that users have a favorite artist to whom they listen to, on average, almost three times as frequently as their second favorite artist. This roughly follows the 80-20 rule where a major portion of the listening counts (nearly 60% in this case) is concentrated in the top 20% of each user’s artists.

![Figure 4.2. Percentage of Total Listen Counts by Rank](image-url)
4.2 Construct Operationalization

4.2.1 Dependent Variables

4.2.1.1 Consumer Preference Similarity: Similarity on Weighted Listen Counts

Consumer preference similarity is one of the core dependent variables at both the ego-network and cluster levels. To measure the level of preference similarity between a user and a friend, we calculate the Euclidean distance between the user’s weighted listen counts and the friend’s weighted listen counts (we explain weighted listen counts in the next paragraph). Calculating Euclidean distance (D) is a commonly used method to measure similarity in prior literature (e.g., (O’Reilly et al. 1989; Wagner et al. 1984). It provides a score that describes the dissimilarity between an individual’s profile and a group (Jackson et al. 1991; Riordan and Shore 1997). In this study, since the similarity is between two users, we modify the formula to suit our operationalization purpose. Equation (1) shows the calculation of the Euclidean distance between two users. Since Euclidean distance is a measure of dissimilarity, it inversely measures the similarity on weighted listen counts of the same artists. \(X_{ik}\) and \(X_{jk}\) in the formula represent the \(i\)th user’s and the \(j\)th user’s weighted listen counts for the \(k\)th artist. The Euclidean distance is the square root of the sum of the differences between \(X_i\) and \(X_j\) on all artists (17,632 in total).

\[
D_{ij} = \sqrt{\sum_{k=1}^{n} (X_{ik} - X_{jk})^2}, \quad n = 17,632, \quad i, j \in (1,1,812)
\]

\(D_{ij}\): the Euclidean distance between the \(i\)th and \(j\)th consumers

\(X_k\): the weighted listen counts on the \(k\)th artist

Taking the Euclidean distance between the weighted listen counts (i.e., \(X_{ik}\) and \(X_{jk}\)) better reflects differences in music preferences than using the absolute listen count. For example, if

---

12 Ego network is composed with the focal consumer and all her directly connected friends in the social network (Wasserman and Faust 1994).

13 Euclidean distance is the most widely adopted and the easiest measure for similarity. Other similarity measures (e.g., cosine distance) have the same indications as the Euclidean distances.
both users listened to Artist A 10 times (absolute listen counts are the same for both -10 times), the Euclidean distance between the two users on Artist A equals to zero, if the distance is calculated based on absolute listen counts. This implies that the two users have the same preference level on Artist A which, may is likely not the case. Let’s consider the following scenario. If one user has 1,000 total listen counts (i.e., she also listens to a lot of other artists) while the other user has only 20 total listen counts, the 10 listen counts then reflect different preference levels on Artist A between the two users. This is because the former user spends only 1% of her time (10/1000) listening to Artist A while the other one uses 50% of her time (10/20) listening to Artist A. The 1% and 50% capture the weighted listen counts for the two users for Artist A. In this case, it is likely that the latter user prefers Artist A more than the former user does. Therefore, to reflect the preference similarity in a more accurate way, we take a user’s total number of listen counts into consideration (i.e., the weighted listen counts). Since Euclidean distance measures the dissimilarity between two consumers, a large number means low preference similarity between the two.

At the ego-network level, average preference similarity is the average value of Euclidean distances between the focal user and all her friends (that is, a pairwise Euclidean distance between a focal user and each of her direct connections is first calculated and then these are averaged across all of the focal user’s direct connections)\(^\text{14}\); while at the cluster level, average preference similarity in the cluster is the average value of Euclidean distances between all pairs of users in the same cluster.

\(^{14}\) Traditionally, ego is removed from the calculations when considering ego-network attributes. Given the unique dependent variable (the similarity between ego and alters) in this dissertation, however, we cannot remove ego from calculation. Otherwise, the dependent variable cannot be calculated.
4.2.1.2 Disparity in Preference Similarity and Preference Diversity: Gini coefficient

To show the diversity of preferences of a specific user at the individual level and the variance of preference similarities at the ego-network level, we use the Gini coefficient as a measure. The Gini coefficient is a widely used index to measure distribution equality (e.g., (Brynjolfsson et al. 2011; Oestreicher-Singer and Sundararajan 2012; Sen 1976). It is based on the Lorenz curve, which plots the cumulative share of Euclidean distance against the cumulative sample share.

![Figure 4.3. Lorenz Curve](image)

Figure 4.3 and Equation (2) show how the Gini coefficient is calculated. A is the area above the Lorenz curve while below the diagonal line and B is the area below the Lorenz curve.

$$Gini\ Coefficient = \frac{A}{A+B} \quad (2)$$

Taking the Gini coefficient (i.e., the disparity of preference similarity) at the ego-network level as an example, the X-axis in Figure 4.3 is the proportion of the sample with Euclidean distances less than or equal to a specific distance level (the distance levels are sorted in ascending order in advance). The Y-axis is the cumulative Euclidean distance towards the sum of all Euclidean distances between the focal consumer and her friends. The diagonal line in Figure 4.3 represents
the perfect equality where the Gini coefficient equals to 0 (A = 0, therefore A/(A+B) = 0). It means that the focal user has the same number of friends at every level of Euclidean distance. In other words, the set of pairwise preference dissimilarities (Euclidean distances) between the focal consumer and her friends is heterogeneously distributed (implying that the focal consumer is more similar to some of her friends and less similar to others). If the Gini coefficient changes as a function of the number of friends in one’s social network (i.e., there is a negative relationship between the Gini coefficient and number of friends), it would provide support for hypothesis H4b. That is, as the number of friends increases, because of limitations to human information processing (Cowan 1988; Payne 1982), the focal consumer cannot process all recommendations from all friends, but instead, she will selectively accept recommendations from a few friends. Figure 4.4 illustrate the extreme cases when Gini coefficient is high and low at the ego-network. As Figure 4.4. shows, a high Gini coefficient indicates that the user shares the same level of preference similarity with the majority of her friends (in this case, the user shares a level of similarity of 1 with four out of her seven friends).

![Figure 4.4. Gini Coefficient (high vs. low) at the Ego-network Level](image)
At the individual level, the Gini coefficient measures the preference diversity of an individual Last.fm user across artists. That is, the Gini coefficient is calculated from the distribution of the individual’s weighted listen counts across all listened artists. A high Gini coefficient indicates that the user’s preferences are very concentrated on a few artists (e.g., she has high weighted listen counts for only a few artists) while a low Gini coefficient indicates that the user’s preferences are diversified (i.e., she has similar weighted listen counts across artists). Figure 4.5 illustrates the extreme cases of when the Gini coefficient is high (90% of the weighted listen count is from one listening to one artist while 10% from all other artists) and low (all artists have approximately 30% of the weighted listen count) at the individual level.

4.2.2 Independent Variables

The independent variables of interest in this study are the network structural measures: degree centrality, betweenness centrality, closeness centrality, and density. Among the four variables, the three centrality measures are applicable to all three levels of analysis while the density measure is only applicable to the cluster level in this study. Table 4.1 provides the
detailed operationalization of these variables. Figure 4.6 shows two examples of ego networks with different levels of centrality measures.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Definition</th>
<th>Operationalization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preference Similarity</strong></td>
<td>Ego-network level: The extent to which a focal user’s preferences match with her friend’s preferences</td>
<td>The average Euclidean distance between the focal consumer and her friends (calculated as the average of all pairwise Euclidean distances between a consumer and her friends). This gives a measure of dissimilarity between a focal user and her friends.</td>
</tr>
<tr>
<td></td>
<td>Cluster level: The extent to which consumers’ preferences overlap with others’ preferences in a cluster.</td>
<td>The average Euclidean distance among all consumers in the cluster (calculated as the average of all possible pairwise Euclidean distances in the cluster). This gives a measure of dissimilarity in consumer preferences in the cluster.</td>
</tr>
<tr>
<td><strong>Disparity of Preference Similarity</strong></td>
<td>The extent to which a focal consumer is unequally influenced by her connections.</td>
<td>Gini coefficient, which measures the variance of preference similarity across a focal user and her connections in the ego-network.</td>
</tr>
<tr>
<td><strong>Preference Diversity</strong></td>
<td>The extent to which an actor has diverse preferences</td>
<td>Gini coefficient, a measure of preference diversity of an individual Last.fm user.</td>
</tr>
<tr>
<td><strong>Independent Variables</strong></td>
<td>Degree Centrality</td>
<td>The number of direct connections a focal user (or a cluster) has. In the case of users, this measures the number of friend connections the focal Last.FM user has on Last.FM. In the case of clusters, it measures the number of other clusters with which the focal cluster is directly connected.</td>
</tr>
<tr>
<td></td>
<td>Betweenness Centrality</td>
<td>The number of times a consumer (or a cluster) occurs on a geodesic path between two other consumers (or clusters).</td>
</tr>
<tr>
<td></td>
<td>Closeness Centrality</td>
<td>Closeness centrality refers to the extent to which an actor is closely connected to all the other actors in the social network (Freeman 1979). The total length of shortest paths to other users (or clusters).</td>
</tr>
<tr>
<td></td>
<td>Density</td>
<td>Density refers to the proportion of possible connections that are actually present in the cluster (Wasserman and Faust 1994). The number of existing connections (friendships) divided by the maximum possible connections (friendships) among all consumers in an ego-network (or a cluster).</td>
</tr>
</tbody>
</table>
4.3 Identifying Clusters

Given the connections among consumers in the social network, we are able to identify clusters that are composed of tightly connected consumers. A common foundation for most clustering algorithms is to maximize the possibility of “random walks”\(^\text{15}\) within the same cluster (Van Dongen 2000). This means that if information starts at an actor and randomly travels to another connected actor, it is more likely to stay within a cluster than travel to another one.

The clustering algorithm used in this study is the Markov Clustering algorithm. This is one of the most efficient and accurate methods to divide a large network into non-overlapping clusters (Enright et al. 2002). This is an iterative algorithm that maximizes the probabilities of random walks within clusters and minimizes those between. Eventually the whole social network will be decomposed to different clusters with minimum connections between them (Van Dongen 2000, 2008). Figure 4.6 presents two clusters with different cluster level characteristics (i.e., cluster centralities and density). After identifying different clusters, we treat each cluster as a single node in the network to calculate its structural characteristics (i.e., degree centrality, 

---

\(^\text{15}\) Random walks means that the information can travel from actors to actors with no constrains on their traveling paths.
betweenness centrality, closeness centrality, and density). The dependent variable at the cluster level is the average preference similarity among all users in a cluster.

Figure 4.7. Two Examples of Clusters
CHAPTER 5
DATA ANALYSIS AND RESULTS

This chapter presents the data analysis to assess the hypotheses developed in Chapter 3. Descriptive statistics and correlations between the variables are provided first, followed by tests of our hypotheses.

5.1. Descriptive Statistics

We used UCINET Version 6.361 (Borgatti et al. 2002) to calculate the social network structural characteristics – centrality measures for all three levels of analysis and density for the cluster level of analysis. Our dependent variables are based on the Euclidean distance and the Gini coefficient. Euclidean distance measures preference dissimilarity between a Last.FM user and her connections. It is calculated using Formula 4(1) in Chapter 4. The Gini coefficient is used to measure an individual’s preference diversity at the individual level and the disparity of preference similarity at the ego-network level, following the Formula 4(2) in Chapter 4. Descriptive statistics for the variables are presented in Table 5.1 (individual level), Table 5.2 (ego-network level), and Table 5.3 (cluster level).

<table>
<thead>
<tr>
<th>Table 5.1. Descriptive Statistics at the Individual Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>Preference Diversity$^{(a)}$</td>
</tr>
<tr>
<td>Degree Centrality</td>
</tr>
<tr>
<td>Betweenness Centrality$^{(b)}$</td>
</tr>
<tr>
<td>Closeness Centrality$^{(b)}$</td>
</tr>
<tr>
<td>N</td>
</tr>
</tbody>
</table>

Notes:
(a) Preference diversity is measured by the Gini coefficient.
(b) Betweenness and closeness centralities are normalized.
Table 5.2. Descriptive Statistics at the Ego-network Level

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Preference Similarity (a)</td>
<td>0.12</td>
<td>0.66</td>
<td>0.31</td>
<td>0.09</td>
</tr>
<tr>
<td>Preference Disparity (b)</td>
<td>-0.05</td>
<td>1.00</td>
<td>0.60</td>
<td>0.19</td>
</tr>
<tr>
<td>Degree Centrality</td>
<td>1.00</td>
<td>119.00</td>
<td>13.42</td>
<td>17.26</td>
</tr>
<tr>
<td>Betweenness Centrality (c)</td>
<td>0.63</td>
<td>19.86</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Closeness Centrality (c)</td>
<td>1.81</td>
<td>1.95</td>
<td>1.91</td>
<td>0.02</td>
</tr>
<tr>
<td>Ego-network Density</td>
<td>0.00</td>
<td>1.00</td>
<td>0.21</td>
<td>0.25</td>
</tr>
<tr>
<td>Individual Preference Diversity</td>
<td>1.10</td>
<td>14.51</td>
<td>2.54</td>
<td>1.12</td>
</tr>
<tr>
<td>N</td>
<td>1586</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: (a) Average preference similarity is the average of Euclidean distances
(b) Preference disparity is measured by the Gini coefficient
(c) Betweenness and closeness centrality are normalized
(d) The number of observations is different from the one for the individual level analysis due to missing values in preference disparity and ego-network density (i.e., users with only one friend have no variance in preference similarity).

Table 5.3. Descriptive Statistics at the Cluster Level

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Preference Similarity (a)</td>
<td>0.15</td>
<td>1.04</td>
<td>0.33</td>
<td>0.13</td>
</tr>
<tr>
<td>Degree Centrality</td>
<td>2.00</td>
<td>183.00</td>
<td>57.78</td>
<td>33.85</td>
</tr>
<tr>
<td>Betweenness Centrality</td>
<td>0.03</td>
<td>1027.76</td>
<td>85.67</td>
<td>140.24</td>
</tr>
<tr>
<td>Closeness Centrality</td>
<td>33.74</td>
<td>82.33</td>
<td>50.36</td>
<td>6.70</td>
</tr>
<tr>
<td>Density</td>
<td>0.00</td>
<td>1.00</td>
<td>0.63</td>
<td>0.32</td>
</tr>
<tr>
<td>Number of Users in Cluster</td>
<td>2.00</td>
<td>120.00</td>
<td>8.01</td>
<td>15.14</td>
</tr>
<tr>
<td>N</td>
<td>207</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: (a) Average preference similarity is the average of Euclidean distances in the cluster

Individual level social network characteristics, take into account the entire social network. As Table 5.1 shows, a Last.FM user is on average directly connected to 13.42 friends (i.e., degree centrality is 13.42). Due to the fact that the social network is quite large, the raw data for betweenness (i.e., the number of times a use lies between two other users in the social network) and closeness centralities (i.e., the reverse of the total distance to all other consumers in the social network) become very large and hard to interpret. Therefore, as recommended (Everett and Borgatti 2005; Freeman et al. 1991), we normalized both the betweenness centrality and closeness centrality in this study. The normalization process is done by UCINET by default (Borgatti et al. 2002). The dependent variable preference diversity, which is measured by the Gini coefficient, has an average value of 0.45. A Gini coefficient of 1 represents equal
distribution of listen counts across the artists the individual listens to (i.e., the highest level of preference diversity). A value of .45 indicates that the Last.FM users in our sample, on average, tend to have concentrated preferences on some artists.

The ego-network level only takes into account the individual (ego) and her direct connections (alters). Therefore, since degree centrality and Euclidean distance are calculated based on the focal individual and her direct connections, these variables are the same for the ego-network level of analysis (in Table 5.2) as for the individual level of analysis (in Table 5.1). Betweenness centrality is different at the ego-network level because the scope of calculating is the ego-network, and not the whole social network as at the individual level. Similar to the individual level betweenness centrality, the raw data for the betweenness centrality is very large and is hard to interpret. As is typical in such cases, we normalized this measure in our study. Closeness centrality is not applicable at the ego-level of analysis because the distance between the focal user (ego) and everybody else in the ego network (alters) is 1 and, therefore, there is no variance in this variable at the ego-level of analysis. Individual preference diversity is the dependent variable we calculated at the individual level. We included it in the ego-network level of analysis as a control variable because if a user’s preferences are diverse it is more likely that her preferences will be similar to other users’ preferences (i.e., lower Euclidean distance). For example, if a user has a diverse set of preferences (e.g., listens to 100 artists), the likelihood of matching other users’ preferences becomes higher as compared to a user who has a limited set of preferences (e.g., only listens to 5 artists).

Using the Markov Clustering algorithm, we divided the whole social network into a total of 207 clusters. These clusters have at least two users and a maximum of 120 users. The Euclidean distance at the cluster level measures the average distance among all users within a
cluster. As shown in Table 5.3, users in a cluster have an average Euclidean distance of 0.33. Each cluster is on average connected to 57.78 other clusters (degree centrality) and lies 85.67 times on the shortest paths between two other clusters (betweenness centrality). As before, UCINET normalized the distances between clusters before calculating the closeness centrality. The average density of a cluster is 0.63, which indicates that nearly two thirds of the possible connections among users in a cluster are connected, which means that Last.FM users are well connected among themselves in a cluster. We use the number of users in the cluster as the control variable at the cluster level of analysis. This is because clusters vary in the number of users they contain. A cluster with more users can potentially have a lower level of average preference similarity (assuming no users have identical set of preferences). Therefore, we need to control for the effect of cluster size on average preference similarity at the cluster level.

Tables 5.4, 5.5, and 5.6 show the variable correlations at each of the three levels of analysis. At the individual level, both degree centrality and closeness centrality (but not betweenness centrality) have significant negative correlations with the Gini coefficient (i.e., the reverse measure of preference diversity). This indicates that a user with many connections or with short distances to others tends to have high preference diversity (i.e., low Gini coefficient). Degree centrality and closeness centrality are highly correlated (0.82 at a significance level of 0.01), which indicates that a user with many connections also has short distances to all other users in the social network. Though betweenness centrality is highly correlated with degree and closeness centralities (.68 and .55 respectively), it does not significantly correlate with our dependent variable (the Gini coefficient).
### Table 5.4. Variable Correlations at the Individual Level

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Preference Diversity</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Degree Centrality</td>
<td>-0.21** (0.00)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Betweenness Centrality</td>
<td>-0.04 (0.09)</td>
<td>0.68** (0.00)</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>4. Closeness Centrality</td>
<td>-0.17** (0.00)</td>
<td>0.82** (0.00)</td>
<td>0.55** (0.00)</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

*Note:* (a) Numbers in parentheses are the p-values
(b) Degree centrality and betweenness centrality measures are log transformed.

### Table 5.5. Variable Correlations at the Ego-network Level

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Average Preference Similarity</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Preference Disparity</td>
<td>-0.25** (0.00)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Degree Centrality</td>
<td>0.30** (0.00)</td>
<td>-0.74** (0.00)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Betweenness Centrality</td>
<td>0.07** (0.01)</td>
<td>-0.45** (0.00)</td>
<td>0.64** (0.00)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Closeness Centrality</td>
<td>0.14** (0.00)</td>
<td>-0.27** (0.00)</td>
<td>0.15** (0.00)</td>
<td>0.08** (0.00)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Ego-network Density</td>
<td>-0.10** (0.00)</td>
<td>0.17** (0.00)</td>
<td>-0.18** (0.00)</td>
<td>-0.17** (0.00)</td>
<td>-0.31** (0.00)</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>7. Individual Preference Diversity</td>
<td>0.70** (0.00)</td>
<td>-0.24** (0.00)</td>
<td>0.26** (0.00)</td>
<td>0.02 (0.39)</td>
<td>0.09** (0.00)</td>
<td>-0.03 (0.21)</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

*Note:* (a) Numbers in parentheses are the p-values
(b) Degree centrality and betweenness centrality measures are log transformed.
Table 5.6. Variable Correlations at the Cluster Level

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Average Preference Similarity</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Degree Centrality</td>
<td>0.20* (0.01)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Betweenness Centrality</td>
<td>0.09 (0.28)</td>
<td>0.80** (0.00)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Closeness Centrality</td>
<td>0.26** (0.00)</td>
<td>0.77** (0.00)</td>
<td>0.62** (0.00)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Density</td>
<td>-0.15 (0.06)</td>
<td>-0.44** (0.00)</td>
<td>-0.41** (0.00)</td>
<td>-0.58** (0.00)</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>6. Number of Users in Cluster</td>
<td>0.03 (0.72)</td>
<td>0.59** (0.00)</td>
<td>0.75** (0.00)</td>
<td>0.58** (0.00)</td>
<td>-0.56** (0.00)</td>
<td>1.00</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Note: Numbers in parentheses are the p-values

At the ego-network, all degree centrality, betweenness centrality, and closeness centrality have significant positive correlations with the Euclidean distances in the ego-network, which indicates that a user with high centrality in the social network is likely to be less similar to her connections (i.e., high Euclidean distance). Degree centrality also has a significant negative correlation with the Gini coefficient (the reverse measure of preference disparity). This indicates that a user with many connections has a high disparity of preference similarity across her connections. Degree and betweenness centrality are significantly and highly correlated (0.64), which means that a user with more connections also tends to have a high betweenness centrality in the ego-network. An individual’s preference diversity has significant correlations with both average preference similarity (0.70) and preference disparity (-0.24). Such significant correlation implies that individual preference diversity does have impacts on both the average and the variance of the preference similarity between a user and her connections. As a result, in the analyses that test our hypotheses, we control for its effects on our dependent variables.

At the cluster level, both degree centrality and closeness centrality have significant positive correlations with Euclidean distance (0.20 and 0.26 respectively). They are also highly
correlated with each other (0.77). Interestingly, we do not see a significant correlation between Euclidean distance and density (-0.15 with a p-value of 0.06) and Euclidean distance and betweenness centrality (0.09 with a p-value of 0.28). There is also no significant correlation between the number of users in a cluster and the average preference similarity in the cluster, which indicates that the size of the cluster does not have a significant impact.

All the above correlations are consistent in direction with our hypothesized relationships between preference similarity and social network characteristics. We next formally test our hypotheses through regression analyses. Given the high correlations among some of our centrality variables, we ran two different types of analyses: first, univariate regressions to show the effects of each variable separately (along with control variables) so that one can see individual effects; and second, multivariate regressions that include all variables for that level of analysis so that one can see the joint effects. In this manner, one can better understand possible effects of multi-collinearity. In addition, we conduct F-tests to compare univariate models with full models, and compare full models with alternative models at each analysis level. In this way, we can determinant whether a variable can significantly increase the explanatory power in the model. If one variable explains a significant portion of the dependent variable’s variance (i.e., significant F-statistic), though it may not have a significant coefficient in the full model regression (due to multicollinearity), we will conclude that it has a significant impact on the dependent variable.

5.2 Hypothesis Testing

5.2.1. Individual Level Results

Preference diversity is the dependent variable at the individual level of analysis. We used the Gini coefficient as a reverse measure of this variable. In other words, a high Gini coefficient
indicates low preference diversity. The three social network structural characteristics whose effects we examined in our hypotheses are degree centrality (H1), betweenness centrality (H2), and closeness centrality (H3). We examined their separate as well as joint (Model 5 and full model) effects on an individual’s preference diversity. To represent the non-linear relationship between degree centrality and preference diversity, we add the logged degree centrality in the model. The nonlinear relationship (in log curve) of the independent variable is appropriate when the non-linear relationship posited suggests that as the independent variable increases the dependent variable increases but at a slower rate.

Table 5.7. Model Results at the Individual Level

(a) Univariate Models and The Full Model

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Full Model</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree Centrality</td>
<td>-0.01**</td>
<td>-0.01*</td>
<td></td>
<td>-0.13**</td>
<td>(0.00)</td>
<td>4.86</td>
</tr>
<tr>
<td>Log(Degree Centrality)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.07</td>
<td>(0.11)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.81</td>
<td></td>
</tr>
<tr>
<td>Betweenness Centrality</td>
<td></td>
<td>-0.04</td>
<td></td>
<td></td>
<td>0.21**</td>
<td>(0.00)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.86</td>
<td></td>
</tr>
<tr>
<td>Closeness Centrality</td>
<td></td>
<td></td>
<td></td>
<td>-0.10**</td>
<td>-0.34</td>
<td>(0.68)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.08</td>
<td></td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.04</td>
<td>0.07</td>
<td>0.01</td>
<td>0.01</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>F-test against Full Model</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>F-test for Model 1 vs. Model 2</td>
<td>0.005**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(b) Individual Variable's Contribution to Explained Variance: Full Model vs. Full Model excluding Individual Variable

<table>
<thead>
<tr>
<th>Change in SSE by excluding the variable in the full model</th>
<th>Degree Centrality</th>
<th>Betweenness Centrality</th>
<th>Closeness Centrality</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-test Significance</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.62</td>
</tr>
</tbody>
</table>

*significant at the 0.05 level (2-tailed).
**significant at the 0.01 level (2-tailed).
***significant at the 0.001 level (2-tailed).

Note: Dependent variable is individual’s preference diversity (Gini coefficient)

From Table 5.7 and looking at the individual regressions (Model 1 – 4), we can see that both degree centrality and closeness centrality have significant impacts on an individual user’s preference diversity, but that betweenness centrality does not. Respectively, the results indicate that if a user has more connections (higher degree centrality), or she is closer to all other users (higher closeness centrality) in the social network, her preferences are likely to be more diverse (lower Gini coefficient). However, we can see from Model 4’s F-test in table 5.7 (b), that closeness centrality does not explain more variance of the dependent variable, above and beyond that explained by degree centrality and betweenness centrality. In addition, the non-linear relationship between degree centrality and preference diversity has a higher explanatory power (adjusted $R^2 = 0.07$) than the linear model (adjusted $R^2 = 0.04$). By adding the logged variable in the model, the new model (Model 2) explains significantly more variance than the linear model (Model 1). Therefore, we conclude that a non-linear curve better describes the relationship between degree centrality and preference diversity. This means that as the number of connections increases, the average influence from each connection decreases (i.e., the marginal diversifying impact of adding a new connection gets weaker).
The results of the full model are more difficult to interpret due to multi-collinearity. The presence of multi-collinearity is evident by the fact that: 1) all centrality measures are highly correlated with each other (a range in 0.55 – 0.82); 2) the variance inflation factor (VIF) scores for degree centrality and closeness centrality are close to 4. A VIF higher than 4 indicates multi-collinearity (O’Brien 2007; Hair et al. 1995). When a model’s $R^2$ is not high, as in this study, some researchers even suggest the VIF should be smaller than 2.5, in order to clear possible multi-collinearity problems (Allison 1998); and 3) the sign of the beta coefficient for betweenness centrality changes in the full model. A flipped sign is an indicator of potential issues with multi-collinearity (Allison 1998; Berry and Feldman 1985). As we see in the correlation table and in the separate regression model, betweenness centrality has a negative (though not significant) correlation with the dependent variable but it has a positive relationship in the full model. As a result, due to multi-collinearity, we are not able to disentangle the separate effects of each variable in the full model and must rely on the separate analyses to interpret the results. Thus, based on the univariate analyses, we conclude that only one of our three hypotheses (H1) at the individual level is fully supported by the data. Though we observe a significant relationship between closeness centrality and preference diversity, closeness centrality does not significantly increase the explained variance in the model. This might be due to multicollinearity (i.e., the variance explained by closeness centrality overlaps with the variance explained by degree centrality given that the two variables are correlated at .82). Therefore, our H2 is not supported.
5.2.2 Ego-network Level Results

At the ego-network level, we examine two different dependent variables – average preference similarity (mean of pairwise Euclidean distances between a focal user and all her direct connections) and the disparity of preference similarity (the variance of preference similarity). We discuss the models and results for each dependent variable respectively. Table 5.8 summarizes both the separate and full model results at this level. Individual preference diversity (the dependent variable in our previous analyses), is the control variable in all models at this analysis level. This is to control the possibility that a user will be more similar on average to her connections if she has more diversified preferences.

5.2.2.1 Results on Average Preference Similarity

As the comparison between Model 1 and 2 in Table 5.8 (a) shows, the non-linear relationship (logged curve) between degree centrality and average preference similarity explains significantly more variance than the linear model. The upward increase in Euclidean distance, which is indicated by the positive sign, shows that degree centrality decreases the average preference similarity between users and their connections. At the same time, the curve becomes flattened as degree centrality becomes large, as indicated by the log relationship. This means that when degree centrality is low (i.e., a focal user has a few connections), the social recommendation system has a stronger heterogenizing effect on the average preference similarity as compared to the effects when degree centrality is high. Therefore, H4a (p < 0.05) is supported.

Betweenness centrality also has a significant non-linear impact on a user’s average preference similarity. That is, as a user’s betweenness centrality increases, the social recommendation system will decrease her average preference similarity to her connections and the effect gets weaker as the user’s betweenness centrality increases. However, betweenness
Table 5.8. Model Results at the Ego-network Level

**(a) Univariate Models and The Full Model**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Control Model</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Full Model</th>
<th>VIF</th>
<th>H4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>DV1: Euclidean Distance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree</td>
<td>0.01**</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.68)</td>
<td>(0.00)</td>
<td>(0.68)</td>
<td>(0.68)</td>
<td>(0.68)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log(Degree)</td>
<td>0.01**</td>
<td></td>
<td></td>
<td>0.01**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td></td>
<td></td>
<td>(0.00)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Betweenness</td>
<td>0.02**</td>
<td></td>
<td></td>
<td></td>
<td>0.01**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td></td>
<td></td>
<td>(0.00)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closeness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.03**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.00)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ego-network Density</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.03**</td>
<td></td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Variable:</td>
<td>0.40**</td>
<td>0.38**</td>
<td>0.37**</td>
<td>0.39**</td>
<td>0.38**</td>
<td>0.38**</td>
<td>0.38**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual Preference</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td></td>
<td>1.2</td>
</tr>
<tr>
<td>Diversity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R-square</td>
<td>0.48</td>
<td>0.51</td>
<td>0.51</td>
<td>0.50</td>
<td>0.50</td>
<td>0.54</td>
<td>0.58</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>F-test against Control</td>
<td>N/A</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.002**</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-test against Full Model</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Comparison between</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1 and Model 2</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td></td>
<td>N/A</td>
</tr>
</tbody>
</table>
(b) Individual Variable's Contribution to Explained Variance: Full Model vs. Full Model Excluding Individual Variable

<table>
<thead>
<tr>
<th>Change in SSE by excluding the variable in the full model</th>
<th>Degree Centrality</th>
<th>Betweenness Centrality</th>
<th>Closeness Centrality</th>
<th>Ego-Network Density</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>29.76</td>
<td>0.33</td>
<td>11.57</td>
<td>604.21</td>
</tr>
<tr>
<td>F-test Significance</td>
<td>0.001***</td>
<td>0.57</td>
<td>0.001***</td>
<td>0.001***</td>
</tr>
</tbody>
</table>

*significant at the 0.05 level (2-tailed).
**significant at the 0.01 level (2-tailed).
***significant at the 0.001 level (2-tailed).

Note: (a) DV1 is a reverse measure of average preference similarity
(b) DV2 is a reverse measure of preference similarity disparity

centrality does not significantly increase the explained variance in the model (insignificant F-test in Table 5.8 (b)). This is due to the high correlation between betweenness can degree centrality (0.64, as shown in Table 5.5), where the variance explained by betweenness centrality overlaps with the variance explained by degree centrality. Therefore, our H5 is not supported. Similarly, closeness centrality has a significant positive relationship with Euclidean distance (supporting H6) suggesting that the cumulative influences from the network have a heterogenizing effect on preference similarity. Finally, ego-network density has a negative relationship with Euclidean distance indicating that the more connected the ego-network is the more homogeneous the preferences within the ego-network (supporting H7). In addition, we can observe from Table 5.8 (b) that excluding ego-network density from the full model will greatly increase the standard error (by 604.21, which is much larger than the change of excluding other variables). This indicates that ego-network density is a better predictor than centrality measures when investigating the average preference similarity in the ego-network.

Though closeness centrality is not significant in the full model, we do see it explains unique variance of our dependent variable (i.e., significant F-test in Table 5.8 (b)). Given closeness centrality is significant in the univariate analysis and it explains unique variance of our
dependent variable, we conclude our H6 is supported. That is, if a user is closer to all others in the social network, the focal user is less likely to be similar to her connections because she is easily influenced by all the heterogeneous preferences in the social network.

5.2.2.2 Results on Disparity of Preference Similarity

We used the Gini coefficient to measure the disparity of preference similarity between a focal user and her connections (i.e., the variance of preference similarity). Table 5.8 (H4b) shows the univariate regression results for disparity of preference similarity. From the table, we can see a significant negative relationship between degree centrality and the Gini coefficient. Figure 5.1 shows two examples of preference distributions and Lorenz curves with two levels of degree centrality (low and high).

![Lorenz Curves](image)

<table>
<thead>
<tr>
<th>Degree centrality = 3,</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gini coefficient = 0.38</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Degree centrality &gt; 85,</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gini coefficient = 0.14</td>
</tr>
</tbody>
</table>

Figure 5.1. Examples of Lorenz curve when degree centrality is low or high

From Figure 5.1 we can see that when a user has only a few connections (e.g., has only three connections), her preference similarities to connections are mostly high (low Euclidean distance). That is, overall, a user is very similar to all her connections. As the number of

---

16 We use degree centrality > 85 in order to have a similar number of observations as in degree centrality = 3, and the degree centrality is large enough.
connections becomes larger (e.g., larger than 85 as in the Lorenz curve on the right), we can see that the Lorenz curve shifts toward the diagonal line and thus the Gini coefficient becomes lower. In this case, the focal user is very similar to some of her connections, those at low Euclidean distances, and dissimilar to others (those at high Euclidean distances). Thus, one’s level of similarity with her connections is not concentrated at high levels, but rather it is more equally distributed across all preference similarity levels. In other words, she shares high similarity with some connections and low similarity with others. This pattern is consistent with our hypothesis – as degree centrality increases (i.e., number of Last.FM friends increases) a user is only influenced by a few of these connections (resulting with high levels of preference similarity) and weakly or not influenced by others (resulting with moderate or low levels of preference similarity). Thus, H4b is supported.

5.2.3. Cluster Level Results

Table 5.9 summarizes the results of hypotheses at the cluster level of analysis. Since there is no cognitive constraint at this level, we posit linear relationships between degree centrality and betweenness centrality (that is, we do not log transform these variables in the regressions). Results of the separate analyses indicate a negative linear relationship between degree centrality and average preference similarity in the cluster showing a heterogenizing effect. Similarly, as a cluster’s closeness centrality increases, the average Euclidean distance within the cluster increases significantly indicating heterogenizing effects on average preference similarity, supporting H8. We do not observe a significant relationship between betweenness centrality and the average preference similarity in the cluster. Therefore, H7 is not supported. Finally, cluster density has a significant negative relationship with the average Euclidean distance in the cluster.
This means that as the cluster’s density increases, a social recommendation system will homogenize user preferences in the cluster (supporting H9).

Table 5.9. Model Results at the Cluster Level

(a) Univariate Models and The Full Model

<table>
<thead>
<tr>
<th>Variables</th>
<th>Control Model</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Full Model</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree Centrality</td>
<td>0.01**</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.44)</td>
<td>(0.72)</td>
<td>(0.03)</td>
<td>(0.01)</td>
<td>(0.04)</td>
<td></td>
</tr>
<tr>
<td>Betweenness Centrality</td>
<td>0.001</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.72)</td>
<td>(0.03)</td>
<td>(0.01)</td>
<td>(0.04)</td>
<td>(0.86)</td>
<td></td>
</tr>
<tr>
<td>Closeness Centrality</td>
<td>0.01**</td>
<td>0.01*</td>
<td>0.01*</td>
<td>0.01*</td>
<td>0.01*</td>
<td>0.01*</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>-0.04*</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.86)</td>
<td>(0.86)</td>
<td>(0.86)</td>
<td>(0.86)</td>
<td>(0.86)</td>
<td></td>
</tr>
<tr>
<td>Control Variable:</td>
<td>0.001</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01*</td>
<td>0.00</td>
<td>-0.01</td>
<td>2.9</td>
</tr>
<tr>
<td>Users in the cluster</td>
<td>(0.79)</td>
<td>(0.11)</td>
<td>(0.37)</td>
<td>(0.04)</td>
<td>(0.41)</td>
<td>(0.14)</td>
<td></td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.04</td>
<td>0.00</td>
<td>0.06</td>
<td>0.02</td>
<td>0.02</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>F-test against</td>
<td>N/A</td>
<td>0.001***</td>
<td>0.004**</td>
<td>0.001***</td>
<td>0.02*</td>
<td>0.001***</td>
<td></td>
</tr>
<tr>
<td>Control Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-test against</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td></td>
</tr>
<tr>
<td>Full Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) Individual Variable’s Contribution to Explained Variance: Full Model vs. Full Model Excluding Individual Variable

<table>
<thead>
<tr>
<th></th>
<th>Degree Centrality</th>
<th>Betweenness Centrality</th>
<th>Closeness Centrality</th>
<th>Cluster Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in SSE by</td>
<td>0.69</td>
<td>0.52</td>
<td>5.26</td>
<td>4.22</td>
</tr>
<tr>
<td>excluding the variable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in the full model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-test Significance</td>
<td>0.41</td>
<td>0.47</td>
<td>0.03*</td>
<td>0.04*</td>
</tr>
</tbody>
</table>

*significant at the 0.05 level (2-tailed).
**significant at the 0.01 level (2-tailed).
***significant at the 0.001 level (2-tailed).

Note: (a) Dependent variable is the average preference similarity in the cluster

The full model in Table 5.9 shows the multivariate analyses for the cluster level variables. From Table 5.9 (b), we observe that both degree centrality and betweenness centrality
do not significantly increase the explained variance in the full model. Therefore, even though degree centrality has a significant impact on the average preference similarity in the univariate analysis, the variance explained by degree centrality overlaps with the explained variance by closeness centrality or density. Betweenness centrality has no significant effect at the cluster level since it is non-significant both in the separate analysis and in the full model. Therefore, our H8 and H9 are not supported. Finally, while the homogenizing effect of density was supported in the separate analysis, it is not significant in the full model. This is likely due to multi-collinearity - closeness centrality and density are correlated at -0.58 and the full model results in high VIF scores (Table 5.9). To test further, we removed closeness centrality from the full model. Degree centrality became significant while density remained non-significant (density and degree centrality are correlated at -0.44). However, the significant F-test for the full model comparison with and without density indicates that density can explain a unique portion of our dependent variable’s variance. Therefore, we conclude that H11 is supported as density has a significant effect in the univariate model and it significantly increases the explained variance in the full model. In summary, the results of the separate and joint analyses lead us to conclude that degree and closeness centralities are significant predictors of preference similarity in a cluster, with closeness centrality being the dominant predictor.

5.3 Summary

Social network structural characteristics do have impacts on consumer preferences but their relative effects differ across the three levels of analysis. Further, while a user’s centrality is an important structural characteristic, different types of centrality measures are significant or have a dominant effect at different levels of analysis. For example, it is interesting to observe that at the two different levels of analysis that use the same dependent variable, we have different
dominant predictors\textsuperscript{17} – at the ego-network level, degree centrality is the dominant predictor of average preference similarity among the focal consumer and all her connections; while at the cluster level, closeness centrality is the dominant predictor of average preference similarity within a cluster.

To better understand why different network characteristics, and especially different centrality measures, have different effects we categorize the social network structural characteristics we examined in terms of bonding and bridging ties (Putnam 2000) and use this as the basis of discussing possible reasons for the different effects observed in this study. This discussion is included in the next chapter. Below we briefly revisit bonding and bridging ties as we summarize the results.

Degree centrality, closeness centrality, and network density describe bonding ties in a social network. Bonding ties refer to two or more people being tightly connected with each other (Hampton 2011; Putnam 2000). Bonding tends to increase homogeneity within the group of people. On the other hand, betweenness centrality reflects bridging ties in the social network. Bridging ties, also called weak ties (Burt 1992; Granovetter 1983), refer to connections that link one network to another. Bridging ties are the sources of new information, knowledge, and resources and typically result in heterogeneity within a network.

Table 5.10 summarizes our findings. Two observations arise from the table. First, as we discuss in the next chapter, our results point to the importance of bonding (rather than bridging) ties to the effects of a social recommendation system on preference similarity. Second, different structural characteristics have different effects at three levels of analysis. At the individual level, degree centrality is the better predictor that explains a user’s preference diversity. At the ego-

\textsuperscript{17} We do not include individual level analysis in this discussion because the dependent variable at the individual level (preference diversity) is different from the dependent variables at the ego-network and the cluster level (preference similarity).
network level, degree centrality and density have significant impacts on a user’s average preference similarity to her connections. Degree centrality also has a significant impact on the variance of the preference similarity (similarity disparity). At the cluster level, closeness centrality is the best predictor of the average preference similarity. We discuss these two observations in the next chapter.

<table>
<thead>
<tr>
<th>Bonding Characteristics</th>
<th>Individual</th>
<th>Ego-Network</th>
<th>Cluster Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preference Diversity</td>
<td>Average Preference Similarity</td>
<td>Average Preference Similarity</td>
</tr>
<tr>
<td>Degree Centrality</td>
<td>Supported (H1)</td>
<td>Supported (H4a, b)</td>
<td>Not Supported (H8)</td>
</tr>
<tr>
<td>Closeness Centrality</td>
<td>Not Supported (H3)</td>
<td>Supported (H6)</td>
<td>Supported (H10)</td>
</tr>
<tr>
<td>Density</td>
<td>N/A</td>
<td>Supported (H7)</td>
<td>Supported (H11)</td>
</tr>
<tr>
<td>Bridging Characteristic</td>
<td>Betweenness Centrality</td>
<td>Not Supported (H2)</td>
<td>Not Supported (H5)</td>
</tr>
</tbody>
</table>
CHAPTER 6

DISCUSSIONS AND CONCLUSION

This chapter discusses the results presented in Chapter 5 and presents the dissertation’s contributions to theory and practice. Possible limitations and future research directions are also discussed.

6.1 Discussion of Results

This dissertation aims at enhancing our understanding of the effects of social recommendation systems, which are becoming increasingly prevalent among e-vendors and other social media websites such as Last.FM. Relying on theories on social networks and human information processing, the study shows that the effects of social recommendation systems are not identical for all users in a social network. Rather, effects are dependent upon the social network’s structural characteristics. This finding is consistent across the three different levels of analysis (individual, ego-network, and cluster) that we examined in the dissertation. Furthermore, our findings suggest that across all three levels of analysis higher levels of social network centrality are associated with more heterogeneous preferences; while higher levels of network density are associated with more homogeneous preferences.

6.1.1. Individual Level: Effects on User Preference Diversity

At the individual level, the dependent variable was an individual user’s preference diversity reflecting the extent to which users have broad (rather than narrow and focused) sets of preferences. Results suggest that higher degree and closeness centralities are associated with
more diversified Last.fm user’s preferences (H1 and H3 are supported), but the relationship does not hold for betweenness centrality (H2 is not supported). That is, bonding ties between a user and others in the network, reflected by the number of other users to whom the individual is connected (degree centrality) and distance from all other users imply more diverse user preferences than those of users at the border of the network. We posit that this is due to the heterogeneous influences (in the form of social recommendations) from other users. However, given that betweenness centrality has non-significant effects and degree centrality has dominant effects, the nature of centrality is also consequential to the effects of social recommendation systems.

There are several possible reasons for why closeness has a relatively less dominant effect than degree centrality and why betweenness centrality has no significant effect. Firstly, closeness centrality is a measure of the total distance to all other users in a social network. As we hypothesized in Chapter 3, the shorter the distance, the easier it is for influence to reach the focal user, and thus to diversify the user’s preferences. However, due to the nature of social recommendation systems, before all those influences reach the focal user, they need to be accepted by the user’s direct connections and incorporated into their own preference set. As depicted in Figure 6.1, for C to influence A, C needs to influence B first. Only when B has incorporated some of C’s preferences, can those preferences be recommended to A and thus influence A’s preferences. Those C’s preferences that finally reach A, however, are part of B’s preferences which have already been captured by another predictor – degree centrality, which reflects the direct influences A receives.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
</table>

Figure 6.1. The Indirect Influence
Secondly, betweenness centrality is an index of potential for control of communication (Freeman 1979). In the context of Last.FM, users do not have a controlling power on what to display or what not to display on their music listening history. Their friends can browse all their listened music freely. The role of “control” in the process of preference influence does not apply and this likely explains the insignificant effect of betweenness centrality.

Thirdly, from a statistical point of view, degree centrality, closeness centrality, and betweenness centrality are highly correlated with each other (all are above 0.6). The variance explained by closeness and betweenness centrality overlaps with the variance explained by degree centrality. This also reduces the statistical importance of closeness and betweenness centrality at the individual level.

In addition, our findings suggest a non-linear relationship such that as a user’s degree centrality increases, the social recommendation system’s effects become weaker. We attribute the non-linear relationship to the user’s limited cognitive capability – that is, as the number of friends in a user’s social network increases, she is not able to process all recommendations from all her friends (Cowan 1988; Payne 1982). Therefore, the marginal effect on preference diversity of adding more connections declines as degree centrality increases. This suggests that size of the social network is consequential to preference diversity up to a point, at which diminishing returns set in.

6.1.2 Ego-Network Level: Effects on Preference Similarity

At the ego-network level, we investigated the social recommendation system’s effects on the average preference similarity between a focal user (ego) and her friends; and the effects on the disparity of preference similarity. Both the average preference similarity and the disparity of preference similarity are influenced by a user’s structural network characteristics. Specifically,
degree centrality (i.e., the number of connections) has a significant non-linear and negative effect on average preference similarity (as Euclidean distance is a reverse measure of similarity); while it has a significant negative effect on disparity of preference similarity (H4a and H4b, supported). Betweenness centrality (i.e., the number of different groups with which a user is connected – and an indicator of bridging ties) has a significant non-linear and negative effect on average preference similarity, however, its explained variance overlaps with the explained variance by degree centrality. Therefore, it does not increase the explanation power if we put betweenness centrality in the model (H5 is not supported). Closeness centrality, which is a reverse measure of total distance to all other users, does not have a significant effect on average preference similarity (H6 is not supported). Besides centrality measures, we also examine an aggregate level structural characteristic – network density. At the ego-network level, density has a significant and positive impact on the average preference similarity between the focal user and all her friends supporting H7. In sum, degree centrality, betweenness centrality, and ego-network density have significant effects on average preference similarity in an ego-network. Thus both bonding (degree centrality and density) and bridging (betweenness centrality) ties are consequential at the ego network level. As at the individual level, due to cognitive constraints these effects become weaker when the focal user has more connections and as the user’s betweenness centrality increases.

Similar to the rationale presented for individual level results, the effect of closeness centrality may have been captured through the effect of degree centrality. This may explain the non-significant effect of closeness centrality in the full model. Though betweenness has a significant effect in the univariate model, this is likely caused by the high correlation with degree centrality (0.64). Betweenness centrality by itself does not explain more variance in our
dependent variable. This is in fact consistent with the rationale provided at the individual level – since users do not control which of their listening preferences others view in the context of Last.FM, betweenness centrality, as an index of information control, is not expected to have a significant effect in the process of preference influence.

We examined the variance of preference similarity to better understand the effects of cognitive constraints on preference similarity (H4b). We had posited that there are two possible mechanisms that explain how the marginal effect of degree centrality on preference similarity decreases. One mechanism is that as the number of one’s connections increases the user will uniformly reduce the number of recommendations from each connection she browses. This would imply that the disparity (variability) in preference similarity between a user and her connections is low (she shares approximately the same level of common preferences with all her connections). An alternative mechanism is that the user will focus on recommendations from a small group of her connections (and will share a high preference similarity with these) and largely ignore recommendations from the rest of her connections. This would result in high disparity of preference similarity. Results provide support for the second mechanism, that is, even though users may have many connections, which is normally the case in online social networks, they will not evenly put their effort and time browsing recommendations from every connection they have. Instead, they only focus on recommendations from a few connections. These people could be their close friends or experts with high credibility (e.g., in the case of Last.FM music masters who have expertise in music).

6.1.3 Cluster Level: Effects on Preference Similarity

At the cluster level, users in a cluster are less similar to one another when the cluster is closer to other clusters in the social network (H10, cluster closeness centrality). The model
comparisons reveal that closeness centrality is the most influential measure among all three centrality measures. Though degree centrality has a significant effect on average preference similarity, it does not explain more variance if we include closeness centrality in the model (H8 not supported). The non-significant relationship between betweenness centrality and the preference similarity at the cluster level (no support for H9) suggests though centrality is an important structural characteristic for the effects of the social recommendation system, the type of centrality is of consequence – just like at the individual and ego-network levels of analysis. Finally, based on F-test indicating cluster density’s additional explanatory power on the dependent variable, cluster density appears to have a small effect on preference similarity (H11 supported).

The cluster level analysis is quite similar to the individual level analysis, if we equate the average preference similarity within a cluster to the preference diversity “within” an individual user. Both measures are reflections of homogeneity within a specific unit. Therefore, not surprisingly, we have similar findings at the cluster level as we have at the individual level – bonding characteristics (i.e., degree centrality and closeness centrality) are the dominant predictors while bridging characteristics (betweenness centrality) are not significant. There are two distinct differences, however, between cluster and individual findings: first, a cluster does not have cognitive constraints. Thus, degree centrality’s effect is linear in models at this level. The effects of social recommendation system at the cluster level will not become weaker if a cluster is connected to more clusters; second, closeness centrality’s explanation power is greater than degree centrality’s at this level.

In the context of Last.FM, closeness centrality is more appropriately viewed as a measure of indirect effects (Borgatti 2005; Valente and Foreman 1998) and not as a measure of efficiency.
(given that preference flows in the network between two users do not necessarily take the shortest path). At the cluster level, since closeness centrality is more influential than degree centrality (which measures immediate direct effects), preference similarity within the cluster is influenced by preferences in the broader social network rather than by the cluster’s immediate neighborhood.

6.2 Contributions

6.2.1 Theoretical Contributions

This study makes important contributions to the body of IS research on recommendation systems and extends this research stream in several ways. First, this is the first study of which we are aware that focuses on social recommendation systems, a new type of recommendation system which is becoming very prevalent in online environments. The effects of these systems on users and how these differ from the effects of other types of recommendation systems are not yet well-understood. Results of the current study are a step in this direction.

Second, as suggested by Arazy et al. (2010), theory-driven design methodology that links kernel theories and system designs is very useful in information system development. Using social theories to guide information systems designs, helps researchers focus on the key antecedents of the desired outcome. As a very new type of recommendation system, social recommendation systems still have many attributes that are not fully understood by researchers and system designers. This study, relying on kernel theories in the social network domain, reveals a key antecedent of social recommendation system’s effects – a user’s (or cluster’s) social network structural information. In other words, social recommendation system’s effects differ depending on a user or cluster’s structural characteristics (e.g., centrality). This challenges
the assumption that recommendation systems have the same effects on all users once these systems are implemented. Thus, researchers have to take into account a user’s network position when examining the effects of a social recommendation system; while system designers have to consider a user’s structural information when generating recommendations for her.

Third, this study contributes to the debate on inconsistent effects of recommendation systems on user preference similarity. By distinguishing the type of recommendation systems used in the studies (content-based or collaborative filtering) and the levels of analysis (individual level or aggregate levels), the study presents clear patterns for the effects of each type of recommendation systems at each level. Thus, our review and synthesis of the literature shows that different types of recommendation systems have distinct effects on user preference similarity and that this effect may vary depending on the level of analysis. That is, even the same type of recommendation system (e.g., collaborative filtering) can have opposite effects at different levels of analysis. Most prior studies on recommendation systems’ effects (not just studies on user preferences) do not describe the type of recommendation system used in their study. Given the different effects that different types of recommendation systems may have, generalizability of these findings may be problematic. As such, our study shows the importance of presenting the type of recommendation systems and the focal level of study in research and taking these characteristics into account when assessing generalizability of findings.

Fourth, we observed non-linear effects of social network characteristics on user preferences. This reflected the amount of information received by the user from her connections and the cognitive limitations that the user has in terms of processing all these recommendations. Recommendation overload and cognitive constraints are not unique to social recommendation systems but rather they can be concerns for the design of traditional content-based and
collaborative filtering recommendation systems. The introduction of non-linear effects of recommendation systems creates new insights and has strong theoretical as well as practical implications, as we discuss in the next section. We discussed and tested the mechanisms via which users deal with recommendation overload in social recommendation systems. Different mechanisms are likely applicable to other types of recommendation systems and are worthy of further exploration.

Finally, the majority of prior studies on recommendation systems mainly focus on user’s perceptions (e.g., trust, perceived usefulness, perceived ease of use) and recommendation systems’ effects on decision making (e.g., recommendation accuracy, required cognitive effort) (e.g., (Al-Natour et al. 2006; Kumar and Benbasat 2006; Tam and Ho 2006). As recommendation system’s ultimate purpose is to provide accurate recommendations, user preference is a key factor that cannot be neglected, especially in the context of e-commerce (Fleder et al. 2010; Pariser 2011). In this study, we examined how a social recommendation system, with different social network structures, might influence a user’s preferences and a group of users’ preference similarity.

More broadly, our study contributes to the social networks literature. The literature has broadly assumed that information flowing through the social network tends to homogenize actors’ (i.e., people in the social network) opinions and preferences (Marsden and Friedkin 1993; Festinger 1954, 2001). This assumption is largely true in a relatively small social network (e.g., a dyadic relationship or a small group of people). With a large enough amount of information exchange among actors, and allowing for the exchange process to continue for a long enough period of time, actors tend to form cohesive opinions. In this study, however, what we observe are diversifying effects in the social network. This is due to two primary reasons. First, the
number of actors’ connections is quite large as is often the case with online social networks. Second, as a key attribute for online social networks, most tie strengths among actors are likely to be weak. Weak ties are considered an indication of diversity in information received by a focal actor (Granovetter 1973). In sum, this study reveals that online social networks have several unique characteristics as compared to traditional offline social networks (e.g., large number of connections, weak ties, etc.). As a result, we observe diversifying effects of social recommendations in online social networks as opposed to cohesion and convergence in offline social networks.

6.2.2 Practical Contributions

With the widespread use of online social network platforms, social recommendation systems provide a new way for e-vendors to utilize social network information to increase product sales and provide users with a better online experience. As pointed out by Adomavicius and Tuzhilin (2005), the first step to generate recommendations is to understand users. E-vendors need to understand the effects of recommendations systems and how they change user preferences in order to adjust their campaign strategies or to implement a recommendation system that aligns with their business goals. Therefore, e-vendors will benefit from the findings from this study.

First, this study reveals that social recommendation systems tend to diversify user preferences. The more central a user is in the social network, the higher the diversity of the user’s preferences and the less similar the user’s preferences to those of her direct connections. Similarly, the more central a user cluster (e.g., an online community) in the social network, the more the social recommendation system heterogenizes user preferences within the cluster. This means that e-vendors need to choose a proper strategy when using social recommendation
systems to promote sales. Social recommendation system’s diversifying effects make it suitable for promoting the sales of niche products (i.e., diversifying users’ preferences); while it may not be so effective for enhancing users’ preferences on popular products (i.e., homogenizing users’ preferences). For instance, from this study we know that a user who locates at the center of a social network is likely to possess a diversified set of preferences. Therefore, it may be easier for a center user to be aware of a niche product (i.e., the likelihood of the niche product to match with the user’s preferences is higher) as compared to a user at the border of the social network (she might have a very concentrated set of preferences, implying that the niche product may not be on her preference list). This has implications on what is being recommended to each user.

Second, the non-linear effect of social recommendation systems indicates that the marginal influence a user receives from her connections diminishes as the number of connections increases. This means that e-vendors should not assume a social recommendation system has the same level of effects on every pair of connections. In other words, e-vendors should not expect the user to browse recommendations from all her friends, especially when the user has a large number of connections. Therefore, social recommendation systems need to focus on social recommendation filtering mechanisms that include a user’s key connections to achieve a balance of influence while avoiding recommendation overload.

6.3 Limitations and Future Research Directions

There are two primary views to explain social network’s effects: structural and connection (Borgatti and Foster 2003). The structural view of social networks, which is used in this study, explains common attitudes and beliefs using an actor’s network structural characteristics; while the connection view of social network considers shared attitudes and
behaviors as solely caused by interactions between actors (Borgatti and Foster 2003). In this study, we adopt the centrality and density concepts from the structural view of social networks and the social influence theory from the connection view in order to provide a more holistic view of the social network’s effects. However, our study did not take into account the influence of strength of ties between two actors (e.g., the amount of interaction) on social recommendation systems’ effects. Limited by our data set, when testing the effects of the social recommendation system we assumed that the connections between all users are of equal tie strength. This is in fact the current practice for recommendation systems on the market. However, we expect and do observe that tie strength may play an important role in forming a user’s preferences. As shown in Hypothesis 4b, when a user has many connections, she tends to accept recommendations from only a few of these. These few connections likely have different characteristics than other connections. Though we are not able to test this in the current study, we propose that such differences may lie in the tie strength and type of ties (e.g., friendship ties, advice ties, etc) with the focal user. Future studies can use laboratory experiments to examine the effects of tie characteristics on user preferences and further our understanding of the effects of social recommendation systems.

The findings of this study are also limited to experience products because the social recommendations used in this study are all music and artists. Experience products’ dominant attributes are difficult to assess prior to usage as opposed to search products whose dominant attributes can be fully acquired prior to product usage (Klein 1998; Nelson 1970). A good example of a search product is a computer. Users are more likely to know the nature of the product and its attributes. Social recommendations in this case are merely a technology that makes users aware of products, which might not be very different from content-based or
collaborative recommendation systems. Future studies can set up an experiment to empirically examine social recommendation system’s effects on experience products and search products; and to examine whether different types of recommendation systems have different effects when product characteristics are different.

In addition, most existing studies on recommendation systems (including this dissertation) are cross-sectional studies. One weakness of cross-sectional studies is that it pictures the world in a snap shot. However, user preferences may change over time; their network position may change over time; and their tie strength with others may also change over time. Therefore, it would be very interesting to conduct longitudinal studies and examine possible patterns in social recommendation systems’ effects. Given the large data sets generated in all the social network platforms every day, future studies may soon have good opportunities to utilize those archival data to pursue questions in this direction.

Several other recommendations for future research derive from the study. First, similar to other types of recommendation systems, social recommendation systems have three major goals: to understand user preferences, to generate accurate recommendations, and to persuade users to adopt these recommendations (Adomavicius and Tuzhilin 2005).

The goal of “understanding users” centers around predicting their preferences. Social recommendation systems use friends’ purchases or likings to estimate the focal user’s preferences. Compared to traditional types of recommendation systems (e.g., content-based and collaborative filtering), the question that arises is whether social recommendation systems can do a better job in estimating user preferences. Since social recommendation systems only use friends’ preferences in making the estimation while neglecting the focal user’s past history of preferences, it may be more effective to combine both when estimating user preferences. That is,
e-vendors can use a hybrid of social recommendation systems and content-based recommendation systems as complementary tools when estimating user preferences. In fact, a few prior studies have found that hybrid recommendation systems (i.e., a combination of two or more types of recommendation systems) can lead to higher user trust, perceived usefulness, and satisfaction (Kumar and Benbasat 2006; Xiao and Benbasat 2007). However, implementing both requires more user information collection and more complicated system design, which could introduce negative outcomes (e.g., increased privacy concern, higher cognitive load, reduced perceived ease of use, or implementation costs). Therefore, empirical evidence is needed to assess the efficacy of social recommendation systems vis-à-vis traditional recommendation systems and also examine whether the increased inconvenience to the user and cost to the e-vendor of implementing hybrid solutions is worth making.

The second goal of social recommendation systems is to generate accurate recommendations. Understanding how network characteristics influence the accuracy of social recommendations can be an interesting research direction to pursue in the future. For example, which characteristics are more important – centrality, density, tie strength, etc – when the system tries to generate accurate recommendations? Are there any circumstances that structural characteristics (e.g., centrality) are more important; and circumstances that tie characteristics (e.g., strength) are more important? In addition, as found in this study, when a user has too many connections, she is likely to be selective and only browse recommendations from a few connections. If the system can identify these connections and increase their weight in generating recommendations (e.g., more recommendations from them), we expect the resulting recommendations would be more accurate. Would this, however, come at the expense of “groupthink” and tunnel vision? Future research can examine these questions
The final goal for a recommendation system is to persuade users to adopt the recommendations. Prior literature has identified a few methods that can increase recommendation’s acceptance rate, such as better user interface design (Tam and Ho 2005), provide explanation facilities (Wang and Benbasat 2005, 2007), and present human-like avatar when delivering recommendations (Hess et al. 2009; Qiu and Benbasat 2009). All these features can be applied to social recommendation system as well. Moreover, since social recommendation system has user’s social network information, how we can use some of the social network information in the persuading process is an interesting question to ask. For example, showing friends’ names under the recommended items, as Amazon.com does currently, could increase the credibility of the recommendations and thus increase the acceptance rate. Future studies, relying on social network and psychology theories, can identify and empirically test more similar features that are influential in the process of persuasion.

6.4 Conclusion

This dissertation investigated the social recommendation system’s effects on user preference diversity and similarity to others in the social network at different analysis levels. Effects vary based on the individual user’s, ego-network’s, and cluster’s social network structural characteristics. Specifically, centrality is positively associated with diversification of user preferences but the type of centrality that has dominant effects varies by level of analysis. Furthermore, preference influence in the social network appears to be primarily based on bonding rather than bridging ties.

The study contributes to the research stream on recommendation systems by providing empirical evidence on the effects of social network recommendation systems and revealing the
importance of structural characteristics when examining these effects. Our synthesis of prior literature suggests that the inconsistent findings on the effects of recommendation systems on user preferences can be resolved by taking into account the type of recommendation system and the level of analysis. Since social recommendation systems are a new phenomenon, both researchers and practitioners have little understanding on their effects. Given that findings on the effects of traditional types of recommendation systems may not generalize to this new type of recommendation system and that social recommendation systems have a set of unique characteristics, additional research is required to enhance our understanding of the phenomenon.
REFERENCE


Mok, D., and Wellman, B. "Did distance matter before the Internet?: Interpersonal contact and support in the 1970s," *Social Networks* (29:3) 2007, pp 430-461.


Van Dongen, S. "Graph Clustering by flow simulation," University of Utrecht, 2000.


