Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2020 Proceedings | Social Computing

Aug 10th, 12:00 AM

Information Networks to Derive Value from Social Media

Pankhuri Malhotra
University of Illinois Chicago, pmalho3@uic.edu

Siddhartha Bhattacharyya
University of Illinois Chicago, sidb@uic.edu

Keran Zhao
University of Illinois Chicago, kzhao23@uic.edu

Information Networks to Derive Value from Social Media

Completed Research

Pankhuri Malhotra
University of Illinois, Chicago
pmalho3@uic.edu

Sid Bhattacharyya
University of Illinois, Chicago
sidb@uic.edu

Keran Zhao
University of Illinois, Chicago
kzhao@uic.edu

Abstract

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. In this paper, the authors introduce a new type of information artifact, implicit brand networks, for obtaining close to real-time estimates of within-industry competition and across-industry complementarities. Statistical examination of the tacit links in the network, using Exponential Random Graph Models from network theory, reveals a mix of network and brand level characteristics responsible for the observed network structure. The paper concludes by discussing the practical applications of the information network, particularly for the automatic extraction of category-specific brand ratings. As information pertaining to category-specific ratings (e.g. sports, tech, luxury etc.) is rarely found in online users' comments, the brand network's ability to automatically reveal such insights, with minimal a-priori assumptions, is a significant contribution of this study.

Keywords

Information networks, social media analytics, brand communities, Twitter.

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 better understanding of these online traces has led to a growing interest in information networks (Oestreicher-Singer et al., 2012; Zhang et al., 2016). 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, customer retention, fraud identification, product adoption and several other business applications (Hill et al., 2006; Fawcett and Provost, 1997)

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). 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 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 where individual nodes represent brands and links between two brands represent common consumer engagement on Facebook (Zhang et al., 2016). Similar to product networks, links within a brand network reflect aggregated preferences of a large number of users; thereby providing 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 (Jackson, 2010).

With their inherent ability to condense the interest space of millions of digitals users to a reduced form repression 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)

In this paper, we derive implicit networks from social media to reveal statistical knowledge on online market structures and automatically extract category-specific brand insights. We first condense the massive digital interest space of millions of brand followers on Twitter into an information network of interconnected brands based on common user activity. We then apply network analysis algorithms, together with social selection models, to identify statistically significant co-interest patterns among brands. Brand networks on social media may arise due to a number of factors ranging from specific user interests, brand characteristics, and other endogenous network phenomena such as transitivity and popularity. To accurately identify the underlying latent mechanisms driving co-interest between brands, we use social selection models, Exponential Random Graph models (Snijders et al., 2006), from network theory. 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 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. To the best of our knowledge, this is one of the first few studies that focuses on revealing statistically valid co-interest consumer patterns not only within brands of the same category but also across-category. Linking to previous literature, cross-category associations are known to be crucial for coordinated promotions, embedded premiums and positioning strategies (Henderson and Arora, 2010); 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.

The second section of the paper looks closely at the within-category competition by zooming into a single industry. By exploiting the audience's interests of a brand across categories, we introduce a fully-automated method for estimating brand ratings along a given category – luxury, tech, sports, travel, etc. The resultant category-specific ratings help assess what the brand stands for and determine the brand's potential for future growth in category extensions (Cutright, 2013). They also serve as important measures to assess brand image fit during co-branding decisions. Methodologically, the task of inferring category-specific brand insights from online user-generated content is not a straightforward text mining process. Specifically, the information pertaining to the audience interests of a brand across other brands (and categories) is rarely found in users comments on a brand's fan page. This substantially limits the data available for analysis and hampers one from use any text mining algorithm to infer any category-specific insights. The brand network provides an effective solution to this problem by relying on a brand's social connections on Twitter to infer category-specific brand insights.

The core contribution of this work is to introduce a new information artifact, implicit brand networks, for deriving statistical insights on online market structures and automatically infer brand ratings in a close to real-time setting. The results allow brands to track their own audience interests as well as those of competitors and serve as the basis for strategy discussions. From a technical perspective, compared to extant text mining approaches that rely on extensive manual inputs, this network-based approach is unsupervised and does not require a priori assumptions on the underlying data. Specifically, by relying on a brand's social connections on Twitter as a critical source of information, implicit brand networks eliminate the need of complex text mining algorithms that are prone to biased, heavily context-specific and not generalizable across social platforms (Tang and Guo, 2013). In the next section, we discuss related literature in the area and show how our work contributes to the growing field of information networks and social media.

Background

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. Instead, the links forming the network are more tacit - an outcome of shared preferences. Sundarajan 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 (Sundarajan et al., 2013). 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 (Sundarajan 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) 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).

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. These probabilistic models allow inferences about whether certain observed network structures are more likely to occur than expected by chance (Snijders et al., 2006). 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).

Brand Networks as 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). As Kannan and Sanchez (1994) note, 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), website co-visitations patterns (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. The focus on narrowly defined categories is largely due to limited data availability and hard-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 competitions 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. There have been numerous examples of successful cross-category associations in the past, including coordinated promotions, embedded premiums and co-branding advertisements (Washburn, 2000). 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. Our study provides a new methodological tool to highlight these statistically valid brand associations, not one defined by the management but one perceived by the direct interests of the digital consumers.

Methodology and Results

Building Information Networks

Previous literature has established the importance of followership data on Twitter as well as its relationship to brand image (Culotta and Cutler, 2016). In this study, we collect the brand followers for the 535 most active brands on Twitter, as given by the social media directory fanpagelist.com. To prevent bots (or spam users) from influencing the network analysis measures, we conduct thorough audits of the brand accounts on SparkToro. Brands having greater than 30% percentage of spurious followers (inactive accounts, over-sharing, URL issues, and propaganda-related accounts) are excluded from the analysis. Overall, the cleansed data includes follower IDs for more than 100M brand followers across multiple categories – Dining, Retail, Automotive, Technology, Airlines, Lodging, and others. Additionally, we collect information on brand engagement (number of tweets released by the brand) and brand age (year when the brand page was founded on Twitter). The next step is to generate the information network based on aggregated interests of users across brands. An edge between two brands is created if they share common user interests. If F_i and F_j represent the list of Twitter users following brands b_i and b_j respectively, then a link 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:

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

The original brand network is almost fully connected with a density of 0.97. The range of common followers varies from few hundred to millions. Though it is possible to work with networks having wide heterogeneity in edge weights, valuable information may be lost due to redundancy generated by the overwhelming number of small connections (Serrano et al., 2009). Further, as links based on two few users may not indicate significant connectivity, extracting the truly meaningful edges is the next logical step. We employ an information filtering algorithm, Disparity Filter (Serrano et al., 2009), to identify the statistically relevant edges in the network. The final descriptive statistics of the filtered network, given α = 0.05, are given in Table 1.

Property	*Meaning*	*Value*
Number of nodes	Number of brands	535
Number of edges	Number of edges	16573
Density	Ratio of number of edges present to the maximum number of edges possible. Value ranges from 0 to 1.	0.12
Average degree	On average, the number of connections a brand exhibits.	61
Maximum degree	Maximum number of connections a brand exhibits.	512 (Starbucks)
Minimum degree	Minimum number of connections a brand exhibits.	3 (Tag Heur)

Table 1. Descriptive Statistics of the Network at α =0.05

Deriving Statistical Insights on Information Networks

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? 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 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. In the next subsection, we employ generative models, from social network analysis, to study the implicit connections (or aggregate user choices) responsible for the observed network structure.

Estimation of Model Effects

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. 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, transitivity 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. 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 (Culotta and Cutler, 2016), we use a set of 535 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. 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 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*

Consumers develop a variety of brand-to-brand associations that subsequently result in co-branding opportunities for firms (Washburn, 2000). 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 as one assumed by management but one perceived through data on consumers' direct interests.

Do certain across-category pairs form more links than others?

4) *Brand Engagement*

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 a high level of engagement (number of tweets) form 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. 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?

Model Results

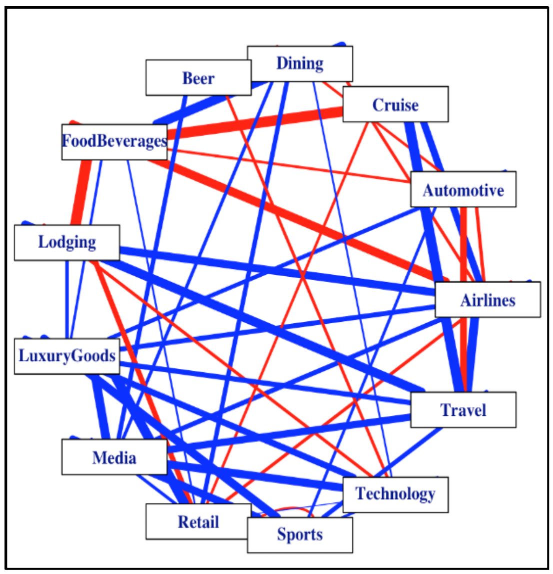
The dependent variable in the ERGM model is the presence of links (or consumer co-interest) among the brands in the network. Our key independent variables are – dyadic covariates (within-category effects – homophily, across-category effects – heterophily, brand engagement) and structural effects (edges and popularity effect). The results of the ERGM models are given in Table 2.

				Model 1	*Model 2*
Network Effect	*Edges*			-1.99***	-3.26***
	Popularity Effect			-0.03	-0.03
Brand Effects	*Brand Engagement*				0.70***
	Within-Category ***Homophily***				2.18***
	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 shown visually in Figure 1.			...
	AIC			102544	85728

Table 2. ERGM Estimation Results

Model 1 only includes structural effects without any brand level characteristics. The significant negative coefficient for the 'edges' parameter implies that edges between brands occur rarely and 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 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). 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 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.



Note: The blue lines between category-pairs represent positive likelihood and red lines represent a negative likelihood for edge formation.

Figure 1. Between-Category Effects

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 2, 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 1, where blue lines between category pairs represent positive likelihood and red lines represent a negative likelihood for edge formation. Examples of positively linked category-pairs include lodging and airlines, technology and sports, cruise and travel, etc. Some of the brand pairs involved in these between-category links are Reebok and Strike-Force Energy, Travelocity and Autralian Open, Hilton and Royal Caribbean, etc. This is an important finding for brand managers given the importance of between-category complementarities for coordinated promotions and co-branding opportunities.

Finally, 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. 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. Overall, results from the ERGM analysis show that the implicit brand-brand network is a valid information artifact, offering meaningful insights to social media managers.

Applications of the Information Network

In the next section, we delve deeper into a single sub-market, automotive in this case, to illustrate this potential of brand networks in automatically generating category-specific brand ratings.

Extracting Category-Specific Brand Insights

We now outline the approach for estimating close to real-time estimates of brand ratings along a given category. In the most generic form, this unsupervised approach, by utilizing brand network as the main information artifact, requires only a single input from the user – the brands of interest. Then for every brand of interest, the algorithm computes the weighted degree centrality across the nine main categories – luxury, sports, food, travel, beer, technology, retail, media and automotive. Weighted degree centrality, also known as node strength, is calculated as $S_i = \sum_j^N w_{ij}$, where w_{ij} is the weighted link between two brands i and j, summed over all N connections of brands. All weighted links between the focal brand and category of interest are Jaccard normalized so that any category with an overwhelming number of followers (such as luxury) does not dominate the measures.

The perceived strength of a brand along a given category, say sports, is evaluated based on the extent to which its fans overlap with accounts belonging to the sports category. It is to be noted that some of these perceived categorical strengths, such as the ones for luxury, sports, and technology, indirectly translate to brand perceptions. Brands that share high co-interest with either of the aforementioned categories are likely to be perceived closer to these categories than those who do not. These category-specific brand insights indirectly arise from the audience's co-interests with certain categories and, in a way, reflect the perception of a brand in the audience's minds. For instance, in Figures 2, we uncover the interests of Mercedes' audience across different categories and show how its brand perceptions are different from those of Chevrolet. While the audience of Chevrolet is primarily interested in cars and media, the audience of Mercedes is engaged with a multiple luxury, travel and sports brands.

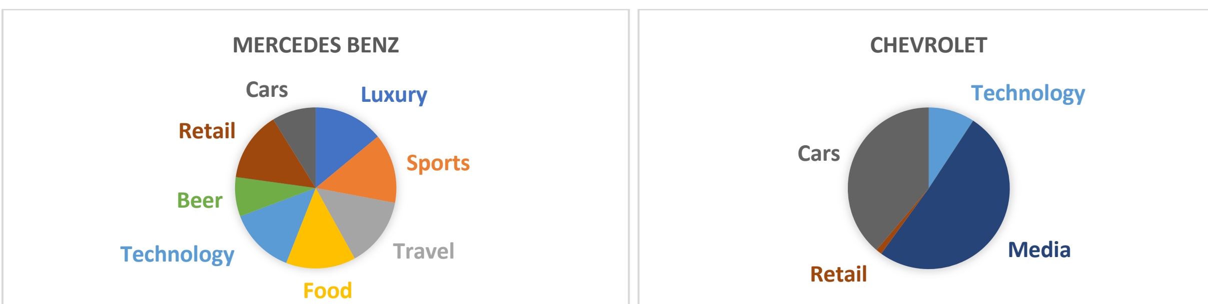


Figure 2. Category-specific Perceptions for Mercedes Benz (left) and Chevrolet (right)

Similar analyses for all car brands are presented in the form of a heat map in Figure 3, where rows correspond to brands and columns correspond to category-specific ratings. All column values have been normalized in the range of 0-1, with larger values associated with darker colorings in the heatmap. We notice Mercedes's rating to be high on multiple categories – sports, luxury, travel, etc. Similarly, we notice the audience of Toyota to be primarily interested in auto brands, and not engage with brands across categories. Another interesting observation to note is that cars sharing a high interest with sports brands are also the ones sharing high overlap with 'technology' brands.By highlighting these categorical brand associations in a clear, yet, subtle manner, brand networks provide an effective solution to managers for assessing their past as well as future co-branding strategies.

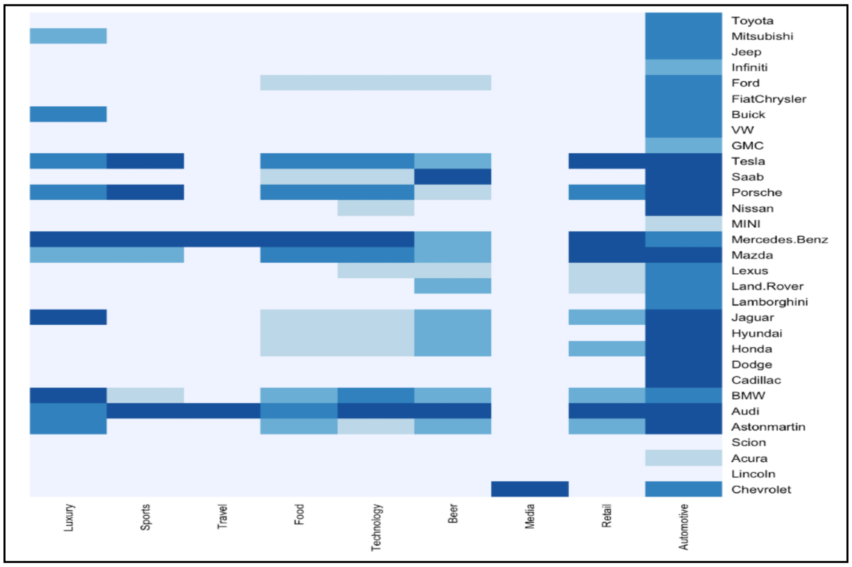


Figure 3. Brand Perceptions for the Automotive Industry

Discussion

Oestreicher-Singer and Sundararajan (2012) have previously shown that information embedded in implicit networks has social and economic impacts. A related study in information networks by Zhang et al. (2016) leverages implicit brand networks on social media for identifying target audiences for a focal brand. Unlike conventional social network studies, the aforementioned works utilize new types of digital artifacts, namely implicit networks, for describing the interest space of millions of digital users in a reduced manageable format. The links in the implicit network reflect aggregated preferences of a large number of consumers across a large number of brands (or products); thus, creating a new kind of interconnected entities, which one might approximate for an "economic network" (Jackson, 2010). In this study, we attempt to answer a new question, namely, how do implicit information networks on social media form and how one can use them to automatically infer brand ratings. By employing social selection models on the Twitter information network, this study helps to reveal statistically valid co-interest patterns between brands, within and across categories. Even though leveraging user co-interest patterns on social media is a common business practice, understanding the brand characteristics leading to these patterns is still an understudied topic. This study fills the remaining gap in this burgeoning literature and encourages future researchers to test newer effects in the ERGM model.

The second contribution of this study is to automatically infer category-specific brand insights without relying on text content on social media. Apart from the negatively skewed nature of user generated content on social media, it is impossible to find information related to brand perceptions (that is luxury, sports, tech, etc.) in individual user comments. This network-based approach provides an efficient and scalable solution to this problem by relying on the co-followership patterns on Twitter. Studies have consistently confirmed the relevance of brand followers on social media and their ability to capture the interests of digital users. Thus, by mining the followers of brands on social media, we create a new information construct that has both methodological as well as practical implications.

The theory underlying information networks is still nascent and merits further investigation to provide more valuable insights on the digital ecosystem. As part of our ongoing research, we are investigating the market definition of brand categories and how further division into subcategories can provide more nuanced information into potential co-branding strategies. Further, user–brand relationships on digital platforms can arise due to a number of factors, including the demographics of users, brand's marketing strategy, and outside events. Though the current information network does not account for these unexplained factors, we encourage future research to test these effects and reveal meaningful insights if substantial differences are observed. Consistency of the overall brand network structure over different social platforms (e.g., Twitter, Facebook, and Instagram) will provide additional validity on the inter-brand relationships and lay the foundation for using implicit networks as a robust tool for inferring audience interests across categories. As part of our future research, we are conducting dynamic network analysis on the inter-brand relationships and investigating the shift of brand ratings over time. This information can be used for multiple business scenarios, including targeted marketing, deconstruction of online consumer behavior, and demand prediction. Overall our paper demonstrates the value of information networks on social media by highlighting both statistical and descriptive insights on brand-to-brand relationships. We hope that the methods introduced in this paper lay the foundation for future research in the area of information networks and social media.

REFERENCES

Akaike, H. 1998. "Information Theory and an Extension of the Maximum Likelihood Principle," in *Selected Papers of Hirotugu Akaike*, pp. 199–213.

Byshkin, M., Stivala, A., Mira, A., Robins, G., and Lomi, A. 2018. "Fast Maximum Likelihood Estimation via Equilibrium Expectation for Large Network Data," *Scientific reports,* (8:1), pp. 1-11.

Culotta, A., and Cutler, J. 2016. "Mining Brand Perceptions from Twitter Social Networks," *Marketing science,* (35:3), pp. 343-362.

Cutright, K. M., Bettman, J. R., Fitzsimons, G. J., and Thomas, R. D. 2013. "Putting Brands in Their Place: How a Lack of Control Keeps Brands Contained," *Journal of Marketing Research,* (50:3), pp. 365-377.

Fawcett, T., and Provost, F. 1997. "Adaptive Fraud Detection," *Data mining and knowledge discovery*, (1: 3), pp. 291-316.

Henderson, T., and Arora, N. 2010. "Promoting Brands across Categories with a Social Cause: Implementing Effective Embedded Premium Programs," *Journal of Marketing,* (74:6), pp. 41-60.

Hill, S., Provost, F., and Volinsky, C. 2006. "Network-Based Marketing: Identifying Likely Adopters via Consumer Networks," in *Statistical Science*, (21:2), pp. 256-276.

Jackson, M. O. 2010. "Social and Economic Networks," *Social and Economic Networks*.

Joiner, C., and Loken, B. 1998. "The Inclusion Effect and Category-Based Induction: Theory and Application to Brand Categories," *Journal of Consumer Psychology,* (7:2), pp. 101-129.

Kannan, P. K., and Sanchez, S. M. 1994. "Competitive Market Structures: A Subset Selection Analysis," *Management Science,* (40:11), pp. 1484-1499.

Kim, J. Y., Howard, M., Cox Pahnke, E., and Boeker, W. 2016. "Understanding Network Formation in Strategy Research: Exponential Random Graph Models," *Strategic management journal,* (37: 1), pp. 22-44.

Laroche, M., Habibi, M. R., Richard, M. O., and 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), pp. 1755-1767.

Netzer, O., Feldman, R., Goldenberg, J., and Fresko, M. 2012. "Mine Your Own Business: Market-Structure Surveillance through Text Mining," *Marketing Science,* (31:3), pp. 521-543.

Oestreicher-Singer, G., and Sundararajan, A. 2012. "The Visible Hand? Demand Effects of Recommendation Networks in Electronic Markets," *Management science,* (58: 11), pp. 1963-1981.

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*, pp. 707-716. 2009.

Ringel, D. M., and Skiera, B. 2016. "Visualizing Asymmetric Competition among More than 1,000 Products Using Big Search Data," *Marketing Science,* (35:3), pp. 511-534.

Robins, G., Pattison, P., Kalish, Y., and Lusher, D. 2007. "An Introduction to Exponential Random Graph (P*) Models for Social Networks," *Social networks,* (29:2), pp. 173-191.

Serrano, M. Á., Boguñá, M., and Vespignani, A. 2009. "Extracting the Multiscale Backbone of Complex Weighted Networks," *Proceedings of the national academy of sciences,* (106:16), pp. 6483-6488.

Snijders, T. A. B., Pattison, P. E., Robins, G. L., and Handcock, M. S. 2006. "New Specifications for Exponential Random Graph Models," *Sociological methodology,*(36:1), pp. 99-153.

Sundararajan, A., Provost, F., Oestreicher-Singer, G., and Aral, S. 2013. "Information in Digital, Economic, and Social Networks," *Information Systems Research,* (24:4), pp. 883-905.

Tang, C., and Guo, L. 2013. "Digging for Gold with a Simple Tool: Validating Text Mining in Studying Electronic Word-of-Mouth (EWOM) Communication," *Marketing Letters,* (26:1), pp. 67-80.

Washburn, J. H., Till, B. D., and Priluck, R. 2000. "Co-Branding: Brand Equity and Trial Effects," *Journal of Consumer Marketing*.

Zhang, K., Bhattacharyya, S., and Ram, S. 2016. "Large-Scale Network Analysis for Online Social Brand Advertising," *MIS Quarterly* (40:4), pp. 849–868.