Information Networks to Derive Value from Social Media

BY

PANKHURI MALHOTRA
Bsc. Physics., Miranda House, India, 2013
Msc. Business Analytics, University of Manchester, United Kingdom, 2015

THESIS
Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in Management Information Systems
in the Graduate College of the
University of Illinois at Chicago, 2020
Chicago, Illinois

Defense Committee:

Siddhartha Bhattacharyya, Chair and Advisor
Yuheng Hu
Vijay Kamble
Jennifer Cutler, Northwestern University
Kumar Mehta, George Mason University

DEDICATION

This work is dedicated to my parents and to those who stood by me to achieve this milestone.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Siddhartha Bhattacharyya, for his continuous support and guidance in shaping this dissertation. His insight and knowledge into the subject matter helped me in designing this research and achieving this huge milestone in my life. I would also like to thank my dissertation committee, Dr. Yuheng Hu, Dr. Vijay Kamble, Dr. Kumar Mehta and Dr. Jennifer Cutler for their insightful comments and reviews, which incented me to broaden my research question from various perspectives.

My sincere gratitude goes to Dr. Ali Tafti who never hesitated to provide support and guidance, even during some of his very busy days. Finally, I cannot fully express how thankful I am to my father, Sanjay Malhotra, who supported me during my PhD, and at all times; to my mother, Taruna Malhotra, my motivator, who is always there for me; and to close friends and family for their unwavering support and encouragement.

- Pankhuri Malhotra

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ERGM	Exponential Random Graph Model
AMT	Amazon Mechanical Turk
DP	Disparity Filter
GloSS	Global Statistical Significance Filter
MCMC-MLE	Markov Chain Monte Carlo maximum likelihood estimation
c-c	category – category
b-c	brand – category
b–b	brand – brand
GMID	Global Market Information database
UGC	User-generated content

SUMMARY

The rise in electronic interactions has made information networks ubiquitous. Correspondingly, research across multiple domains has begun to acknowledge the social and economic value of these networks for business decision-making. One such type of information networks are implicit brand networks that provide a direct digital window into the "interest space of millions of brand fans" on social media. Unlike conventional social networks that involve direct interaction between the participating entities, links within a brand network are implicit and arise due to the co-followership activity of digital users.

In this dissertation, we derive brand networks from social media to provide statistical knowledge on online market structures and automatically infer brand associations over time. First, we employ Exponential-Random-Graph-Models from network theory to examine the tacit information contained in brand-brand links and reveal a mix of network and brand level characteristics responsible for the observed co-followership patterns. Some of the significant effects include homophily based on category, cross-category interactions between certain pairs (such as Automotive-Sports, Travel-Restaurants, Apparel-Personal Care) and frequency of a brand's engagement with online fans. While most existing studies focus on within-category competition, the notion of examining cross-

SUMMARY (continued)

category effects have largely been overlooked. This dissertation aims to bridge the gap by highlighting statistically relevant co-interest patterns between brands of the same as well as different categories.

After the statistical significance of the network links is established, we introduce an automated scalable approach for studying the asymmetric cross-category associations of brands over time. A new construct, brand transcendence, is defined that captures the diverse associations of brands onto new categories. Overall, the use of network-based constructs allows managers to visualize the transcending brand associations at three different levels : category-category, brand–category and brand-brand, depending on their business objectives. Further, as user-brand relationships on social media change over time, we compare the results of the brand network across two time periods – 2017 and 2020. The analysis helps to visualize the fluctuating brand associations over time and investigate its impact on co-branding opportunities. Managers can also study the network patterns over time to evaluate the effectiveness of their marketing campaigns or assess the impact of external events on their brand's image in consumers' minds.

SUMMARY (continued)

Third, we propose an automated approach for mapping the relative positioning of competing brands in a single framework. Perceptual maps, obtained from the network, help uncover the competitive landscape of brands along the dimensions: centrality and distinctiveness. Compared to extant data mining approaches that rely on substantial human intervention, this unsupervised automated approach lets practitioners study the relative positioning of their brand not only against a set of common competitors but against any other brand in the ecosystem; thus, uncovering a broader picture on both within-industry competition and across-industry complementarities. To investigate the usefulness of our proposed methodology, we validate the findings from our automated approach against external survey ratings and conduct extensive robustness checks to ensure reliability of underlying Twitter data.

Other practical applications of the brand network are also discussed. First, competitor analysis helps to identify the closest competitors of a brand and uncover differential associations of each brand in the competing group. Second, employing standard clustering algorithms on the network allows managers to identify communities of brands with similar user preferences; this provides an alternate view of consumer segments based on direct data on their diverse interests

I. INTRODUCTION

The rise of information technologies has made digital networks increasingly prevalent. As many would agree, the social and economic impact of such networks is expected to surpass the effects caused by the widespread adoption of IT in the past decade (Sundararajan et al. 2013). More than ever, the increased availability of massive amount of digital trace data and tremendous potential of networks to gain a better understanding of these online traces has led to a growing interest in information networks (Oestreicher-Singer et al. 2012; Zhang et al. 2016; Lazer et al. 2009). Most existing research on information networks focusses on a particular network type, that is, *social* networks where individuals/entities explicitly interact with one another. Digital information embedded in social networks has proven to benefit organizations in a number of ways, including targeted marketing (Hill et al. 2006), customer retention (Dasgupta et al. 2008), fraud identification (Fawcett and Provost 1997), product adoption (Godes and Mayzlin 2009) and several other business applications.

Another type of information network is the *economic network* where links are established by the shared economic interests between entities, such as a co-purchase network of products (Oestreicher-Singer et al. 2012). Nodes here are products that happen to be purchased frequently. Unlike social networks, a link in a product network does not explicitly reflect a node's decision to voluntarily connect with others; instead, it implicitly reflects the aggregated preferences of a large number of consumers (i.e. their co-purchasing patterns on Amazon) (Sundararajan et al. 2013). Another variant of this type of implicit network is a brand network, introduced by Zhang et al. (2016), where individual nodes represent brands and links between two brands represent common consumer engagement on Facebook. Similar to product networks, links within a brand network reflect aggregated preferences of a large number of users and provide a direct model to identify target users for online brand advertising.

These new types of digital artifacts rely on implicit connections to reflect consumers' interests and provide a rather novel view of “information in networks” as opposed to traditional social networks; thus, creating a new kind of interconnected entities, which one might approximate for an “economic network” (Jackson 2010). Moreover, with their inherent ability to condense the vast digital interest space of millions of users into a reduced form which is more amenable for research and business application purposes, implicit information networks have started to garner increasing attention from researchers across domains (Sundararajan et al. 2013; Levina and Arriaga 2014).

In this dissertation, we leverage these new types of digital networks, particularly implicit brand networks, to focus on three key aspects of brand management – a) *statistical analyses of online market structures* b) *eliciting brand associations* and c) *brand positioning.* Individual nodes in the network represent brands, and a weighted link between two brands represents the strength of common consumer co-interest. The implicit network, thus, reveals brand-to-brand associations based on common consumer activity. By incorporating users' interests across a broad ecosystem of brands, brand networks help uncover a broader picture on both within-category competition and across-category complementarities. Having relevant knowledge about cross-category brand associations is crucial for a range of business applications such as coordinated promotions, embedded premiums, co-branding and brand extensions (Cutright, 2013; Xiao and Lee, 2014); however, there is little or no evidence on identifying these cross-category effects using current digital approaches. Our proposed methodology helps to fill this void in the literature. The cross-category branding insights, revealed through the network, serve as important measures to assess brand fit during co-branding decisions. They also help to determine the brand's potential for future growth in category extensions.

While most existing research in IS focusses on descriptive and predictive properties of information networks (Oestreicher-Singer et al. 2012; Zhang et al. 2016), statistical analyses of the generative features of information networks have largely been ignored. Information networks, despite being highly valuable, are still a new and understudied topic in IS literature and merit further investigation in terms of - what drives link formation between entities?

Addressing this knowledge area, we first employ generative models, from social network analysis, to investigate the implicit connections responsible for the overall network structure. To accurately identify the underlying latent mechanisms driving co-interest between brands, we use social selections models, in particular, Exponential Random Graph models (Snijders et al. 2006). Among all current methods for modeling relational data, exponential random graph models (ERGMs) are generally known to be " *The most promising class of statistical models for expressing structural properties of social networks observed at a given moment in time*" – Byshkin et al.(2018). With their ability to address dependency as well as stochasticity among network ties, social selection methodologies provide inherent modeling advantage over extant regression models (Kim et al. 2016). Traditional regression models unrealistically assume that the entities are independently distributed – an assumption violated in network data, and also the very information we intend to capture for explaining the brand-brand associations. The results of the ERGM analysis reveal a mix of network and individual level brand characteristics responsible for the formation of links between brands; thereby also disclosing a set of latent brand characteristics that users determine while co-following brands on social media.

In the second section on eliciting brand associations, we demonstrate the brand network's ability to act as an effective business intelligence tool to deliver insights into potential co-branding opportunities in a close to real time setting. Besides the obvious tangible associations of a brand that relate to its physical attributes (e.g. mileage of a car, horsepower), there are another set of intangible associations that the go beyond the functional attributes of a brand and even transcend categories. Brand associations pertaining to categories, despite being vital for various managerial decisions (such as co-branding, licensing and brand extensions), are an understudied topic in IS and marketing literature. In this work, we provide a novel methodological tool to managers for studying the category-specific associations of their brands, as well as those of their competitors and allies, in a timely manner. Asymmetry among brand pairs is taken into account to calculate the directed networks weights that reveal which brands are more likely to benefit from co-branding alliance. It is also taken into account to calculate the 'transcendence' of a brand onto a new category. The directional associations of a focal brand into a new category essentially reveal what percentage a brand's fans are interested in a new category.

Further, the use of network-based constructs allows visualizing the co-branding opportunities at three different levels: brand-brand (b-b), brand – category (b-c) , category-category (c-c). The different level of analyses helps to answer why certain co-branding opportunities are more viable than others based on the audience interests of the focal brand. The analysis is conducted over time to study the shift in brand transcendence and investigate its impact on the co-branding opportunities. If critical associations to certain brands (or categories) have waned, this helps bring timely intelligence for managers to identify the problem and take action. Similarly, if new associations have formed during the course of time, it provides information on potential co-branding alliances. Managers can also study the network patterns over time to evaluate the effectiveness of their marketing campaigns or assess the impact of external events on their brand's associations in consumers' minds.

The third section of the dissertation outlines a big data approach for inferring brand positioning using implicit brand-to-brand networks. Understanding brand positioning is a crucial area of research in marketing. Traditional survey approaches, typically used for inferring brand positioning, can be cumbersome and costly. Other digital approaches relying on online user generated content and browsing history have been known to suffer from potential limitations such as biased content, substantial manual intervention and privacy regulations. Addressing this issue, we propose a network-based approach for inferring brand positioning using social connections of a brand on Twitter. We first extract the transcendence matrix of competing brands to highlight the perceived associations of brands into different categories. Then, connections within and across categories are studied separately to calculate brand perceptions. Brands that share strong consistent associations within category and are perceived to be *central*. Similarly, the across-category associations of a brand, in the transcendence vector, are used to discern the *distinctiveness* of brands in consumers' minds.

Unlike Dawar and Bagga (2015), who rely on traditional surveys to infer a brand's position in terms of centrality and distinctiveness, our automated approach provides an efficient and inexpensive way to create similar perceptual maps using publicly available social media data. Second, we provide a highly generalizable method that allows researchers to go beyond the numerical centrality-distinctiveness values and uncover the underlying graph structure that reveals how one's brand position is different from others. The inclusion of network derived measures allows researchers to assess a brand's position not only against a set of prespecified exemplar accounts, but to any other brand (or set of brands) in the ecosystem. To validate the effectiveness of our methodology, we compare the brand ratings from our automated approach with directly elicited survey ratings. We conduct the survey through Amazon Mechanical Turks, AMT, which has proven to a reliable source for conducting social science research (Buhrmester et al., 2011; Mason and Suri, 2012). We find consistently high correlations, $r > 0.8$ with $p < 0.001$, among all standard marketing demographic categories (gender, age and income).

Overall, the central theme of our work is to investigate the usefulness of brand networks for business decision-making, provide in-depth validation checks to support our claims and outline directions for future research. A research agenda, as shown in Figure 1, guides this dissertation. We start by discussing the research context, including relevant studies in the extant literature, limitations of previous approaches and how we overcome them in this dissertation. Integrating literature from three disciplines - IS, marketing and network science, we explain how implicit brand-brand networks incorporate consumer perceptions across a broad brand ecosystem and help address key issues in brand management. Moving to the methods section, we first define brand-brand networks, describe how they are generated from historical consumer-brand interactions on Twitter and lay out important network statistics. This is followed by ERGM analysis to understand the network connections and reveal significant brand characteristics responsible for the co-followership patterns.

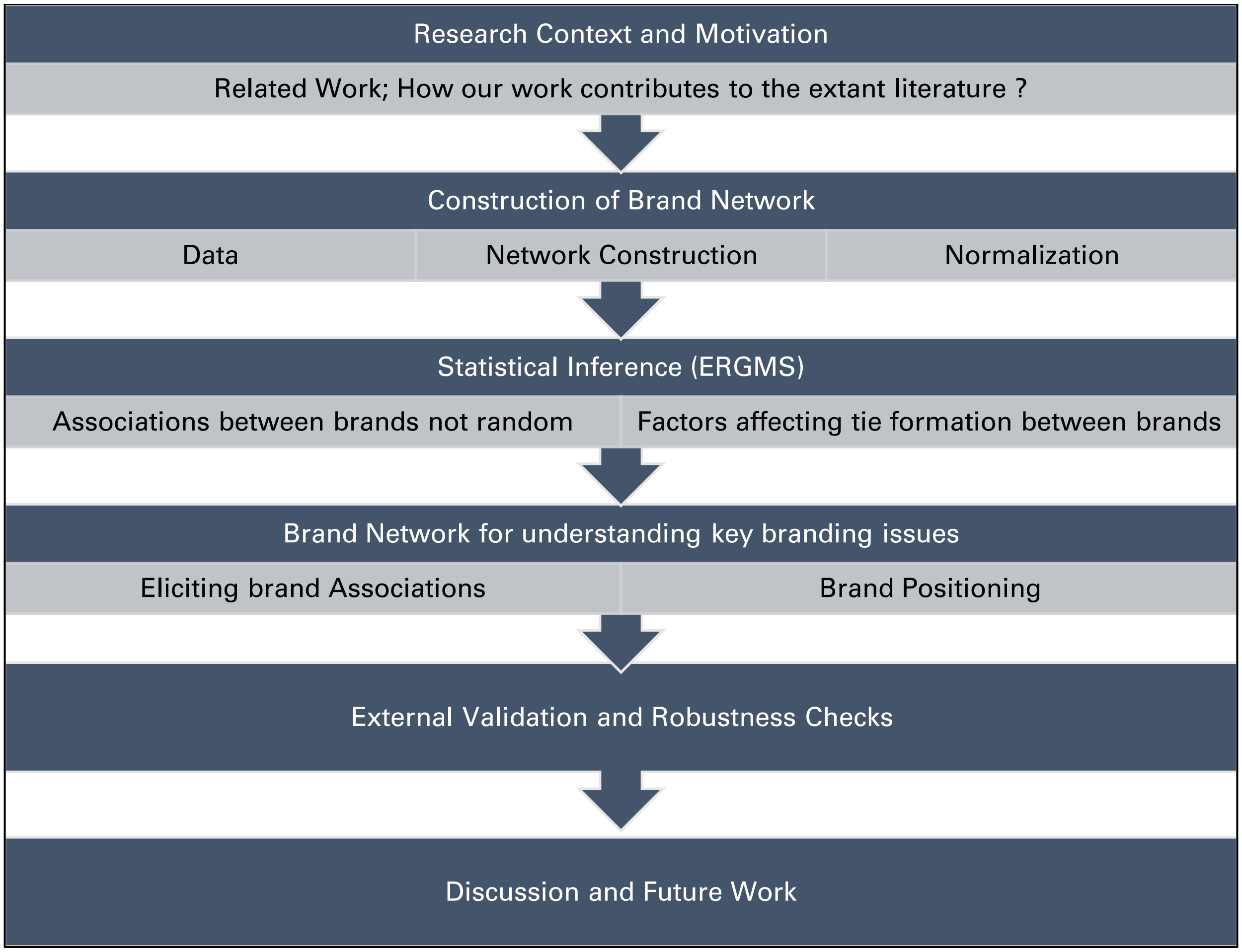


Figure 1. Research Agenda

Once the statistical significance of the connections has been established, we discuss how informational value can be derived from brand networks on social media. We first demonstrate the network's ability to elicit asymmetric cross-category brand associations over time. A new construct, transcendence, is defined to capture the diverse associations of brands onto new categories. Then we elaborate on how brand network variables, derived from the transcendence matrix, can reflect relative positioning of competing brands in consumers' minds. Specifically, we show how perceptual maps aid in visualizing brand positioning by creating a competitive landscape of brands along the two dimensions – centrality and distinctiveness. Other practical applications of the brand network are also discussed. First, competitor analysis helps to identify the closest competitors of a brand and uncover differential associations of each brand in the competing group. Second, employing standard community detection algorithm on the sub-network helps to reveal segments of brands with similar user interests. We discuss our validation methodology and provide various robustness checks to ensure consistency of results. Finally, we conclude by summarizing the implications of our work, noting its limitations, and suggesting avenues for future research.

II. LITERATURE REVIEW

a) Business Value of Social Media

The ongoing data deluge has fundamentally altered the decision-making process (Wu et al. 2019). There is strong piece of evidence on the positive impact of large-scale data and its value for generating considerable returns for firms (Brynjolfsson and McElheran 2019). More recently, the argument has been extended to data external to firms, specifically referring to information extracted from social media (Luo et al. 2013; Wu, Hitt and Lou 2019). More than 80% of the US population has a social media profile (Edison Research Triton Digital 2017), indicating the tremendous opportunity for firm owners to harness this digital information for strategic decision making.

Brand presence on social media is no longer a want, but a need (Forbes, 2018). If we solely rely on numbers, Facebook is the most popular social media website with over 1.52 billion active daily users compared to 326 million active users on Twitter (Data provided by Facebook and Twitter, 2018). Though the popular microblogging site, Twitter, lacks in overall monthly users compared to Facebook, it makes up in other areas that are crucial for businesses. Findings of Smith et al (2012) show user-brand interaction on Twitter to be more individualistic than that on Facebook. They find Twitter followers, compared to those on other social media platforms, to be more interested in hearing from the brand itself than from other online users. Second, according to Twitter, more than 70% of people who follow a brand on Twitter do so to get updates on latest products and discounts (Twitter, 2013). Research (Kim and Ko, 2012; Etter and Plotkowiak, 2011) shows that companies rely on Twitter's ability to help shape brand image by accessing a massive brand community through low-cost tweets.

Initial research on brand communities had a geographic constraint with main focus on the physical presence of the admirers of a brand (McAlexander et al, 2001; Muniz and O'Guinn, 2001). With the emergence of new technologies and advent of web 2.0, brands transcend geography and so do communities. This results in formation of brand communities on social media, that are a subset of a broader concept – 'virtual' or 'online' brand communities (Laroche et al, 2012). Marketers have found that brand communities established on social media lead to value creation (shared consciousness, brand use, brand loyalty) and engagement among community markers (Laroche et al, 2012). Overall the combination of social media and brand community (known as social media-based brand communities) allows inherent advantages to companies, such as vast reach, low cost and high communication efficiency (Kaplan and Haenlein, 2010). The digital footprints of consumers on these platforms create a rich data source for branding purposes such as deconstruction of consumer behavior, campaign automation and targeted advertising. Also, with the rise of 'big data' technologies businesses are able to access and gather this limitless consumer information on web without having the need to worry about storage and processing capabilities. Leveraging large scale online data for consumer research, helps overcome challenges faced by traditional survey and focus group-based methods that are known to be expensive and cumbersome, and also limited in terms of reach across consumers and brands.

Social network-based communities are now being used to promote brands (Fournier and Lee 2009), with online brand followers proving to be more loyal and committed than ever (Jahn and Kunz 2012). Regardless of whether the fans choose to post content or not, the ‘mere virtual presence’ of followers on a brand’s fan page has shown to reflect and impact ‘brand image’ (Kuksov et al. 2013). This line of reasoning is further supported by survey research (Adobe’s Social Intelligence Report 2014) that shows that there is a real monetary value to having Twitter followers; approximate revenue per visit from Twitter is $0.62.

When users join brand communities on Twitter, it provides evidence of their voluntary affiliation with that entity (Culotta and Cutler, 2016). Survey research has shown that users follow brands on social media with the intention of purchasing and knowing more about their favorite brands (Forbes, 2018). This user-brand association on social media, established through followership activity, can be interpreted as an expression of affinity (Kuksov et al, 2013;

Naylor et al, 2012). Alternatively, one can also view this relationship through the lens of homophily - a concept in which people tend to associate with those who are similar to them in socially significant ways (McPherson et al, 2001). This line of argument is further supported by studies in consumer research (Berger and Heath 2007; Childers and Rao 1992; Escalas and Bettman 2003) that show a strong relationship between brand image and characteristics and identities of a brand's followers.

Exploiting the social structure of a brand's online community, Culotta and Cutler (2016) find the information contained in co-followership data to be a reliable source for inferring attribute-specific brand perceptions. A strong correlation between their automated approach and directly elicited survey ratings demonstrates the potential of mining Twitter brand communities to enhance online decision support for managers.

More recent research in IS uses the number of Twitter followers as a measure to understand consumer interest in a company and find it highly valuable (He et al. 2017). Hoffman and Fodor (2010) find the number of followers to be one of the most useful metrics for gauging brand engagement and awareness on micro-blogging platforms such as Twitter. Mining the social structure of a brand's follower base on social media can help capture useful information from all brand fans, regardless of whether they create little or no user content (Culotta and Cutler 2016).

b) Information in Networks

Exploiting the ongoing digital data explosion, constructing and analyzing implicit networks for scaling business research has garnered increasing attention from researchers (Provost et al. 2009; Zhang at el; 2016). Specifically, in the area of audience selection, Zhang et al. (2016) show that implicit networks, established through aggregated interests of people on Facebook, are useful for online brand advertising. A related idea is used in Provost et al. (2009) for inferring brand affinity from co-visitation patterns on social network pages.

Unlike conventional social network studies, digital networks of this kind do not involve direct interaction between the participating entities (Arriaga and Levina 2008). Instead, links forming the network are more tacit - an outcome of shared preferences. Sundararajan et al. (2013) point out the importance of these tacit connections as “information” relevant for decision making, an idea that has been previously studied under the domain of collaborative filtering. Other potential advantages of implicit networks include the ability of these digital artifacts to condense the high dimensional preference space of millions of digital consumers into a reduced form, which is more amenable for research and managerial purposes (Sundararajan et al. 2013).

While most existing research in IS focusses on descriptive and predictive properties of information networks (Oestreicher-Singer et al. 2012; Zhang et al. 2016), statistical analyses of the generative features of information networks have largely been ignored. Generative models have the ability to explain what constitutes the tacit connections in the network and whether these connections arise due to randomness or specific consumer choices (Kim et al. 2016). The literature on network inference (Robins et al. 2007; Chatterjee and Diaconis 2013) is particularly well suited for understanding questions on the structural properties of information networks. Network inference can help explain how consumer choices, leading to links in the implicit network, reflect deeper latent constructs such as specific node characteristics (Sundararajan et al. 2013). In fact, researchers are beginning to see that network inference approaches, with the capability of handling statistical dependencies, can provide a better understanding than traditional data inference techniques (Martens and Provost 2011).

Prior research in management and IS has typically employed regression models such as probit or logit to understand network formation among firms (Chung et al. 2000; Gulati and Gargiulo 1999). However, for a number of reasons, these methods can be problematic, including the strong independence assumptions violated by network data (Kim et al. 2016). Past studies have attempted to overcome these problems by employing correction techniques such as clustering of standard errors and elaborate weighting methodologies (Stuart 1998; Barnett 1993; Gulati 1995). Even after accounting for these issues, standard regression models fail to incorporate local network effects such as triadic relations and dyadic influences (e.g., homophily between firms due to similar attributes) in the model (Contractor et al. 2006).

In this paper, we employ a class of social selection models, in particular Exponential Random Graph Models (ERGM), to confirm the statistical relevance of brand-brand associations and to identify the factors driving network formation. This family of models was first introduced in the mid-80's (Frank & Strauss, 1986) and later popularized in the 90's by a variety of researchers (Snijders et al, 2006). Recently ERGMs have been adopted by a small, yet growing, number of studies in social science discipline to study resource complementarity (Lomi and Pallotti 2012), inter-organizational dependence (Howard et al. 2017) and strategic alliances (Ghosh et al. 2016).

These probabilistic models allow inferences about whether certain observed network structures are more likely to occur than expected by chance. For instance, links in networks may either arise due to endogenous structural effects (e.g., triadic closure – if a node has strong ties to two neighbors, then these neighbors must have at least a weak tie between them) or exogenous actor level attributes (e.g., homophily – birds of a feather flock together). Such models can thus be valuable for understanding the local social processes responsible for the observed network structure (Snijders et al. 2006). In this study, the use of ERGM model helps to reveal a mix of network and individual level brand features responsible for the formation of links between brands, correspondingly disclosing a set of latent brand characteristics that users determine while co-following brands on social media.

c) Market Structures

Studies involving market structure analysis focus on uncovering what brands (or products) are perceived to be similar (or dissimilar) in consumers' minds; and help shape strategic decisions such as tracking competition, identifying substitutes, pricing and product re-designing (Kannan and Sanchez 1994; Urban et al. 1984). Typically, two types of digital footprints have been tapped for market structure analysis: User-generated content, UGC, (Lee and Bradlow, 2011; Netzer et al, 2012) and internet browsing history (Kim et al, 2011; Damangir et al, 2018). Keeping the richness in quality of UGC aside, text mining of user-generated content requires extensive manual tuning, with most of the analyses being heavily context specific and not applicable across platforms (Culotta and Cutler, 2016). Das and Chen (2007) compare several sentiment classification algorithms and obtain accuracy ranging from 25%-40% for out of sample validation. The heavy noise in text data arises due to a very small percentage of internet users choosing to post content online (Gao et al, 2015). Analysis of these limited opinions can lead to biased results.

Another potential drawback with this method includes bias in information as majority of online users tend to post in extreme bimodal situations – either very satisfied or very dissatisfied with the brand (Goes et al, 2014). Even after balancing the polarity in the data, extracting meaningful brand associations from online user-generated content is not a straightforward text mining process. Specifically, the information pertaining to the association of a brand across category dimensions, such as luxury, sports, travel, technology etc., is rarely found in individual users' comments. This substantially limits the data available for analysis and hampers the use of any text mining algorithm to infer specific brand associations for market structures.

Apart from UGC, many academic fields including computer science, information sciences and marketing examine the internet browsing behavior to study brand preferences of individuals (Moe, 2003; Ringel and Skiera, 2016). Though current big data technologies allow easy collection of clickstream data, such information may suffer from potential privacy limitations (Bucklin and Sismeiro, 2009). Tracking accurate user activity is also difficult with increasing number of users browsing on different platforms (e.g. web or mobile). Overcoming these major data concerns, brand networks provide a new privacy-friendly source for uncovering brand co-considerations patterns among consumers minds. Brand to brand relationships, extracted from aggregated preferences of large number of Twitter users, capture consumer co-interest in a rather novel manner.

Kannan and Sanchez (1994) note that most market structure models aim to identify submarkets where within-group interaction is stronger than the across-group competition. A majority of studies in this area have looked into brand switching data (Kannan and Sanchez 1994; Erdem 1996), website co-visitations patterns (Kim et al. 2011; Ringel and Skiera 2016) and brand co-mentions (Netzer et al. 2012) to uncover relationships, typically within a single category. While most existing studies focus on within-category competition, the notion of looking into cross-category effects has largely been overlooked. Elrod (2002) believes that the focus on narrowly defined categories is largely due to limited data availability and tough-to-scale modeling techniques.

Today, even with the ready availability of big data and advanced computing power, extant modeling approaches struggle to identify drivers of market structures that reflect both competition as well as complementarity. This work aims to bridge this gap by exploiting network inference models to highlight the statistical relationships between brands of the same as well as different categories. The brand network, itself, is viewed as a market structure of brands, arising due to co-interest patterns of digital users across categories. To the best of our knowledge, this is one of the first few studies to draw statistical analyses on online market structures that span across categories.

d) Importance of Cross-Category Brand Associations

Consumers regularly develop categorical associations to understand brands and to make related choices (Joiner and Loken, 1998). Categories are socially constructed boundaries that segment the market space into groupings that are perceived to be similar along certain criteria (Brexendorf and Keller, 2017). Traditionally, boundaries have been defined by the portfolio of products, or categories, for which consumers would find the given brand relevant (Brexendorf and Keller, 2017). These boundaries are not immutable; instead, they are elastic and stretchable. Cutright et al (2013) mention that consumer perceptions of categories are pliable and influenced by a variety of personal and contextual factors.

As such, not only firm-controlled, but also competitor-driven external marketing activities define a brand's associations and limit its scope in people's minds (Keller and Lehmann, 2006). Particularly, activities related to people, places and other related brands may impact the perception of a brand in consumers' minds (Keller, 2003). For example, by sponsoring the New Zealand All Blacks rugby team, Adidas got access to the desirable sports brand associations and a new target audience for its product range (Motion et al, 2003). As consumers regularly update their categorical associations about brands (Joiner and Loken, 1998), it crucial to identify these associations in a timely and cost-effective manner (Batra et al, 2010).

The marketing literature acknowledges the importance of categorical associations for determining extensions, licensing and co-branding deals (Keller, 2003; Ratneshwar et al, 1993); however, little has been done empirically to address this understudied, yet crucial, area of research. Our study provides a new methodological tool to understand cross-category brand associations, not as one identified by management apriori but as one perceived by the direct activity of the users on social media. Furthermore, depending upon the marketing goals, these brand associations can be visualized at three levels: category – category (c–c), brand – category (b–c) and brand – brand (b–b). The different levels of analyses help to answer questions on why certain co-branding opportunities are more viable than others based on audience interests of the focal brand

e) Social Network Analysis for Branding

While extensive research has been done in the area of brands and networks separately, few have employed network analysis tools to understand key branding issues. Table I presents the unique positioning of our research compared with previous network studies on branding. Focusing mainly on within-category connections, Netzer et al (2012) create competitive market structures of car brands. Using survey approaches, Dawar and Bagga (2015) obtain within-category maps for centrality and distinctiveness. Ringel and Skiera (2016) develop mapping methods to visualize large market structures within a single category (LED-TV products). In contrast to ours, none of these studies focus on asymmetric cross-category brand associations as way to identify co-branding opportunities.

Table I. Comparison with Previous Studies

	Netzer et al (2012)	Ringel and Skiera (2016)	Culotta and Cutler (2016)	This study
Objective	Create competitive structure maps using text mining and network analysis	Develop mapping methods for visualizing complex market structures.	Propose a methodology for inferring attribute-specific brand perceptions	1) Statistical analysis of generative features of a brand network 2) Inferring asymmetric cross-category brand associations over time 3) Perceptual Mapping
Data	Co-mentions on online discussion form	Website co-visitation patterns	Co-followership data on Twitter	Co-followership data on Twitter over time.
Output	Market structures (within a category)	Mapping solutions for large complex market structures (within a category for 1000+ products)	Perceptual maps for a set of pre-defined exemplars.	1) Cross-category association maps at three different levels: category-category, brand-category and brand-brand 2) Perceptual Maps
Asymmetry	No	Yes	No	Yes
Segmentation	Yes	Yes	No	Yes
Competitor analysis	No	No	No	Yes
Dynamics	No	No	No	Analysis presented over time.

Typically involving analysis within a single category, these studies use brand co-mentions (Netzer, 2011) and website co-visitation patterns (Ringel and Skiera, 2016) as a proxy for similarity or association among brands. In our case, relying on followership data offers three modelling advantages – 1) Eliminates the use of complex text mining algorithms that can be highly biased due to polarized content on social media (Goes at al 2014) 2) Avoid clickstream data which has many limitations ranging from issues in tracking accurate user activity to strict data privacy regulations (Bucklin and Sismeiro, 2009; Malhotra et al, 2004) 3) Effectively scale to large sets of brands and resulting network structures that reflect the activity of diverse group of users.

The use of co-followership data to infer brand co-associations follows the recent work of Culotta and Cutler (2016). While Culotta and Cutler (2016) seek to derive perceptual attribute ratings from Twitter followership data, the focus of our work is different – statistical analyses of online market structures and deriving asymmetric cross-category brand transcendence over time. Further, unlike Culotta and Cutler (2016), our large-scale network approach does not require any supervised knowledge on exemplars and uses categorical affiliations of brands to infer brand perceptions. Specifically, the inclusion of network derived measures allows us to assess a brand's position not only against a set of prespecified exemplar accounts but considering any other brand in the ecosystem; revealing both within-category competition and across-category complementarity. Our method is also highly transparent, allowing one to go beyond numerical brand ratings and uncover the underlying dynamic graph structure that reveals how one's brand position is different from others over time.

Further, the extent of brand associations, in terms of within and across category associations, are used to identify *central* (C) and *distinctive (D)* brands in each category. To effectively map the brand positions along the C-D scale, standard perceptual techniques are used. Since the early 70's, perceptual maps have been the analytical foundation for examining relative positioning of competitive brands (Green et al. 1989; Shocker and Srinivasan 1979). Henderson's (1998) work on brand associative networks highlights the importance of brand-brand associations, and their application to understanding perceptual market structures. When a brand manager studies a network of brands that are competitors in a product category, network centrality helps in identifying brands that consumers perceive as core in a category (Henderson, 1998). Utilizing this theoretical framework, we use perceptual maps to plot network centrality measures that implicitly reflect a brand's position. With traditional survey methods involving substantial trade-offs between completeness, feasibility and cost (Aaker, 1996), secondary social media data offers an efficient and inexpensive way to generate near real-time estimates of brand ratings.

III. NETWORK CONSTRUCTION

The key contribution of this dissertation is to provide an automated and scalable approach for brand management using information networks obtained from large-scale user brand interactions on Twitter. In this section, we introduce implicit brand–brand networks, describe how they are generated from historical consumer-brand interactions on Twitter and lay out important network statistics. Our proposed framework for building a brand network is composed of three main phases. The first phase consists of collection of data on top brands. Using Twitter as our platform for analysis, we collect followership data from online brand communities across two time periods – 2017 and 2020. The second phase is to generate the weighted brand network based on common user activity. We notice the brand networks to be strongly connected with the number of common followers ranging from a few hundred to more than a million. As connections based on too few users may not indicate significant affinity, we extract the significant edges using a statistical filtering algorithm. Rather than using an arbitrary threshold for removing edges below a certain weight, we use the Disparity Filter algorithm (Serrano et al, 2009) which identifies significant edges while preserving the multiscale nature of our network. The third phase is to normalize the brand networks and prevent large brands from dominating the analyses. Detailed robustness checks are conducted for each method to ensure consistency of results across a range of functions.

a) Data

Drawing from the notion that the social signal of 'who follows a brand' provides a strong reflection of brand image (Naylor, 2013; Kuksov et al, 2013; Culotta and Cutler, 2016), we use a set of 507 brands' Twitter accounts as the basis for our analysis. The most active Twitter brand accounts are selected based on the followership data on social media directory - fanpagelist.com. Twitter's public API is used to collect the list of followers for the year 2017 and 2020. We manually verify that all Twitter accounts correspond to the official brand and discard any brands having less than 1000 followers. Overall, the data consists of brands from major categories - Luxury, Retail, Automotive, Sports, Technology, Dining, Food, Lodging, Travel, Cruise, airlines and Beer. Each brand is assigned to a specific category based on the basic (or superordinate) category level analyses[1] (Loken and Ward, 1990). To prevent bots (or spam users) from influencing the network analysis measures, all Twitter brand accounts included in the analysis have gone through a manual audit on the

[1] Superordinate and subordinate category level analyses – The superordinate category is the highest umbrella category containing diverse exemplars with low degree of class inclusion (for instance, beer and dining). Going one level deeper, the subordinate level contains exemplars that are comparable across specific attributes (for instance, fruity beers and American dining). In our work, we use the basic (or superordinate) category levels with comparisons across brands and not specific product types.

audience intelligence website SparkToro[2]. Brand accounts with unusually high spurious activity are removed from the analysis. The nature of our approach (i.e. leveraging co-followership data), also helps to minimize the effect of any spurious followers on our final results. As Culotta and Cutler (2016) mention, by aggregating millions of social links between users and brands on Twitter, one is able to avoid the noise arising from spurious follower connections; and generate meaningful brand insights based on co-preferences of a large number of digital users. The normalization approach, discussed later, further helps to negate the effect of any fake followers on our final measures.

[2] Sparktoro is a software company that provides intelligence reports on Twitter brand accounts. Their algorithm is designed to identify accounts that fall into one or more of the following buckets – spam accounts, bot accounts, propaganda accounts and inactive accounts .

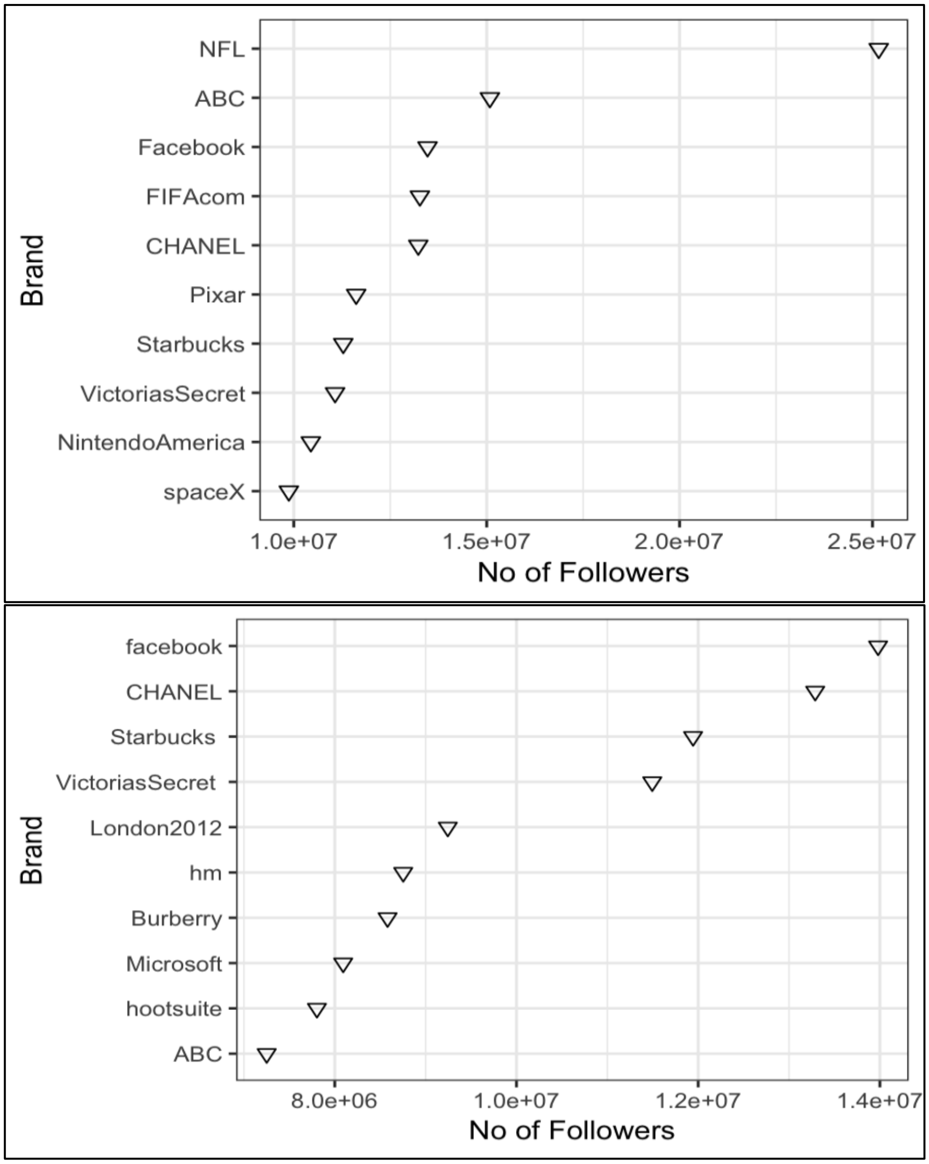


Figure 2. Top 10 Twitter brand accounts by number of followers in the year 2020 (top) and 2017 (bottom)

Figures 2 shows the top-followed Twitter accounts for the years 2017 and 2020 with brands like NFL, ABC, Facebook and Chanel leading the charts. The distribution of followers across categories is also shown in Figure 3 (below). We notice brands belonging to the luxury and sports categories to have the largest followers, followed by airlines, apparel and restaurants. Though, on an average, the number of followers has increased from 2017 to 2020, the beer and airlines category show a mild drop in numbers.

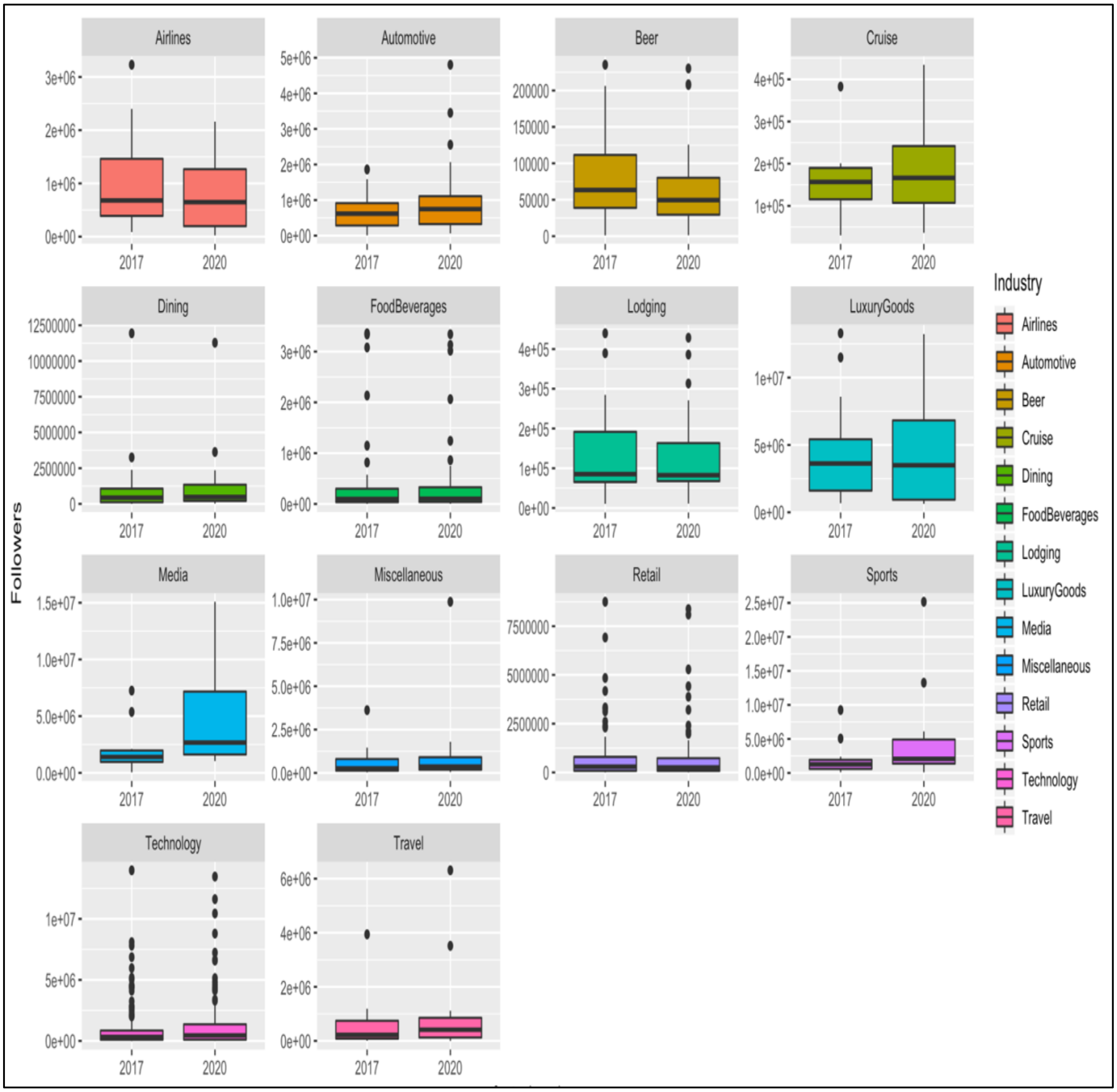


Figure 3. Number of followers per category over time

Within brands of the same category, the size of brand communities can also vary over a large range, as seen in Figure 4. For example, in the automotive industry, the size of brand communitues ranges from a few thousands (Lincoln, Infiniti and Acura) to more than a million followers (e.g., Mercedes, BMW, Porsche and Audi). In the following section on network generation, we employ a normalization procedure to account for the varying brand community sizes.

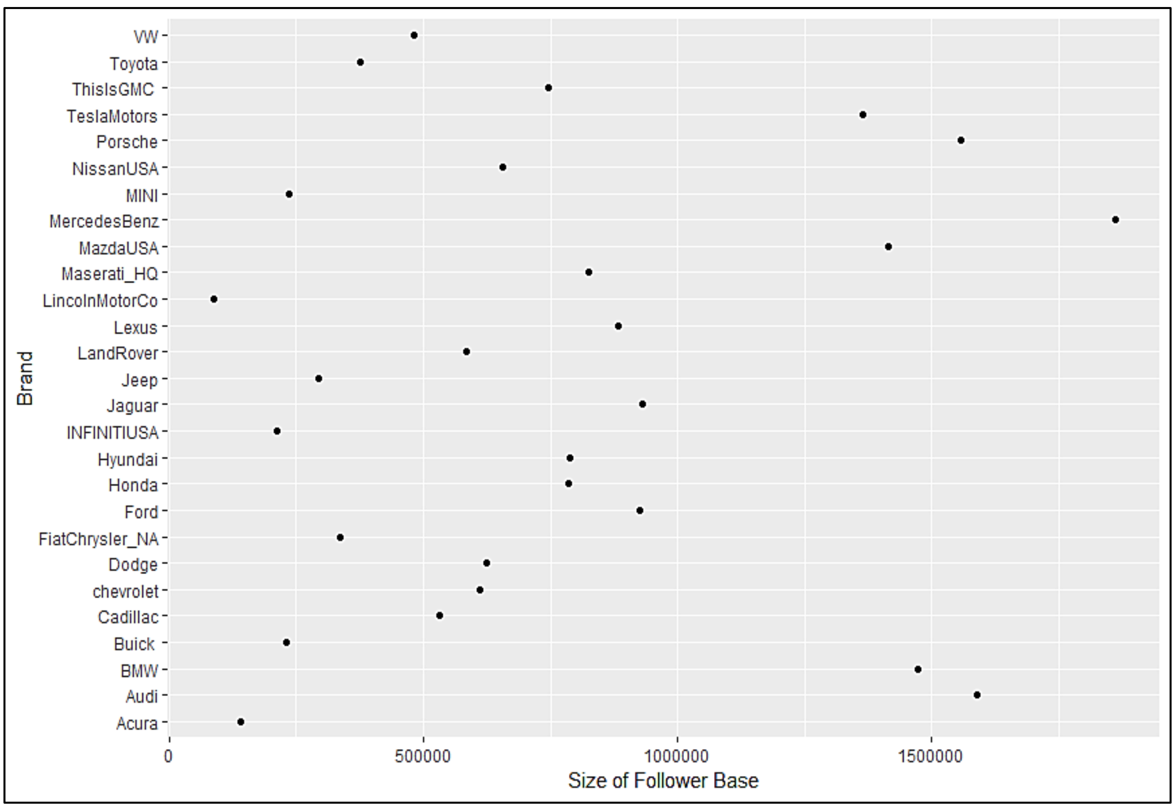


Figure 4. Distribution of followers in the automotive industry

b) Network Generation

The next step is to extract the common followers between all brand pairs. The raw brand network is defined as $< b_i\ b_j, w_{ij} >$ where b_i and b_j are the individual brands (or nodes) and w_{ij} represents the common followers between any two brand pairs b_i and b_j. If F_i and F_j represent the list of Twitter accounts following brands b_i and b_j, then an edge between two nodes is created if and only if $F_i \cap F_j > 0$. Alternatively, the edge list can be represented as a weighted adjacency matrix A_{ij} where:

$$A_{ij} = \begin{cases} w_{ij} & \text{if brand i and brand j are connected} \\ 0 & \text{otherwise} \end{cases}$$

Need for Network Filtering

Overall, two brand networks are extracted – one for 2017 and the other for 2020. We find the original brand networks to be highly dense with common followers between almost all pairs of brands. The range of common followers varies from a few hundred to more than a million users. While it is generally possible to work with such dense networks, valuable information may be lost due to redundancy generated by the overwhelming number of connections (Serrano et al, 2009; Radicchi et al, 2011). Further, connections based on too few followers may not indicate significant connectivity. Given this wide heterogeneity in raw edge weights (that is, number of common followers) extracting the truly relevant brand-brand connections is the next logical step.

A common way to extract the relevant network structure is through applying a global threshold which would simply remove the edges with weights below an arbitrary cut-off. This, however, can destroy the multi-scale properties of the brand network. Instead we use Disparity Filter (Serrano et al, 2009), a filtering algorithm for multiscale networks, to obtain a reduced but meaningful representation of the network. The Disparity Filter extracts the relevant backbone of a complex network by identifying the statistically significant edges with respect to a null model. The statistically relevant connections are the ones that satisfy -

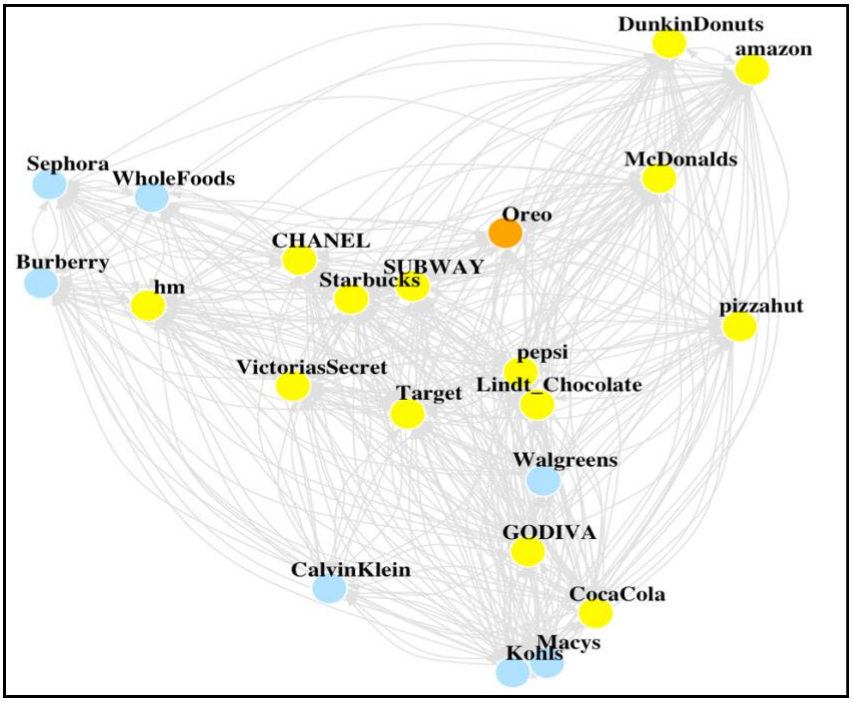
$$\alpha_{ij} = 1 - (k-1)\int_0^{p_{ij}} (1-x)^{k-2} dx < \alpha$$

$p_{ij} = \frac{w_{ij}}{\sum_j w_{ij}}$ are the normalized weights. For a certain of nodes of degree k, under the null hypothesis the normalized weights p_{ij} are uniformly distributed across k-1 points in the interval [0,1]. All edges α_{ij} that represent a statistical deviation from the null model, $\alpha_{ij} < \alpha$, are identified as statistically significant. In this way small brands, with low number of common followers, are not ignored during network reduction. We use the commonly specified significance level α = 0.05 to extract the important connections in the brand networks. The basic descriptive statistics of the filtered networks are given in Table II.

Table II. Descriptive statistics of the filtered networks for the year 2017 and 2020

Property	*Meaning*	*Network statistics for 2017*	*Network statistics for 2020*
Number of nodes	Number of brands	507	507
Number of edges	Number of edges	14743	14834
Density	Ratio of number of edges present to the maximum number of edges possible. Value ranges from 0 to 1.	0.11	0.11
Average degree	On average, the number of connections a brand exhibits.	58	59
Maximum degree	Maximum number of connections a brand exhibits.	481 (Starbucks)	471 (Starbucks)
Minimum degree	Minimum number of connections a brand exhibits.	3 (Tag Heuer)	7 (Timex)

The filtered network for 2017 consists of 507 brands and 14743 edges with an average degree of 58. For 2020, the statistics are similar with the new network consisting of 507 brands and 14834 edges and an average degree of 59. Naturally, while there may be some brand-brand ties that remain intact from 2017 to 2020, there could be others that vanish (or newly form) in 2020 . For example, in Figure 5, we see how the food brand Oreo retains its ties with Coca-Cola, Subway, McDonalds, Lindt Chocolate (and others) from 2017 to 2020, and also forms new ties with Disney and Dominos in 2020.



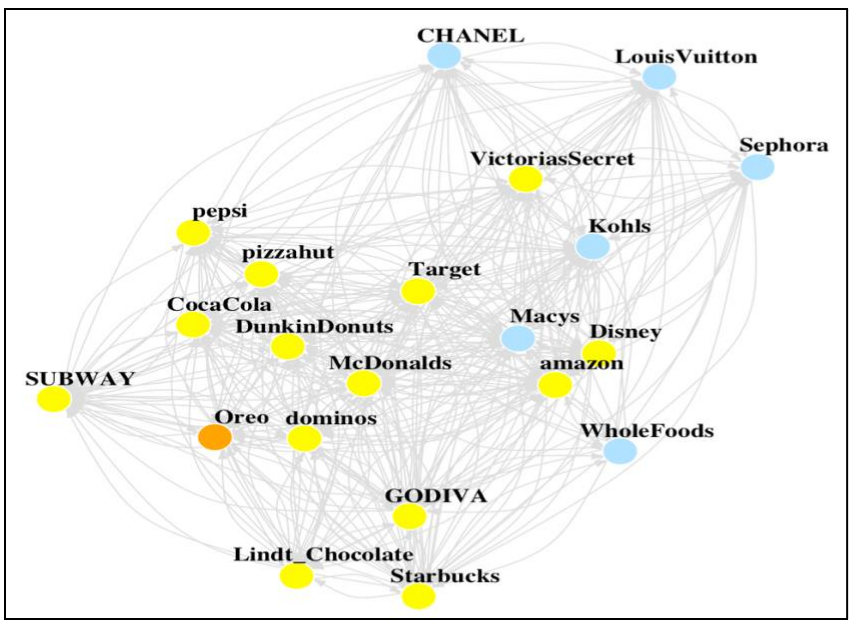


Figure 5. Subgraph showing some (not all) connections of the food brand, Oreo in 2017 (top) and 2020 (bottom).

Brands marked in yellow are directly connected to Oreo while the ones marked with blue are at a 2-degree separation from it.

Sensitivity to Network Pruning

While we have used Disparity filter to obtain a reduced but meaningful representation of the original brand network, there are few other information filtering algorithms (global threshold and Global Statistical Significance (GloSS) filter) available in the network science literature. Table III revisits these alternative methods and supports our reasoning for choosing the Disparity Filter. Compared to the Global Threshold technique that destroys the multiscale nature of networks by removing edges below a certain threshold, Disparity Filter and GloSS preserve the heterogeneity in edge weights and do not ignore small nodes (and edge weights). Further, as Radicchi et al (2011) find both GloSS and Disparity filter to produce identical results on application to real-world networks, we continue with the analysis using Disparity Filter.

Table III. Comparison of alternative network filtering algorithms

Global Threshold	Disparity Filter (Serrano et al, 2009)	Global Statistical Significance (GloSS) Filter (Radicchi et al, 2011)
Removes all connections with edge values below a given threshold.	Locally identifies the statistically relevant weights at the node level.	Globally identifies the statistically relevant weights at the edge level.
Destroys the multiscale nature of the network by ignoring edges below a certain scale	Preserves the multiscale nature, small nodes and edges are not ignored.	Preserves the multiscale nature, small nodes and edge values are not ignored.
	On application to real world networks, performance of Disparity and GloSS filter are very similar (Radicchi et al, 2011). Both filters mostly select the same edges, hence the reduced network structures are identical (Radicchi et al, 2011). Global Threshold technique, on the hand, fails to capture the main characteristics of the original network.	

To ensure consistency in information across different significance levels when using the Disparity Filter, we extract another set of reduced networks using α = [0.01, 0.1]. The summary statistics of the networks is given under Table IV. We test the correlation among the reduced networks using Quadratic Assignment Procedure (QAP), which is widely popular in network studies for comparing graphs (Baker and Hubert 1981; Krackhardt 1988; Borgatti et al. 1999). The high correlation values, $r > 0.9$, indicate that the reduced networks are essentially similar in terms of underlying interactions and successfully capture the meaningful brand-brand relationships.

Table IV. Sensitivity to network pruning

Year	Original Network	Red. net (α = 0.01)	Red. net (α = 0.05)	Red. net (α = 0.1)
2017	507 b, 124797 e	507 b, 6909 e	507 b, 14743 e	507 b, 22261 e
2020	507 b, 128271 e	503 b, 6580 e	507 b, 14834 e	507 b, 23249 e

Network summary statistics (#brands b and #edges e) at different significance levels

	Original Network	Red. net (α = 0.01)	Red. net (α = 0.05)	Red. net (α = 0.1)
Original Network	1	0.93	0.98	0.98
Red. net (α = 0.01)		1	0.96	0.95
Red. net (α = 0.05)			1	0.98
Red. net (α = 0.1)				1

QAP Correlation statistics for the original and reduced networks - year 2017. Similar analysis has been conducted for the brand network obtained in 2020 and high correlation is obtained between the original and reduced networks.

Overall, the brand network structure holds consistent across a range of alternative functions and is subjected to robustness checks before deriving any informational value. For the rest of the paper, the filtered network, at α = 0.05, is used for analysis as well as application. In Figure 6, we show the filtered information network, where nodes and links have been colored by brand categories. We use the Fruchterman Reingold layout, a force-directed layout algorithm, that groups nodes based on high inter-connectedness. It is noticeable that brands of the same category tend to be more inter-connected than those who are not. As brand-brand links arise from co-followership, this pattern really means that a large number of Twitter users tend to follow multiple brands of the same category. This observation is further statistically tested using ERGM analysis in Section IV.

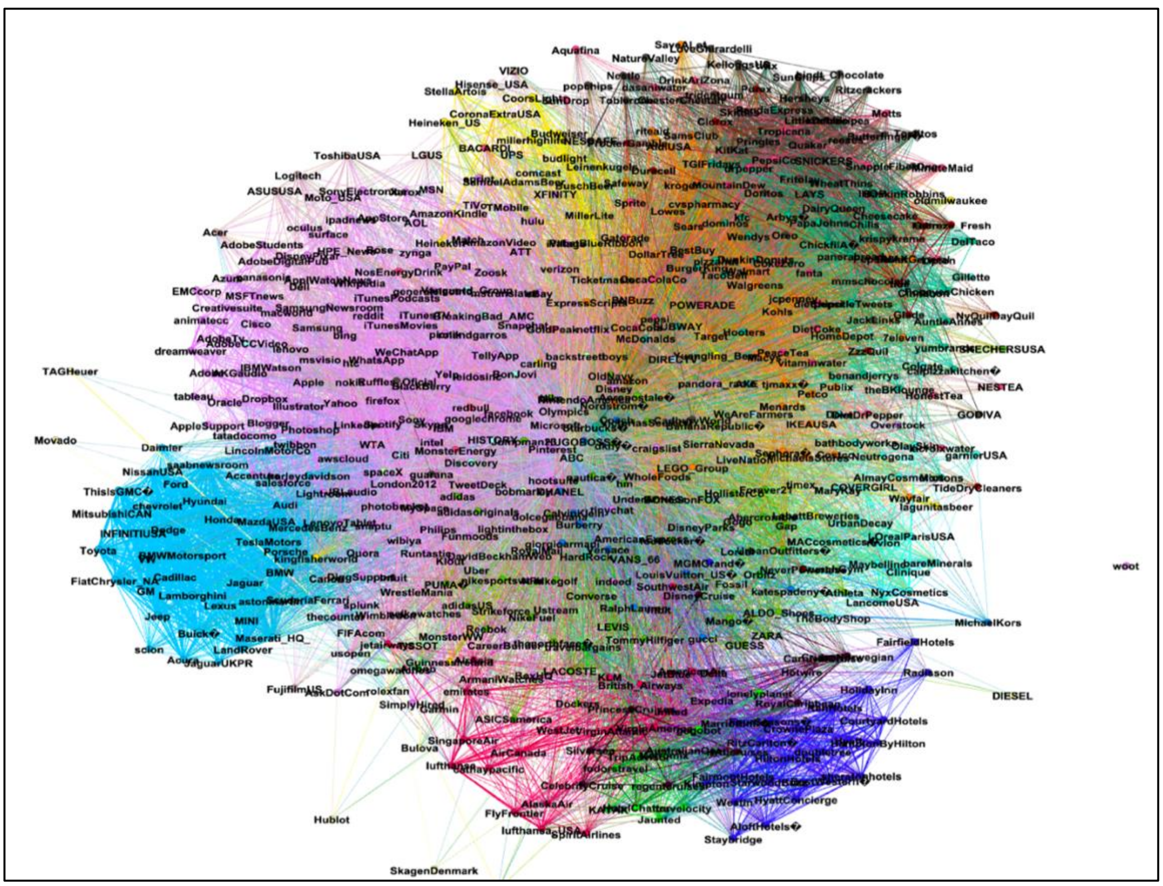


Figure 6. Brand Network for the year 2017

c) Normalization of Edge Weights

As noted earlier, brand community sizes can vary both within and across category. Naturally brands with large brand communities, e.g., Chanel, Microsoft and Starbucks, tend to have more common followers than those with smaller brand communities. Normalization of edge weights is required to account for the varying brand community sizes.

Symmetric Normalization

We first use a popular and empirically successful measure (Pan et al, 2010; Culotta and Cutler, 2016), Jaccard index, to compute the normalized edge weights. The Jaccard Index measures similarity between two sets by dividing their intersection with the union.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

By including the size of brand communities in the denominator (|A| and |B|), the Jaccard index prevents large brands from dominating the network analyses measures. The normalized weights measure the relative strength of consumer co-interest between two brands. Figure 7 shows how normalized edge weights provide a more suitable measure of brand-brand affinity than raw edge scores. Without normalization (top two edges), the edge weight between Burberry and Chanel is higher than that of Honda and Hyundai, even though almost all of

Honda's brand community follows Hyundai. Here, normalization takes care of the popularity bias.

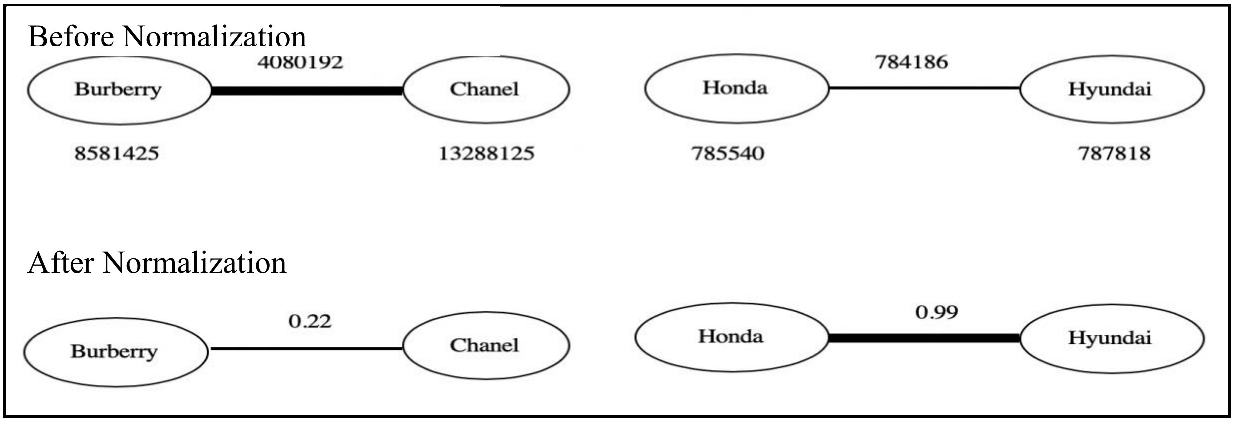


Figure 7. Symmetric Normalization of edge weights

While we choose to normalize the edge weights using Jaccard, there is another method to normalize weights in a network, i.e. conditional probability. We, next, normalize the raw edge weights using conditional probability and investigate how sensitive the results are to these choices in Appendix A.

Asymmetric Normalization

Association asymmetry occurs when the degree of association between any two brands is not the same i.e. the associations from A to B may not always equal the associations from B to A (Lei, Dawar and Lemmink, 2008; DeSarbo and Grewal, 2007).). Ignoring the directionality of brand associations can lead to incorrect estimates about consumer brand knowledge. Relating this directly to marketing activities, Farquhar and Herr (1993) suggest that brand-building activities should focus on strengthening the outgoing directional associations from the brand to others, whereas brand-leveraging activities should focus on the incoming directional associations to the brand from others.

We observe many cases of associative asymmetry in our brand network and use conditional probability to account for such scenarios. For instance, in figure 8, we notice that Nike and Adidas share a noticeably high number of common followers in 2020 (close to 930000). Though this number is evidently very high, the nature of relationship between the two brands may not be so direct. From Adidas's perspective, almost all of its users are interested in Nike and the outgoing directional strength is close to 1. But for Nike, with less than half of its users interested in Adidas, the outgoing directional strength is comparatively much lower, 0.1. Incorporating directionality in the network helps to uncover this crucial piece of information, that is not revealed in a simple undirected weighted network.

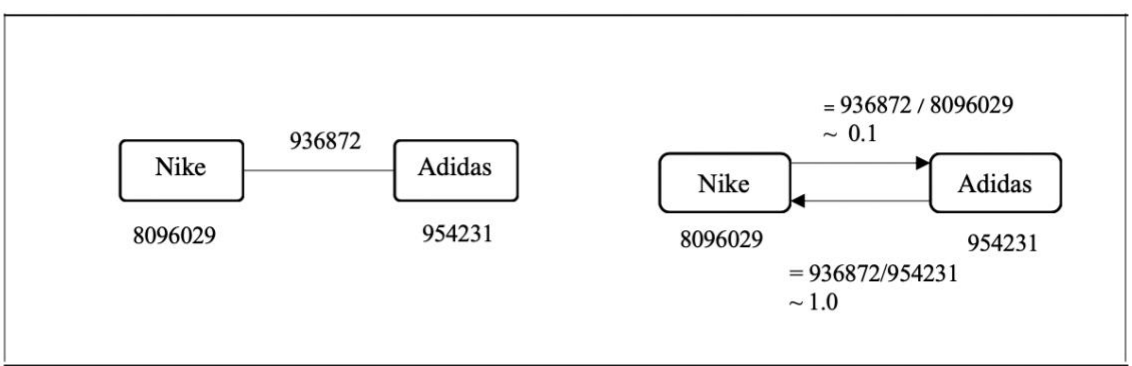


Figure 8. Associative Asymmetry

Mathematically, the conditional probability measure calculates the normalized links between any two brands A and B as –

$$P(A \cap B) = \frac{|A \cap B|}{|A|}$$

where the numerator $|A \cap B|$ is the number of common followers between brand A and B, and denominator $|A|$ is the number of followers of the focal brand. The normalization approach also helps to negate the effect of any fake followers on our final measures. The common followers are divided by the number of followers of the focal brand. This ensures that any inflation in the numerator, A∩B, caused by fake accounts is negated by the denominator, number of followers of A.

IV. STATISTICAL ANALYSES OF THE INFORMATION NETWORK

While extensive research has been done in the area of brands and networks separately, not many have employed information networks, in particular implicit networks, for brand management. Previous studies on brand associative networks outline the importance of brand associations, and show how well-known descriptive techniques in graph theory can be used to leverage these associations for application purposes (Henderson et al. 1998; John et al. 2006). Popular techniques in graph theory that aim to describe structural features of networks include degree distributions, centrality, communities, assortativity and others (Robins et al. 2007). In a lot of cases, however, one might need to go beyond these descriptive features and learn a well-fitting model that best explains the local processes responsible for network formation.

Information networks, despite being highly valuable, are still a new and understudied topic in IS literature and merit further investigation in terms of - what drives link formation between entities? Oestreicher-Singer and Sundararajan (2012) have previously shown that information embedded in implicit networks has social and economic impacts. They analyze the predictive information contained in a co-purchase network on Amazon and find demand spillovers across products. A related study in information networks by Zhang et al. (2016) leverages implicit brand networks on social media for audience selection framework. While existing studies focus on descriptive and predictive properties of information networks (Oestreicher-Singer et al. 2012; Zhang et al. 2016), statistical analyses of the generative features of information networks have largely been overlooked. Generative models have the capability to explain the formation of implicit links in the information network; thereby highlighting the significant brand features that users determine while co-following brands on social media. By employing social selection models on the Twitter information network, this study helps to reveal statistically valid co-interest patterns between brands of the same as well as different categories. Even though leveraging user co-interest patterns on social media is a common business practice, understanding the brand characteristics leading to these observed co-interest patterns is still an understudied topic. This study fills the remaining gap. In the next subsection, we employ generative models to study the implicit connections (or aggregate user choices) responsible for the observed network structure.

a) Estimation of Model Effects

Social selection models, from network theory, are particularly well-suited to this kind of problem where tie formation is an inter-dependent process influenced by both nodal attributes and endogenous network effects (Kim et al. 2016). We focus on a class of p* models, called Exponential Random Graph Models (ERGMs), to examine the multiple interdependent social processes responsible for the brand network formation. The purpose of ERGM, in a nutshell, is to build a stochastic model that captures the generative features of the observed brand network. Conceptually, the observed network is treated as one realization from a set of possible network outcomes. Our goal is to identify the plausible mechanisms responsible for the implicit connections between brands. Since ties between brands arise from the aggregated interests of Twitter users, the ERGM model essentially reveals what drives co-interest between two brands. Mathematically, ERGMs take the following form –

$$P(Y = y) = \left(\frac{1}{k(\theta)}\right) \exp\{\theta\ g(y)\}$$

where y is the observed network and Y denotes possible network realizations. The term $g(y)$ is a vector of network statistics responsible for link formation, for example, homophily[3] , transitivity[4] or other nodal features. Here θ denotes the vector of unknown coefficients corresponding to $g(y)$, and is estimated using Markov Chain Monte Carlo maximum likelihood estimation (MCMC-MLE) procedures (Robins et al. 2007). The normalizing factor $k(\theta)$ is calculated by summing up $\exp\{\theta\ g(y)\}$ over all possible network configurations. The iterative procedure simulates a distribution of random graphs using MCMC, to get refined values of the model parameter θ (Robins et al. 2007).

3 Homophily – Similarity breeds connection. See McPherson et al (2001)

4 Transitivity - Similar to its mathematical cousin, transitivity posits that if *a* chooses *b* as a friend and *b* chooses *c* as a friend, then *a* will choose *c* as a friend. See Holland and Leinhardt (1977)

To reveal the factors driving co-interest (essentially, links) between brands in the given information network, we formulate the following research questions.

1) Edges

Drawing from the notion that the social signal of 'who follows a brand' provides a strong reflection of brand image (Naylor 2013; Kuksov et al. 2013; Culotta and Cutler 2016), we use a set of 507 Twitter brand accounts as a basis for analysis. In the information network, two brands are connected if followers of one brand are also interested in the other brand. As connections based on too few common followers may not indicate significant connectivity, the use Disparity Filter on the original fully connected network helps to identify the statistically relevant links. The ERGM analysis is conducted on this filtered network, extracted using α = 0.05. Our first aim is to establish that edges in the network are restricted to specific pairs of brands and do extend across any arbitrary pairs. In other words, they are formed due to genuine consumer co-interest arising from complementary brand features or marketing programs.

Do edges in the network tend to form across arbitrary of brands?

2) Homophily

Cognition theorists (Shachar et al. 2000) have found the tendency among people (or entities) to associate with those who are similar to them in socially significant ways (birds of a feather flock together). This relationship between similarity and association, commonly known as the principle of homophily, has been widely popular in sociology, social network analysis, and computational social sciences (McPherson et al. 2001). Homophily in terms of links between brands of the same category would mean that users tend to follow multiple brands of the same industry. For a given brand, this means that your Twitter fans are 'informed' or 'avid' consumers of the market considering they are also following other brands of the same industry. In the context of brand networks, we investigate:

Are brands of the same industry more likely to connect than others?

3) Cross-Category Effects

Brand associations are an important determinant of brand equity (Aaker 1991; Keller 1993). Consumers develop a variety of brand-to-brand associations that subsequently result in co-branding opportunities for firms (Washburn 2000). More so, recent research on cross-category associations highlights the importance of complementary promotions, embedded premiums and joint positioning strategies for brands and firms (Leeflang and Parreño-Selva 2012). The implicit brand associations in the network reflect aggregated preferences of users across categories and show how some category pairs attract more common interest than others. For example, high across category links between

Airbnb and FIFA or Nike and Red Bull are not just outcomes of mere chance, but possibly a result of advertising and future co-branding opportunity for firms. Such brand knowledge can help managers identify potential target audiences, not one assumed by management but one perceived through data on consumers' direct interests. This is an important question for researchers, given ample evidence on the importance of cross-category associations lately.

Do certain cross-category pairs have more links than others?

4) Brand Engagement

Social media platforms provide an excellent source for businesses to build and foster relationships with consumers. As an increasing number of consumers choose to affiliate with their favorite brands on social media, online brand communities have received a lot of attention in current years. Marketers have found that brand communities established on social media lead to value creation (shared consciousness, brand use, brand loyalty) and engagement among community markers (Laroche et al. 2012). As most users follow brands with the intention of knowing more about the product and ongoing sales (Vision Critical 2013), the extent of brand engagement (tweets released by a brand) may impact a user's intention to join or leave a brand's fan page.

Do brands with high level of engagement (number of tweets) have more links than others?

5) Popularity Effect

Many real-world networks including the Internet and social networks are characterized by the popularity effect, called Preferential Attachment, whereby the more connected a node is, the more likely it is to receive additional links. Intuitively, heavily linked nodes represent well-known entities, with a lot of associations. Other nodes in the network are more likely to form relations with these highly connected nodes rather than relatively less visible nodes. This phenomenon is sometimes called the Matthew Effect (or rich get richer effect). Extending our analysis on preferential attachment to brand networks, we would like to investigate if brands with many links tend to form more links. The absence of this effect would imply that brand-brand connections develop more from marketing efforts and genuine user choice than existing popularity in the network.

Do more connected brands have a higher probability of forming new links?

b) Model Results

The dependent variable in the ERGM model is the presence of links (or consumer co-interest) among the brands in the network. Our key independent variables are – dyadic covariates (within-category effects – *homophily*, across-category effects – *heterophily*, brand engagement) and structural effects (edges and popularity effect). The results of the ERGM models are given in Table V.

The convergence plots and goodness of fit statistics are included in Appendix B and C.

Table V. ERGM Estimation Results

	Variables		Model 1	Model 2
Network Effect	Edges		-1.99***	-3.26***
Network Effect	Popularity Effect		-0.03	-0.03
Brand Effects	Brand Engagement			0.700***
Brand Effects	Within-Category **Homophily**			2.18***
Brand Effects	Between-Category **Heterophily**			
	Airlines	Automotive		-0.99**
	Airlines	Beer		-0.57
	Automotive	Beer		-0.62*
	Airlines	Cruise		2.04**
	Airlines	Travel		2.56**
	Automotive	Travel		-2.03**
	Beer	Travel		-1.40
	………………………… Remaining interaction effects between categories are shown visually in Figure 4. All numeric values included in Appendix B.			……………………..
	AIC		102544	85728

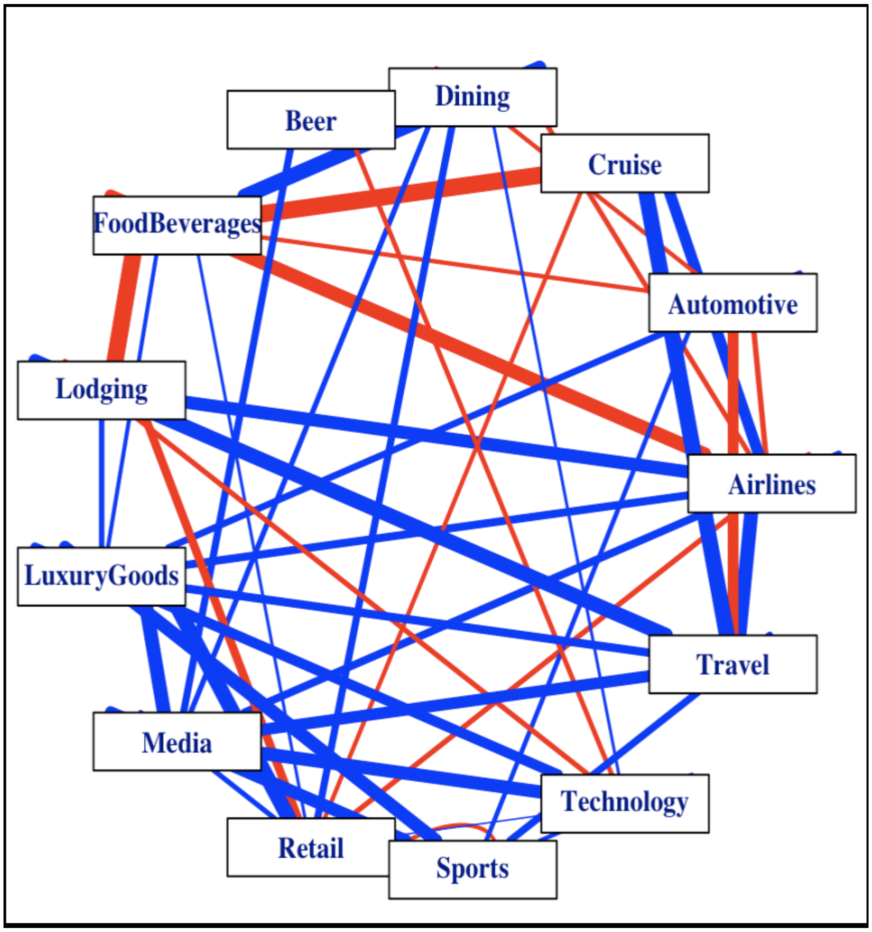


Figure 9. Between-Category Effects.

The blue lines between category-pairs represent positive likelihood for edge formation and red lines represent negative likelihood for edge formation. Detailed values of the coefficients (and standard errors) are mentioned in Appendix B.

Model 1 only includes structural effects without any brand level characteristics. The significant negative coefficient for the 'edges' parameter implies that edges between brands do not extend across arbitrary pairs. The probability of the formation of edges between brands is $= exp^{-1..99}/(1 + \ exp^{-1.99})$ = 12%; and this 12% corresponds to the density of the observed brand network. In a nutshell, the negative edge coefficient confirms that co-interest between brands (or edges) is not observed at random; but only occurs among specific brand pairs. Next, we include the popularity effect to test if popular brands (that is, one with many connections) tend to form more new links than others. The coefficient is not significant; implying the absence of any such effect. Thus, brand-to-brand connections develop more from marketing efforts and personal user choices than existing brand popularity in the network. In Model 2, we include the brand level characteristics, along with structural effects, to test whether the combined terms provide a better model fit. In general, smaller Akaike's Information Criterion (AIC) values mean better model fit (Akaike 1998; Kim et al. 2016). The AIC of model 2 is substantially lower than that of model 1, suggesting that both brand level characteristics and structural effects are important in explaining the observed information network.

In model 2, the significant positive coefficient for brand engagement means that brands who engage more with their fans on social media tend to form more links than others. Since links between brands arise from common followership, this means high levels of brand engagement also relate to an increase in brand followers. This affirms the relevance of brand management on social media and justifies the increased resources that brand owners invest in managing fan communities on social media.

Next, the significant positive coefficient for the nodematch parameter 'within-category' shows support for homophily. The log-odds of brands of the same category forming links are +2.18. As links arise from common followership, this really means that brands of the same category are more likely to attract common followers than others. The fact that consumer co-interest in brands is significantly linked with a category is an indication of 'informed' users who have interest in specific categories (or markets) on Twitter. Examples of strongly linked brands within the same category include Porsche and Tesla in automotive, Carnival and Norwegian in cruise, Loreal and Revlon in retail, etc.

Moreover, to identify any significant between-category interactions across complementary brands, we include the nodemix parameter for all brand pairs. The between-category terms in the nodemix parameter capture the heterophilous relationships between brand pairs of different categories (for example, consumer co-interest between automotive and beer brands). As with any standard regression technique, we include a base category corresponding to the pairings that should not be included. In our case, the base category is 'miscellaneous". As shown in Table 5, a positive significant coefficient for any category pair (under between-category effect) reflects an increased likelihood of consumer co-interest between the respective industries. For visual clarity, the same results are also presented in figure 9, where blue lines between category pairs represent positive likelihood and red lines represent a negative likelihood for edge formation. Categories linked with blue lines attract significantly high consumer co-interest on social media and offer meaningful insights to brand owners for identifying potential target audience. They can also point to previous joint-advertising efforts between the respective category pairs. Examples of strongly linked category-pairs include lodging and airlines, technology and sports, cruise and travel, etc. Some of the brand pairs involved in these cross-connected links are Reebok and Strike-Force Energy, Travelocity and Australian Open, Hilton and Royal Caribbean, etc. Given the importance of between-category complementarities for coordinated promotions and co-branding opportunities, this is an important finding for brand managers.

Overall, results from the ERGM analysis show that the implicit brand-brand network is a valid information artifact, arising from specific user interests on Twitter. Even though leveraging user co-interest patterns across categories is a common business practice, the idea of revealing statistically relevant complementary category pairs is an understudied topic in both IS and marketing literature. The ERGM results offer meaningful insights to social media managers and show how some category pairs attract more user interest than others. Furthermore, though the current analysis relies on superordinate categories (i.e. luxury, technology, retail etc.), the inter-category associations can be readily obtained for subordinate categories (i.e. watches, sportswear, packaged foods etc.) depending upon the marketing objectives. The assignment of brands to categories is flexible and can be changed by simply re-labeling the attributes in the network.

Keller (1993) suggests that entire categories (e.g. automotive), not just individual brands, have their own personality space formed by user stereotypes. Despite the theoretical evidence on categories exhibiting their own associations, little work has been done in this area. The ERGM analysis provides a unique way to highlight these category-category associations based on the indirect interests of social media users. Specific branding insights from the ERGM analysis reveal potential target audience for focal brands and also help to quantify the effectiveness of previous advertising strategies. In conclusion, the results show that consumer co-interest in brands on social media is driven by specific brand characteristics and highlight several managerial implications in the area of targeted marketing, complementary branding, and online consumer behavior.

V. APPLICATIONS OF THE BRAND NETWORK

In this section, we discuss the usefulness of brand networks for studying two core areas of brand management – eliciting asymmetric brand associations and mapping brand positioning. The first section on eliciting brand associations focuses on revealing cross-category insights for co-branding opportunities. By incorporating directionality into the network edges, asymmetric relations among brand pairs are revealed that help determine which brand will benefit the most from a cobranding alliance. A new construct, transcendence, is defined that captures the transcending associations of a brand onto new categories. The construct is refined further to separate a brand's own idiosyncratic associations from its category average. The new measure (net transcendence) is more informative than the original value of transcendence as it ignores the cross-category associations that are generic to the category and identifies those that are truly intrinsic to the brand itself. Moreover, to compare the transcendence of a brand with its competitors, a transcendence matrix is defined where rows represent brands and columns represent transcendence across different categories. The matrix provides a comprehensive view into the competitive landscape by highlighting the position of brands and their competitors in different categories.

As user-brand relationships on social media change over time, the paper compares the results of the brand network across two time periods – 2017 and 2020. The analysis helps to visualize the fluctuating brand associations over time and investigate its impact on co-branding opportunities. Managers can study the network patterns over time to evaluate the effectiveness of their marketing campaigns or assess the impact of external events on their brand's associations in consumers' minds. If critical associations to certain brands (or categories) have waned, this helps bring timely intelligence for managers to identify the problem and take action. Similarly, if new associations have formed during the course of time, it provides information on potential emerging co-branding opportunities.

The second section on positioning describes how certain brand constructs obtained from the transcendence matrix can reflect the relative positioning of competing brands in consumers' minds. Perceptual maps, obtained from the network, help uncover competitive landscape of brands along the dimensions: *centrality* and *distinctiveness.* Our core contribution, here, is a new methodological tool that not only provides an efficient and scalable way to infer a brand's position, but also provide a granular assessment of why that positioning occurs. For instance, disentangling the links of Mercedes's brand community across categories shows how the brand is distinct from others in the same category. Unlike Dawar and Bagga (2015), who rely on traditional surveys to infer a brand's position in terms of centrality and distinctiveness, our automated approach allows researchers to go beyond the numerical centrality-distinctiveness values and study the exact underlying graph structure of the brand's associations and reveal what makes them distinct from others. With an average correlation of 0.7 with directly elicited survey ratings, this large-scale data driven approach provides a reliable means to automatically infer brand position in a timely manner.

a) Eliciting Brand Associations

The study of brand associations is crucial for understanding consumer brand knowledge (Keller, 2003). Given that it is imperative for managers to know what consumers feel about their brand as well its association with other brands (Henderson et al, 1998), it is crucial to identify methods that convey this information in a timely and cost-effective manner. Typically, brand associations have been studied under the umbrella of "consumer-associative networks" that attempt to map the brand impressions stored in consumers' minds (Aaker, 1996; Henderson et al, 1998). Broadly falling under the category of mental or memory models, these networks consist of nodes that can represent a brand (McDonalds), a product (Big Mac), or an attribute (quality); and links between nodes suggest an association in consumers' minds (Aaker, 1996). Built on the notion that "*consumer perceptions of brands and market structures are more* *important than a-priori managerial statements of intended brand strategies",* brand associative networks have been widely popular among marketers and psychologists for a long time (Henderson, 1998;Henderson et al. 2002; Teichert and Schöntag, 2010). In most of these studies, consumer-brand associations are collected through surveys and focus groups, which are limited to few sets of brands and consumers. While having input from a broader base of consumers is desirable, recruiting and maintaining a large set of survey participants can lead to cost and feasibility constraints for managers (Aaker, 1996).

With the advent of web 2.0, researchers have shifted to more digital approaches, particularly text mining of user-generated content, to study brand associations. A potential drawback with this method includes bias in information as majority of online users tend to post in extreme bimodal situations – either very satisfied or very dissatisfied with the brand (Goes et al, 2014). Even after balancing the polarity in the data, extracting meaningful brand associations from online user- generated content is not a straightforward text mining process. Specifically, the information pertaining to the association of a brand across category dimensions, such as luxury, sports, travel, technology etc., is rarely found in individual users' comments. This substantially limits the data available for analysis and hampers the use of any text mining algorithm to infer specific brand associations. Recent advances in social network analysis open doors to a wide range of out-of-box solutions that can scale well beyond conventional methods used in brand management.

In this section, we introduce a new, scalable approach for inferring brand associations using implicit brand networks on social media. Implicit networks, with their inherent ability to condense the vast digital interest space of millions of users into a reduced form representation, provide a rather novel view into the digital ecosystem and have started to garner increasing attention from researchers across domains (Sundararajan et al, 2013; Oestreicher-Singer and Sundararajan, 2012; Zhang et al, 2016). In our study, the connections of a focal brand in the network are used to discern the associations of brands in consumers' minds. Naturally, while some brands are found to possess strong associations within category (viewed as *central brands*), there are others having diverse associations across categories (viewed as *transcendent brands*). Though the idea of generating brand-to-brand associations has been studied before, it has mostly been restricted to brands (or products) within a single category such as camcorders in Kim et al (2011), car brands in Netzer et al (2012) and LED-TVs in Ringel and Skiera (2016).

To the best our knowledge, this is one of the few studies that investigates the cross-category associations of brands as a way to identify co-branding opportunities over time. Having relevant knowledge about cross-category brand associations is crucial for other marketing tasks as well including coordinated promotions, embedded premiums and brand extensions (Cutright, 2013; Xiao and Lee, 2014); however, there is little or no evidence on identifying these cross-category effects using current digital approaches. Our proposed methodology helps to fill this void in the literature. The cross-category brand insights, revealed through the brand network, serve as important measures to assess brand fit during co-branding decisions. They also help to determine the brand's potential for future growth in category extensions.

The paper introduces a new construct, transcendence, defined in the context of a large ecosystem of brands belonging to different categories. Transcendence measures how a brand transcends its own category to connect with others across categories. Keller (2003), in his seminal work on branding, mentions that higher the shared associations between the brand and the new category, greater is the perception of fit. Our study provides a new methodological tool to identify these shared associations between brands and categories using publicly available social media data. The new construct, transcendence, indirectly captures the fit of brands onto new categories; not as one identified by management apriori but as one perceived by the direct interests of users on social media.

By incorporating directionality into the network edges, we are able to capture the asymmetric relations among brand pairs and show which brand will benefit the most from a cobranding alliance. Depending upon the marketing goals, we outline how the cross-category associations can be visualized at three levels: category – category (c–c), brand – category (b–c) and brand – brand (b–b). The different level of analyses helps to inform managers about why certain co-branding opportunities may be more viable than others. Further, as user-brand relationships on social media change over time, we compare the results of the brand network across two time periods – 2017 and 2020.

Overall, our core contribution is a new digital artifact that helps researchers and practitioners avoid marketing myopia by identifying nontraditional branding opportunities that would be otherwise hard to see. We employ a network-analytics approach to calculate a brand's net association as a function of the average associations of its category and its own idiosyncratic associations. The category-specific associations not only help understand the position of brands in consumers' minds, they also help to classify them from other competitors in the field. From a managerial perspective, the brand network requires little human intervention in processing the underlying large-scale user data and infers brand associations in an efficient and cost-effective manner. Furthermore, as our approach relies on publicly available information on social media, it is easily scalable to large set of brands, and resulting network structures reflect preferences of a diverse set of users.

Research Context

A rich literature in cognitive psychology talks about mental models that highlight subconscious associations evoked in consumers' minds when they think of brands (Keller, 1993; Henderson et al. 1998). Often such associations can go beyond the functional attributes of the focal brand and even transcend categories (Batra et al, 2010; Meyvis and Janiszewski, 2004). Supphellen (2000) states how the subconscious nature of these associations can make their elicitation difficult through traditional marketing techniques. Important opportunities for co-branding, brand alliances and brand extensions could be missed because managers are not aware of associations that are relevant to brands in other categories. The brand network provides a novel solution to this problem by relying on a brand's social connections on Twitter to infer category-specific brand associations.

At a high level, our proposed algorithm extracts the perceptual associations of brands by disentangling their audience interests in different categories. While some brands are found to possess strong associations within category, there are others found to violate category norms and have diverse associations across categories. The paper introduces a new construct, transcendence, that measures how a brand transcends its own category to connect with others across categories. Asymmetry among brand pairs is taken into account to calculate the transcending associations of a brand into new categories. The transcendence of a brand along any given category, say sports, is based on the outgoing links that help capture the proportion of a brand's users interested in the new category (sports in this case). The more outgoing associations there are between the brand and a new category, the greater is the perception of fit (Keller, 2003). By highlighting these shared associations between brands and categories in a novel manner, the new construct 'transcendence' indirectly captures the fit of brands onto new categories.

Transcendence of a brand onto a new category can also inform how strongly the brand is positioned with respect to that specific category. Since some of these categories can indirectly translate to perceptions (for example – luxury, sports and technology), brands with strong transcendence onto these categories can be viewed as exhibiting those perceptions. Further, to measure the associations not shared by the overall category (Keller, 2002), we calculate a brand's net transcendence as the deviation of a brand's own idiosyncratic associations from its category average. The new measure (net transcendence) is more informative than the value of raw transcendence as it ignores the cross-category associations that are generic to the category and identifies those that are truly intrinsic to the brand itself.

The category-category associations (e.g. associations emanating from luxury to sports) are another useful piece of information that the brand network indirectly delivers in a novel manner. Previous literature suggests that entire categories (e.g. sports), not just individual brands, have their own personality space formed by user stereotypes (Levy, 1981; Keller, 1993). Despite the theoretical evidence on categories exhibiting their own associations (Keller, 1993; Batra et al, 2010), little work has been empirically done to identify them using publicly available data sources. The brand network provides a unique way to highlight these category-category associations based on the indirect interests of social media users.

Lastly, in addition to transcendence, there are brands that possess strong associations within their own category. These brands can be viewed as *central. Centrality* is defined in terms of the extent to which a brand shares association with other brands in its own category (Carpenter and Nakamoto, 1989; Nedungadi and Hutchinson, 1985). The concept of centrality or typicality bears direct relation to a brand's probability of recall, consideration and choice among consumers' minds (Loken and Ward, 1990). These central brands are the ones that come first in consumers' mind and serve as reference points in their category. It is worth mentioning that the constructs centrality and transcendence are not mutually exclusive; a brand can be strongly positioned in its own category and still transcend onto specific categories. We expand on this notion further in the results sections using examples from the brand network. The construct ‘centrality’ is also discussed in section V.b for creating perceptual maps. In the section that follows, we define our constructs for transcendence using empirical data.

Operationalizing Transcendence

In this section, we outline the steps for obtaining transcendence by exploiting the connections of a brand in the network. For any given brand of choice, the first step is to disentangle its outgoing connections across the main categories – luxury, sports, food, travel, beer, technology, dining, retail, airlines, media and automotive. The outgoing links help to capture the transcending associations of a brand into the new category by calculating the proportion of its brand community interested in the new category.

The algorithm, then, computes the weighted degree centrality of the brand across the different categories. Weighted degree centrality of a given brand along any new category, say luxury, is calculated as the sum of its outgoing weighted links to all luxury brands. The measure is then divided by the number of brands in the new category (luxury, in this case) to prevent categories with large number of brands from dominating the analyses.

1) More formally, the set of brands in the network can be represented as B, where any individual brand $b \in B$. Brand categories, G, are subsets of B, such as G $\subseteq$ B. In general, a brand can belong to multiple categories. The transcendence of focal brand b onto a new category G is evaluated as:

$$t_b^G = \frac{\sum_{k \in G} w_{b,k}}{|G|}$$

where $\sum_{k \in G} w_{b,k}$ is the summation of outgoing weighted links from brand b to all k brands in category G, and $|G|$ gives the number brands in category G.

Considering a set of non-overlapping categories G_1, G_2………G_p , the transcendence of a brand b across p categories is a 1 x p dimensional vector:

$$t_b = [t_b^{G_1} \quad t_b^{G_2} \;\; t_b^{G_3} \ldots\ldots\ldots\ldots\ldots t_b^{G_p}]$$

In the transcendence vector t_b , brand b's association to its own category G_b essentially measures centrality ($t_b^{G_b}$). Higher the strength of this association, the more central the brand is in its own category. The transcendence vector of a brand can also be analyzed with respect to its competitors in the category. The 1 x p dimensional vector t_b can be further extended to a n x p matrix where n rows represent brands $(b_1, b_2 \ldots\ldots\ldots\ldots.. b_n)$, and p columns represent transcendence across the p categories (see Figure 10, below).

$$
\text{Brands}\left\{
\underbrace{
\begin{array}{l}
[t_{b_1}^{G_1} \quad t_{b_1}^{G_2} \quad t_{b_1}^{G_3} \ldots\ldots\ldots\ldots\ldots t_{b_1}^{G_p}] \\
[t_{b_2}^{G_1} \quad t_{b_2}^{G_2} \quad t_{b_2}^{G_3} \ldots\ldots\ldots\ldots\ldots t_{b_2}^{G_p}] \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
[t_{b_n}^{G_1} \quad t_{b_n}^{G_2} \quad t_{b_n}^{G_3} \ldots\ldots\ldots\ldots\ldots t_{b_n}^{G_p}]
\end{array}
}_{\text{Transcendence across p categories}}
\right.
$$

Figure 10. The transcendence matrix, t_b^G, for n brands across p categories.

2) The category-category level associations i.e. associations emanating from one category G_i to another category G_p, are calculated as follows:

$$
t_{G_i}^{G_p} = \frac{\sum_{b \in G_i} t_b^{G_p}}{|G_i|}
$$

where i ≠ p. This measures the average transcendence of brands belonging to one category, say G_i, onto another category G_p. Using examples from the brand network, in Table VI below, we demonstrate how category-category associations are calculated for the automotive industry.

Table VI. Obtaining c-c associations

Brands ($b \in G_i$)	Luxury (G_1)	Sports (G_2)	…….	Travel (G_p)
Audi (b_1)	$t_{b_1}^{G_1}$	$t_{b_1}^{G_2}$	…….	$t_{b_1}^{G_p}$
Mazda (b_2)	$t_{b_2}^{G_1}$	$t_{b_2}^{G_2}$	…….	$t_{b_2}^{G_p}$
BMW (b_3)	$t_{b_3}^{G_1}$	$t_{b_3}^{G_2}$	…….	$t_{b_3}^{G_p}$
Tesla (b_4)	$t_{b_4}^{G_1}$	$t_{b_4}^{G_2}$	…….	$t_{b_4}^{G_p}$
……………	……………	……………	……. .	……………
……………	…………….	……………	……. ..	……………
Honda (b_n)	$t_{b_n}^{G_1}$	$t_{b_n}^{G_2}$	……. ..	$t_{b_n}^{G_p}$
Category average	$t_{G_i}^{G_1} = t_{auto}^{luxury} = \frac{\sum_{b \in auto} t_b^{G_{luxury}}}{\|auto\|}$	$t_{G_i}^{G_2} = t_{auto}^{sports} = \frac{\sum_{b \in auto} t_b^{G_{sports}}}{\|auto\|}$	……. ..	$t_{G_i}^{G_p} = t_{auto}^{travel} = \frac{\sum_{b \in auto} t_b^{G_{travel}}}{\|auto\|}$

3) To separate a brand's own idiosyncratic associations ($t_b^{G_p}$) from its category average ($t_{G_i}^{G_p}$) , we calculate the net transcendence of brand b onto category G_p as follows:

$$t\_net_b^{G_p} = t_b^{G_p} - t_{G_i}^{G_p}$$

where $b \in G_i$ and $i \neq p$; positive values for $t\_net_b^{G_p}$ indicate individual brand's transcendence to be above the category average and negative values indicate individual brand's transcendence to be below the category average. Similar to the raw transcendence vector $t_b^{G_p}$, the net transcendence vector of a brand b across p categories is:

$$t\_net_b = [t\_net_b^{G_1} \quad t\_net_b^{G_2} \quad t\_net_b^{G_3} \ldots\ldots\ldots\ldots\ldots t\_net_b^{G_p}]$$

The 1 x p dimensional vector t_net_b can be further extended to a n x p matrix where rows represent n brands in a category and p columns represent the net transcendence of brands across the p categories (see Figure 11, below). The above matrix provides a more comprehensive view into the competitive landscape of brands by highlighting a brand's own association as well as those of its competitors. In the next section, we discuss our results on cross-category maps and lay out important managerial implications.

$$
\begin{array}{c}
[t\_net_{b_1}^{G_1} \quad t\_net_{b_1}^{G_2} \quad t\_net_{b_1}^{G_3} \ldots\ldots\ldots\ldots t\_net_{b_1}^{G_p}] \\
[t\_net_{b_2}^{G_1} \quad t\_net_{b_2}^{G_2} \quad t\_net_{b_2}^{G_3} \ldots\ldots\ldots\ldots t\_net_{b_2}^{G_p}] \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
[t\_net_{b_n}^{G_1} \quad t\_net_{b_n}^{G_2} \quad t\_net_{b_n}^{G_3} \ldots\ldots\ldots\ldots t\_net_{b_n}^{G_p}]
\end{array}
$$

Figure 11. Net transcendence matrix

Cross-Category Maps

Depending upon the business objective, the category-specific insights, revealed through the brand network, can be visualized at three different levels : category-category (c-c), brand-category (b-c) and brand-brand (b-b). Using examples from the automotive and beer category, we present the results at three levels (c-c, b-c, b-b) and conclude by discussing the practical applications of the brand network. Such analysis can be easily extended to other categories as well. Figure 12 (left) presents the c-c associations of the automotive category in the year 2017. Some of the prominent categories that the automotive fans are interested in include technology, sports, dining and luxury. The pattern is similar for the year 2020, except for the media category where the associations with the automotive category have increased (see Figure 12, right).

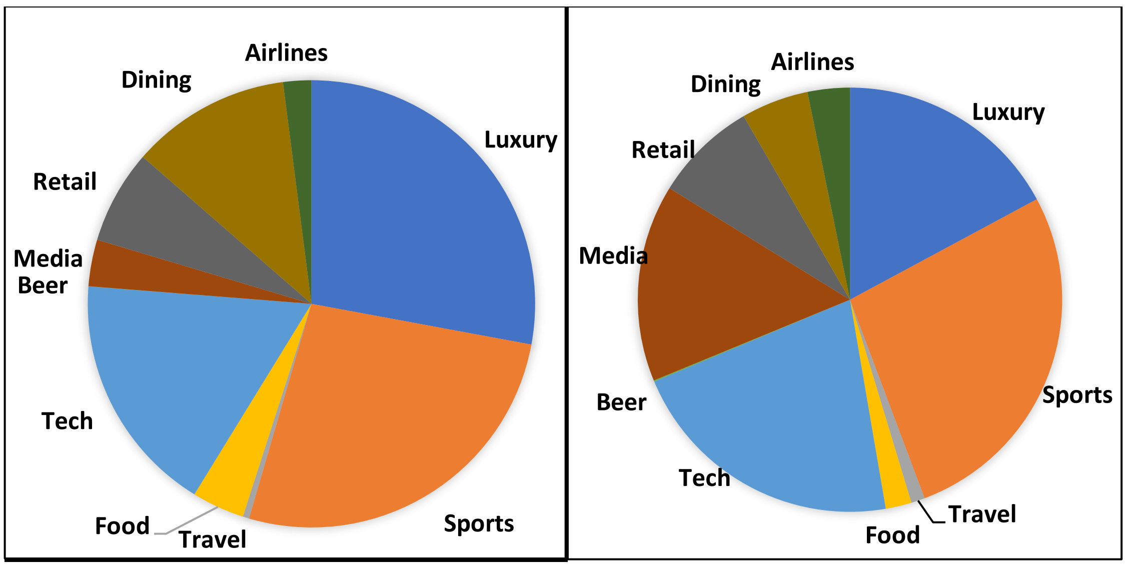


Figure 12. C-C associations of the automotive category in the year 2017 (left) and 2020 (right)

Assessing the c-c associations provides a window into the broader interests of the automotive fans and shows which categories are most viable for co-branding or brand extensions. For instance, the automotive brand Audi has previously successfully extended into the high-tech ski-clothing category where its own category's associations (such as sports, luxury and technology) align well with the target category. Though the current analysis relies on superordinate categories, depending upon the marketing objectives, these inter-category associations can be readily extended to subordinate categories. The assignment of brands to categories is flexible and can be simply changed by re-labeling the attributes in the network.

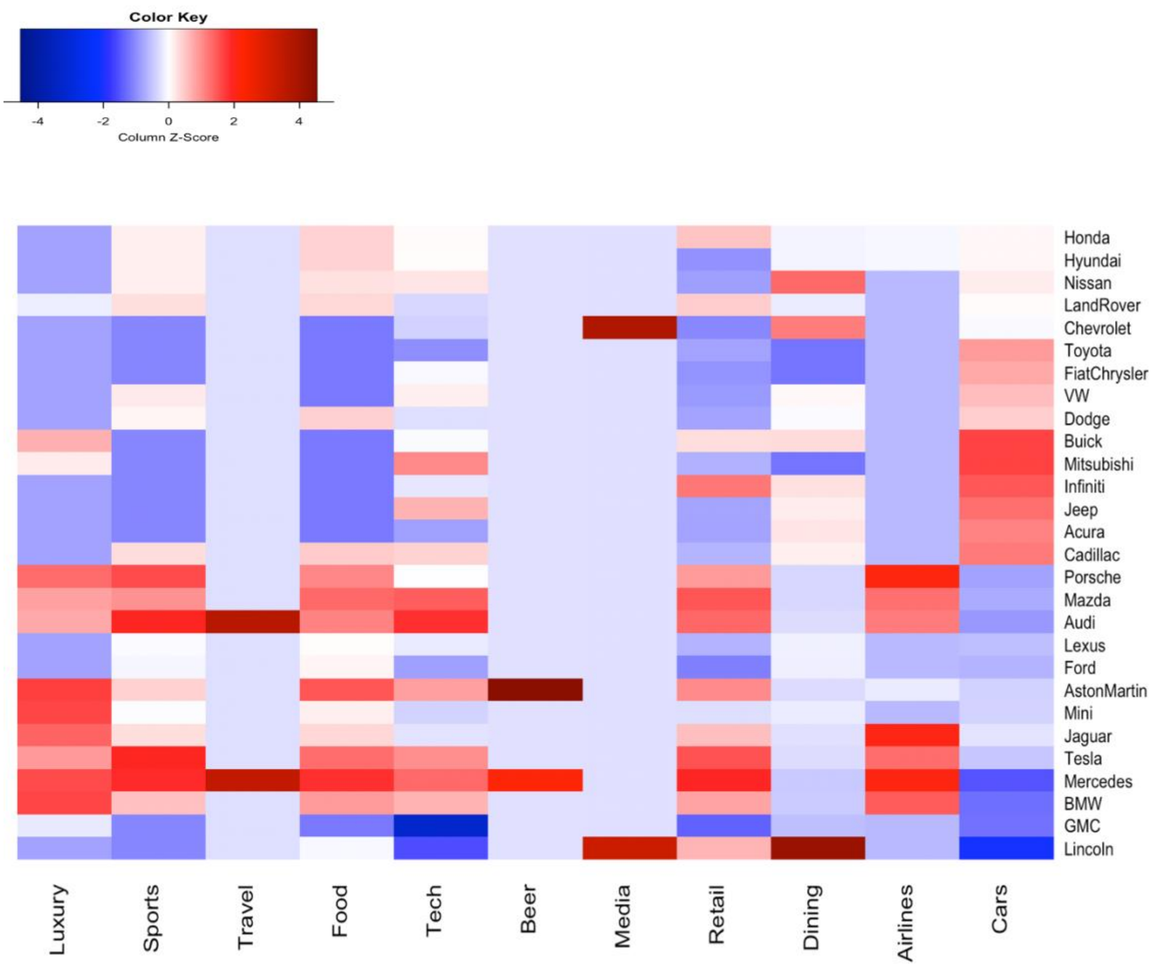


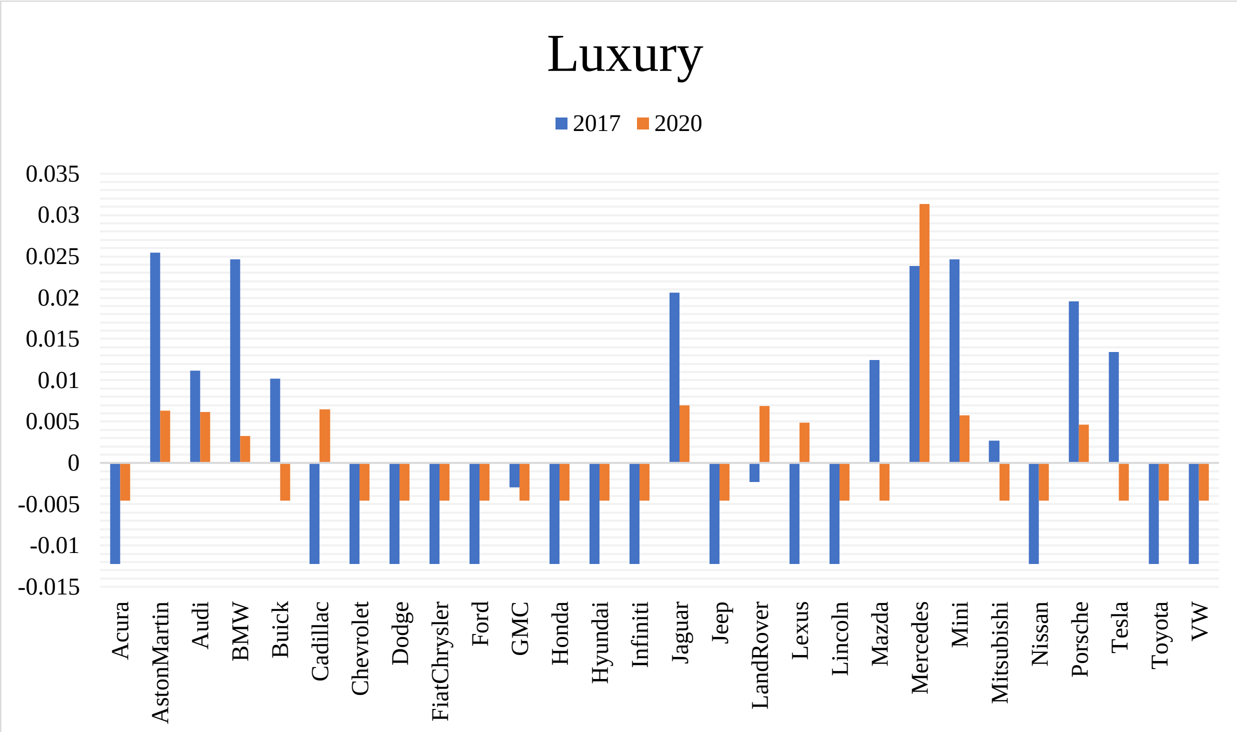
Figure 13 Net transcendence matrix, $t\_net_b^{G_p}$, reflecting brand-category (b-c) associations of the automotive brands (year 2017).

Categories having no associations with the auto brands have been eliminated from the heatmap.

We next study the net transcendence , $t\_net_b^{G_p}$, of brands (b-c) in the automotive category in Figure 13. All column values have been scaled, with positive values associated with red coloring and negative values associated with blue coloring in the heatmap. Few auto brands have positive net transcendence across multiple categories - luxury, food, retail and technology. For example, Mercedes has a broad set of associations across multiple categories, even though its associations with its own category are low. This makes Mercedes high on transcendence but low on centrality. Brand fans of Mercedes engage more with a diverse set of brands in luxury, sports and technology categories than with other auto brands. On the other hand, there are brands like Toyota whose audience is primarily interested in the auto category and not engaged strongly with brands across categories, making it high on centrality but low on transcendence. Overall, looking at the bottom half of the heatmap, we notice that car brands with high net transcendence across categories are generally the ones having lower centrality in their own category, for example Audi, Mercedes, Porsche, Tesla, Jaguar, BMW and Mazda. But certain brands are different - brands like Jeep and Mitsubishi which share strong associations within the auto category and also have moderate cross-category associations into technology and retail. Thus, centrality and transcendence are not strictly mutually exclusive, and a brand can be perceived to be both depending upon its associations in the network.

Striking the right balance between centrality and transcendence can be critical and will depend upon where the brand wants to position itself in consumers' minds. For instance, for brands like Tesla, transcending into new categories (such as Technology, Luxury and travel) may be more important than strengthening their low associations in the car category. Or, on noticing strong current associations with the desired categories, it may seek a greater audience with auto enthusiasts in the future. Mainstream brands like Toyota, on the other hand, may focus less on transcending onto new categories as long as their centrality in the automotive group remains strong.

A brand's associations in the transcendence matrix may reflect its marketing goals and advertising efforts, but those associations aren't static. Brands may, for varied reasons, want to shift their associations to new categories in search of alliances or co-branding opportunities. Central brands may seek to become more transcendent in consumers' minds to gain access to new markets. Similarly, transcendent brands may want to become more central to improve market shares.



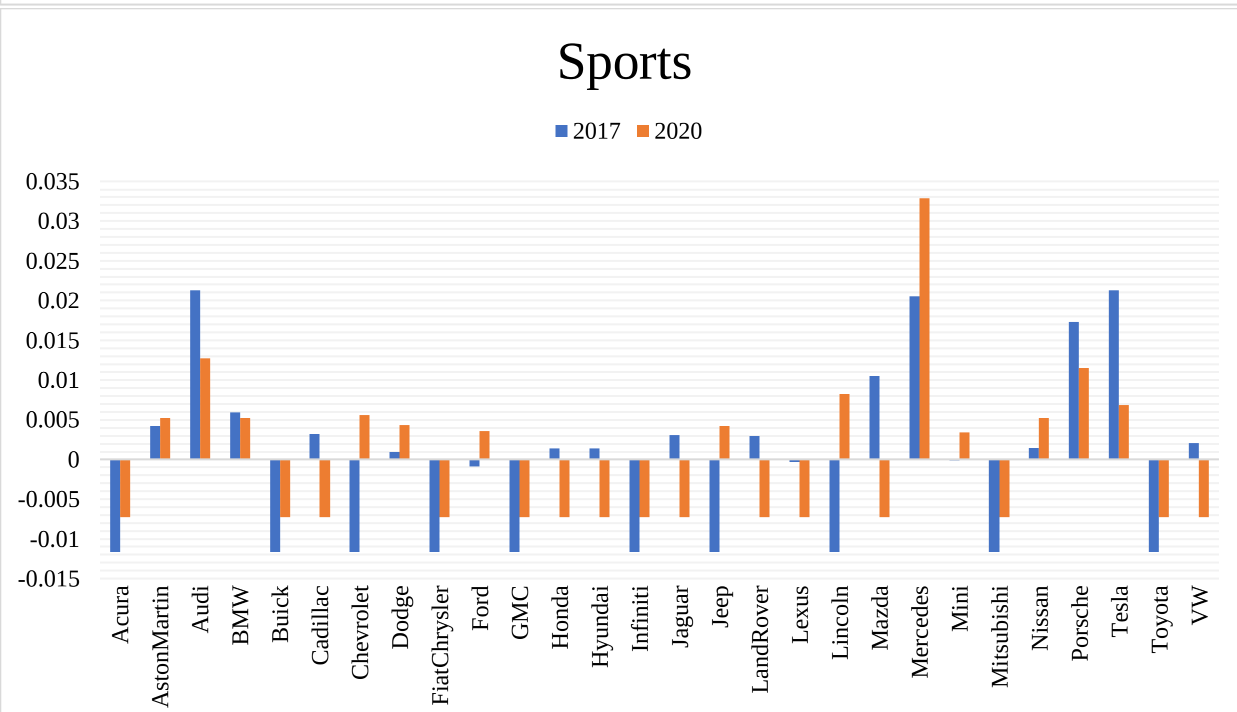


Figure 14. Change in net transcendence over time.

Next, we examine the change in net transcendence of car brands, along luxury and sports categories, over time (see Figure. 14). The dynamic plots for other categories can be analyzed in a similar way. It is interesting to observe, in Figure 14, that the luxury associations have decreased for majority car brands, namely, Aston martin, Audi, BMW, Jaguar, Mini and Porsche. Few brands, on the hand, like Cadillac, Land Rover and Mercedes witness an increase in associations to luxury. The network's ability to highlight this shift in brand's associations over time can be of vital use to managers. It allows them to quantify the impact of their marketing campaigns (or other external events) on their brand's associations. Timely intelligence can help them identify potential problems and take action. The brand network can also help managers take a more detailed look into the issue by uncovering the specific cross-category brand – brand (b-b) associations that may have diminished over the course of time.

Considering transcendence onto the sports category, we notice brands like Chevrolet, Dodge, Ford, Lincoln and Jeep move towards positive net transcendence in 2020. On the other hand, brands like Honda, Hyundai and Lexus move towards negative net transcendence in 2020. The transition to positive (or negative) net transcendence into sports could be caused by a number of reasons – brand's marketing strategy, past co-branding alliances, embedded promotions or other external events. Though the current empirical work does not examine the causes for such shifts in brand positions, it provides mangers with the timely intelligence on the subject. Future marketing studies can use this work as a foundation to further investigate the causes for changes in brand position over time.

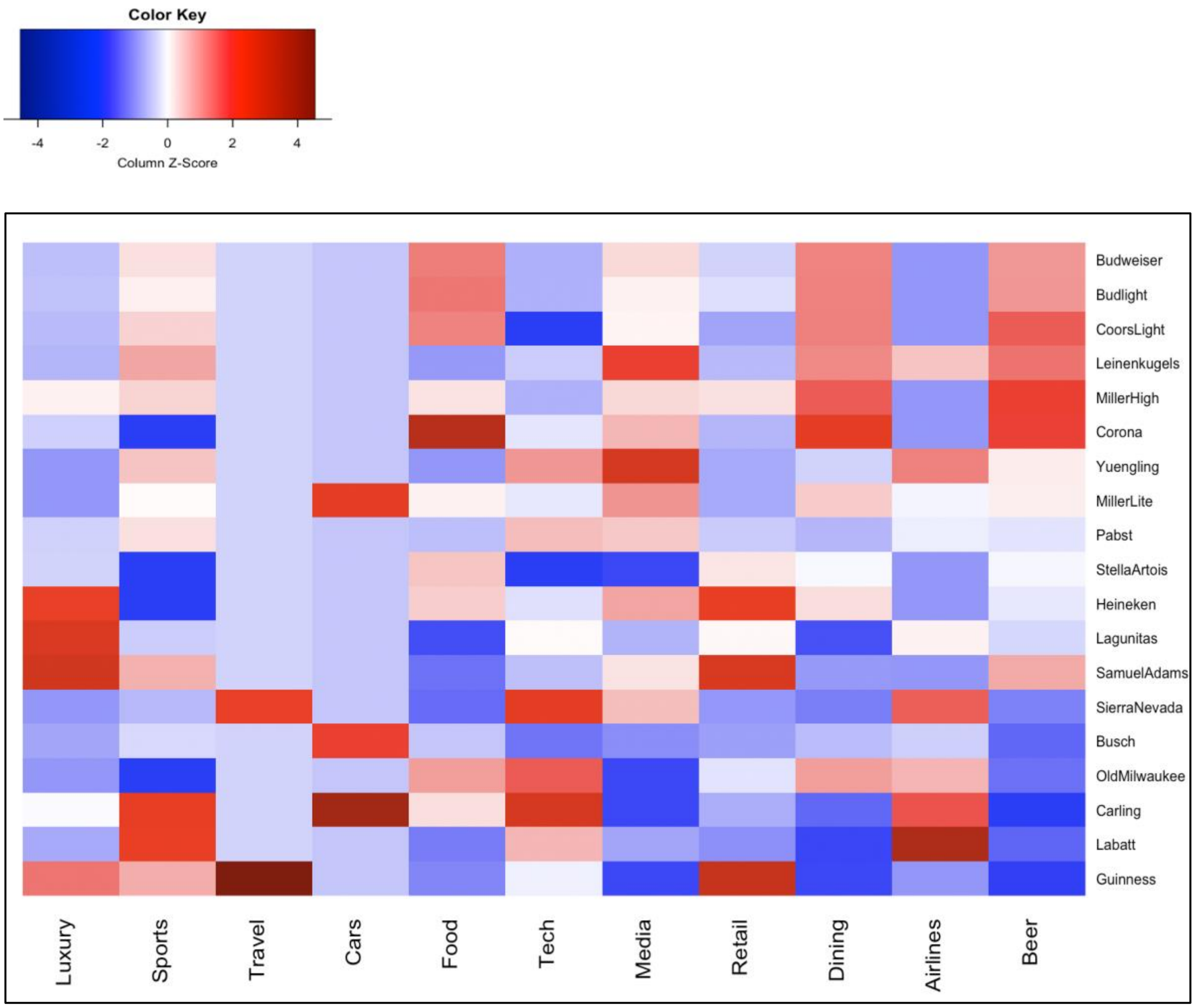


Figure 15. Net transcendence matrix, $t\_net_b^{G_p}$, reflecting brand-category associations of the beer brands (year 2017)

Categories having no associations with the beer brands have been eliminated from the heatmap

In Figure 15, we show the net transcendence , $t\_net_b^{G_p}$, of beer brands along different categories. Brands like Budweiser, Budlight, Corona and Coors share strong associations to the beer category (making them high on centrality) and also transcend onto other major categories like food and dining. Similarly, brands like Guinness and Lagunitas have low associations to the beer category (making them low on centrality) but show stronger associations across-category to sports, retail and luxury. Other brands with low centrality and high net transcendence include Sierra Nevada, which has strong associations to technology and travel categories. These category-specific brand insights arise from the audience's co-interests with certain categories and, in a way, reflect the position of a brand in the audience's minds. For instance, in Figure 15, we uncover the interests of Budlight's audience across different categories and notice how its brand positioning is different from that of Sierra Nevada. While Budlight's audience is primarily interested in beer, food and dining, the audience of Sierra Nevada is heavily engaged with travel and technology brands. Due to this, the former is positioned strongly amongst food enthusiasts, while the latter resonates more with technology and travel fans. In the next section, we discuss other practical uses of the cross-category brand associations, particularly for competitor analysis and segmentation.

Competitor Analysis

One of the key aspects of branding is to create unique brand associations that are not shared by other competitors in the field (Keller et al, 2002). These brand associations do not necessarily have to correspond to the physical or functional attributes of the brand (e.g. look of a car, average mileage etc.) and can be a part of brand intangibles (Keller and Lehmann, 2006). Brand intangibles usually transcend physical attributes and are a common means by which marketers differentiate their brand from others (Kotler and Keller 2006; Keller, 2006).

In this section, we outline how managers can use the brand network to identify the differentiating associations of a brand that are not shared by its closest competitors. First, we find the k-closest competitors of a focal brand (for example, Mercedes) from the brand network. For this, we use outgoing directed weights from Mercedes as a proxy for distance. There are many kinds of distance functions used to calculate the similarity among data points in nearest-neighbor search– Euclidean, Manhattan, Hamming etc. In our case, the outgoing directed weights from the focal brand serve as a proxy for distance. Higher the weight between two brands, smaller the distance between them. Sorting the weights in descending order one can find the k-closest neighbors of Mercedes. For instance, if k = 2, the nearest neighbors of Mercedes are Audi and BMW (see Figure 16)

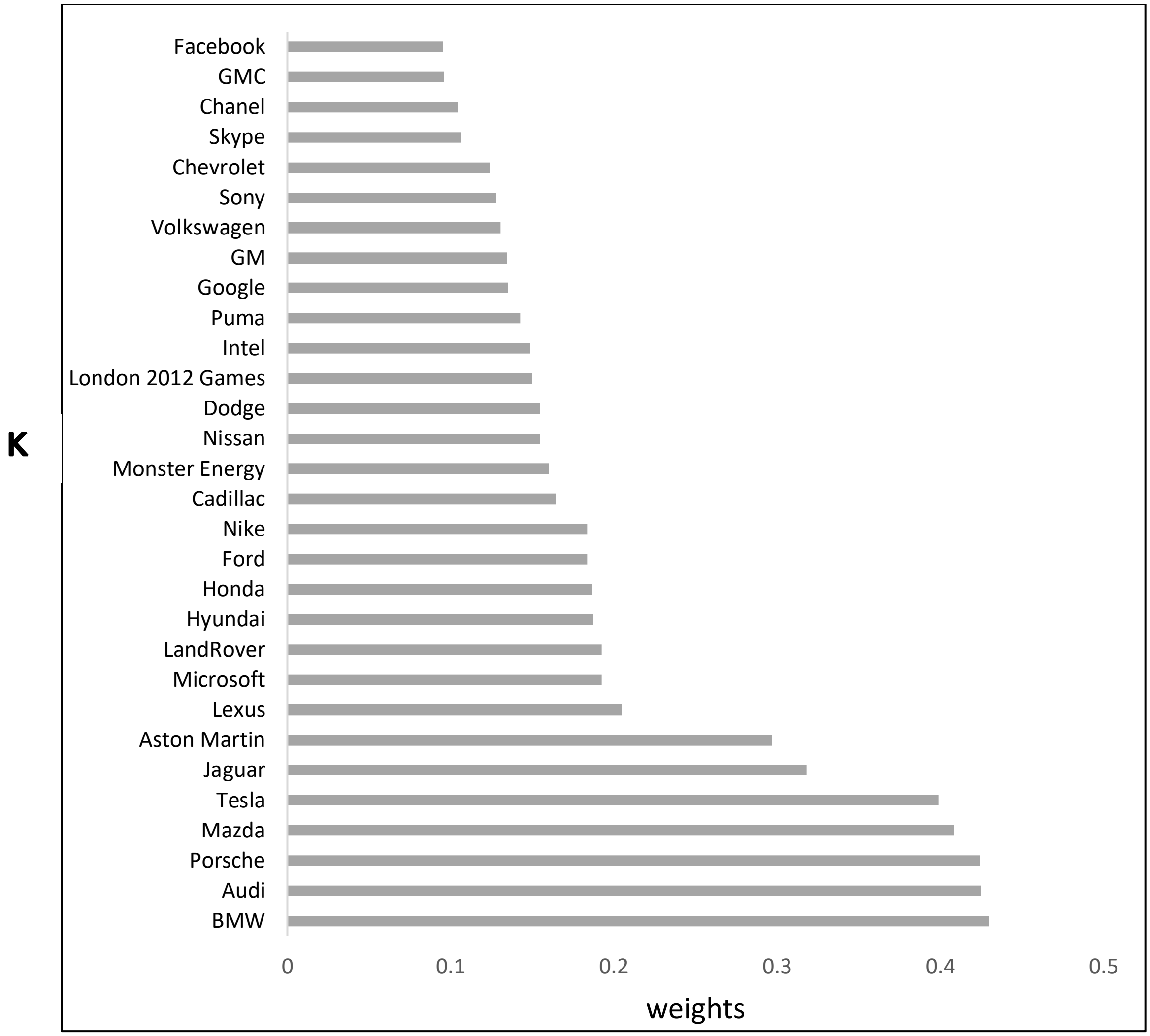


Figure 16. Nearest Neighbor Analysis for Mercedes (K = 30)

Once the k-closest competitors have been identified, one can use the category-specific brand associations from the transcendence matrix to identify a brand's associations not shared by its competitors. Going back to the case of Mercedes, we analyze the brand associations of its 2-closest competitors, Audi and BMW, in Figure 17. Here, all column values have been scaled. While Mercedes has the strongest association to luxury and sports, Audi has a stronger hold in technology category. BMW, on the other hand, has the strongest associations to travel among the three. Other prominent brand associations of Mercedes, not shared by its competitors, include those to beer, food and media. On the centrality construct, we notice BMW and Audi outperform Mercedes with stronger associations to the automotive category. Overall, the such competitor analysis can help brand managers identify potential target audiences (or markets) and obtain precise information on who the competition really is; not as one perceived apriori by management but one based on the direct interest of the brand fans.

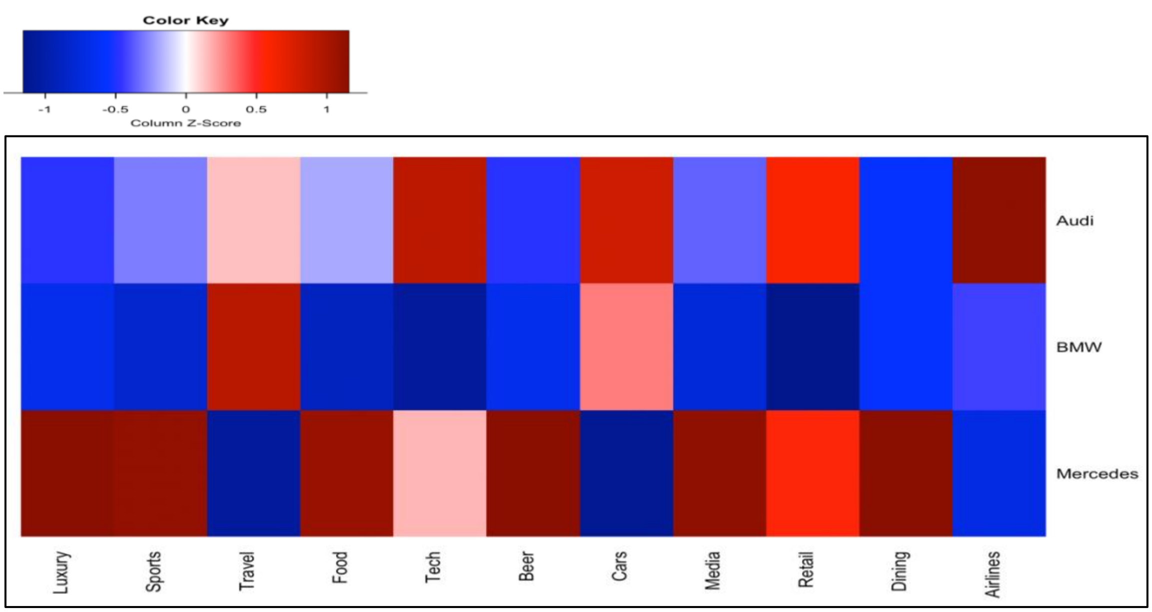


Figure 17.Net transcendence matrix of Mercedes and its 2-nearest neighbors, Audi and BMW.

The current analysis has been conducted for the year 2020 and can be easily obtained other time periods

The competitors can also transcend categories and might be potential candidates for future brand alliances. For example, non-automotive brands such as Microsoft, Nike, Monster Energy Drink, Puma, Google and Chanel are among the top nearest neighbors of Mercedes in the network. Could Mercedes and Chanel co-promote and co-brand, just as BMW and Louis Vuitton have done before? We believe that the different type of results presented by our method (c-c, b-c, b-b) may provide useful input to help answer such questions. Competitor analysis for any other brand can be conducted in a similar manner. In the next section we discuss another practical application of the brand network – Community Detection.

Community Detection – Segmentation

The study of market segmentation is fundamental to much of marketing research (Wind, 1978). To keep up with the increasingly crowded marketplace, managers constantly need to identify market segments that effectively capture who their competitors are, what do they offer and what type of customers do they attract (Ter Hofstede, et al. 1999). In the most generic form, market segmentation is defined as the strategy of dividing a large target market into smaller communities of consumers, businesses or countries that have similar user needs and preferences (McDonald et al. 2003). Another useful application of the brand network is its ability to reveal communities of brands that appeal to common Twitter users. A community membership is identified based on the commonality among user groups, with respect to similar brand preferences or devotion to common brands.

Taking the example of the automotive category, we show how standard community detection algorithms from network science can be used to reveal segments of auto brands. Specifically, we use Walktrap clustering algorithm (Pons et al. 2005), a hierarchical agglomerative method to show how the brand network can reveal segmentation. The algorithm works on the idea of detecting areas of high density within the graph, through a random walk process. The basic idea is that if two brands lie in the same community, the probability of finding the third brand located in the same community, by a random walk process, should almost be the same as for the first two brands. The results of the algorithm are illustrated in Figure 18. To highlight the possible community structures at different hierarchical levels, the results are presented as dendrograms.

For both the years 2017 and 2020, we notice cluster 1 (marked by blue labels) mostly consisting of popular high-end European and British auto brands such Audi, BMW, Porsche, Mercedes and Jaguar and Aston Martin. Similarly, cluster 2 (marked by black labels) mostly consists of mainstream American and Japanese brands - Honda, Hyundai, Ford, Toyota and others. Though the overall community structures remain consistent from 2017 to 2020, there are few brands that show a change in membership over the years. In 2017, it is interesting to see the Japanese mainstream brand Mazda having dense connections with other high-end luxury brands in the bottom cluster. Forbes (2017) mentions how Mazda, by offering premium amenities in its new model 2017-Mazda CX-9 Signature, is considered the most luxurious vehicle that the Japanese brand maker has launched to date. "*Mazda has never been considered a luxury brand, but maybe it's time to reconsider that classification*" - Forbes (2017)[5]. Autoguide also mentions how the brand aims to achieve premium positioning through some of its latest upscale models; its branding strategy being clearly effective as seen through its positioning in the luxury cluster in 2017. The analysis for the year 2020 (both transcendence matrix and community detection), however, shows that Mazda is unable to retain its associations to luxury, and moves to the cluster containing mainstream auto brands. Similarly, we notice the premium brand, Acura, move to the luxury cluster in 2020. Another example is Tesla, which moves to a single-brand cluster in 2020, hinting at its new exclusive position in users' minds that is different from other luxury or mainstream auto brands, arising from its emerging popularity in electric vehicles. Such timely intelligence on the shift of

---

[5] Forbes (2017), *2017 Mazda CX-9 Signature Test Drive And Review: Luxury Without The Brand*, retrieved from - https://www.forbes.com/sites/jasonfogelson/2017/01/23/2017-mazda-cx-9-signature-test-drive-and-review-luxury-without-the-brand/#768ae1927d6c

a brand's cluster membership can help managers track the effectiveness of their marketing strategies over time. The communities reflect the competitive market structures existing in the auto-market and can provide knowledge on the possible co-consideration patterns in users minds.

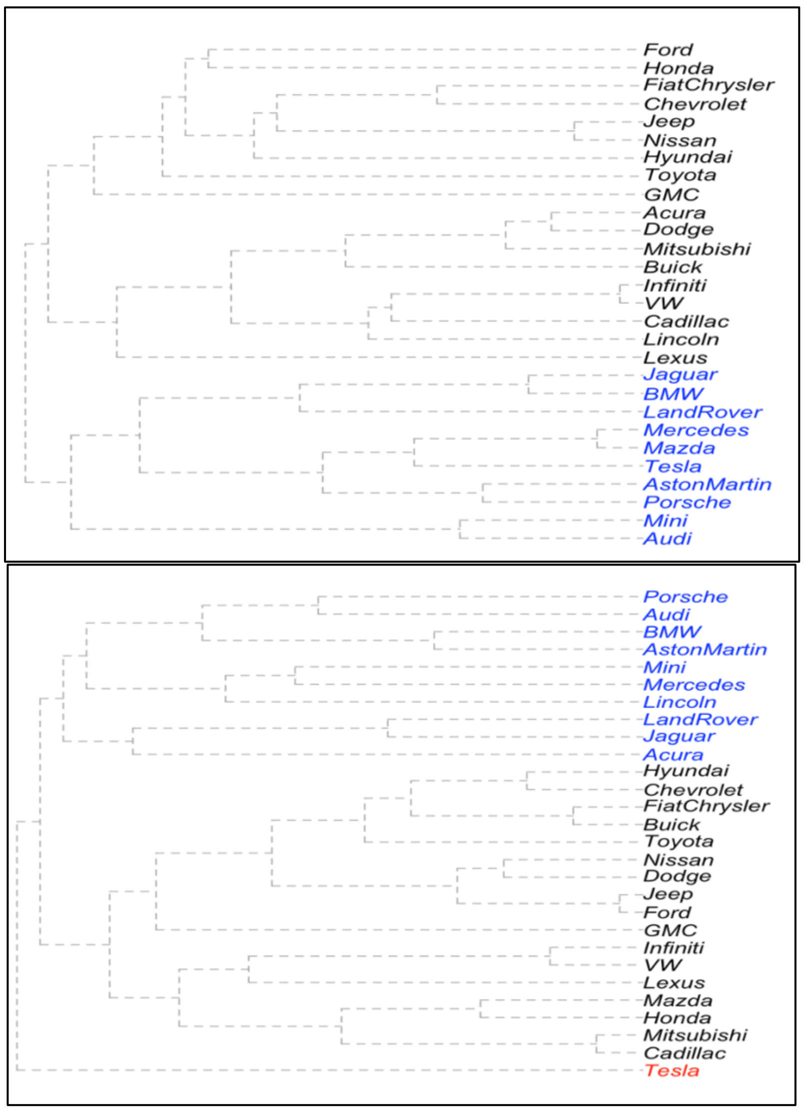


Figure 18. Communities obtained using Walktrap algorithm.

Top dendrogram corresponds to year 2017, bottom dendrogram corresponds to year 2020

Discussion

In the past decade, the importance of effectively eliciting brand associations has become an emerging topic of interest for all brand owners. Besides the obvious tangible associations of a brand that relate to its physical attributes (e.g. mileage of a car, horsepower), there are another set of intangible associations that the go beyond the functional attributes of a brand and even transcend categories. Brand associations pertaining to categories, despite being vital for various managerial decisions (such as co-branding, licensing and brand extensions), are an understudied topic in marketing literature. Some of the greatest brands in the world have defied category norms and transcended their initial market boundaries to achieve much more, for example Apple and GE (Keller, 2014).

Our work provides a novel methodological tool to managers for studying the category-specific associations of their brands, as well as those of their competitors and allies, in a timely manner. Depending upon the marketing objectives, the cross-category associations can be visualized at different levels : category-category, brand-category and brand-brand. We introduce a new construct, transcendence, that measures how a brand transcends its own category to connect with others across categories. Asymmetry among brand pairs is taken into account to calculate the transcending associations of a brand onto new categories. The transcendence of a brand along any given category, say sports, is based on the outgoing links that capture the proportion of a brand's followers interested in the sports category. Another critical avenue of future research could be to analyze different possible co-branding scenarios by studying both incoming and outgoing associations of a brand. Studying the incoming associations to the focal brand from a new category (say, sports) would help answer the question - What percentage of sports fans are interested in me? From a brand's perspective, both questions may be equally important before establishing any co-branding deals, and analyzing them simultaneously would help to create more nuanced insights.

Overall, the analysis is conducted at two time periods to track shifts in brand transcendence; this allows brand managers to assess the effectiveness of their marketing strategies or gauge the impact of external events on their brand's associations in users' minds. The different levels of analyses also help to answer why certain co-branding opportunities are more viable than others. For example, assessing the c-c associations provides a window into the broader interests of automotive fans and shows specific categories (such as technology, sports, dining and luxury) are more viable for co-branding than others. Going one step further, b-c associations capture the transcendence of brands onto new categories and show why certain categories for co-branding opportunities may hold more promise than others. For instance, investigating the transcendence vector of Mercedes, we notice how a large percentage of Mercedes' fans are engaging with luxury, technology, travel and sports – making them as suitable categories for co-branding. Finally, at b-b level, we find brands such as Microsoft, Nike, Monster Energy Drink, Puma, Google and Chanel among the closest neighbors of Mercedes in the network. Clearly Mercedes fans, as revealed through the previous b-c analysis, are interested in technology, sports, and luxury. We believe that the different types of results presented by our method (c-c, b-c, b-b) may help inform managers why certain co-branding opportunities hold more promise than others.

Managers can use brand networks for other marketing goals as well. First, competitor analysis helps to identify the closest competitors of a brand and uncover differential associations of each brand in the competing group. Second, employing standard clustering algorithms on the network allows managers to identify communities of brands with similar user preferences; this provides an alternate view of consumer segments based on direct data on their diverse interests. Both these practical applications of the brand network can be easily implemented by marketing practitioners to get timely estimates on their brand's competitive landscape. In the next section, we discuss the usefulness of brand networks for studying the relative positioning of competing brands using perceptual maps.

b) Mapping Brand Positioning

Understanding brand positioning is a crucial area of research in marketing. According to Keller (2002; 2014), brand positioning is the "*act of designing the company's offering and image to occupy a specific place in the minds of the target market*". In more general terms, positioning describes how a brand is perceived in the minds of consumers, relative to its competitors. Traditional survey approaches, typically used for inferring brand positioning, can be cumbersome and costly. Other digital approaches relying on online user generated content and browsing history have been known to suffer from potential limitations such as biased content, substantial manual intervention and privacy regulations. Our study provides a new, scalable approach for inferring brand positioning using implicit brand-to-brand networks on Twitter. Perceptual maps, drawn from the brand network, map the competitive landscape of brands along two dimensions – centrality and distinctiveness. Managers can use these maps to draw strategic assessments of the market, identify competing brands, evaluate effectiveness of their marketing activities, and redefine branding strategies. Consistently strong correlation between our automated approach and external survey ratings, average $r = 0.7$ and $p < 0.001$, validate the effectiveness of our novel methodology.

Compared to recent network-based approaches that focus on mapping product positioning using large-scale clickstream data (Ringel and Skiera, 2016), our method focusses on a more reliable privacy-friendly data source, that is, followership patterns on Twitter. The use of clickstream data has many limitations ranging from issues in tracking accurate user activity (due to use of multiple devices and widespread use of VPNs) to strict data privacy regulations (Bucklin and Sismeiro, 2009; Malhotra et al, 2004). Further in contrast to Ringel and Skiera's work which focuses on aggregate-level product information contained in a single website, brand networks provide insights into dis-aggregate level information across brands in multiple categories. This information can be useful for analyzing both within-industry competition and across-industry complementarities. With their ability to incorporate consumer interest across a broad brand ecosystem, brand networks provide useful information to 1) identify relative positioning of competing brands, 2) co-branding opportunities across industries 3) identifying target audience and 4) assess the impact of digital branding efforts.

Theoretical Foundation

In terms of theoretical contribution, our work is built on two core areas of marketing literature - brand associative networks (Henderson et al, 1998; John et al, 2006) and brand positioning (Keller, 1993, 2014; Sujan and Bettman, 1989). In Figure 19, we show how our work knits these two research areas in an empirical setting. At a high level, we first extract the transcendence matrix of competing brands to highlight the perceived associations of brands into different categories. While some brands are found to possess strict category associations (Anderson and Spellman, 1995), there are others found to violate category norms and have diverse associations across categories (Brexendorf and Keller, 2017). Here, brands with high shared associations within their own category are viewed as *central. Centrality* is defined in terms of the extent to which a brand shares association with other brands in its own category (Carpenter and Nakamoto, 1989; Nedungadi and Hutchinson, 1985). The concept of centrality or typicality bears direct relation to a brand's probability of recall, consideration and choice among consumers' minds (Loken and Ward, 1990). These central brands are the ones that come first in consumers' mind and serve as reference points in their category, for example, Toyota in cars and Budweiser in beers (Dawar and Bagga, 2015).

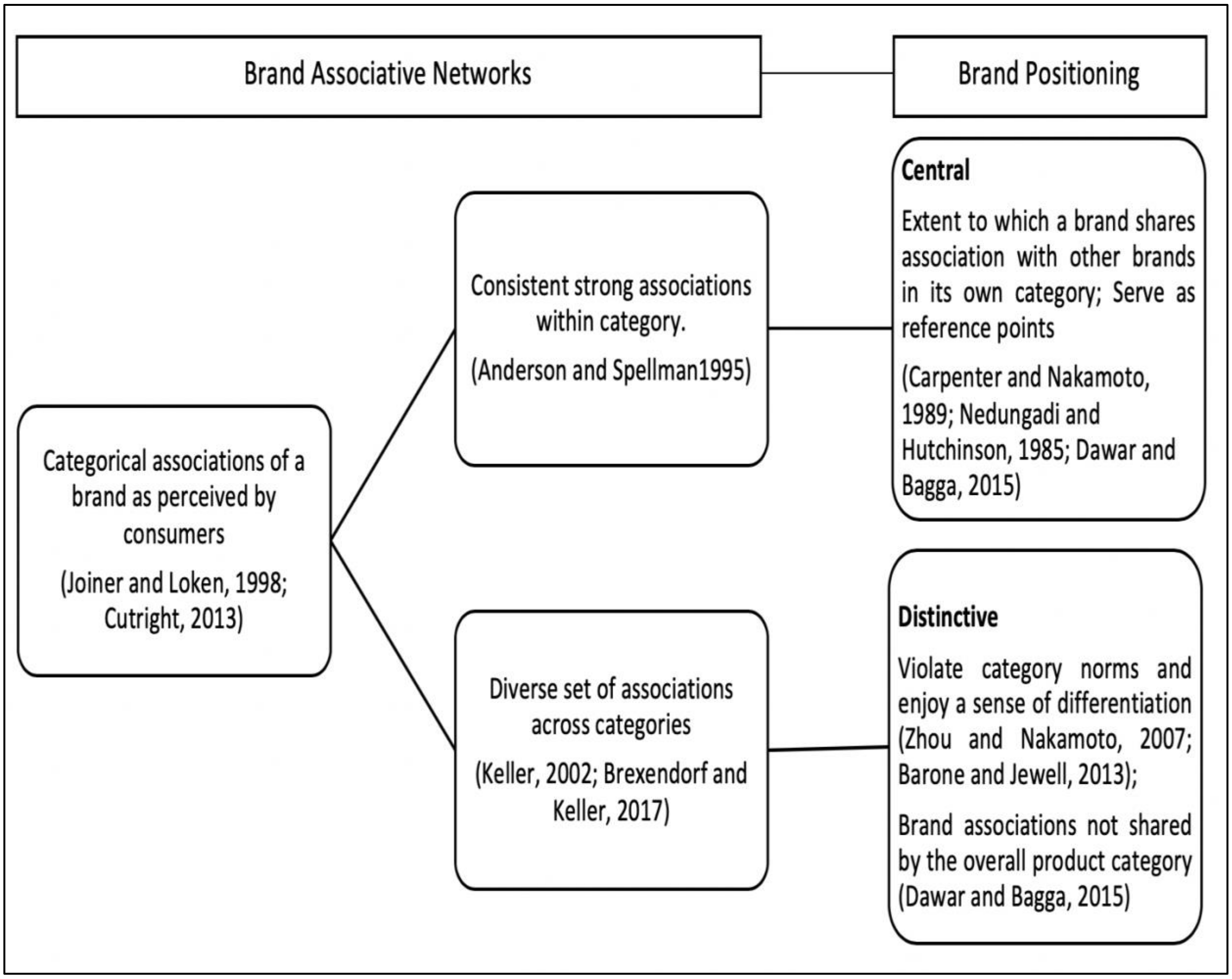


Figure 19. Obtaining Centrality-Distinctiveness Maps

Similarly, the across-category associations of a brand, in the transcendence vector, are used to discern the distinctiveness of brands in consumers' minds. Both the number of connections to different categories as well as the weighted sum of links are used to calculate a brand's distinctiveness. Distinctiveness, a cognitive outcome of differentiation, allows brands to violate category norms and enjoy a unique position in the minds of consumers (Zhou and Nakamoto, 2007; Barone and Jewell, 2013). Across-category links represent unique brand associations not shared by the overall product category and help to differentiate a brand from its competitors (Keller, 2002). Such distinctive brands, like Tesla in cars or Guinness in beers, avoid direct competition from central players in the industry and enjoy a sense of differentiation (Dawar and Bagga, 2015).

Unlike Dawar and Bagga (2015), who rely on traditional surveys to infer a brand's position in terms of centrality and distinctiveness, our automated approach provides an efficient and inexpensive way to create similar perceptual maps using publicly available social media data. Second, we provide a highly generalizable method that allows researchers to go beyond the numerical centrality-distinctiveness values and study the exact overlapping interests of their brand communities. For example, high across-category links of Starbucks with Wholefoods or Hyatt with American Air are not just outcomes of mere chance but an opportunity for future enquiry for all brand owners. Such brand knowledge can help managers identify potential target audiences, not as one assumed by management but one perceived by consumers' direct interests.

One obvious concern facing studies using users' interests from social media source is: to what extent do brand fans on social media represent the general population? Do certain Twitter brands accounts slant towards a specific audience, say younger individuals for instance. Initial inquiries have reported that followership data on Twitter successfully captures brand-attribute perceptions beyond demographic similarities (Culotta and Cutler, 2016; El Gazzar and Mourad, 2012). In our results section, we investigate this issue further by examining the demographic influence of Twitter followers. In the next section, we describe the process for inferring brand positioning using perceptual maps.

Methodology and Results

We propose a stepwise process for inferring brand position, see Figure 20. The first step is to obtain the transcendence vector of a brand b that highlights its associations across different p categories, t_b = $[t_b^{G_1} \quad t_b^{G_2} \quad t_b^{G_3} \ldots\ldots\ldots\ldots\ldots t_b^{G_p}]$. In the second step, the within and across-category associations are analyzed separately to obtain values on centrality and distinctiveness. In the transcendence vector t_b, brand b's association to its own category G_b essentially measures centrality ($t_b^{G_b}$). Brands that have high $t_b^{G_b}$ share strong consistent associations within their own category and are perceived to be central. Similarly, the across-category associations of a brand, in the transcendence vector $t_b^{G_b}$, are used to discern the distinctiveness of brands in consumers' minds. Using Opsahl (2010), both the number of connections to different categories as well as the weighted sum of links are used to calculate distinctiveness, D_i of brand b_i , as follows –

$$D_i = K_i \times \left(\frac{S_i}{K_i}\right)^{\alpha} = K_i^{1-\alpha} \times S_i^{\alpha}$$

where K_i is the number of non-zero category connections of b_i, S_i is the row sum of the transcendence vector and α is the tuning parameter which is set to 0.5. α = 0.5 ensures that both the number of categories, K_i, and weighted sum of links to different categories, S_i, are given equal importance in the formulation. The construct is refined further to separate a brand's own idiosyncratic associations from its category average. This helps to obtain the differential associations of a brand not shared by the overall category.

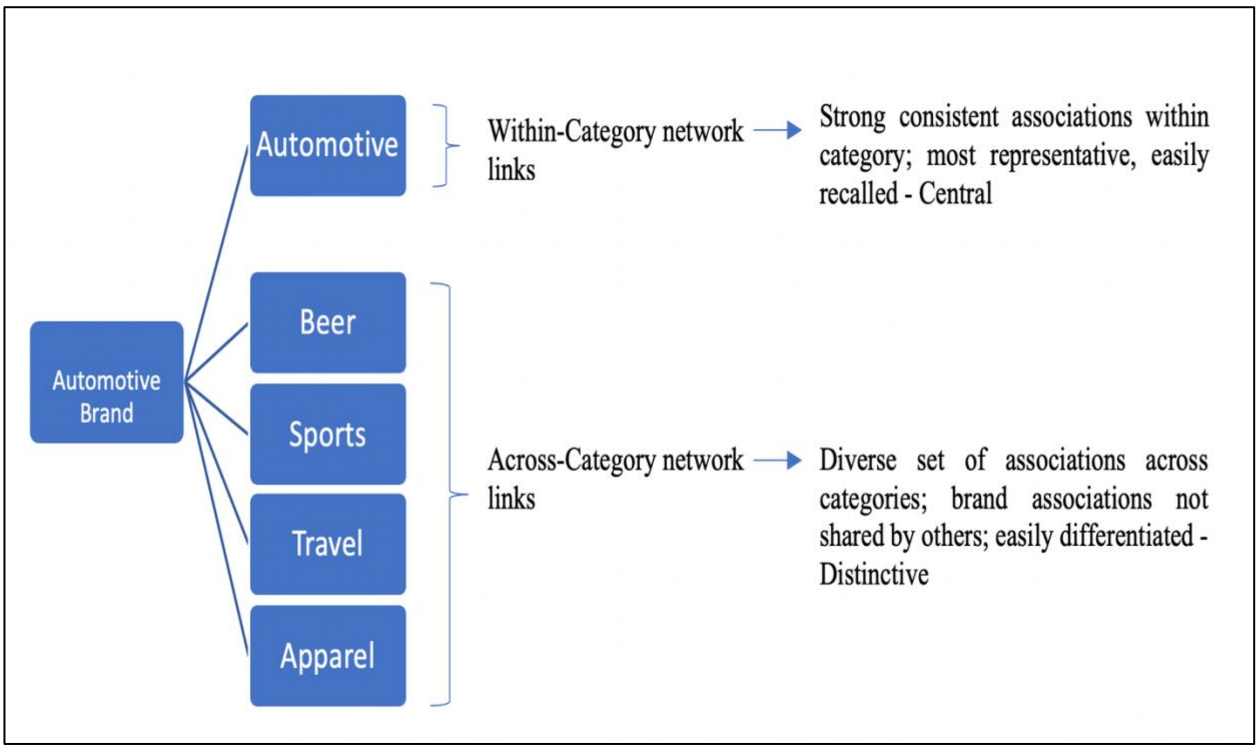


Figure 20. shows the stepwise process for inferring brand position.

The next step is to create perceptual maps along the two dimensions: *centrality* and *distinctiveness.* All values have been scaled from 0-1. Each data bubble corresponds to an individual brand and is sized proportional to its market share. Market share for all brands, except automotive, is collected from Euromonitor Passport (formally known as Global Market Information database or GMID). For the automotive industry, total vehicles sold for each brand are collected from the auto sales data tracking website GoodCarBadCar.net. We start by examining the perceptual map for the beer industry in Figure 21.

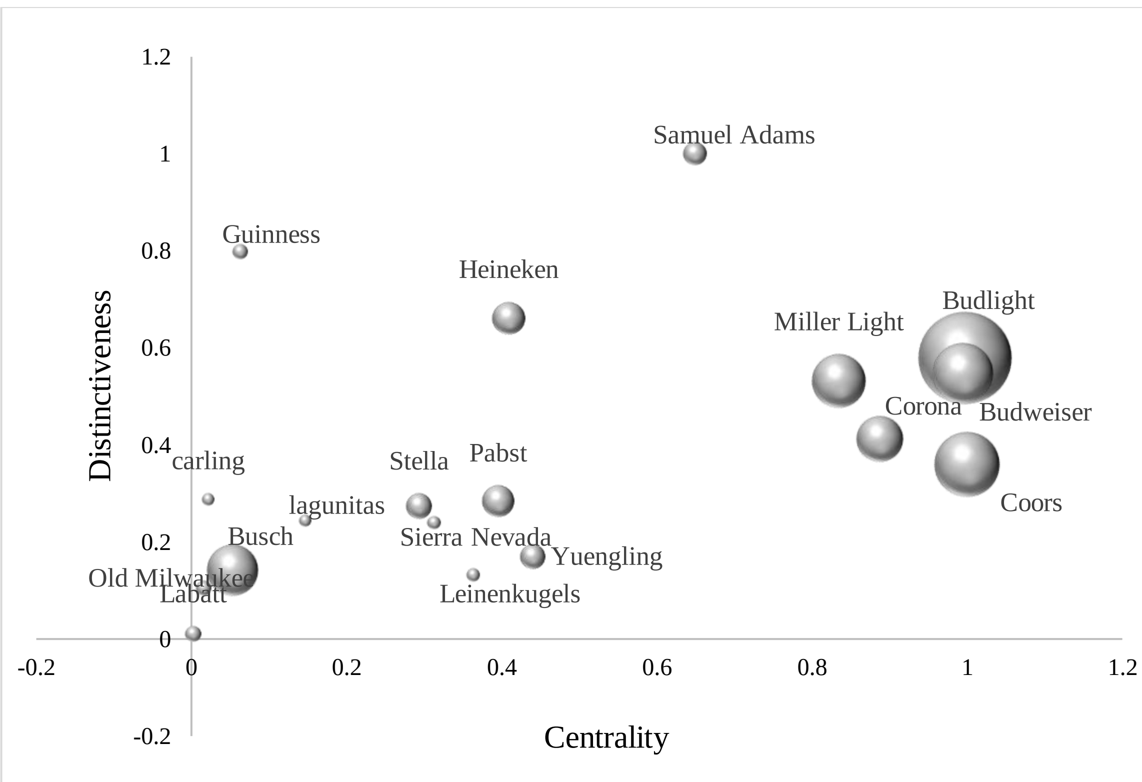


Figure 21. Perceptual map for the beer category

Budweiser, Budlight, Miller, Coors and Corona are positioned at the high centrality zone, accounting for more than 40% of the North American Beer Industry. High values of centrality imply a large overlap of audience interest from other beer brand communities. To examine the underlying network structure leading to these high centrality values, we extract the subgraph of beer brands in Figure 22. Indeed, we notice Budweiser, Budlight, Miller, Coors and Corona to be the core players with associations to almost all other beer brands. Due to their shared popularity among a large group of beer admirers and strong category associations, we call these brands *central*. These brands are most representative of their category and come first to mind to the widest group of people.

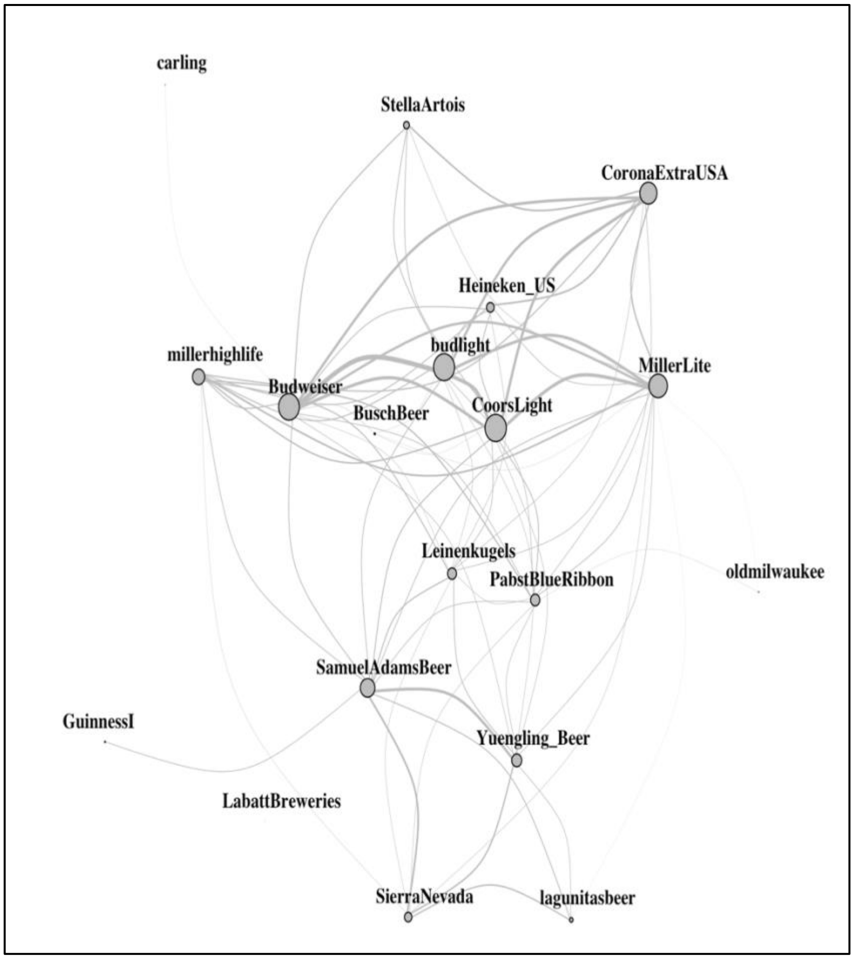


Figure 22. Subgraph of beer brands

Brands with low centrality on the perceptual map, such as Guinness and Carling, are the ones that do not share a lot of connections with other beer brands in the network. In other words, their brand communities do not significantly overlap with that of other beer brands. However, in the case of Guinness, a large proportion of its brand community (more than 90%) overlaps with brands from industries such as Luxury, Apparel/Footwear, Retail and Restaurants, leading to high across-industry network links (see net transcendence matrix in Figure 15). These across-category links represent unique brand associations not shared by the overall category and help to differentiate a brand from its competitors. Therefore, such brands (e.g. Guinness) are perceived to be *distinctive*.

Another interesting example is that of Samuel Adams. The brand shares a large proportion of its brand community with many non-beer brands (leading to high across-industry links) as well as multiple beer brands (leading to moderately high within-industry links); positioning it as highly distinct and also moderately central. There is also a set of brands that rank low both on centrality and distinctiveness. Examples include old Milwaukee, Pabst, Labatt and few others. Low centrality values on both axes means that these brands do not share consumer co-interest with other brands in the network, regardless of the industry type. These brands usually occupy a peripheral position in the network. Figure 23 shows the perceptual map for the automotive industry.

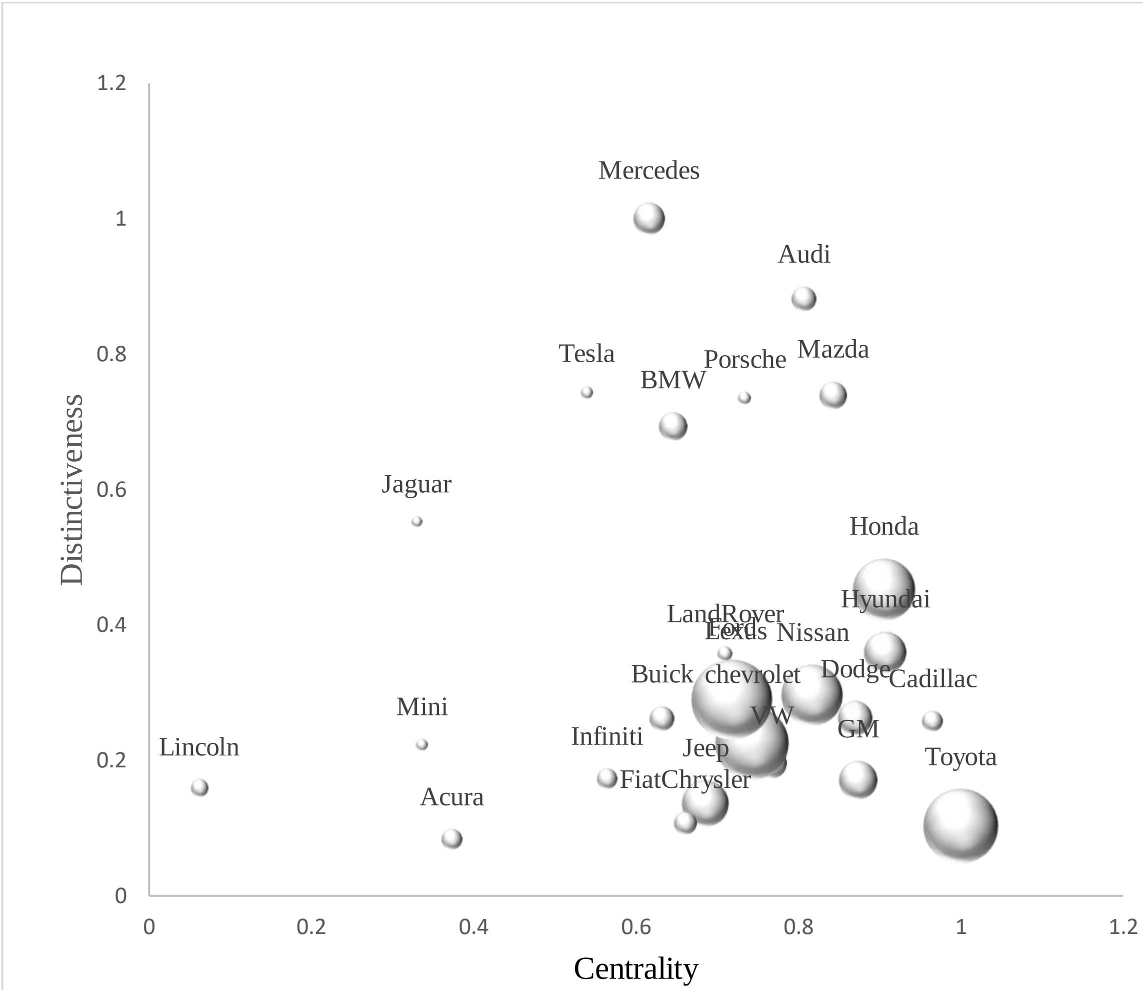


Figure 23. Perceptual Map for the Automotive Category

Generally considered to be conventional brands, Ford, Chevrolet, Toyota and Honda fall in the high centrality zone and account for more than 40% sales. These central brands are the ones that consumers perceive as core in a category. Their low across-industry centrality reflects the fact that they share brand audience with mostly auto brands and no other categories. They may lack on distinctiveness but are very popular among consumers interested in auto. On the other hand, brands such as, Mercedes, Audi, Porsche and Tesla share brand audience with many non-automotive brands – leading to high across-industry scores and a therefore a distinctive brand image. For instance, Mercedes is the most distinctives brand with associations from luxury, airlines, apparel and other non-automotive brand categories (see net transcendence matrix in Figure 13). In Figure 13, we uncover the interests of Toyota's and Mercedes' audience by analyzing the data on what other brands they are interested in. It is interesting to note the contrast between Mercedes' and Toyota's brand communities. Mercedes has a diverse brand audience compared to Toyota's predominantly automotive followers. This speaks of the conventional central image of Toyota among cars in general. Perceptual maps for the other industries can be obtained similarly.

One major concern facing studies using social media data is how online brand fans represent the general customer base of brands. Studies have reported that the use of Twitter brand followers for marketing research has proven to be reliable source for inferring brand perceptions, free from any major demographic biases (Culotta and Cutler, 2016). We investigate this issue further by running two sets of regression analyses, including age and gender of Twitter brand followers, to identify any major demographic factors affecting our final brand perceptions - centrality and distinctiveness. We collect the publicly available demographic information of Twitter brand followers from Zoomph.com for a set of 142 brands. The model (N = 142) is given below –

$$Centrality_{network} = Gender + Age_1 + Age_2 + Age_3 + Age_4;$$

$$Distinctiveness_{network} = Gender + Age_1 + Age_2 + Age_3 + Age_4$$

$Centrality_{network}$ and $Distinctiveness_{network}$ are measures derived from the brand network. Gender is the percentage of a brand's followers that are males. $Age_{1\text{-}4}$ are the percentage of brand followers that are 18-24, 23-38, 39 - 54, and 55 + years old. For both models, we find none of the age variables significant ($p > 0.05$). We do find gender to be significant for both cases, $p < 0.01$, with a slightly more male audience predicting centrality and distinctiveness. However, the adjusted R_2 for models is very small, $r_2 < 0.2$, meaning that the effect is not substantially driving the network measures.

Validation

To validate the effectiveness of our methodology, we compare the brand ratings from our automated approach with directly elicited survey ratings. We conduct the survey through Amazon Mechanical Turks, AMT, which has proven to a reliable source for conducting social science research (Buhrmester et al., 2011; Mason and Suri, 2012). Overall, researchers find AMT to be a valuable subject recruitment tool with demographic characteristics of MTurks being more representative and diverse than student samples typically used in experimental research (Berinsky et al, 2012). Particularly for the purpose of inferring brand image perceptions, Culotta and Cutler (2016) find survey ratings collected on AMT to be free from any potential demographic biases.

Although AMT is generally considered a valid subject recruitment tool, we ask our survey respondents to report information on their income, age and gender to account for any demographic influence in our sample. Participants are required to be located in the United States and be 18+ years of age. A successful completion record of at least 100 prior assignments with 90% acceptance rate is required for all survey respondents. Participants were asked to rate a brand, on a scale of 1 – 5, on centrality and distinctiveness. Brands were grouped by sector and separate surveys, consisting of 250 participants each, were conducted for beer and automotive brands. The brand order was randomized, and attention filters were included to identify any invalid responses. A separate column on brand recognition is provided where respondents can mark brands not familiar to them. Recognition rates varied from 100% for brands like Honda, Toyota and Budlight to 50% for brands like Carling and Lagunitas.

Finally, we computed average survey ratings for each brand along the dimensions, centrality and distinctiveness. We, then, create scatter plots for Survey vs Network measures in Figure 24, 25, 26 and 27. We find the correlations to be sufficiently high with $p < 0.001$ in every case, encouraging the use of brand networks for future marketing research. The average correlation coefficient is 0.7, with the strongest correlation of 0.8 for beer brands along the *distinctive* dimension.

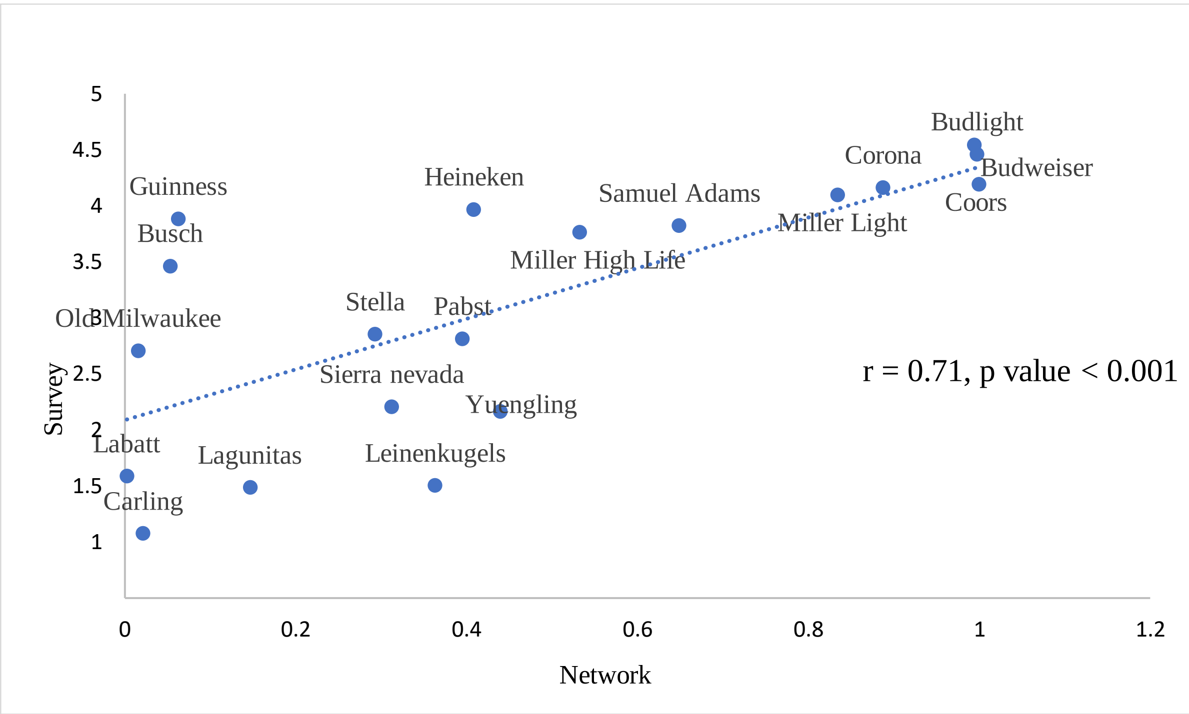


Figure 24 Survey vs Network Correlation for Centrality: Beer brands

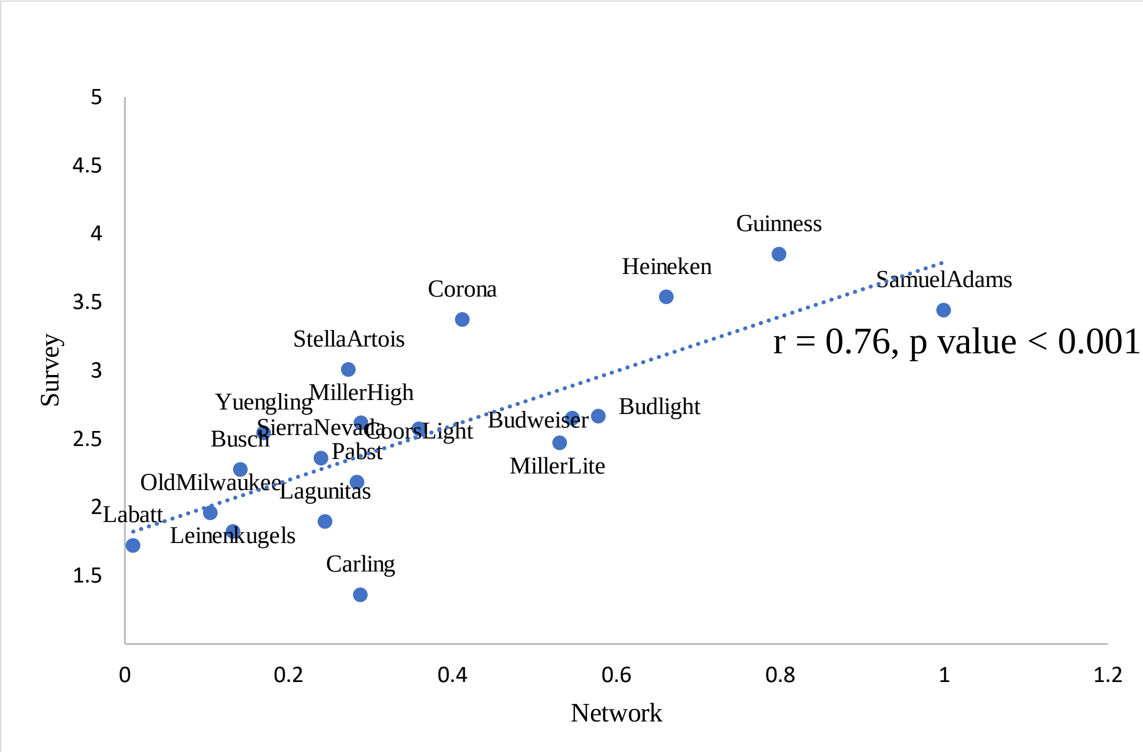


Figure 25. Survey vs Network Correlation for Distinctiveness: Beer brands

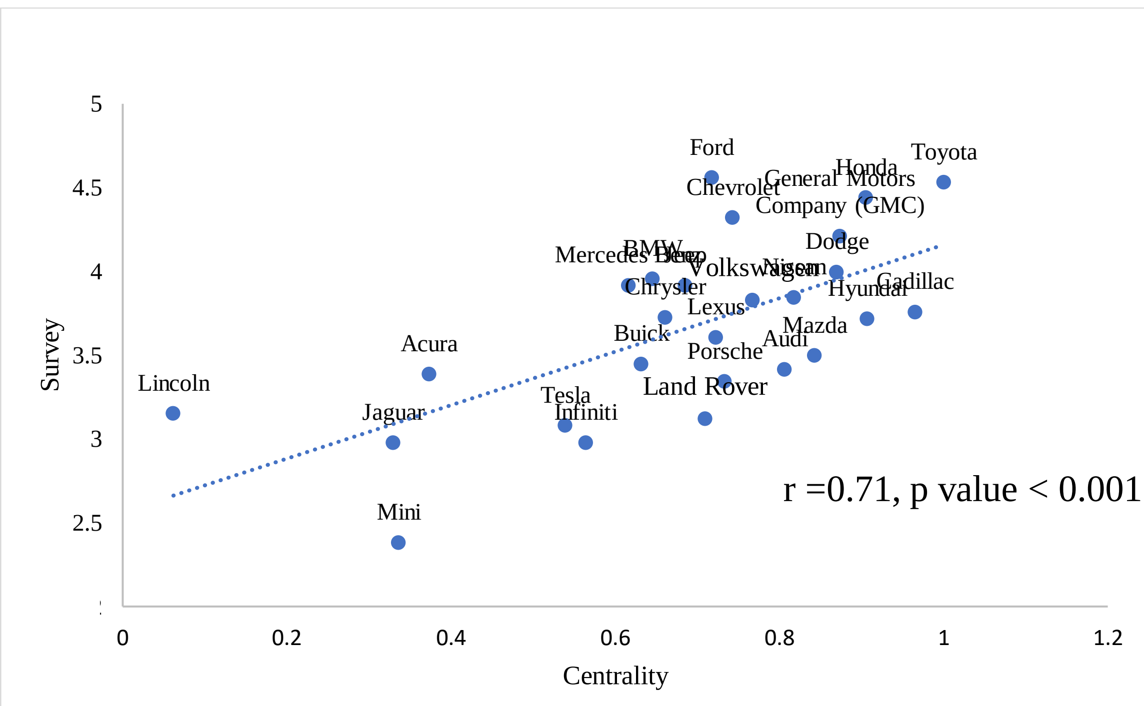


Figure 26. Survey vs Network Correlation for Centrality: Car brands

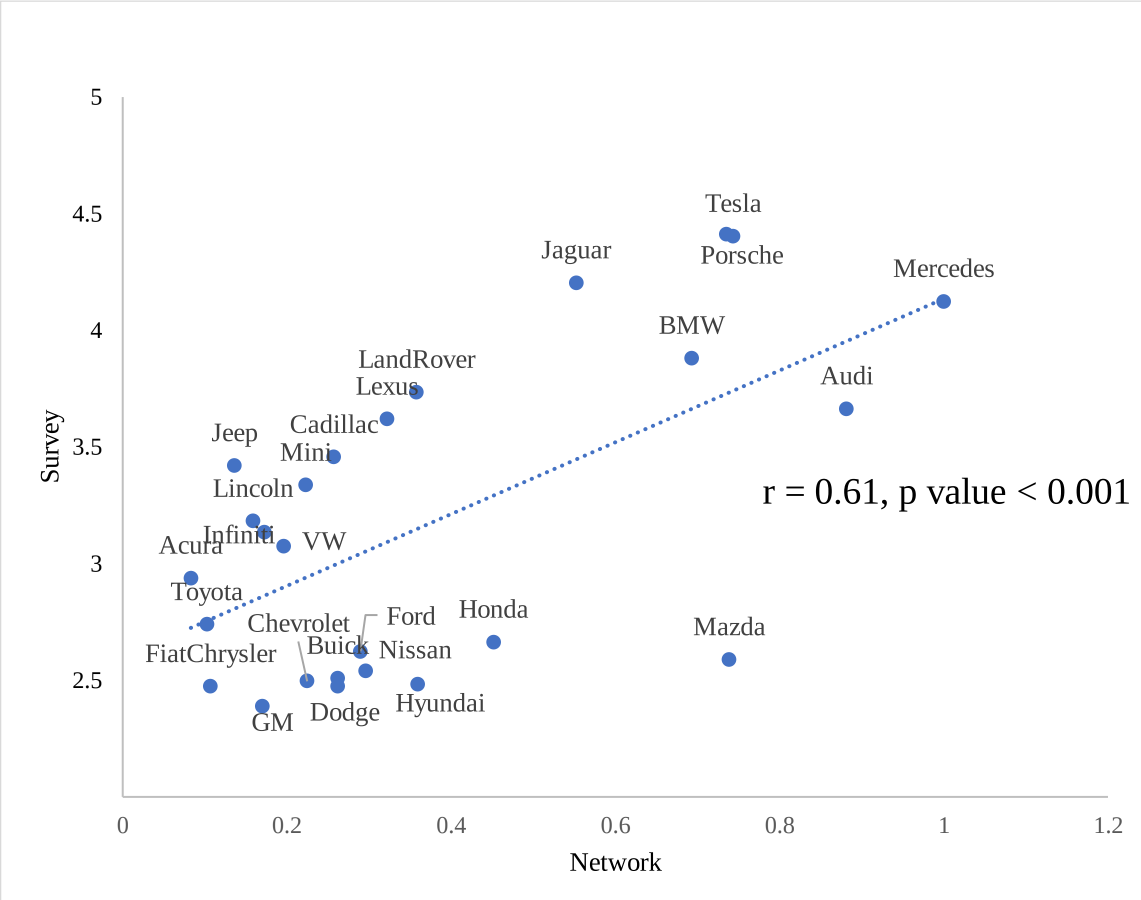


Figure 27. Survey vs Network Correlation for Distinctiveness: Car brands

For the set of beer brands, we notice the scatter points to be evenly distributed along the best line, implying that the high correlation values are not driven by a handful of observations. For the car brands we do notice a slight positive skew, where most brands have low values on distinctiveness and a handful of brands like Tesla, Porsche and Audi have very high distinctive ratings. To ensure that the high correlation is not driven by these few brands, we compute the Spearman Rank coefficient which is less sensitive to extreme observations.

The average Spearman coefficient is 0.7 across all categories, suggesting that the strong association still remains.

We perform additional validation checks to ensure that the demographics of the survey respondents are not influencing their perception on centrality and distinctiveness. Figure 28 shows the demographic characteristics of the survey respondents. The gender distribution is fairly balanced, and the mean sample income lies between 30K – 99K. Majority of the survey respondents are between 25 – 44 years, which corresponds to average age of brand followers on Twitter.

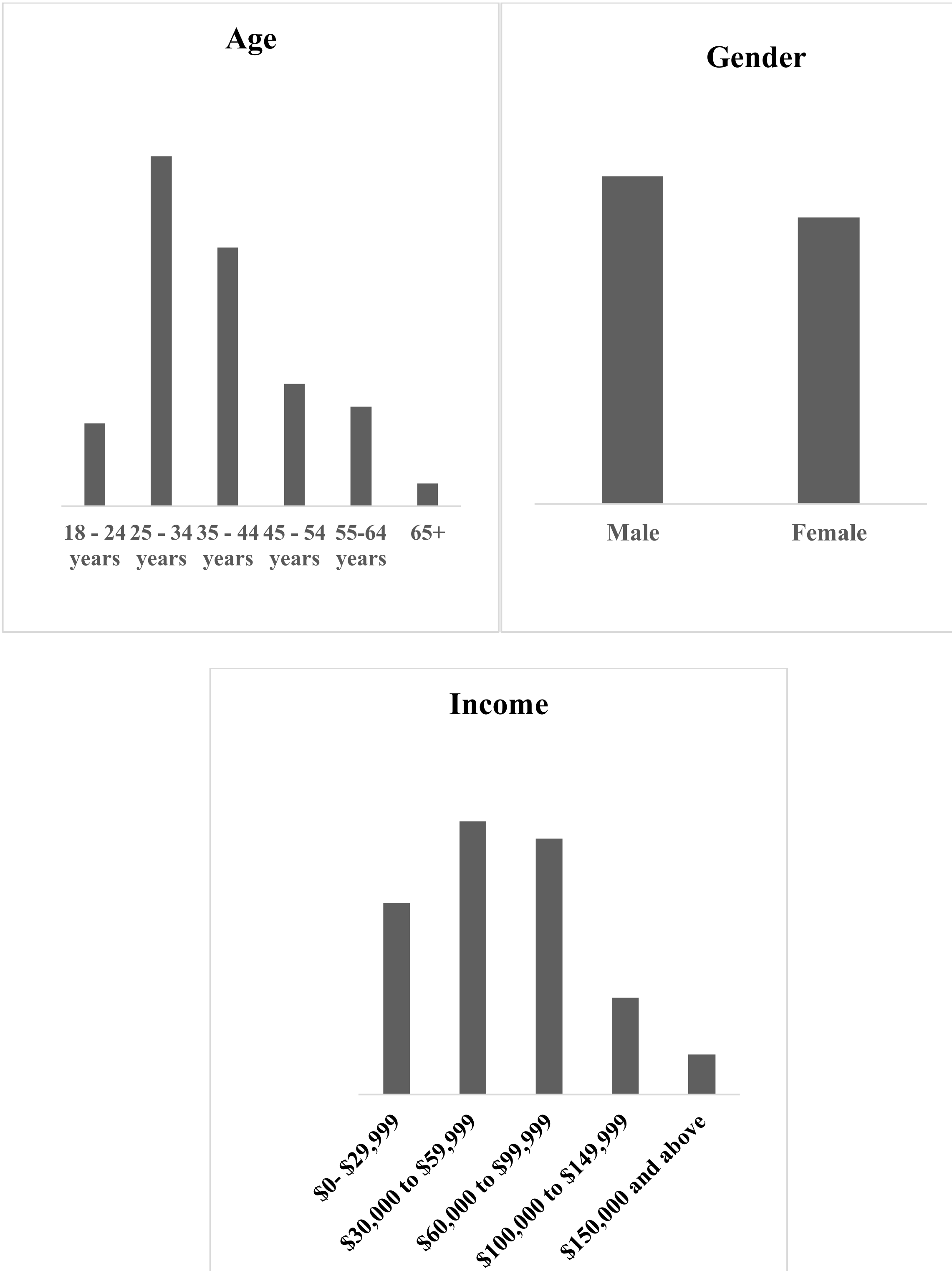


Figure 28. Demographics of survey respondents

We computed the correlations between survey ratings of different demographic categories to identify any major shifts in brand perceptions based on gender, age or income. We find consistently high correlations, average r = 0.8, for all categories (included in Appendix D). This suggests that the survey ratings are not influenced by the age, gender or income levels of the participants. Further, high correlation of the survey ratings with our automated approach prove that the overall brand perceptions on centrality and distinctiveness are free of any demographic biases. Naturally, there may be other sources of biases, such as, both Twitter followers and AMT respondents spend more time online than an average consumer. However, this is beyond the scope of our

Discussion

In the past decade, the importance of effectively mining consumer perceptions to infer brand positioning has become an emerging topic of interest for all brand owners. Typically, marketers have relied on surveys and focus groups to measure how consumers perceive brands as similar or different from their competitors. Even though surveys provide a direct way to capture consumer-brand perceptions, they are limited in terms of reach across consumers and brands. Recruiting and maintaining a large set of survey participants also leads to cost and time constraints for managers.

With the digital media revolution and widespread social media adoption, digital footprints of consumers provide an excellent source for inferring a brand's position. Some of the popular digital traces commonly used in brand management include online user generated content and browsing behavior. Though current big data technologies allow easy access to this type of digital content, such information may suffer from potential limitations such as biased content (Gao et al, 2015), substantial manual intervention (Culotta and Cutler, 2016) and privacy regulations (Bucklin and Sismeiro, 2009).

Addressing this issue, we propose a big data approach for inferring brand positioning using implicit brand-to-brand networks. Unlike social networks, a link in a brand network does not reflect a node's decision to voluntarily connect with others; instead it reflects *aggregated preferences of many Twitter users (i.e. their co-following patterns on Twitter).* Thus, links forming the network are "tacit" - an outcome of shared interests. The key contribution of our work is to introduce these new types of digital artifacts for capturing audience interests across a large ecosystem of brands, without extensive pre-processing and machine learning expertise.

Compared to extant data mining approaches that rely on substantial human intervention, this unsupervised automated approach lets practitioners study the relative positioning of their brand not only against a set of common competitors but to any other brand in the ecosystem. Specifically, the inclusion of network derived measures allows one to go beyond numerical brand ratings and conduct a granular assessment of why that rating occurs. For instance, disentangling the links of Mercedes's brand community across categories shows how the brand is distinct from others in the same category. With an average correlation of 0.7 with directly elicited survey ratings, this large-scale data driven approach provides reliable means to automatically infer brand position.

VI. CONCLUSION AND FUTURE RESEARCH

Your consumers matter, and so do their preferences. Managers may often feel that they know what their brand represents. However, this apriori belief is more reflective of their aspirations from intended marketing strategies, rather than actual consumer beliefs (Henderson et al, 1998). In the past decade, the importance of effectively eliciting brand associations has become an emerging topic of interest for all brand owners. Besides the obvious tangible associations of a brand that relate to its physical attributes (e.g. mileage of a car, horsepower), there are another set of intangible associations that the go beyond the functional attributes of a brand and even transcend categories. Despite being vital for various managerial decisions (such as co-branding, licensing and brand extensions), cross-category associations of brands are an understudied topic in literature. Some of the greatest brands in the world have defied category norms and transcended their initial market boundaries to achieve much more, for example Apple and GE (Keller, 2014).

Information networks, with their inherent ability to incorporate user choices over a large brand ecosystem, provide a novel and scalable approach for inferring key brand associations over time. While traditional methods consider a limited set of brands and consumers, the brand network captures millions of user-brand interactions across hundreds of brands in multiple categories; thus, uncovering a broader picture on both within-category competition and across-industry complementarities. With most of the data collection and network analyses automated, the brand network acts as an effective business intelligence tool to deliver insights into users and brands in a timely manner.

For a brand network to provide reliable branding knowledge, statistical inference on the network is required to establish that associations between brands do not arise from randomness, but from specific user interests. We employ Exponential Random Graph Models (ERGM) from network science to understand the key factors which drive associations between brands. The ERGM model reveals a mix of network and individual level brand characteristics that help explain the formation of links between brands; thereby disclosing a set of latent brand characteristics that users determine while co-following brands on social media. Some of the significant effects include homophily based on category, cross-category interactions between certain pairs (such as Automotive-Sports, Travel-Restaurants, Apparel-Personal Care) and frequency of a brand's engagement with online fans. Linking to previous literature, cross-category associations of brands are known to be crucial for coordinated promotions, embedded premiums and positioning strategies (Henderson, and Arora 2010; Leeflang and Parreño-Selva 2012); however, there is little or no evidence on identifying these cross-category effects based on empirical social media data. Our study helps to bridge this gap in the existing literature.

After the statistical significance of the brand network is established, we introduce an automated scalable approach for eliciting the associations of a brand over time. A new construct, brand transcendence, is defined to capture the diverse associations of a brand onto new categories. By incorporating asymmetry into the transcendence vector, we are able to determine the proportion of a brand's fans interested in a new category. Keller (2003), in his seminal work on brand extensions, mentions that higher the shared associations between the brand and the new category, greater is the perception of fit. Our study provides a new methodological tool to identify these shared associations between brands and categories using publicly available social media data. The new construct, transcendence, measures the fit of brands into new categories; not as one identified by management apriori but as one perceived by the direct interests of users on social media. Overall, the analysis is conducted at two time periods to track changes in net brand transcendence; this allows brand managers to assess the effectiveness of their marketing strategies or gauge the impact of external events on their brand's associations in users' minds.

Furthermore, depending upon the marketing objectives, the results can be visualized at three different levels : category-category, brand-category and brand-brand. The different levels of analyses (c-c, b-c and b-b) help to answer why certain co-branding opportunities are more viable than others. For example, assessing the c-c associations provides a window into the broader interests of automotive fans and shows specific categories (such as technology, sports, dining and luxury) are more viable for co-branding than others. Going one step further, b-c associations capture the transcendence of brands onto new categories and show why certain categories for co-branding opportunities may hold more promise than others. For instance, investigating the transcendence vector of Mercedes, we notice how a large percentage of Mercedes' fans are engaging with luxury, technology, travel and sports – making them as suitable categories for co-branding. Finally, at b-b level, we find brands such as Microsoft, Nike, Monster Energy Drink, Puma, Google and Chanel among the closest neighbors of Mercedes in the network. Clearly Mercedes fans, as revealed through the previous b-c analysis, are interested in technology, sports, and luxury. We believe that the different types of results presented by our method (c-c, b-c, b-b) may help inform managers why certain co-branding opportunities hold more promise than others.

Managers can use brand networks for other marketing goals as well. First, competitor analysis helps to identify the closest competitors of a brand and uncover differential associations of each brand in the competing group. Second, employing standard clustering algorithms on the network allows managers to identify communities of brands with similar user preferences; this provides an alternate view of consumer segments based on direct data on their diverse interests. Both these practical applications of the brand network can be easily implemented by marketing practitioners to get timely estimates on their brand's competitive landscape. The analysis allows managers to track their own audience' interests as well as those of their competitors (or allies) in a fast and inexpensive manner.

Perceptual maps, obtained from the network, help uncover the competitive landscape of brands along the dimensions: *centrality* and *distinctiveness*. To investigate the usefulness of our proposed methodology, we validate the findings from our automated approach against external survey ratings and conduct extensive robustness checks to ensure reliability of underlying Twitter data. This approach, leveraging large scale social media data provides an alternative to traditional survey and focus group-based methods which can be expensive and cumbersome, and also limited in terms of reach across consumers and brands. The core contribution is a new methodological tool that not only provides an efficient way to infer a brand's position, but also give granular insights into why that positioning occurs. For instance, disentangling the links of Mercedes's brand community in the transcendence matrix shows how the brand is distinct from others in the same category.

From a methodological standpoint, the implicit brand networks, introduced in the paper, condense the high dimensional interest space of millions of brand followers to a reduced form representation which is more amenable for research and business application purposes. Compared to extant digital approaches that rely on extensive pre-processing and machine learning expertise, our straightforward automated approach is largely unsupervised and makes minimal assumptions on the underlying data. More specifically, with most of the data collection and network analyses automated, the brand network acts as an effective business intelligence tool to deliver insights into potential co-branding opportunities in a close to real time setting.

Overall, our approach offers a number of benefits to marketers, and also identifies avenues for future research. Though we use Twitter brand communities for our analysis, it will be interesting to compare Facebook and Instagram “fan” relationships. Brand networks on different platforms can vary based on a number of factors such as user demographics, user age, industry or brand’s marketing strategy. Consistency of brand associations across different platforms will provide additional validation findings and reveal meaningful insights if substantial differences are observed. The analyses presented in this paper is limited to brands which maintain active Twitter accounts; while most of the top brands do have active Twitter pages, a few may not. Given the large set of brands considered (~500), this does not impact our analyses and findings. Considering that certain brands may be active on specific social media platforms, it will be interesting to examine combined analyses of multiple brand networks. Varying content on brand pages and different approaches taken by brands may also affect consumers' decisions to follow brands. For instance, we notice a few brands to post general news content rather than actual brand related information. While this appears to occur rarely in our data, it can be useful for future research to identify and adjust for such cases. One could, also, extend the analysis to bi-partite graphs and create massive explicit networks representing user-brand followership on social media. This information can be used for multiple business scenarios including targeted marketing, deconstruction of online consumer behavior and demand prediction. Other potential extensions include enriching the underlying data with demographics of Twitter users to ascertain that inter-brand relationships arise due to genuine user choices and marketing efforts, and not due to demographics of brand followers.

Overall, the aim of our work is to provide useful tools to researchers and practitioners interested in automatically monitoring brand associations and positioning over time. Large scale data focused methods for brand management are relatively new and present many opportunities for future research. We hope the methods introduced in this dissertation serve as a foundation for researchers interested in leveraging implicit brand networks for gaining insights into consumers and brands.

VII. CITED LITERATURE

Arrigara, M., & Levina, N. (2008). Social dynamics in online cultural fields. *ICIS 2008 Proceedings*, 120.

Aaker, D.A. (1991), Managing Brand Equity, The Free Press, New York, NY.

Aaker, D. A. (1996). Measuring brand equity across products and markets. *California management review*,38(3).

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike* (pp. 199-213). Springer, New York, NY.

Anderson, J. R., & Bower, G. H. (1974). A propositional theory of recognition memory. *Memory & Cognition*, *2*(3), 406-412.

Ahluwalia, R. (2008). How far can a brand stretch? Understanding the role of self-construal. *Journal of Marketing Research*, 45(3), 337-350.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: memory retrieval as a model case. *Psychological review*, 102(1), 68.

Adobe Digital Index, The Social Intelligence Report - Q1, 2014, available at: https://www.cmo.com/content/dam/CMO_Other/ADI/SocialIntelligence_Q12014/Q1_2014_social_intelligence_report.pdf (retrieved June 8, 2018).

Baker, F., L. Hubert. (1981). The analysis of social interaction data. *Soc. Methods Res*. 9 339–361

Barnett WP. 1993. Strategic deterrence among multipoint competitors. *Industrial and Corporate Change* 2(2):249–278.

Brynjolfsson, E., & McElheran, K. (2019). Data in Action: Data-Driven Decision Making and Predictive Analytics in US Manufacturing. Available at *SSRN* 3422397.

Byshkin, M., Stivala, A., Mira, A., Robins, G., & Lomi, A. (2018). Fast maximum likelihood estimation via equilibrium expectation for large network data. *Scientific reports*, 8(1), 1-11.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?. *Perspectives on psychological science*, 6(1), 3-5.

Batra, R., Lenk, P., & Wedel, M. (2010). Brand extension strategy planning: Empirical estimation of brand–category personality fit and atypicality. *Journal of marketing* research, 47(2), 335-347.

Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the national academy of sciences*, *101*(11), 3747-3752.

Berger, J., & Heath, C. (2007). Where consumers diverge from others: Identity signaling and product domains. *Journal of Consumer Research*, *34*(2), 121-134.

Brexendorf, T. O., & Keller, K. L. (2017). Leveraging the corporate brand: the importance of corporate brand innovativeness and brand architecture. *European Journal of Marketing*, 51(9/10), 1530-1551.

Boush, D. M., & Loken, B. (1991). A process-tracing study of brand extension evaluation. *Journal of marketing research*, 28(1), 16-28.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *science*, 323(5916), 892-895.

Bucklin, R. E., & Sismeiro, C. (2009). Click here for Internet insight: Advances in clickstream data analysis in marketing. *Journal of Interactive marketing*, 23(1), 35-48.

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *science*, 286(5439), 509-512.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. *Political analysis*, 20(3), 351-368.

Barone, M. J., & Jewell, R. D. (2013). The innovator's license: A latitude to deviate from category norms. *Journal of Marketing*, 77(1), 120-134.

Cutright, K.M., Bettman, J.R. and Fitzsimons, G.J. (2013), “Putting brands in their place: how alack of control keeps brands contained”, *Journal of Marketing Research*, Vol. 50 No. 3, pp. 365-377.

Chatterjee, S., & Diaconis, P. (2013). Estimating and understanding exponential random graph models. *The Annals of Statistics*, *41*(5), 2428-2461.

Culotta, A., & Cutler, J. (2016). Mining brand perceptions from twitter social networks*. Marketing Science*, 35(3), 343-362.

Carpenter, G. S., & Nakamoto, K. (1989). Consumer preference formation and pioneering advantage. *Journal of Marketing research*, 26(3), 285-298.

Cutright, K.M. (2012), “The beauty of boundaries: when and why we seek structure in consumption”, *Journal of Consumer Research*, Vol. 38 No. 5, pp. 775-790.

Cutright, K.M., Bettman, J.R. and Fitzsimons, G.J. (2013), “Putting brands in their place: how a lack of control keeps brands contained”, *Journal of Marketing Research*, Vol. 50 No. 3, pp. 365-377.

Childers, T. L., & Rao, A. R. (1992). The influence of familial and peer-based reference groups on consumer decisions. *Journal of consumer research*, *19*(2), 198-211.

Contractor, N. S., Wasserman, S., & Faust, K. (2006). Testing multitheoretical, multilevel hypotheses about organizational networks: An analytic framework and empirical example. *Academy of Management Review*, *31*(3), 681-703.

Chung, S., Singh, H., & Lee, K. (2000). Complementarity, status similarity and social capital as drivers of alliance formation. *Strategic management journal*, *21*(1), 1-22.

Das, Sanjiv R, Mike Y Chen. 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science* 53(9) 1375{1388.

Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., Nanavati, A. A., & Joshi, A. (2008, March). Social ties and their relevance to churn in mobile telecom networks. In *Proceedings of the 11th international conference on Extending database technology: Advances in database technology* (pp. 668-677). ACM.

Dawar, N., & Bagga, C. K. (2015). A better way to map brand strategy. *Harvard Business Review*, *93*(6), 90-97

Damangir, S., Du, R. Y., & Hu, Y. (2018). Uncovering patterns of product co-consideration: a case study of online vehicle price quote request data. *Journal of Interactive Marketing*, 42, 1-17.

DeSarbo, W. S., & Grewal, R. (2007). An alternative efficient representation of demand-based competitive asymmetry. *Strategic Management Journal*, 28(7), 755-766.

El Gazzar, N., M. Mourad. 2012. The effect of online communication on corporate brand image. *International Journal of Online Marketing*.

Elrod, Terry, Gary J. Russell, Allan D. Shocker, Rick L. Andrews, Lynd Bacon, Barry L. Bayus, et al. (2002), "Inferring Market Structure from Customer Response to Competing and Complementary Products," *Marketing Letters*, 13 (3), 221–32.

Erdem, T. (1996). A dynamic analysis of market structure based on panel data. *Marketing Science*, *15*(4), 359-378.

Etter, M., Plotkowiak, T., & Stanoevska-Slabeva, K. (2011). CSR Communication Strategies for Twitter: Microblogging as a Tool for Public Relations.

Escalas, J. E., & Bettman, J. R. (2003). You are what they eat: The influence of reference groups on consumers' connections to brands. *Journal of consumer psychology*, *13*(3), 339-348.

Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the american Statistical association*, 81(395), 832-842.

Fawcett, T., & Provost, F. (1997). Adaptive fraud detection. *Data mining and knowledge discovery*, *1*(3), 291-316

Fournier, S., & Lee, L. (2009). Getting brand communities right. *Harvard business review*, *87*(4), 105-111.

Forbes (2018) Social Media Is Increasing Brand Engagement and Sales. Accessed January 2018, https://www.forbes.com/sites/tjmccue/2018/06/26/social-media-is-increasing-brand-engagement-and-sales/#167e26217cb3

Farquhar, P. H., & Herr, P. M. (1993). The dual structure of brand associations. *Brand equity & advertising: Advertising's role in building strong brands*, 263-277.

Gao, G. G., Greenwood, B. N., Agarwal, R., & McCullough, J. (2015). Vocal minority and silent majority: How do online ratings reflect population perceptions of quality?. *Mis Quarterly*, 39(3), 565-589

Goes, P. B., Lin, M., & Au Yeung, C. M. (2014). "Popularity effect" in user-generated content: Evidence from online product reviews*. Information Systems Research*, 25(2), 222-238.

Ghosh, A., Ranganathan, R., & Rosenkopf, L. (2016). The impact of context and model choice on the determinants of strategic alliance formation: Evidence from a staged replication study. *Strategic Management Journal*, 37(11), 2204-2221.

Green, P. E., Carmone, F. J., & Smith, S. M. (1989). *Multidimensional scaling: Concepts and applications*. Allyn and Bacon.

Godes, D., & Mayzlin, D. (2009). Firm-created word-of-mouth communication: Evidence from a field test. *Marketing science*, 28(4), 721-739.

Gulati, R., & Gargiulo, M. (1999). Where do interorganizational networks come from?. *American journal of sociology*, *104*(5), 1439-1493.

Gulati R. 1995. Social structure and alliance formation patterns: a longitudinal analysis. *Administrative Science Quarterly* 40(4): 619–652.

Holland, P. W., & Leinhardt, S. (1977). A method for detecting structure in sociometric data. In *Social Networks* (pp. 411-432). Academic Press.

He, S., Rui, H., & Whinston, A. B. (2017). Social media strategies in product-Harm Crises. *Information Systems Research*, *29*(2), 362-380.

Hill, S., Provost, F., & Volinsky, C. (2006). Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science*, 21(2), 256-276.

Hoffman, D. L., & Fodor, M. (2010). Can you measure the ROI of your social media marketing?. *MIT Sloan Management Review*, *52*(1), 41.

Howard, M. D., Withers, M. C., & Tihanyi, L. (2017). Knowledge dependence and the formation of director interlocks. *Academy of Management Journal*, 60(5), 1986-2013.

Henderson, G. R., Iacobucci, D., & Calder, B. J. (1998). Brand diagnostics: Mapping branding effects using consumer associative networks. *European Journal of Operational Research*, 111(2), 306-327.

Henderson, T., & Arora, N. (2010). Promoting brands across categories with a social cause: Implementing effective embedded premium programs. *Journal of Marketing*, *74*(6), 41-60.

Jackson MO (2010) *Social and Economic Networks* (Princeton University Press, Princeton, NJ).

Joiner, C. and Loken, B. (1998), "The inclusion effect and category-based induction: theory and application to brand categories", *Journal of Consumer Psychology*, Vol. 7 No. 2, pp. 101-129.

John, D. R., Loken, B., Kim, K., & Monga, A. B. (2006). Brand concept maps: A methodology for identifying brand association networks. *Journal of marketing research*, *43*(4), 549-563.

Jahn, B. and Kunz, W. (2012) How to transform consumers into fans of your brand. *Journal of Service Management* 23(3): 344–361

Kannan, P. K., & Sanchez, S. M. (1994). Competitive market structures: a subset selection analysis. *Management Science*, *40*(11), 1484-1499.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53, 59–68.

Kim, J. B., Albuquerque, P., & Bronnenberg, B. J. (2011). Mapping online consumer search. *Journal of Marketing Research*, 48(1), 13-27.

Kim, J. Y., Howard, M., Cox Pahnke, E., & Boeker, W. (2016). Understanding network formation in strategy research: Exponential random graph models. *Strategic management journal*, 37(1), 22-44.

Kotler, P., K. L. Keller. 2006. *Marketing Management*, 12th ed. Prentice-Hall, Upper Saddle River, NJ.

Kim, A. J., & Ko, E. (2012). Do social media marketing activities enhance customer equity? An empirical study of luxury fashion brand. *Journal of Business research*, *65*(10), 1480-1486.

Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social networks*, *10*(4), 359-381.

Keller, K. (2003). Brand Synthesis: The Multidimensionality of Brand Knowledge. *Journal of Consumer Research, 29*(4), 595-600. doi:10.1086/346254

Keller, K. L., Sternthal, B., & Tybout, A. (2002). Three questions you need to ask about your brand. *Harvard business review,* 80(9), 80-89.

Keller, K. L., & Lehmann, D. R. (2006). Brands and branding: Research findings and future priorities. *Marketing science*, 25(6), 740-759.

Kuksov, D., Shachar, R., & Wang, K. (2013). Advertising and consumers' communications. *Marketing Science*, *32*(2), 294-309.

Lomi, A., & Pallotti, F. (2012). Relational collaboration among spatial multipoint competitors. *Social networks,* 34(1), 101-111.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., ... & Jebara, T. (2009). Computational social science. *Science*, *323*(5915), 721-723.

Levy, S. J. (1981). Intepreting consumer mythology: a structural approach to consumer behavior. *Journal of marketing*, 45(3), 49-61.

Levina, N., & Arriaga, M. (2014). Distinction and status production on user-generated content platforms: Using Bourdieu's theory of cultural production to understand social dynamics in online fields. *Information Systems Research*, 25(3), 468-488.

Luo, X., Zhang, J., and Duan, W. 2013. "Social media and firm equity value," *Information Systems Research* (24:1), pp 146-163.

Leeflang, P. S., & Parreño-Selva, J. (2012). Cross-category demand effects of price promotions. *Journal of the Academy of Marketing Science*, *40*(4), 572-586.

Loken, B., & Ward, J. (1990). Alternative approaches to understanding the determinants of typicality. *Journal of Consumer Research*, *17*(2), 111-126.

Lei, J., Dawar, N., & Lemmink, J. (2008). Negative spillover in brand portfolios: Exploring the antecedents of asymmetric effects. *Journal of marketing*, 72(3), 111-123.

Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research*, 48(5), 881-894.

Laroche, M., Habibi, M. R., Richard, M. O., & Sankaranarayanan,R. (2012). The effects of social media-based brand communities on brand community markers, value creation practices, brand trust and brand loyalty. *Computers in Human Behavior*, 28(5), 1755-1767.

Manning, Christopher D, Prabhakar Raghavan, Hinrich Schutze. 2008. *Introduction to information retrieval* , vol. 1. Cambridge: Cambridge University Press.

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information systems research*, 15(4), 336-355

McDonald, M., Christopher, M., & Bass, M. (2003). Market segmentation. In *Marketing* (pp. 41-65). Palgrave, London.

Martens, D., & Provost, F. (2011). Pseudo-social network targeting from consumer transaction data. *Faculty of Applied Economics, University of Antwerp, Belgium*.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods*, 44(1), 1-23.

Motion, J., Leitch, S., & Brodie, R. J. (2003). Equity in corporate co-branding: the case of Adidas and the All Blacks. *European Journal of Marketing*, 37(7/8), 1080-1094.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, *27*(1), 415-444.

McDonald, M., Christopher, M., & Bass, M. (2003). Market segmentation. In *Marketing* (pp. 41-65). Palgrave, London.

Meyvis, T., & Janiszewski, C. (2004). When are broader brands stronger brands? An accessibility perspective on the success of brand extensions. *Journal of Consumer Research*, 31(2), 346-357.

McAlexander, J. H., Schouten, J. W., & Koenig, H. F. (2002). Building brand community. *Journal of marketing*, 66(1), 38-54.

Muniz, A. M., & O'guinn, T. C. (2001). Brand community*. Journal of consumer research*, 27(4), 412-432.

Nedungadi, P., & Hutchinson, J. (1985). The prototypicality of brands: Relationships with brand awareness, preference and usage. *ACR North American Advances*.

Netzer, Oded, Roneb Feldman, Jacod Goldenberg, Moshe Fresko. 2012. Mine your own business: Marketstructure surveillance through text mining. *Marketing Science* 31(3) 521{543.

Naylor, R. W., Lamberton, C. P., & West, P. M. (2012). Beyond the “like” button: The impact of mere virtual presence on brand evaluations and purchase intentions in social media settings. *Journal of Marketing*, *76*(6), 105-120.

Oestreicher-Singer, G., & Sundararajan, A. (2012). The visible hand? Demand effects of recommendation networks in electronic markets. *Management science*, *58*(11), 1963-1981.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social networks*, 32(3), 245-251.

Pons, P., & Latapy, M. (2005, October). Computing communities in large networks using random walks. In *International symposium on computer and information sciences* (pp. 284-293). Springer, Berlin, Heidelberg.

Pan, Y., Li, D. H., Liu, J. G., & Liang, J. Z. (2010). Detecting community structure in complex networks via node similarity. *Physica A: Statistical Mechanics and its Applications*, *389*(14), 2849-2857

Provost, F., Dalessandro, B., Hook, R., Zhang, X., and Murray, A. 2009. "Audience Selection for On-Line Brand Advertising: Privacy-Friendly Social Network Targeting," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Ringel, D. M., & Skiera, B. (2016). Visualizing asymmetric competition among more than 1,000 products using big search data. *Marketing Science*, 35(3), 511-534.

Radicchi, F., Ramasco, J. J., & Fortunato, S. (2011). Information filtering in complex weighted networks. *Physical Review E*, 83(4), 046101.

Ratneshwar, S., Pechmann, C., & Shocker, A. D. (1996). Goal-derived categories and the antecedents of across-category consideration. *Journal of Consumer Research,* 23(3), 240-250.

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social networks*, *29*(2), 173-191.

Serrano, M. Á., Boguná, M., & Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. *Proceedings of the national academy of sciences*, *106*(16), 6483-6488.

Shocker, A. D., & Srinivasan, V. (1979). Multiattribute approaches for product concept evaluation and generation: A critical review. *Journal of Marketing Research*, *16*(2), 159-180.

Stuart TE. 1998. Network positions and propensities to collaborate: an investigation of strategic alliance formation in a high-technology industry. *Administrative Science Quarterly* 43(3): 668–698.

Sundararajan, A., Provost, F., Oestreicher-Singer, G., & Aral, S. (2013). Research commentary—information in digital, economic, and social networks. *Information Systems Research*, 24(4), 883-905.

Smith, AndrewN., Eileen Fischer, and Chen Yongjian (2012), "How Does Brand-Related User-Generated Content Differ across YouTube, Facebook, and Twitter?," *Journal of Interactive Marketing*, 26 (2), 102–13.

Snijders, T. A., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological methodo*logy, 36(1), 99-153.

Shachar, R., & Emerson, J. W. (2000). Cast demographics, unobserved segments, and heterogeneous switching costs in a television viewing choice model. *Journal of Marketing Research*, 37(2), 173-186.

Supphellen, M. (2000). Understanding core brand equity: guidelines for in-depth elicitation of brand associations. *International Journal of Market Research*, 42(3), 1-14.

Sujan, M. and Bettman, J.R. (1989), "The effects of brand positioning strategies on consumers' brand and category perceptions: some insights from schema research", *Journal of Marketing Research*, Vol. 26 No. 4, pp. 454-467

Tang, C., & Guo, L. (2015). Digging for gold with a simple tool: Validating text mining in studying electronic word-of-mouth (eWOM) communication. *Marketing Letters*, *26*(1), 67-80.

Ter Hofstede, F., Steenkamp, J. B. E., & Wedel, M. (1999). International market segmentation based on consumer–product relations. *Journal of Marketing Research*, *36*(1), 1-17

Teichert, T. A., & Schöntag, K. (2010). Exploring consumer knowledge structures using associative network analysis. *Psychology & Marketing*, *27*(4), 369-398.

Uzzi, B., & Spiro, J. (2005). Collaboration and creativity: The small world problem. *American journal of sociology*, 111(2), 447-504.

Urban, Glen L., P. L. Johnson, John R. Hauser. 1984. Testing competitive market structures. *Marketing Science* 3(2) 83{112.

Vision Critical (2013), "From Social to Sale: 8 Questions to Ask Your Customers" [White Paper], Vancouver, BC, Canada, http://www.visioncritical.com/social2sale.

Wind, Y. (1978). Issues and advances in segmentation research. *Journal of marketing research*, 15(3), 317-337.

Wu, L., Hitt, L., & Lou, B. (2019). Data Analytics, Innovation, and Firm Productivity. *Management Science.*

Xiao, N., & Lee, S. H. M. (2014). Brand identity fit in co-branding. *European Journal of Marketing*.

Yildirim, P., Gal-Or, E., & Geylani, T. (2013). User-generated content and bias in news media. *Management Science*, *59*(12), 2655-2666.

Zhou, K. Z., & Nakamoto, K. (2007). How do enhanced and unique features affect new product preference? The moderating role of product familiarity. *Journal of the Academy of Marketing Science*, 35(1), 53-62.

Zhang, K., Bhattacharyya, S., & Ram, S. (2016). Large-Scale Network Analysis for Online Social Brand Advertising. *Mis Quarterly*, *40*(4).

VIII. APPENDICES

APPENDIX A

Sensitivity to Edge Weights

We use Jaccard Similarity (Pan et al, 2010 ; Culotta and Cutler, 2016) and conditional probability (Ringel and Skiera, 2016) to compute normalized brand-brand links in the network. In this section, we analyze the alternative weighting choices, available in the network literature, to ensure consistency of our final results.

Jaccard Coefficient. By including the size of brand communities in the denominator (|A| and |B|), the Jaccard index prevents large brands from dominating the network analyses measures (Pan et al, 2010). It measures similarity between two sets of brand communities as the intersection of the two sets divided by their union -

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Cosine Similarity. Treating the follower lists as vector representations, the cosine similarity (Manning, 2008) between two brand communities, A and B, is calculated as follows –

$$C\,(A,B) = \frac{|A \cap B|}{\sqrt{|A|} \times \sqrt{|B|}}$$

The effect of the denominator is to normalize the size of the brands, and thus prevent large brands from dominating the analyses, as in Zhang et al (2016).

Conditional Probability. Capturing local competitive asymmetry between node pairs is critical when one node exerts a higher relative weight than other. For instance, 63% of B's fans follow A, but only 50% of A's fans also follow B in figure below.

Ringel and Skiera (2016) propose conditional probability as an effective way to capture local asymmetric relationship between product pairs. This intuitive measure is calculated as follows -

$$P(A|B) = \frac{|A \cap B|}{|B|}$$

In contrast to the symmetric relations (i.e. undirected edges) produced by Jaccard and Cosine similarity, conditional probability produces asymmetric (i.e. directed) edges.

Table VII and VIII display the correlation values for the centrality measures (both within and across category) calculated using the alternative weighting schemas. Consistently high correlation suggest that our main metrics hold across a wide range of edge weighting functions, thus confirming the validity of our approach.

Table VII. Correlations for Within-category Centrality calculated for alternative weighting schemas

	Jaccard	Cosine	Conditional Probability
Jaccard	1	0.99	0.84
Cosine		1	0.86
Conditional Probability			1

Table VIII. Correlation for cross-category centrality calculated for alternative weighting schemas

	Jaccard	Cosine	Conditional Probability
Jaccard	1	0.98	0.65
Cosine		1	0.60
Conditional Probability			1

APPENDIX B

Estimates of the nodemix parameter

The between-category terms, capturing the heterophilous relationships between all category pairs, are shown in Table IX below. Keeping the base category interaction to be between – miscellaneous x miscellaneous, a positive coefficient represents a positive likelihood of consumer co-interest between the respective categories.

Table IX. Coefficients and standard errors of the 'nodemix' parameter

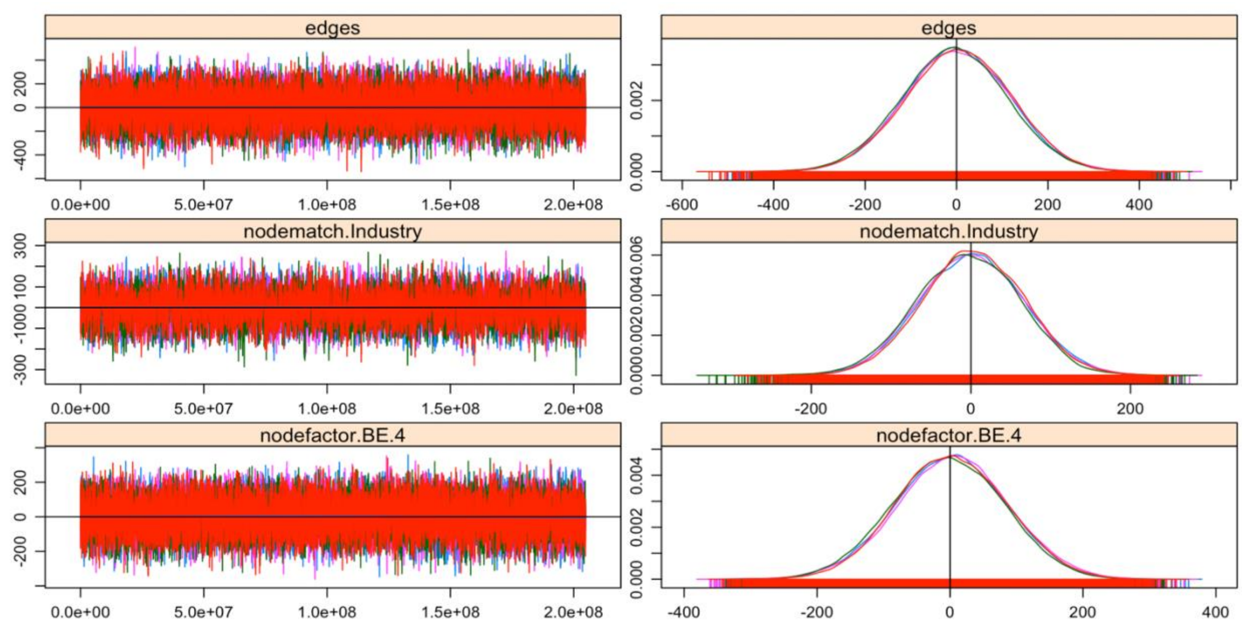
A	B	estimate	std.error
Airlines	Automotive	-0.99**	0.23
Airlines	Beer	-0.57	0.25
Automotive	Beer	-0.62*	0.23
Airlines	Cruise	2.04**	0.17
Airlines	Dining	-0.86**	0.20
Automotive	Dining	-0.71**	0.17
Beer	Dining	1.40**	0.11
Cruise	Dining	-0.10	0.26
Airlines	FoodBeverages	-2.27**	0.30
Automotive	FoodBeverages	-0.70**	0.13
Beer	FoodBeverages	0.23	0.12
Cruise	FoodBeverages	-2.75**	0.74
Dining	FoodBeverages	2.49**	0.06
Airlines	Lodging	2.09**	0.11
Cruise	Lodging	0.28	0.29
Dining	Lodging	-0.48	0.19
FoodBeverages	Lodging	-3.15**	0.50
Airlines	LuxuryGoods	1.29**	0.17
Automotive	LuxuryGoods	1.10**	0.16
Beer	LuxuryGoods	0.51	0.27
Cruise	LuxuryGoods	0.14	0.47
Dining	LuxuryGoods	0.43	0.19
FoodBeverages	LuxuryGoods	0.75**	0.14
Lodging	LuxuryGoods	1.01**	0.20
Airlines	Media	1.27**	0.20
Automotive	Media	-0.68	0.36
Beer	Media	1.29**	0.23
Cruise	Media	1.05*	0.37
Dining	Media	0.96**	0.19
FoodBeverages	Media	-0.01	0.21
Lodging	Media	0.82**	0.25

LuxuryGoods	Media	2.55**	0.23
Airlines	Retail	-0.89**	0.12
Automotive	Retail	-0.27*	0.09
Beer	Retail	0.04	0.10
Cruise	Retail	-0.78**	0.20
Dining	Retail	1.31**	0.06
FoodBeverages	Retail	0.57**	0.06
Lodging	Retail	-1.52**	0.17
LuxuryGoods	Retail	2.70**	0.07
Media	Retail	0.87	0.11
Miscellaneous	Retail	0.13	0.10
Retail	Retail	-0.67**	0.03
Airlines	Sports	0.23	0.22
Automotive	Sports	0.84**	0.16
Beer	Sports	0.45	0.24
Cruise	Sports	-0.71	0.64
Dining	Sports	0.06	0.20
FoodBeverages	Sports	-0.43	0.19
Lodging	Sports	-0.78	0.36
LuxuryGoods	Sports	2.21**	0.19
Media	Sports	2.07**	0.24
Miscellaneous	Sports	0.19	0.27
Retail	Sports	0.34*	0.11
Airlines	Technology	-0.17	0.09
Automotive	Technology	0.20*	0.07
Beer	Technology	-0.84**	0.14
Cruise	Technology	-0.34	0.16
Dining	Technology	0.55**	0.07
FoodBeverages	Technology	-0.15	0.07
Lodging	Technology	-0.87**	0.13
LuxuryGoods	Technology	1.69**	0.08
Media	Technology	2.12**	0.08
Retail	Technology	0.26**	0.05
Sports	Technology	0.71**	0.09
Airlines	Travel	2.56**	0.13
Automotive	Travel	-2.03**	0.59
Beer	Travel	-1.40	0.59
Cruise	Travel	2.93**	0.22
Dining	Travel	-0.52	0.28
Lodging	Travel	2.42**	0.14
LuxuryGoods	Travel	1.34**	0.25
Media	Travel	1.86**	0.26
Retail	Travel	-0.46*	0.16
Sports	Travel	1.23**	0.24
Technology	Travel	0.23	0.11

APPENDIX C

Goodness of fit

Parameters of the ERGM model are estimated using Monte-Carlo Markov Chain (MCMC) procedures. The advantage of MCMC – MLE approach is that the parameter estimates are generated using pseudo maximum likelihood, by repeatedly comparing the simulated graph distributions graphs with the actual data. In Figure 29, we show the MCMC-MLE convergence plots for all the parameters. For the left plot, the x-axis represents simulations and the y-axis represents parameter values. We notice good mixing for all the parameters as the MCMC sample statistics bounce randomly around the observed values. For the plot on right, difference between the observed and simulated values follow a roughly bell-shaped distribution centered around 0, providing a statistical evidence on the good model fit.



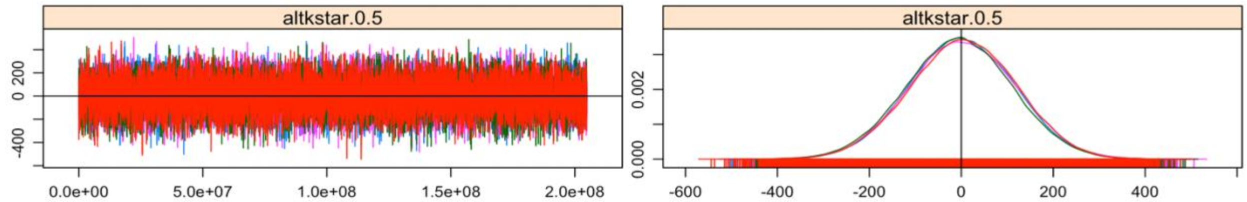
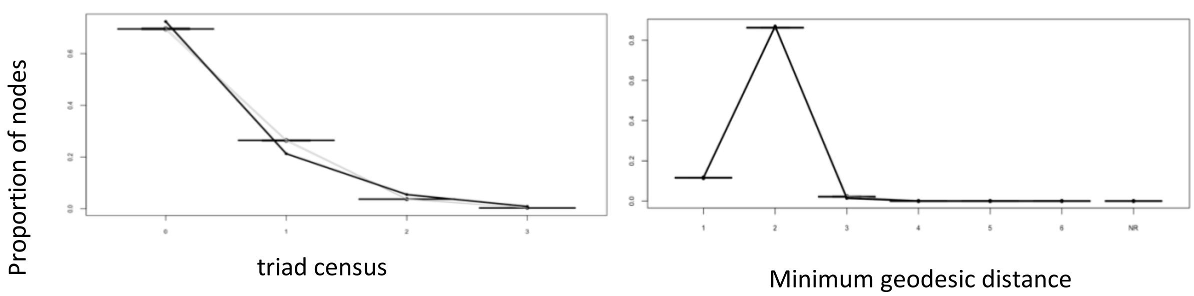


Figure 29. Convergence Plots for all the ERGM parameters.

The convergence plots for the ‘nodemix’ parameters have not been shown due to space limitations but can be made available on request.

In Figures below, we employ visual tests for goodness of fit to examine the match between predicted and observed network structure. The black dots (and lines) represent the observed brand network and grey dots (and lines) represent the simulated networks from the fitted ERGM model. If the black lines fall close enough to the grey lines, the model fit is good. We notice the black line fall close enough to the grey lines, which means the model fit is good.



APPENDIX D

Correlation of survey ratings across demographic categories

	Distinctiveness	
	Female	*Male*
Female		0.97

Pearson Correlation for survey ratings per demographic category: Beer Category

	Centrality	
	Female	*Male*
Female		0.99

144

Centrality					
	18 - 24 years	*25 - 34 years*	*35 - 44 years*	*45 - 54 years*	*55-64 years*
18 - 24 years	1				
25 - 34 years	0.83	1			
35 - 44 years	0.86	0.92	1		
45 - 54 years	0.89	0.93	0.93	1	
55-64 years	0.64	0.79	0.81	0.83	1

Distinctiveness					
	18 - 24 years	*25 - 34 years*	*35 - 44 years*	*45 - 54 years*	*55-64 years*
18 - 24 years	1				
25 - 34 years	0.91	1			
35 - 44 years	0.85	0.91	1		
45 - 54 years	0.84	0.91	0.95	1	
55-64 years	0.61	0.77	0.88	0.88	1

Centrality					
	$0- $29K	*$30K to $59K*	*$60K to $99K*	*$100K to $149K*	*$150K and above*
$0- $29K	1				

$30K to $59K	0.98	1			
$60K to $99K	0.99	0.98	1		
$100K to $149K	0.99	0.97	0.98	1	
$150K and above	0.91	0.94	0.94	0.91	1

Distinctiveness					
	$0- $29K	*$30K to $59K*	*$60K to $99K*	*$100K to $149K*	*$150K and above*
$0- $29K	1				
$30K to $59K	0.92	1			
$60K to $99K	0.91	0.94	1		
$100K to $149K	0.87	0.93	0.95	1	
$150K and above	0.76	0.85	0.85	0.86	1

Pearson Correlation for survey ratings per demographic category: Automotive Category

Centrality		
	Female	*Male*
Female		0.96

Distinctiveness		
	Female	*Male*
Female		0.94

Centrality					
	18 - 24 years	*25 - 34 years*	*35 - 44 years*	*45 - 54 years*	*55-64 years*
18 - 24 years	1				
25 - 34 years	0.75	1			

35 - 44 years	0.69	0.88	1		
45 - 54 years	0.61	0.91	0.93	1	
55-64 years	0.46	0.73	0.91	0.90	1

Distinctiveness					
	18 - 24 years	*25 - 34 years*	*35 - 44 years*	*45 - 54 years*	*55-64 years*
18 - 24 years	1				
25 - 34 years	0.85	1			
35 - 44 years	0.80	0.94	1		
45 - 54 years	0.77	0.92	0.94	1	
55-64 years	0.78	0.95	0.97	0.95	1

Centrality					
	$0- $29K	*$30K to $59K*	*$60K to $99K*	*$100K to $149K*	*$150K and above*
$0- $29K	1				
$30K to $59K	0.93	1			
$60K to $99K	0.90	0.92	1		
$100K to $149K	0.89	0.92	0.93	1	

$150K and above	0.85	0.84	0.92	0.91	1

Distinctiveness

	$0-$29K	*$30K to $59K*	*$60K to $99K*	*$100K to $149K*	*$150K and above*
$0-$29K	1				
$30K to $59K	0.95	1			
$60K to $99K	0.94	0.92	1		
$100K to $149K	0.97	0.94	0.95	1	
$150K and above	0.92	0.89	0.96	0.92	1

VITA

PANKHURI MALHOTRA

University of Illinois at Chicago
Department of Information & Decision Sciences
E-mail: pmalho3@uic.edu
Phone - 3122566666

EDUCATION

PhD, Information Systems, University of Illinois at Chicago *2015 – summer 2020*

Academic interests – Social Network Analysis, Business Analytics, Digital Networks, Information Systems Strategy, Machine Learning, Big Data Analytics, Digital Marketing, Databases, Statistical Modelling, Econometrics, Causal Inference.

Thesis – The Power of Online Brand Networks
Committee – Siddhartha Bhattacharyya, Yuheng Hu, Vijay Kamble – University of Illinois, Chicago
Jennifer Cutler – Northwestern University
Kumar Mehta – George Mason University

MSc Business Analytics, University of Manchester *2014 –2015*

Thesis – Advanced Predictive Modelling for a telecom client of Convergys, Manchester (UK)

Coursework - Mathematical Programming and Optimization, Social Media and Web Analytics, Financial Modelling, Data Analytics, Business Statistics, Information Systems Strategy, Simulation and Risk Analysis

BSc Physics Honors, Miranda House, University of Delhi *2010 – 2013*

Software Knowledge – R, Python, AWS, Hadoop, Rapid Miner, SAS, SPSS, Stata, SQL, Excel.

ACADEMIC WORK EXPERIENCE

Courses Taught: Business Data Mining.

Graduate Teaching Assistant: Data Mining for Business and Economics, Advanced Data Mining, Statistics for Management, Information Systems Strategy, Introduction to MIS.

Future Teaching Interests: Business Analytics, Machine Learning, Statistical Modelling, Social Network Analysis

Research Assistant: Data Mining, Machine Learning, Network Science

SERVICE TO COMMUNITY

- Reviewed papers for *Information Systems Research, KDD, ICIS, AMA, AAAI and others.*
- Session organizer for *INFORMS, WITS, AMA*

CURRENT RESEARCH

- **Malhotra, P.**, & Bhattacharyya, S. (2018, June). Large Scale Online Brand Networks to Study Brand Effects. In *Workshops at **Thirty-Second AAAI Conference on Artificial Intelligence***.

- **Malhotra, P.**, Bhattacharyya, S., (2019). "Online Brand Networks: A new approach to brand positioning" – 3rd round revision at ***Journal of Marketing***

- **Malhotra, P.**, Bhattacharyya, S., Zhao, K. (2019) "Information Networks for Market Structure Insights" Under review ***Management Science***

- **Malhotra, P.**, Bhattacharyya, S., (2020) "Information flows in Knowledge Networks" – *Submission to **Management Science (by winter 2021)***

- Zhao, K., Lu, Y., **Malhotra, P.**, (2020) "The Role of Content Similarity in Audience Engagement: Evidence from Live Streaming Platforms" *Submission to **Information Systems Research (by winter 2021)***

- **Malhotra, P.**, Bhattacharyya, S., (2019) "Branding in the age of social media: A comparison of user-brand relationships in the digital ecosystem" – Work in Progress - Data Analysis Phase

INDUSTRY WORK EXPERIENCE

Analytics Intern
Convergys (Manchester, United Kingdom) *June 2015 – August 2015*

Consultant, Global Operations Division
Child Rights and You (New Delhi, India) *2013 – 2014*

SELECTED CONFERENCE TALKS

- Malhotra, P., Bhattacharyya, S., "Network Analysis Techniques to Study Branding Effects." Workshop on Information technology and Systems (WITS 2016), Dublin, Ireland.
- Malhotra, P., Bhattacharyya, S., "Large Scale Online Brand Networks to study Branding Effects" AAAI – Workshop on Artificial Intelligence and Marketing Science 2018, New Orleans, USA; AMA Winter Academic Conference 2018, New Orleans, USA.
- Malhotra, P., Bhattacharyya, S., "Examining drivers of brand performance using online brand networks" The Interactive Marketing Research Conference 2018, Amsterdam, Netherlands.
- Malhotra, P., Bhattacharyya, S., Xu, M., "Brand Ecosystems from Large-Scale Data on Social Media Brand Communities" AMA Summer Academic Conference 2018, Boston, USA.
- Malhotra, P., Bhattacharyya, S., "Mining Consumer Perceptions: The Role of Networks in Branding" Informs Marketing Science Conference 2019, Rome, Italy.
- Malhotra, P., Bhattacharyya, S., "Mining Consumer Perceptions: The Role of Networks in Branding" INFORMS Annual Meeting 2019, Seattle, USA.

AWARDS AND AFFLIATIONS

- University of Illinois at Chicago Doctoral Scholarship and CBA Doctoral Fellowship *2015 – Present*
- UIC Business Graduate Assistantship *2015 – Present*