The Workshops of the Thirty-Second
AAAI Conference on Artificial Intelligence

Inferring Brand Knowledge from Online Consumer Associative Brand Networks

Minghong Xu, Siddhartha Bhattacharyya
Information and Decision Sciences
College of Business and Administration
University of Illinois at Chicago
mxu29@uic.edu, sidb@uic.edu

Abstract

Consumer perceptions of a brand is key to brand value. Traditional methods rely on time-consuming and expensive approaches like surveys and focus groups to elicit such perceptions; these are also very limited in terms of reach across consumers and brands. This paper examines consumer associative brand networks inferred from large-scale data on consumers' engagement across a broad collection of brands. It presents a statistical analyses of brand networks, to help determine whether such networks obtained from large-scale data on consumer co-interest in social media provide a valid and reliable source of brand insights and knowledge.

Introduction

Consumer perceptions of a brand is key to brand value (Keller, 1993; Fournier, 1998), and understanding and managing how consumers perceive their brands is a priority for brand managers. Consumer perceptions of brands and market structure are considered to be more important than stated brand strategies (Henderson et al. 1998). However, such consumer perceptions and the associations they make to the focal brand can be difficult to identify, requiring the use of carefully designed questionnaires, surveys and focus groups (Aaker, 1996; Henderson et al. 1998); such approaches also capture the views of only a restricted set of consumers, and on limited number of brands. Keller (2016) notes the potential for leveraging "the vast, abundant data sources now available online" for empirical studies on brands. Online social media, for example, enable a whole system of interactions of users, products, brands and firms, and present opportunities for collecting unsolicited word of mouse. Such platforms provide large scale data on consumers' interactions with varied brands, offer new way to capture consumer perceptions, and can help obtain brand knowledge and market structure insights.

In this paper, we study a consumer associative brand network inferred from large-scale data on Facebook activity trails over a time frame of four years. Two brands are linked if consumers of one brand are also interested in the other brand, as reflected in the number of overlapping users. A common set of consumers with shared interest in two brands indicates "similar features" or "related perceptions" (Culota and Cutler, 2016). This type of consumer associative brand connections may be valuable to marketers and brand managers in evaluating advertising efforts, understanding brand positioning and making marketing decisions. A recent paper by Zhang et al. (2016) demonstrates that such brand networks provides useful insights on audience selection for advertising. Culotta and Cutler (2016) use associations between a focal brand and specified exemplar brands to mine brand perceptions from brand-followership data on Twitter. Malhotra and Bhattacharyya (2016) suggest the use of brand networks to study brand positioning. This paper presents a statistical analyses of brand networks, to help determine whether such networks obtained from large-scale data on consumer co-interest across brands in social media provide a valid and reliable source of brand insights and knowledge.

The network of associations between brands, derived from users' interactions with brands on social media, arises through a range of factors, including brands' marketing efforts, user choices and personal interest, social network influences, and can incorporate noise. Statistical analysis of the network is necessary, to establish that observed associations between brands are not a result of randomness, to examine the factors that play significant roles in formation of brand-to-brand associations, and to determine if the network is stable over time and can thereby provide consistent brand knowledge. For this purpose, we utilize

Exponential Random Graph Models (ERGMs) (Hunter et al, 2008) to analyze the brand network. ERGMs are a family of random graph models that draw statistical inference on the processes influencing the formation of an observed network structure. A network is considered as being realized through a set of random variables relating to these processes. The models use a distribution over the variables, and maximum likelihood estimates for these are obtained based on fit with the observed network data.

We consider statistical inference for the formation of links between brands in the network, which carries key information on which brands are associated based on consumer co-interest. Using EGRMs and drawing on Keller's (1993) widely used brand knowledge framework, we examine statistically valid insights that a brand network can provide. Specifically, we address the following research questions: (1) What type of brand knowledge does the brand network provide? (2) Does brand network provide consistent brand knowledge? (3) What are the key processes and effects that inform the connections between brands?

We find that the observed network is not the result of random associations between brands, but arises from consumer co-interest across specific sets of brands. This brand network can thus provide useful insights on brand awareness and brand image. By analyzing the observed network over different time periods, another key finding is that the network remains consistent over time; thus, insights obtained from the brand network can be expected to remain useful for future practice.

Brands can vary by the size of their consumer communities, with a few brands attracting higher number of users. Such large brands, given their broader user base, tend to carry more connections with other brands. A key question then is whether these few 'popular' brands dominate the network, with their connections to other brands expanding over time. Our analysis shows that the network is not dominated by a few large brands; this further emphasizes the validity of insights drawn from the network. Another key finding is that consumers' engagements with brands display strong reciprocity. Thus, if a brand A's consumers are engaged with another brand B, the reverse is also likely. This is interesting given the fact that consumer community sizes vary for different brands. In addition, incorporating brand category as a node attribute reveals interesting differences between brands. For example, celebrity brands tend to attract more in-links from other brands, and fewer out-links, which highlights the potential of celebrity sponsored advertisements.

The paper is organized is as follows: Section 2 briefly reviews related work, Section 3 describes the brand network, Section 4 discusses brand knowledge and corresponding hypotheses, and Section 5 presents the model results. Section 6 notes conclusions and future work.

Related Work

In a seminal paper, Keller (1993) proposed the widely accepted framework on Customer-Based Brand Equity, and noted that "perhaps a firm's most valuable asset for improving marketing productivity is the knowledge that has been created about the brand in consumers' minds from the firm's investment in previous marketing programs". Perceptual maps have been a standard method for comparative analyses of brands based on consumer perceptions on specific attributes (Hauser and Koppelman, 1979). Consumer perceptions are typically elicited through time-consuming and expensive approaches using surveys and focus groups (Aaker, 1996) which are also very limited in terms of reach across consumers and brands. Social media marketing represents one of the latest trends in marketing practice and research (Ashley and Tuten, 2015; Swaminathan, 2016), and Keller (2016) notes the potential of large-scale data from digital platforms for empirical study of brands.

Recent work has considered the mining of online customer review data for brand perceptions (Krawczyk and Xiang, 2016). In an approach similar to ours, Culotta and Cutler (2016) mine brand followership data from Twitter to determine brand perceptions based on consumer co-interest with certain exemplar brands. Our work is different in that we mine large-scale data on consumer engagement with brands on Facebook to construct a brand network depicting connections across a broad set of brands, and then conduct a statistical analysis of this network. Zhang et al. (2016) found such a Facebook brand network useful for audience targeting.

Network analysis to study branding effects was suggested by Henderson et al. (1998), where they examined small brand associative networks based on brand perceptions elicited from 46 individuals on seven sports car brands. Our work presents a scalable approach to develop a brand associative network considering a large number of brands and consumers. This paper is the first to conduct statistical analyses of brand networks, to examine the factors which play a significant role in the formation of associations between brands, and to establish whether such networks present a reliable source of brand knowledge and insights.

Statistical methods have been developed in recent years to analyze social and other networks. Among them, ERGMs (Hunter et al., 2008; Handcock et al., 2008) are state of the art models which allow generalization beyond the restrictive dyadic independence assumptions of the earlier p_1 models (Holland and Leinhardt, 1981). EGRMs have been applied to understand the underlying processes that drive the formations of various social networks. Our study utilizes ERGMs on quasi-social networks, which are aggregated from the consumer-brand affiliation data.

Consumer Associative Brand Network

We consider brand networks derived from longitudinal data recording users' activities on Facebook brand pages. Activities are of three types: a post which is an initial message of a thread of discussions, a comment which follows a post, and a like which expresses an attitude towards a post. We consider users who have repeated activities (more than once, to reduce false positives) on a brand page as engaged users of a brand. We consider brand networks constructed using yearly data from 2012 to 2015.

Each node in the brand network is labeled by a "brand", such as CBSSports, VinDiesel etc. A node represents the consumer base of a focal brand on Facebook. For the analysis in this paper, we consider the top 100 brands from the "most talked about brands" in different categories on fanpagelist.com. Appendix I lists these 10 celebrity, 10 college, 10 lodging, 15 media, 15 retail, 10 service, 10 sports and 20 technology brands.

Edges are established based on the shared engaged users between brands. Directed edges are used since the impact of 10 common users is different for a brand with 100 engaged users than for a brand with a community of 1000. A directed edge from a brand A to a brand B represents percentage of brand A's consumers who are also engaged with brand B:

$$edge\ A \rightarrow B\ = \frac{\#common\ engaged\ users\ of\ A\ and\ B}{\#\ engaged\ users\ of\ A}$$

Given the focus of this study on whether brands are connected through common users, we consider binary ties. To analyze factors driving network formation at different tie strengths, we use two threshold levels to define binary edges: medium (0.001) and low (0.0001).

Besides node size, we are also interested in users' engagement patterns with the brand as described by their level (number) of activities. Keller (2016) suggests the brand pyramid as a conceptual tool to consider consumer's varying levels of engagement with a brand. For this, we calculate the number of user activities on one brand page at different quantiles across all 100 brands. The activity level quantiles provide a way to visualize the brand pyramid, and are found to remain consistent over years. We adopt a Pareto Principle (roughly 80% of the effects come from 20% of the cause), and use the 80% quantile of nine activities (which means among all the user-brand engagements, 20% of them have more than nine activities) as an indication of active engagement. Specifically, users having as least nine activities on a brand page are defined as the active users of the brand. The active user percentage is used as a brand covariate in the statistical analyses. Brand category is another feature considered.

Brand Knowledge and Hypotheses

Keller (1993) defines consumer-based brand knowledge through two components: brand awareness and brand image. We argue that the brand network derived from social media data can provide additional metrics for measuring brand knowledge in these two components. With a large set of brands and their associations, the network enables a broader view of brand image and awareness for a focal brand, and comparisons to similar brands or competitors. To evaluate the significance and reliability of brand knowledge and insights from the network, we examine the mechanisms underlying network formation, i.e., the effects that can explain the connections between brands. The links between brands may arise from brand features, previous marketing efforts, and user choices, which are valuable inputs for brand managers in decision making. Observed brand connections may also result from the nature of user interactions with social media platforms.

To obtain useful brand knowledge, the information contained in the brand network should be consistent over time. If the network is observed to be unstable, we cannot rule out the possibility that the network is formed by randomness. In addition, the knowledge obtained from the network may not be useful for future practice if brands' associations with other brands change significantly over time. Thus we test:

Hypothesis 1. The brand network is consistent over time.

Brand knowledge is what consumers conceive about the brand from the firm's previous marketing programs and brand promotions (Keller 1993). Social media platforms provide easy access to a broad selection of brands. An interesting question is whether consumers' engagements with brands and the resulting connections between brands are from users' specific interest and brands' marketing practices, or arise purely because of social media's open nature and ready access. This is examined in:

Hypothesis 2. Edges tend to form across arbitrary pair of brands.

The first component of brand knowledge, brand awareness, relates to the likelihood that a brand name will come to mind and the ease with which it does so. At the basic level, brand awareness can be indicated by the node size of the brand compared to other brands in a comparative group. Besides relative node size, a brand's in-degree also indicates brand aware-ness. In social networks, in-degree often reflects a node's "popularity" in the relation defined by the edges, like friendship or advisorship. Similarly, in brand networks, a high in-degree indicates that consumers of multiple other brands are aware of the focal brand. For example, VinDiesel and EntertainmentWeekly have the highest in-degrees in the 2015 brand network, indicating that these brands have high awareness. The "rich get richer" phenomenon is often observed in social networks, where the number of in-coming edges to a node substantially increases the likelihood of additional in-coming edges. An important question to investigate is whether awareness is driven primarily by marketing practices or whether brands having high-awareness can expect increased awareness driven by this "rich-get-richer" effect. Thus we formulate the third hypothesis to test whether there is a similar tendency in brand network, whereby high awareness brands tend to increase their awareness among consumers:

Hypothesis 3. Brands with many in-links tend to receive more in-links.

The second component of brand knowledge, brand image, refers to the set of associations attached to the brand that consumers hold in memory. Edges in the brand network, both in-coming or out-going, can indicate brand image. For example, Levis is linked with some college brands, implying that Levis has a "young consumers", "casual" image. Another example is FoxSports, which is connected with many lodging brands, indicating a connection with travelers.

We are interested in the underlying process that drives the formation and dissolution of such brand connections. For this, we consider two network effects, homophily and reciprocity, which have been noted to drive network formation in various social networks. Homophily suggests that brands having similar characteristics will be connected with each other. In the brand network, homophily may arise from consumers' co-engagement in similar brands; we examine homophily through brands being in the same category. Reciprocity implies that if brand A's consumers are engaged with brand B, then brand B's consumers are also engaged with brand A. The analysis on reciprocity is important, given that weights on directed edges relate to the sizes of the focal brands, and that brand sizes can vary widely. These network effects are examined through:

Hypothesis 4. Brands tend to connect with brands from the same category.

Hypothesis 5. Brand connections tend to be mutual.

We also propose two brand level features which may explain brand connections. They are active user percentage, and brand category. These are posed respectively in the two research questions:

Question 1. Does a brand's active user percentage affect the brand's in- and out-links?

Question 2. Does a brand's category affect the brand's in- and out-links?

Model and Results

We use ERGMs to model the formation and dissolution of ties in the brand network, and statistically test our hypotheses. ERGMs are a family of statistical models for analyzing data on social and other networks. The purpose of using ERGM is to describe parsimoniously the local selection forces that shape the global structures in a network, and to specify the processes which gives rise to the observed network (Hunter et al., 2008; Handcock et al., 2008). ERGMs express the probability of an observed network y as:

$$Pr(Y = y) = \left(\frac{1}{k}\right) exp\left\{\sum\nolimits_{A} \eta_A g_A(y)\right\}$$

where (i) the summation is over all effects A included in the model which may explain the formation of ties; (ii) η_A is the parameter corresponding to the effect A; (iii) $g_A(y)$ is the network statistic corresponding to the effect A; $g_A(y)$= 1 if the structure of the effect is observed in the network y, and is 0 otherwise; (iv) k is a normalizing quantity which ensures that the expression gives a proper probability distribution.

Five network structural effects with two brand-level features (active user percentage and brand category), are used in the models: *Edges* which models the general tendency of the network to have ties; *Mutual* which models the tendency towards reciprocal ties; *Popularity* which models whether high in-degrees lead to higher in-degrees; *EdgeCovariatePrevYear* which models the correlation between edges in the current year and those in the previous year; *NodeMatch on category* which models the tendency for nodes of the same category to have ties. Table 1 shows the ERGM models and compares the networks in 2015 to those in 2013. For each target network, we fit three types of models: the basic model with the five local structures; the active user (ActiveU) model which further includes active user percentage, along both a node's incoming links (*ICovActUP*) and outgoing links (*OCovActUP*); and the type model (Type) which adds binary indicators for different categories, for both incoming links and outgoing links (for example, *ILCelebrity* and *OLCelebrity*).

Edges' covariance with those of previous year (2014 or 2012) is positively significant in different years regardless of different tie weight thresholds taken. The presence of an edge in the previous year highly increases the likelihood that an edge will be present in the current year, thus indicating a strong stability effect. Hypothesis 1 is supported, which means the brand network is consistent over time, and can thus provide reliable insights and stable measures of brand knowledge.

The Edges effect is significantly negative in all models, which shows that the tendency for edges to expand and to extend to any pair of brands is negative. Given this general tendency of brand networks to be sparse, the observed ties between brands can be taken to arise from effects such as brand features or marketing programs, and reflect genuine and significant consumer choice. Thus we can reject hypothesis 2 with 99.9% confidence.

	2015 Medium Threshold			2013 Medium Threshold			2015 Low Threshold			2013 Low Threshold		
	Basic	ActiveU	Type	Basic	ActiveU	Type	Basic	ActiveU	Type	Basic	ActiveU	Type
Edges	-4.547 *** (0.181)	-4.917 *** (0.240)	-5.051 *** (0.316)	-4.637 *** (0.217)	-4.343 *** (0.292)	-5.418 *** (0.374)	-3.453 *** (0.098)	-3.150 *** (0.128)	-3.820 *** (0.190)	-4.284 *** (0.121)	-4.275 *** (0.151)	-4.718 *** (0.179)
Mutual	2.197 *** (0.162)	2.206 *** (0.174)	3.167 *** (0.230)	3.331 *** (0.174)	3.476 *** (0.185)	3.652 *** (0.192)	0.883 *** (0.087)	0.901 *** (0.093)	2.031 *** (0.126)	2.462 *** (0.099)	2.487 *** (0.098)	3.142 *** (0.127)
Popularity	0.031 ** (0.010)	0.031 *** (0.010)	0.015 (0.014)	-0.010 (0.012)	-0.011 (0.013)	-0.005 (0.015)	0.035 *** (0.003)	0.035 *** (0.003)	0.029 *** (0.005)	0.018 *** (0.003)	0.018 *** (0.003)	0.011 * (0.004)
EdgeCov 2014/2012	5.346 *** (0.139)	5.396 *** (0.142)	5.624 *** (0.173)	5.577 *** (0.140)	5.532 *** (0.152)	5.549 *** (0.146)	4.018 *** (0.085)	4.043 *** (0.091)	4.359 *** (0.108)	3.807 *** (0.790)	3.79 2*** (0.079)	3.693 *** (0.087)
Node Match Category	0.425 * (0.172)	0.453 ** (0.156)	0.530 ** (0.173)	0.5312 ** (0.164)	0.505 ** (0.158)	0.516 ** (0.172)	0.371 *** (0.094)	0.360 *** (0.102)	0.417 *** (0.105)	0.532 *** (0.094)	0.526 *** (0.097)	0.525 *** (0.098)
ICovActUP		0.222 (0.660)			-4.249 *** (0.959)			0.293 (0.321)			-0.251 (0.411)	
OCovActUP		1.873 * (0.682)			2.282 ** (0.845)			-2.145 *** (0.367)			0.133 (0.487)	
ILCelebrity			1.031 ** (0.351)			1.220 ** (0.401)			0.842 *** (0.173)			1.139 *** (0.196)
OLCelebrity			-1.301 *** (0.367)			-0.824 * (0.366)			-1.757 *** (0.217)			-1.057 *** (0.212)
ILCollege			0.993 ** (0.343)			0.573 (0.408)			0.435 * (0.175)			-0.033 (0.208)
OLCollege			-1.453 *** (0.352)			0.065 (0.318)			-1.451 *** (0.187)			0.370 * (0.164)
ILMedia			0.925 ** (0.333)			1.297 *** (0.379)			0.905 *** (0.176)			1.062 *** (0.197)
OL.Media			-1.259 *** (0.344)			-0.636 * (0.308)			-1.812 *** (0.190)			-0.822 *** (0.171)
ILRetail			0.350 (0.325)			0.410 (0.391)			0.403 ** (0.154)			0.831 *** (0.187)
OLRetail			0.003 (0.275)			-0.082 (0.309)			0.321 * (0.156)			-0.579 ** (0.176)
ILService			-1.057 ** (0.382)			1.399 *** (0.395)			-0.584 ** (0.193)			-0.101 (0.213)
OLService			1.963 *** (0.235)			-0.038 (0.331)			1.947 *** (0.157)			0.732 *** (0.169)
ILSports			0.774 * (0.370)			0.625 (0.403)			0.614 *** (0.168)			0.738 *** (0.202)
OLSports			-1.153 *** (0.370)			0.325 (0.299)			-1.765 *** (0.208)			0.077 (0.184)
ILTech			0.231 (0.315)			0.284 (0.383)			0.274 (0.160)			0.168 (0.183)
OLTech			-0.098 (0.257)			-0.080 (0.298)			-0.403 * (0.158)			0.222 (0.155)

Significance codes: 0 '*' 0.001 '**' 0.01 '*' 0.05**

Table 1. Exponential Random Graph Models on Medium and Low Threshold Networks

The Popularity effect is significant in 7 out of the 12 models, but with very small estimates (close to 0). Hypothesis 3 is not supported. The tendency of brands having high-awareness to derive increased awareness driven by a "rich-get-richer" effect is very weak. This implies that brand awareness develops more from marketing efforts and consumer interest than through existing 'popularity'.

The homophily effect is significant in all the models, and Hypothesis 4 is supported. Two brands being in the same category almost doubles the likelihood that an edge will occur between these two brands. Users tend to engage with several different brands in the same category, which indicates that consumers on social media are aware of not only the basic functions of, or needs fulfilled by, the product category, but also the differences between multiple brands in the category.

Reciprocity is positive and significant across models, and Hypothesis 5 is supported. The co-engagement relation is thus reciprocal. Despite the fact that brand sizes vary, it is interesting to find that consumers' co-engagement in brands tends to be mutual.

The coefficient values (and their significance) for active user percentage vary between years and over different thresholds. Its covariance with in-links ranges from significantly negative to slightly positive, while its covariance with out-links swings between negative and positive values. Thus, in answer to Research Question 1, a focal brand's active user percentage does not show a definite relation to its in-links or out-links. However, the positive and significant values of covariance between active user percentage and out-links in medium threshold networks indicate the potential for stronger out-links from brands having higher proportion of active users. This can be a useful consideration for advertising and customer targeting purposes.

Lastly, on Research Question 2, being in certain categories does affect the brand's in-links and out-links. Celebrity and media brand indicators have positive relations with the brand's in-links and negative relations with their out- links. The same effect is significant for college brands in 2015, but not significant in 2013. The service brand indicator is seen to have an opposite effect, with a negative relation with the brand's in-links and a positive relation with out-links.

We perform MCMC model diagnosis on all models and ascertain goodness of fit on in-degree and out-degree distributions. Given space limitations, we omit details. Appendix 2 shows diagnostics on the category model for the 2015 medium threshold network. The model diagnoses show that all the estimates are sufficiently converged.

Conclusions and Future Work

This paper examines consumer associative brand network inferred from large volumes of user activity data on Facebook brand pages. We develop statistical models on the brand network to help analyze the underlying process of brand network formation, and estimate different network effects and brand features that drive the formation of brand associations. Drawing upon Keller's (1993) widely used framework of brand knowledge, we show how the brand network can provide useful insights for marketing and brand managers.

Brand associations obtained from large scale data on consumer's online engagement with brands present new opportunities. Our findings help establish the potential of such data and brand networks for future research on varied issues of importance to marketing and brand managers. Our work also highlights the value of ERGMs for statistical inference on such networks.

The presented study is based on 100 highly talked about brand on Facebook. Larger brand networks will be analyzed in future studies. Weighted edges reflecting tie strengths between brands is another topic of continuing research. Other related ongoing work examines clustering on brand networks to obtain consumer segmentations based on common interest across brands.

References

Aaker, D. 1996. Measuring brand equity across products and markets. *California Management Review* 38 (3): 102–120.

Ashley, C. and Tuten, T. 2014. Creative Strategies in Social Media Marketing: An Exploratory Study of Branded Social Content and Consumer Engagement. *Psychology & Marketing* 32, 1 (Jan 2015), 15–27.

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

Fournier, S. 1998. Consumers and their brands: Developing relationship theory in consumer research. *Journal of consumer research*, 24(4), 343-373.

Hauser, J.R. and Koppelman, F.S. 1979. Alternative perceptual mapping techniques: Relative accuracy and usefulness. *Journal of marketing Research* (1979): 495-506.

Holland, P.W. and Leinhardt, S.1981. An Exponential Family of Probability Distributions for Directed Graphs. *Journal of the American Statistical Association*, 76(373), 33–65.

Handcock, M.S., Hunter, D.R., Butts, C.T., Goodreau S.M. and Morris, M. 2008. Statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data. *Journal of Statistical Software* 24, 1, 1548–7660.

Henderson, G.R., Iacobucci, D. and Calder, B.J. 1998. Brand Diagnostics: Mapping Branding Effects using consumer associative networks. *European Journal of Operational Research* 111, 306-327.

Hunter, D.R., Handcock, M.S., Butts, C.T., Goodreau, S.M. and Morris, M. 2008. Ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. *Journal of Statist. Software* 24, 3.

Keller, K.L. 1993. Conceptualizing, Measuring, and Managing Customer-Based Brand Equity. *Journal of Marketing*, 57, 1-22.

Keller, K.L. 2016. Reflections on customer-based brand equity: perspectives, progress, and priorities. *AMS Review*, 6:1-2, 1-16.

Krawczyk, M. and Xiang, Z. 2016. Perceptual mapping of hotel brands using online reviews: a text analytics approach. *J Information Technology & Tourism*, 16(1), 22-43.

Malhotra, P. and Bhattacharyya, S. 2016. Network Analysis Techniques to Study Branding Effects. *Workshop on Information Technology and Systems* (WITS-2016), Dublin.

Swaminathan, V. 2016. Branding in the Digital Era: New Directions for Research on Customer-Based Brand Equity. *AMS Review*, 6, 33–38

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

Appendix I. 100 Brands list

Brand	Category	Brand	Category	Brand	Category	Brand	Category
VinDiesel	Celebrity	ExcaliburLasVegas	Lodging	kohls	Retail	WTA	Sports
Ludacris	Celebrity	PalazzoLasVegas	Lodging	zappos	Retail	ATPWorldTour	Sports
JustinBieber	Celebrity	montecarlovegas	Lodging	Zara	Retail	SportsIllustrated	Sports
rihanna	Celebrity	circuscircus	Lodging	Levis	Retail	PGATour	Sports
wakaflocka	Celebrity	BestWestern	Lodging	dickssportinggoods	Retail	WrestleMania	Sports
JoelOsteen	Celebrity	abcnews	Media	Underarmour	Retail	today	Technology
neymarjr	Celebrity	Snoopy	Media	lululemon	Retail	InstagramEnglish	Technology
GeorgeLopez	Celebrity	foxsports	Media	quiksilver	Retail	Intel	Technology
DonaldTrump	Celebrity	NBCNews	Media	Forever21	Retail	yahoo	Technology
KimKardashian	Celebrity	WorldNewsTonight	Media	timberland	Retail	lenovo	Technology
tamu	Colleges	nytimes	Media	citi	Services	SamsungMobile	Technology
Harvard	Colleges	newyorker	Media	uber	Services	aol	Technology
spartans.msu	Colleges	ign	Media	LiveNation	Services	AppStore	Technology
NYU	Colleges	time	Media	AncestryUS	Services	Google	Technology
UCBerkeley	Colleges	usatoday	Media	AmericanExpressUS	Services	xbox	Technology
uflorida	Colleges	entertainmentweekly	Media	directv	Services	PlayStation	Technology
MITnews	Colleges	ESPN	Media	FarmersInsurance	Services	netflixus	Technology
universityof-michigan	Colleges	CBSSports	Media	VisaUnitedStates	Services	Zoosk	Technology
YaleUniversity	Colleges	Vogue	Media	hrblock	Services	DisneyPixar	Technology
PrincetonU	Colleges	CBSNews	Media	24HourFitness	Services	HP	Technology
hilton	Lodging	Amazon	Retail	NFL	Sports	Skype	Technology
MGMGrand	Lodging	hm	Retail	NHL	Sports	ATT	Technology
Riuhoteles	Lodging	walmart	Retail	olympics	Sports	xfinity	Technology
LuxorLasVegas	Lodging	REI	Retail	wimbledon	Sports	pinterest	Technology
wynnlasvegas	Lodging	nike	Retail	sportingnews	Sports	googlechrome	Technology

Appendix 2. Goodness of Fit

We perform the MCMC model diagnostics on all the 12 models and analyze the goodness of fit on in-degree distribution and out-degree distributions. Given space limitations, we present diagnostic result for the category model on 2015 medium threshold network. All the estimates are sufficiently converged and our model fits the observed data well.

The model diagnostics are used to ascertain convergence of the MCMC processes in the model estimations. The plots in Figure A2.1 indicate change of model statistics during the last iteration of the MCMC procedure. For each model statistic, the left hand side plot gives the change of the statistic with iterations and the right hand side plot is a histogram of the statistic. Both plots are normalized, so the observed data locate at 0. The models are considered to have converged if the MCMC sample statistics bounce randomly around the observed values, and the difference between the observed and simulated values of the sample statistics have a roughly bell-shaped distribution centered at 0.

The Goodness-of-Fit test checks how well the estimated model captures certain features of the observed network, especially, how well it reproduces the observed network properties that are not in the model. We do this by fitting on network statistics which are not in the model, and are essential in describing the networks. For this, we consider the in-degree distribution and out-degree distribution, and compare the observed values in the original network to the distribution of values we get in simulated networks. In the plots in Figure A2.2, the bold solid lines represent the observed values. The dashed lines represent the simulated values, with the light gray curves representing the range in 10th and 90th quantiles, and boxplots showing the median and interquartile range. Model fit is considered good if the observed values are largely within the ranges, and are close to the medians.

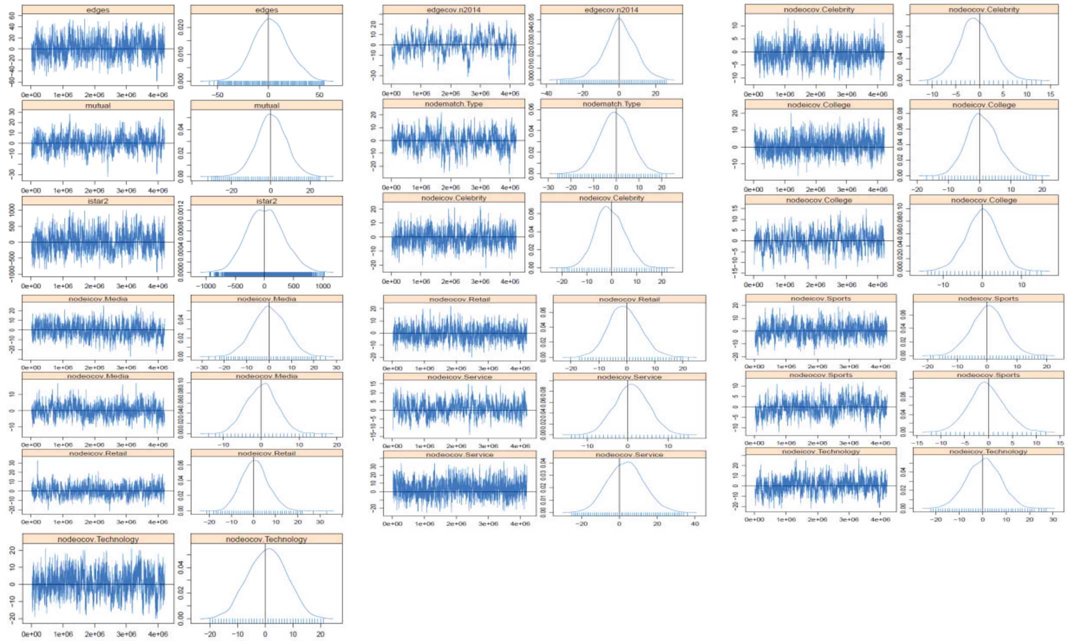

Figure A2.1: Convergence of parameters

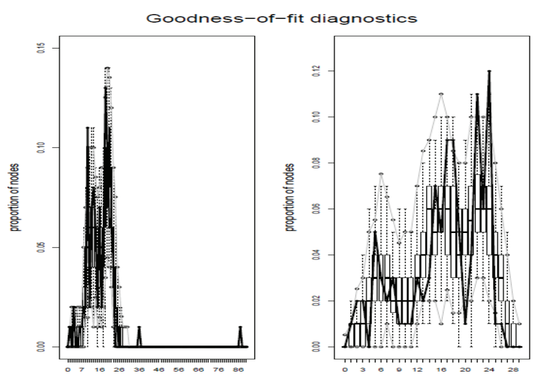

Figure A2.2: Goodness of fit