

Harvesting Brand Information from Social Tags

Social tags are user-defined keywords associated with online content that reflect consumers' perceptions of various objects, including products and brands. This research presents a new approach for harvesting rich, qualitative information on brands from user-generated social tags. The authors first compare their proposed approach with conventional techniques such as brand concept maps and text mining. They highlight the added value of their approach that results from the unconstrained, open-ended, and synoptic nature of consumer-generated content contained within social tags. The authors then apply existing text-mining and data-reduction methods to analyze disaggregate-level social tagging data for marketing research and demonstrate how marketers can utilize the information in social tags by extracting key representative topics, monitoring common dynamic trends, and understanding heterogeneous perceptions of a brand.

Keywords: social tags, user-generated content, brand associative networks, text mining, topic modeling

Online Supplement: <http://dx.doi.org/10.1509/jm.16.0044>

The advent of user-generated content has revolutionized the art and science of marketing research by making available a significant amount of online data that reflect consumers' opinions, attitudes, and preferences for products, services, and brands. Marketing scholars have proposed many methods to obtain insights on brand perceptions and competitive market structure by mining search data (e.g., Ringel and Skiera 2016), microblog data (e.g., Culotta and Cutler 2016), online reviews (e.g., Lee and Bradlow 2011; Tirunillai and Tellis 2014), and posts on an online discussion forum (Netzer et al. 2012). At the same time, marketing practitioners have obtained brand perceptions and brand associative networks by mining online discussions and posts about brands (e.g., Nielson's Brand Associative Map, McKinsey's Brand Navigator).

In this article, we focus on user-generated social tags associated with brands and propose an approach to analyze the large set of brand associations obtained from social tags for marketing research. Users largely employ social tags in social media platforms to organize and discover content in line with its topical relevance. Many forms of online content can be tagged: for instance, web links (links to a video, photo, blog post, or news article) are tagged in Delicious; images and photos are tagged and shared on Pinterest, Instagram, and Facebook; videos are tagged on YouTube; and tweets are tagged (via hashtags) on Twitter. Mining social tags provides marketing researchers with unique opportunities to understand

brand associations that are directly and explicitly mentioned by individual users/consumers.

Despite the popularity of the use of social tags, there is comparatively little research on what marketing researchers can learn from social tags. A recent study by Nam and Kannan (2014) shows that social tags contain significant informational value in understanding brand associative network and competitive market structure and predicting firm performance. Although Nam and Kannan demonstrate the value of social tagging data, they leave several key issues unexplored. First, it is not clear whether social tags produce new, different insights relative to existing approaches in eliciting brand associations. Such a comparative analysis would highlight the relative advantages and limitations of using social tags for marketing research. Second, Nam and Kannan's study focuses on aggregate-level information contained in a large set of social tags. They aggregated and summarized more than 7,000 tags obtained from the millions of individual tagging activities into brand-level metrics such as valence and competitive overlap of brand associations. Although those metrics provide a useful summary of high-level brand information, they do not provide insights into disaggregate-level information in brand associations and individual users. The richness inherent in the disaggregate-level tagging data can enable marketing managers to exploit (1) user-level disaggregate brand association information (e.g., How can managers identify and describe heterogeneous perceptions on the brand? What representative topics describe the brand perceptions of the distinct user segments, and what insights can these topics provide on improving overall brand perceptions?) and (2) temporal disaggregate brand information (e.g., Which brand associations are dynamically correlated, and how do they contribute to the evolving brand image?). We focus on these hitherto unexplored issues to highlight the value of social tags for marketing research.

Hyoryung Nam is Assistant Professor of Business, University of Washington Bothell (e-mail: hnam1@uw.edu). Yogesh V. Joshi is Associate Professor of Marketing (e-mail: yjoshi@rhsmith.umd.edu), and P.K. Kannan is Ralph J. Tyser Professor of Marketing Science (e-mail: pkannan@rhsmith.umd.edu), Robert H. Smith School of Business, University of Maryland. Dominique Hanssens served as area editor for this article.

Our goal in this article is to present an approach to collect and analyze social tagging data for marketing research and show how marketing managers can derive useful insights from social tags. We present an application of existing text-mining and data-reduction methods to understand consumers' brand perceptions reflected in social tagging data. Our research employs data reduction models such as latent Dirichlet allocation (LDA) topic modeling (e.g., Blei 2012; Blei, Ng, and Jordan 2003; Griffiths and Steyvers 2004; Puranam, Narayan, and Kadiyali 2017; Tirunillai and Tellis 2014), dynamic factor analysis (DFA; e.g., Du and Kamakura 2012; Zuur et al. 2003), and clustering analysis to help managers efficiently process a large volume of qualitative information. We show how social tags can be used to infer the major latent topics underlying consumers' categorization of tags associated with a brand and to understand heterogeneous perceptions of a brand with the emergence of new content related to the brand; in addition, we identify tags' latent factors on the basis of their correlations over time and show how the factors evolve dynamically.

Table 1 summarizes the unique positioning of our research relative to recent approaches developed for processing brand information obtained from large amounts of data generated by consumers. As Table 1 shows, this article proposes a method for obtaining brand associative network structure by employing the information contained in tagging data, which has been underexplored in marketing compared with search data (e.g., Ringel and Skiera 2016), microblog data (e.g., Culotta and Cutler 2016), and data from online discussion forums (e.g., Netzer et al. 2012). Our approach provides unique and distinctive insights first and foremost because our approach uses brand associations that are directly and explicitly stated by users/consumers to determine brand associative structure and brand position. In contrast, in extant research, brand associative strengths are indirectly inferred from the similarity between brand followers and exemplar account followers (Culotta and Cutler 2016), product positions are indirectly inferred from search and comparison patterns (Ringel and Skiera 2016), and brand associations are discovered by an automatic keyword extraction algorithm (Netzer et al. 2012). Second, this article proposes a methodology to analyze and visualize a large set of brand perceptions/attributes. In contrast, Ringel and Skiera (2016) focus on visualizing positioning of a large set of products, and Culotta and Cutler (2016) focus on deriving social perceptions of three brand attributes for a large set of products. Our study shows how to analyze a large set of qualitative brand perceptions/attributes (>1,000 tags) and acquire managerial insights for marketing strategies.

Given that our main objective is to illustrate how marketing managers derive new and distinctive insights from the information contained in social tags, this article is structured as follows. We first lay out the conceptual foundation that establishes the appropriateness of using social tags to capture brand associations. We then describe the procedures to obtain brand associations from social tags and discuss the similarities and distinctiveness of our proposed social tag-based approach compared with existing approaches to generate brand maps, including text mining and customer interviews. Subsequently, we illustrate how existing text-mining methods and data-reduction methods can help marketers process and monitor

the tag information using perceptual brand maps by deriving representative topic distributions of distinct consumer segments and extracting common dynamic trends. We conclude with a discussion of the applications of social tags in specific settings and the associated managerial implications and outline areas for further research.

Background for Social Tags

Many social media platforms employ social tags (e.g., hashtags in Twitter, pins in Pinterest, geotags in diverse social media platforms; for examples, see Web Appendix A). In these social tagging platforms, users interpret, categorize, and summarize a large volume of content, including textual data (e.g., reviews, blogs, microblogs, news articles) and nontextual data (e.g., images, videos, songs). Tags may include descriptive words (e.g., "news," "style," "tips"), identifiers for the brand and subbrand (e.g., "mac," "iphone," "ipod," "iwatch"), identifiers for the category (e.g., "computers," "smartwatch"), and descriptors of how consumers think and feel about the focal brand (e.g., "cool," "inspiration").

Social media platforms employ social tagging systems primarily because (1) tags enhance the customer experience by promoting user convenience in content discovery and content categorization and (2) tags enable the platform to track and manage user content through user-generated topical categories. Interpreting the meaning of social tags is contingent on the type of objects being tagged (textual posts, images, photos, etc.) and users' motivations for social tagging. Table 2 shows the taxonomy of existing social tagging systems in (1) content management platforms and (2) microblogs and social network platforms. The differences in tagging systems mainly arise from their design (Huang, Thornton, and Efthimiadis 2010).

In content management platforms, social tags primarily serve as a tool to manage a collection of content. Representative examples include social bookmarking platforms (e.g., Delicious, Tumblr, reddit) and content curation platforms (e.g., Pinterest, Last.fm). In these platforms, users build and manage a collection of content and search, discover, and share content using social tags (e.g., Ames and Naaman 2007; Gilbert et al. 2013; Strohmaier, Körner, and Kern 2010). Thus, the tags associated with brands/products in these platforms provide insights into (1) how online users interpret and perceive content associated with brands/products, (2) how a brand is grouped together with competing brands, and (3) how potential customers construct the consideration set.

Microblogs and social network platforms (e.g., Facebook, Twitter, Instagram) employ tagging systems to help users categorize, search, monitor, and participate in discussions based on user-defined tags. Because posts in microblogs and social network platforms have a short life, tags in these platforms compared with tags in content curation platforms are more often about temporally relevant information and emergent topics, which may appear and disappear quickly (Huang, Thornton, and Efthimiadis 2010; Teevan, Ramage, and Morris 2011). Thus, investigating the trends of tags in these platforms enables marketing managers to monitor emergent topics, track engagement in each topic, and efficiently capture customer views through keywords associated with brands. In addition, geotags

TABLE 1
Comparison with Previous Studies to Analyze Brand Information from a Large Amount of Data

	Netzer et al. (2012)	Nam and Kannan (2014)	Ringel and Skiera (2016)	Culotta and Cutler (2016)	This Study
Objective	To visualize competitive market structure using text mining on big data	To investigate the informational value of the customer-based brand equity derived from consumer tagging data in firm valuation	To understand asymmetric competition in the product categories	To propose a new methodology to infer attribute-specific brand ratings based on the similarity between exemplar accounts and brand follower accounts	To propose a new approach to analyze a large set of brand attribute information obtained from user-generated tagging data for marketing research
Data	169 car models and 1,200 attributes	60 brands with 7,000+ attributes	1,000+ products with no attributes	200 brands with 3 attributes	7 brands with 6,000+ attributes
Sources	Online discussion forum	Social tags	Search data from a product comparison website	Twitter	Social tags
Brand association elicitation	Text mining algorithms	User-generated social tags	N.A.	Predefined by researchers	User-generated social tags
Output	Visualization of market structure of brands	Customer-based brand equity metrics	Visualization of market structure of products	Social perception score on an attribute of brands	Aggregate and disaggregate brand perception map
Segmentation	K-means clustering	No	Multilevel Louvain	No	LDA parameterized Gaussian finite mixture model
Dynamics	No	Yes	No	No	Yes
Dynamic dimension reduction	No	No	No (dimension reduction)	No	Yes (DFA)
External validation	• Purchase data • Survey	• BCM • Search data • Blog data	Survey	Survey data	BCM (survey)

Notes: N.A. = not applicable. BCM = brand concept map.

TABLE 2
Overview of Social Tagging Systems

	Content Management Platforms	Microblogs and Social Network Platforms
Examples	<ul style="list-style-type: none"> • Social bookmarking platforms (Delicious, Tumblr, reddit) • Content curation platforms (Pinterest, last.fm) 	<ul style="list-style-type: none"> • Facebook • Twitter • Instagram
Why do users use tags?	<ul style="list-style-type: none"> • To build, manage, and share a collection of content based on topical relevance • To search and discover content using others' tags. 	<ul style="list-style-type: none"> • To describe and share the gist of social communication based on topical relevance (geolocation, members, type of events, emotions, and thoughts) • To search, monitor, and generate the trending topics
Insights for marketing managers?	<ul style="list-style-type: none"> • To understand users' interpretations and perceptions about content • To capture customers' consideration set • To understand competitive market structure 	<ul style="list-style-type: none"> • To monitor engagement and participation in emergent topics • To capture the voice of customers • To understand the context of brand usage experience

that reveal geolocation information of posts on social media platforms provide geographical details of user experiences related to a brand.

From a user perspective, the key motivations for social tagging fall into two categories: content classification and content description (see also Strohmaier, Körner, and Kern 2010). Researchers have found that people create different types of keywords depending on their motivations for tagging. For instance, when people intend to categorize content, they are more likely to use high-level attributes as tags; yet when people intend to describe content, they are more likely to use contextual attributes as tags (Strohmaier, Körner, and Kern 2010). In addition, each type of motivation could be driven by self-oriented needs (e.g., organization of content for one's own reference), social communication (e.g., information sharing and opinion expression regarding the content with other users), or a combination of both (see also Ames and Naaman 2007). Table 3 summarizes the key motivations for social tagging.

Whatever the user's motivation, the process of social tagging involves interpreting the gist of the content by relating the content to concepts organized in the user's memory and subsequently describing, categorizing, or communicating that

content. Therefore, tags reflect not only the content that is tagged but also a succinct representation of the user's knowledge structure—that is, his or her mental representation of related concepts. Thus, one can view social tags as the outcome of categorization, description, or communication of content filtered through the lens of a person's knowledge structure—that is, an individual-specific, thoughtful interpretation of content.

Figure 1 presents an illustrative example of social tagging process for two users, using one of the most popular web links tagged with Apple on Delicious: Guy Kawasaki's blog post on what he learned from Steve Jobs. User A's social tags, based on her previous tagging activities, indicate that she has been interested in corporate strategy (e.g., recommended_reading, guide, strategy, casestudy, businessmodel, marketing, innovation) related to various information technology firms (e.g., Yahoo, Microsoft, Google, Apple). This previous interest might lead her to engage with Kawasaki's blog post and tag it with "innovation," "guide," "recommended_reading," and "Apple." In contrast, User B's previous tagging history indicates that she has been primarily into technology, management, and development; thus, she describes the same blog post with tags such as "product," "management," and "stevejobs." As such, each user associates different keywords with an object/content depending

TABLE 3
Motivations for Creating Social Tags

	Description	Characteristics of Tags
Content Classification/Categorization		
Self-oriented motivation	Building own system of classification of content	High-level attributes
Social communication	Helping others find content in the category (discover content) and have a better categorization system (discover tags)	High-level attributes
Content Description		
Self-oriented motivation	Presenting the gist of the content (especially useful when the content is not textual; e.g., images, music, products)	Contextual attributes
Social communication	Creating trending topics and participating in discussions on trending topics	Contextual attributes

FIGURE 1
Illustration of the Social Tag Creation Process

A: Blog Post

Guy Kawasaki

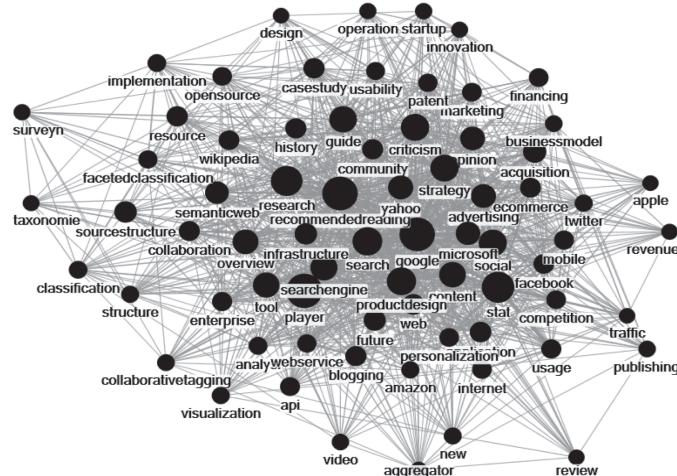
Add to circles

Guy Kawasaki Oct 9, 2011 (edited) - Public
(Sat01) What I Learned From Steve Jobs

Many people have explained what one can learn from Steve Jobs. But few, if any, of these people have been inside the tent and experienced first hand what it was like to work with him. I don't want any lessons to be lost or forgotten, so here is my list of the top twelve lessons that I learned from Steve Jobs.

Experts are clueless.

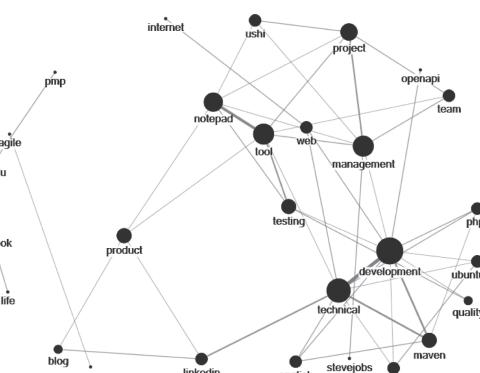
B: User A's Social Tags Before Seeing the Article



C: User A's Social Tags for the Article



D: User B's Social Tags Before Seeing the Article



E: User A's Social Tags for the Article



on her interest, motivation, and knowledge structure, which may be reflected in previous tagging activities.

From the perspective of a brand, an individual user's tags associated with a collection of documents and online content related to the brand can provide useful information regarding the user's knowledge structure associated with this brand. Each document or piece of content can be viewed as a brand-related stimulus, to which the user responds with tags that reveal partial information of his or her knowledge structure about the brand. In addition, as users encounter new content online over time, this may also interact with, change, or shape their knowledge structure and category schema, through either categorization or analogy (Buchanan, Simmons, and Bickart 1999; Griffiths, Steyvers, and Tenenbaum 2007). Tags created to categorize or describe this content can potentially reveal such dynamics. Similarly, differences in tags that different people use for the same content can provide insights regarding the extent of heterogeneity in users' knowledge structures about a focal brand. This renders tags ideal for exploring how users believe, think, feel, and reason about brands.

In summary, social tags can provide a rich and heterogeneous interpretation of a brand across and within individuals over time as their knowledge structure is shaped by their encounters with stimuli (content, interactions, and experiences; Nam 2012; Nam and Kannan 2014). As such, these tags provide a rich associative structure reflective of how consumers relate to brands. Thus, social tags can serve as effective input for constructing brand associative maps.

An Overview of Our Method

Figure 2 outlines our method to collect and analyze brand information from social tagging data. It consists of four stages: (1) data collection, (2) association elicitation, (3) aggregation and visualization, and (4) analysis of disaggregate-level data. For the sake of illustration, we choose Apple as our focal brand.

Step 1: Data Collection

Data collection starts with specifying the set of content in a social tagging platform in a specified time frame. For instance, we consider social tags from Delicious, a top 500 global website in terms of traffic rank (Alexa.com 2011), with a three-month global Alexa traffic rank of 252 (March 13, 2011, through June 12, 2011). Thus, this social bookmarking website is widely accessed and fairly representative in terms of gathering a broader of opinions from online users. Step 1 in Figure 2 shows examples of bookmarks tagged with "Apple" and other keywords by different users.

To collect bookmarks relevant to Apple, we first specify the start list of social tags and then collect the set of content tagged with those social tags. The start list of social tags can be constructed by drawing on a combination of (1) brand name and subbrands, (2) tags frequently linked to a brand, (3) tags frequently linked to competitors, and (4) predefined keywords identified from previous surveys and/or consumer interviews. Researchers can specify the set of competitors using (1) a predefined set of competitors by marketing managers; (2) an external source such as Standard Industrial Classification (SIC) code, Hoover's database, or Google Finance;

and (3) co-occurrence patterns of social tags across brands in the tagging networks (i.e., brands that have many of the same tags associated with them as the focal brand). In the data used for our analysis, we identify the set of Apple's competitors using multiple data sources: other firms with the same four-digit SIC code, Hoover's classification of competitors, Google Finance, viewing history provided by Yahoo Finance, and the tagging co-occurrence structure on Delicious. We include as competitors companies that appear at least three times in these five sets.¹ These include Microsoft, Google, Blackberry, Nokia, Dell, and HP. This set can be easily redressed or expanded as desired by the focal brand manager. Once the tagging data associated with the start list can be collected, the data can be aggregated on any given time frame (hourly, daily, weekly, monthly, quarterly, or yearly basis). Depending on managerial objectives, an appropriate time window can be specified.

Step 2: Association Elicitation

The second step is to identify a set of relevant core social tags for the focal brand. Step 2 of Figure 2 illustrates the association elicitation process. This goal is achieved by evaluating all tags associated with a brand using predefined metrics and identifying tags that score the highest on these metrics. To do so, we propose the use of the following metrics.

Associative strength metrics. Core associations for the focal brand can be identified on the basis of the associative strength between a brand and a tag. A key metric for capturing associative strength between a brand and a tag is "co-occurrence volume," measured as the number of times a brand is linked to a tag through a bookmark. Thus, the co-occurrence volume of tag j with brand i for a given time window t , $N_t(B_i, T_j)$ is defined as the volume of bookmarks linked to both brand i and tag j during the time window t . In other words, this metric captures how many brand i -tag j pairs are created during time window t .

An alternative way for capturing associative strength is to scale the co-occurrence volume, $N_t(B_i, T_j)$, by the bookmark volume linked to the brand and the bookmark volume linked to the tag, as specified in Equation 1. This metric, "scaled co-occurrence volume," measures the cosine distance between each brand and each tag; prior research has found this metric to be useful in capturing the similarity between two tags in Delicious (Robu, Halpin, and Shepherd 2009).

$$(1) \quad SN_t(B_i, T_j) = \frac{N_t(B_i, T_j)}{\sqrt{N_t(B_i)N_t(T_j)}},$$

where $N_t(B_i)$ is the volume of bookmarks linked to brand i during time window t , and $N_t(T_j)$ is the volume of bookmarks linked to tag j during time window t . Note that there are other alternatives for capturing the associative strength between a brand and a tag based on assigning different weights to each social tag. For instance, one could use the weight based on

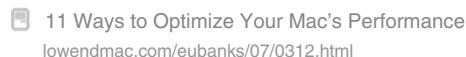
¹Here, we employ both external criteria such as SIC code and Hoover's and consumer-driven criteria such as viewing history from Yahoo Finance and tagging co-occurrence structure on Delicious. Researchers can further consider obtaining a snowball sample of competitors based on tagging structure.

FIGURE 2
Illustrative Description of Social Tag-Based Marketing Research

A: Step 1—Collection of Social Tags Associated with Brand


'You've got to find what you love', Jobs says
news-service.stanford.edu/news/2005/june15/jobs-061505.html

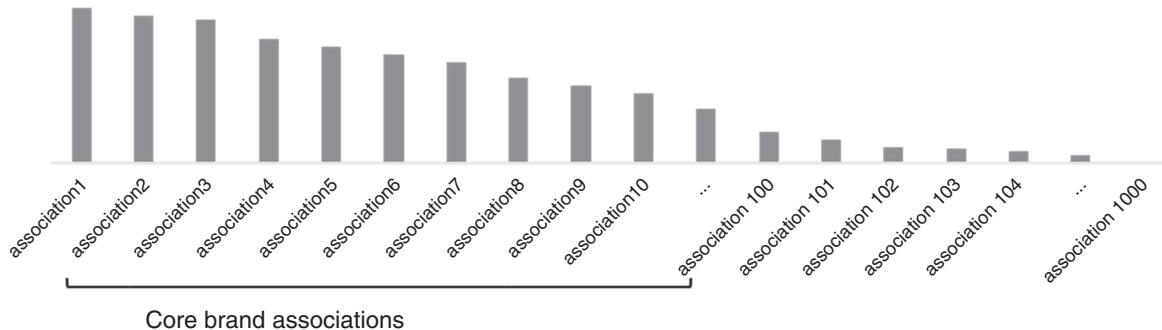
inspiration stevejobs speech life apple


11 Ways to Optimize Your Mac's Performance
lowendmac.com/eubanks/07/0312.html

mac osx tips optimization apple

B: Step 2—Association Elicitation

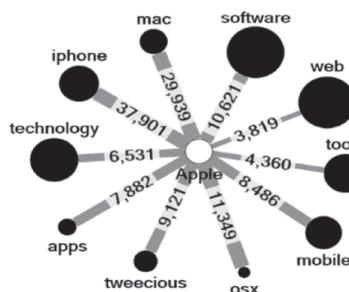
Associative strength of keywords with Apple



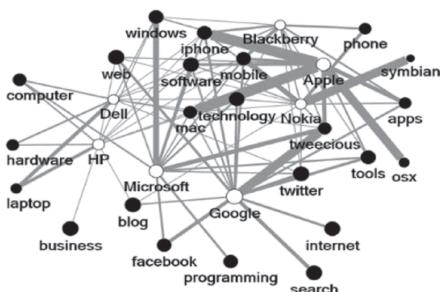
Core brand associations

C: Step 3—Aggregation and Visualization

Brand-Centric Map for Apple^a



Multibrand Map for Apple and Competitors^b

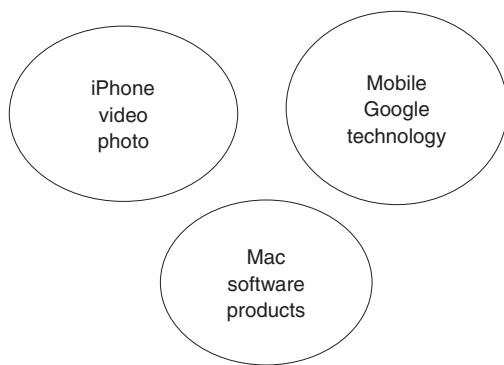


Brand Metrics

- Social attention
- Valence of associations
- Breadth of associations
- Competitiveness

D: Step 4—Analysis of Disaggregate-Level Data

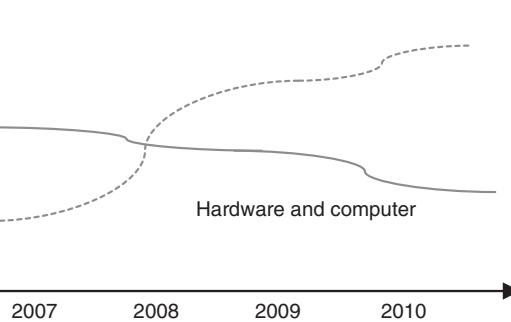
Discovering Representative Topics for Customer Segments



Discovering Dynamic Trends in Brand Associations

Mobile and technology

Hardware and computer



^aWe created the map using Delicious bookmark data generated in 2009. The size of the circle is proportional to the volume of bookmarks linked to each keyword, and width of the link is proportional to the co-occurrence volume of two keywords with Apple, which is stated in the number on each link.

^bWe created social tag maps using Delicious bookmark data generated in 2009 using the Fruchterman–Reingold graph algorithm. The size of the node is proportional to the volume of bookmarks linked to each keyword, and the opacity of the link is proportional to the co-occurrence volume of a keyword with each brand.

TABLE 4
Decision Rules to Select the Set of Relevant Brand Associations

A: Number of Tags to Explain 95% and 90% of Co-Occurrence Volume							
% Co-Occurrence Volume Explained	Apple	Blackberry	Dell	Google	HP	Microsoft	Nokia
95%							
Number of tags	2,254	1,647	1,972	2,578	2,105	2,012	1,727
% of tags	37%	40%	48%	39%	43%	36%	41%
Minimum co-occurrence volume	31	5	4	79	6	28	6
90%							
Number of tags	1,430	969	1,258	1,685	1,300	1,298	1,057
% of tags	24%	23%	30%	26%	27%	23%	25%
Minimum co-occurrence volume	65	12	7	161	12	59	13
B: Percentage of Co-Occurrence Volume Explained by Decision Rules							
Decision Rule	Apple	Blackberry	Dell	Google	HP	Microsoft	Nokia
N(B, T) > 5							
Number of tags	4,502	1,632	1,517	5,737	2,219	3,840	1,843
% of tags	74%	39%	37%	87%	45%	69%	44%
% explained	99.4%	94.9%	92.3%	99.8%	95.5%	99.2%	95.6%
N(B, T) > 10							
Number of tags	3,703	1,064	953	5,210	1,420	3,139	1,235
% of tags	61%	26%	23%	79%	29%	57%	29%
% explained	98.6%	91.0%	86.4%	99.6%	91.0%	98.3%	91.7%

Notes: Tables are based on social tags created for each brand in 2009. Similar tables can be created using the scaled-volume and weighted-volume metrics.

the number of other tags mentioned in each bookmark or the weight based on the order of social tag mentioned in each bookmark. In the data we analyze, we find alternative weighted metrics to be highly correlated with the co-occurrence metric. (For additional discussion on alternative associative strength metrics, see Web Appendix B.)

The required number of tags. The set of relevant brand associations can be obtained by specifying the level of explanatory power desired. Depending on the desired amount of information, a marketing manager can flexibly choose different cutoffs. For instance, Table 4, Panel A, shows the number of tags needed to explain 95% and 90% of co-occurrence volume of all tags linked to each brand based on bookmark data generated in 2009. For instance, 95% of co-occurrence volume of tags linked to Apple can be explained with 2,254 tags (37% of all tags linked to Apple) and the minimum co-occurrence volume of these tags is 31. The choice of the set of relevant associations can be further complemented by specifying a prerequisite level of co-occurrence for each tag (e.g., co-occurrence volume greater than 5 and 10). Table 4, Panel B, presents the percentage of co-occurrence volume explained by multiple decision rules. For instance, once a researcher selects associations whose co-occurrence volume is greater than 10, (s)he can explain 98.6% of co-occurrence volume of Apple with 3,703 tags (61% of all tags) and 91% co-occurrence volume of Blackberry with 1,064 tags (26% of all tags).

Step 3: Aggregation and Visualization

In Step 3, the goal is to combine the associations discovered in the elicitation stage into a holistic description of the brand. Visualization of brand associative networks helps meet different

managerial objectives. For instance, a brand-centric map can present the key associations for the focal brand, or for each competing brand. Distinct associative maps can be created using either all keywords, only valenced keywords (that capture both positive and negative sentiment), or only descriptive, neutral keywords. Step 3a in Figure 2 shows the various social tags associated with Apple during 2009. The size of a node represents the volume of keywords generated, and the width of the edge represents the associative strength between two nodes, proportional to the co-occurrence volume of the two keywords. Although Step 3a does not consider the intertag relationships, one can build brand associative network with intertag relationships. Such associations can also be represented in a multi-brand map to highlight interconnected associations across brands. Step 3b of Figure 2 presents the associative networks for multiple brands (i.e., the focal brand and its top competitors: Apple, Blackberry, Dell, Google, HP, Microsoft, and Nokia). From this figure, a manager can gain insight regarding relative positions of each brand on a network of keywords of interest. To further derive a spatial representation of the competitors in the market, existing methods to construct a perceptual map—such as multidimensional scaling (e.g., DeSarbo et al. 1996; Shugan 1987), Bayesian models of graph formation (e.g., Hui, Huang, and George 2008), or correspondence analysis (e.g., Carroll, Green, and Schaffer 1986)—can be employed.

The aggregated brand information can be further explored with the brand metrics potentially related to the diagnostic value of brand assets (Nam and Kannan 2014). These metrics can capture dynamics within social attention generated by a brand; the richness, valence, and dispersion of brand associations; and

the competitiveness of a brand (for more discussions about brand metrics based on social tags, see Web Appendix C). Although the information obtained from these aggregated social tags provides insights on brand health and brand equity, such aggregation does not fully harvest rich, qualitative information contained in social tags. For example, the aggregated metrics (1) do not provide insights on which keywords are more frequently associated together, (2) do not identify which keywords are dynamically more correlated and thus move together in the social tagging platform, and (3) ignore the possibility that there could be distinct clusters of consumers who interpret the content differently. Thus, to harvest a large volume of qualitative information, it is critical to overcome these challenges by employing tools that analyze disaggregate-level data.

Step 4: Analysis of Disaggregate-Level Data

Step 4's goal is to understand and interpret disaggregate-level information contained in social tags by applying existing text mining techniques and data reduction methods. We focus on two levels of disaggregate information in social tagging data: user-level disaggregate brand association information and temporal disaggregate brand association information. The first challenge is how to obtain a qualitative summary of the associative relationship of a large volume of brand associations created by individual users. We show how existing LDA models can help identify the representative underlying topics that capture the brand perceptions of distinct user segments through the associative relationships between more than 1,000 user-generated keywords associated with a brand. These insights are key to improving the over-all brand perceptions by homing in on specific segments with perceptions that are not aligned with the core brand perceptions and examining how these perceptions could be improved through appropriate communication campaigns or marketing plans directed at these segments. The subsequent challenge is to investigate temporal disaggregate brand association information and discover which associations are more frequently paired with a brand over time. We apply DFA and show how it provides insights into dynamic relationships between brand associations. By doing so, we shed light on how the brand image is evolving over time and correlate it with specific events. Thus, analyzing disaggregate-level information in social tagging data provides a richer view of brand associative structure compared with aggregate-level social tag metrics.

Comparing the Social Tag-Based Approach with Existing Approaches

In this section, we evaluate the suitability of the use of social tags for discovering brand associations by comparing the brand associations elicited from our approach with those elicited from existing approaches. We discuss the key differences of the proposed social tag-based approach with the two most common existing approaches: primary data-based approach and text mining. Then, we discuss whether the proposed social tag-based approach elicits new, distinct brand associations relative to existing approaches. The analysis suggests that the social

tag-based approach can serve as a complementary method for eliciting brand associations. We also discuss the scope and limitation of the proposed social tag-based approach.

Existing Approaches to Elicit Brand Associations

Existing approaches vary in terms of the nature and richness of information contained in the data, as well as in the resources and expertise required for successful implementation. The differences mainly arise from the elicitation process of core brand associations and the nature of the collected data. Next, we discuss the similarities, complementarities, and differences of the following select established approaches: Zaltman's metaphor elicitation technique (ZMET; e.g., Zaltman 1997; Zaltman and Coulter 1995), brand concept maps (BCMs; e.g., John et al. 2006; Joiner 1998), categorization and sorting (e.g., Blanchard, Aloise, and DeSarbo 2016; Blanchard and DeSarbo 2013; Hamilton et al. 2014; Ratneshwar and Shocker 1991), and the more recent text-mining approaches (e.g., Lee and Bradlow 2011; Netzer et al. 2012; Tirunillai and Tellis 2014).

Primary data-based approach. Many prior studies have employed primary data (consumer surveys and interviews) to elicit brand associations. We primarily review the three existing popular methods: ZMET, BCMs, and categorization and sorting. The primary assumption of the ZMET approach (Zaltman 1997; Zaltman and Coulter 1995) is that a significant portion of consumers' thoughts and knowledge is stored in a nonverbal form and cannot be fully elicited with verbal communication. Thus, ZMET employs in-depth personal interviews using qualitative techniques such as Kelly's repertoire grid, laddering exercises, and verbal/nonverbal cues (e.g., images during the elicitation stage) to understand the core associations linked to a topic. Although ZMET can help identify deep, unconscious thoughts and feelings related to a brand by using multiple qualitative approaches as well as both verbal and nonverbal aspects of a consumer's behavior, this process is quite challenging to implement and often involves close interactions with only a few consumers. The elicitation stage is highly time and labor intensive (e.g., seven to ten days for subjects to collect visual images and two-hour in-depth, one-on-one interviews to obtain an individual brand map). Accessibility is another issue for ZMET because it requires interviewers with expertise in qualitative elicitation techniques, thus raising administrative costs.

The BCM method (e.g., John et al. 2006; Joiner 1998; Novak and Gowin 1984), in contrast, employs more structured procedures to elicit core associations, map the associations, and synthesize individual maps into consensus maps. To elicit core associations, BCM utilizes prior consumer research as well as input from the brand management team. Then, through one-on-one interviews, the researchers create individual concept maps primarily drawing on identified core associations. The final consensus map is developed on the basis of the aggregated frequency of the individual maps, revealing a hierarchical associative structure with differential associative strengths. Compared with ZMET, the BCM method is somewhat easier to administer and analyze. In addition, it flexibly accommodates inputs from managers. However, it may

not be adequate for eliciting unconscious feelings and brand associations that need additional in-depth probing.

Researchers have also employed the categorization and sorting method to elicit brand perceptions and category perceptions (e.g., Blanchard, Aloise, and DeSarbo 2016; Blanchard and DeSarbo 2013; Hamilton et al. 2014; Ratneshwar and Shocker 1991). In this approach, respondents are typically asked to sort a set of objects (brands) into categories according to their perceived similarity. Compared with ZMET and BCMs, sorting is a relatively easy task for the respondents and can be completed faster, with less respondent fatigue (Bijmolt and Wedel 1995). However, in this approach, the brand attributes for the basis of the sorting task are predetermined by researchers, and multiple rounds of data collection are required to learn dynamics in brand associations.

Thus, the methods based on primary data face the following challenges: (1) they are labor intensive because they employ qualitative analysis and one-on-one personal interviews, (2) they often require specialized expertise, (3) they are often implemented at specific time periods and are a static representation of brand perception rather than a dynamic brand map over time, (4) they involve small sample sizes and tend to be very expensive if the focal brand tries to obtain a brand map from a larger sample, and (5) they are based on stated brand associations and thus bound by the elicitation techniques.

Text-mining approach. Recent work using text mining has offered promise for addressing some of the problems identified with ZMET and BCM. Text mining is a tool that helps discover patterns in raw text and extract relevant information from textual data. Recent marketing studies employing text-mining tools have created brand-associative networks by automatically identifying keywords from user-generated content such as posts on online user forums or online user reviews (e.g., Lee and Bradlow 2011; Netzer et al. 2012; Tirunillai and Tellis 2014). Here, the elicitation stage consists of multiple steps: cleaning and preparing the text and extracting appropriate information from the text (Netzer et al. 2012). Researchers use a rule-based approach, a machine-learning approach, or a hybrid of these two approaches to extract the information in the text (Netzer et al. 2012). In the aggregation and visualization stage, researchers identify the associative relationships on the basis of the co-occurrence pattern of the identified keywords.

Compared with a primary data-based approach, the brand association elicitation process in text-mining approaches enables researchers to obtain brand associations from a large volume of data with a low level of human labor, given the automation offered during the elicitation process. Thus, brand associations elicited from automatic keyword extraction in a text-mining approach (1) can be automatically acquired, thus cutting down on labor, time, and expertise requirements; (2) can be automatically updated on a real-time basis, thus providing a dynamic, rather than a static, map; (3) are constructed from a larger customer base, with minimal additional costs; and (4) can track an extensive set of associations, including competitors' brand associations. Nevertheless, the keyword extraction process in text mining tends to be computation intensive, as it employs multiple stages of model estimation and data training for the keyword extraction process.

Evaluation of the Social Tag-Based Approach

The proposed social tag-based approach has several advantages over the existing methods. Table 5 summarizes the comparison of the social tag-based approach to existing methods. The major contrast is at the association elicitation stage. Primary data-based approaches employ the consumers' direct input from surveys, interviews, or sorting tasks typically designed to collect brand images/attributes and thus tend to generate more subjective keywords, be more sensitive to sample size, and often depend on researchers' interpretation. Text mining employs automatic keyword extraction algorithms that can be sensitive to the assumptions of a researcher or a marketing manager. In contrast, the social tag-based approach utilizes brand associations directly generated by consumers in association with their interactions with brands. This, however, makes them sensitive to (yet capable of capturing) consumers' biases and potential social influences. In this subsection, we discuss similarities, differences, and complementarities of our proposed social tag-based approach with existing approaches based on two illustrative empirical studies.

The social tag-based approach versus the primary data-based approach. We investigated whether the social tag-based approach elicits new, distinct brand associations compared with one of the primary data-based approaches, the BCM approach. To obtain a BCM for Apple, we conducted one-on-one interviews with 23 subjects. Following John et al.'s (2006) methodology of obtaining consensus BCM, we (1) ask subjects "what comes to mind when [they] think about Apple," (2) give them detailed instructions as to how to draw a BCM (using the example presented in John et al. [2006, Figure 2, p. 553]), (3) ask them to draw their own concept map, and (4) draw Apple's consensus BCM using responses from all subjects. Forty-seven associations, mentioned by more than 25% of the respondents, are present in the consensus map. For comparison, a social tag-based brand map for Apple is constructed by drawing on the co-occurrence relationships between each association in the consensus BCM and the brand during a corresponding time window (six-months of social tagging data).

We investigated the correlation between the two metrics from the consensus BCM—the frequency and weighted frequency² of each association—and the co-occurrence volume of the corresponding social tag with the brand. Overall, the frequency of each association in BCM and the co-occurrence volume of the corresponding social tag is significantly correlated ($r = .56, p < .01$), as is the weighted frequency metric of each association in BCM and the corresponding co-occurrence volume metric of social tags ($r = .54, p < .01$). The correlation between BCM and our social tag-based approach is reasonably high given that we compared the social tags of more than 20,000 responses with BCMs obtained from 23 respondents.

²Similar to John et al. (2006), we give level 1 (direct) associations a weight of 3, level 2 (indirect) associations a weight of 2, and level 3 or lower (indirect) associations a weight of 1. The weighted frequency is calculated as the sum of the multiplication of this weight given to each association and the associative strength provided by each respondent.

TABLE 5
Comparison of Methodologies to Obtain Brand Associations

	Primary Data (ZMET, BCM, Sorting)	Text Mining (Lee and Bradlow 2011; Netzer et al. 2012)	Social Tag-Based Approach
Association elicitation	<ul style="list-style-type: none"> In-depth personal interviews Both verbal and nonverbal cues (e.g., photos, images) Prior consumer research Manager's opinions/insights Consumer interview Sorting and categorization task 	<ul style="list-style-type: none"> Elicited by a text-mining tool Rule-based Machine learning Hybrid approach 	<ul style="list-style-type: none"> Directly stated by consumers/online users Available as secondary data
Aggregation and visualization	<p>Participants create a map or visual montage (ZMET) and develop their brand maps in personal interview (BCM)</p> <p>Maps derived based on the perceived similarity between brands (objects) (sorting)</p>	<p>Maps based on elicited product attributes/brand associations from text-mining model</p>	<p>Maps based on tags stated by consumers/users in the absence of a researcher</p>
Richness of information	<ul style="list-style-type: none"> Deep understanding of a brand Unconscious aspects can be revealed Hierarchical associative structure Perceived similarity between brands 	<ul style="list-style-type: none"> Large-scale data Dynamics of associations Competitive intelligence 	<ul style="list-style-type: none"> Large-scale data Dynamics of associations Competitive intelligence Undirected Ability to capture the associations on nontextual data (images, music, etc.)
Limitations	<ul style="list-style-type: none"> Data from few subjects Difficult to track dynamics Difficult to collect large-scale data (data on indirect competitors or large number of brand associations) Difficult to quantify associative strengths 	<ul style="list-style-type: none"> Constrained by algorithmic interpretation Requires human labor for data training 	<ul style="list-style-type: none"> The number of social tags elicited by consumers is limited Sensitive to customers' biases and potential social influences
Costs	High to moderate (qualitative analysis expert required; primary data collection required)	Moderate (multiple stages of text-mining processes)	Low (publicly available and readily accessible)

Although our analysis suggests that brand associations elicited from a social tag-based approach are significantly consistent with brand associations elicited from BCM, several differences exist. We found that evaluative associations (e.g., “cool,” “innovative”) more frequently appeared in BCM, whereas descriptive associations (e.g., “iPod,” “computer”) more frequently appeared in social tags. This is because (1) the question in the survey induces respondents to generate more attitudinal and evaluative associations and (2) respondents tend to think about the brand in a more holistic way when they are given the brand name; in tagging online, users are given specific context and thus tend to think about more details. Despite such differences, we found that the frequency of each association in the consensus BCM and the tag co-occurrence volume for evaluative 30 keywords is highly correlated ($r = .70, p < .01$) and that for descriptive 17 keywords is also highly correlated ($r = .74, p < .01$). We obtained similar results for the weighted frequency of associations in the consensus BCM.³ We conducted similar analyses for the other two brands (Microsoft and Google) and found a similar pattern. In summary, although primary data-based approaches are more likely to reveal subjective, evaluative keywords than the social tag-based approach, there is a significant similarity in the brand attributes obtained through the primary data-approach and a social tag-based approach.

The social tag-based approach versus the keyword extraction process in text mining. We investigated whether the social tag-based approach elicits different brand associations from the keyword extraction process in a text-mining approach. We compared the social tags created on a blog post written by Guy Kawasaki about what he learned from Steve Jobs (see the example in Figure 1) with a text-mining analysis on the same blog post. The blog post contains 83 sentences, 1,143 words, and 6,529 characters. Following the text-mining procedure illustrated in Netzer et al. (2012), we identified 331 distinct keywords after removing the stop words, and these keywords were mentioned 597 times in the blog. Drawing on 70 users’ social tagging activities, we identified 57 distinct keywords, which were mentioned 217 times in social tags. Table 6 illustrates the list of the top keywords identified by a social tag-based approach and text-mining approach.

As we expected, the information entropy⁴ for the keywords identified by the social tag-based approach is lower (3.15) than that identified by text mining (5.51). The distribution of social tags is more concentrated on several representative keywords (the top four most frequently mentioned tags represent 50% of tag usages) than the distribution of keywords identified by text mining on the blog post (the top five most frequently mentioned keywords represent 10% of keyword usages). Social tags are not an accurate reflection of *all* the keywords in the original blog content, which can lead to a potential omission bias;

³The weighted frequency of associations in the consensus BCM and the tag co-occurrence volume for evaluative keywords is also highly correlated ($r = .74, p < .01$), as is that for descriptive keywords ($r = .73, p < .01$).

⁴We defined entropy in line with existing literature (Godes and Mayzlin 2004) as $-\sum_{j=1}^{NT} [N_t(B_i, T_j) / \sum_{k=1}^{NT} N_t(B_i, T_k)] \log [N_t(B_i, T_j) / \sum_{k=1}^{NT} N_t(B_i, T_k)]$.

TABLE 6
Keyword Comparison Between Social Tag-Based Approach and Text-Mining Approach

Keywords	Frequency	%
Social Tag-Based Approach		
stevejobs	43	20.5%
apple	22	10.5%
inspiration	22	10.5%
guykawasaki	18	8.6%
startup	13	6.2%
entrepreneurship	11	5.2%
lesson	9	4.3%
business	7	3.3%
design	7	3.3%
twitter	3	1.4%
wisdom	3	1.4%
Text-Mining Approach		
stevejobs	20	3.4%
people	14	2.3%
big	9	1.5%
player	9	1.5%
apple	8	1.3%
product	8	1.3%
app	6	1.0%
challenge	6	1.0%
hire	6	1.0%
tell	6	1.0%
unique	6	1.0%

nevertheless, they can serve as an efficient filter of the original content. The correlation between the term frequency of social tags and the term frequency of all the keywords (597 distinct keywords) identified by text mining was $.574 (p < .01)$, and the correlation was higher for the keywords for which term frequency in text mining was greater than 1 (116 distinct keywords; $r = .661, p < .01$), greater than 2 (54 distinct keywords; $r = .696, p < .01$), and greater than 3 (36 distinct keywords; $r = .702, p < .01$).

Consumers’ tagging activities are not a mere reflection of the original blog content. Twenty-four percent of social tags (14 out of 57 social tags) were terms used in the original blog text. Seventy-six percent of social tags were never mentioned in original text and represent consumers’ own interpretation of the content (e.g., “inspiration,” “entrepreneurship,” “innovation”), the author information (e.g., “Guy Kawasaki”), and categorization cues (e.g., “business,” “strategy,” “idea”). That is, the social tag-based approach provides an additional layer of insights on the original text because it incorporates consumers’ input. The keywords identified by text mining on a blog post are fixed and objective (only varying by algorithm choice in text-mining approaches and assumptions by researchers and managers), whereas the keywords identified by social tagging on blog posts vary as more consumers create tags. The brand associations discovered in the social tag based-approach are affected by each customer’s own point of view and potential social influence. Thus, the social tag-based approach can effectively capture people’s perceptions on a subject that is filtered through their own experiences and memories. When such social, individual-level bias is not desirable to interpret the

raw text (e.g., when deriving an objective interpretation of the content) or when the raw text includes all the individual-level bias, eliciting brand associations from text mining can be a better solution. However, if a researcher wants to obtain customers' subjective interpretation of the content, wherein customers' biases are incorporated, the social tag-based approach is preferable.

More importantly, we believe that the social tag-based approach can be a complementary tool for association elicitation when training the data for the text mining process. The keyword extraction process in text mining often requires human labor for better adaptation of algorithms to the domain of study (e.g., identifying the list of stop words, identifying important keywords such as brand names and product attributes). To this effect, Netzer et al. (2012) highlight the need to replace the initial training of data in text mining with tagging work using crowd-sourced marketplaces. We believe the use of social tags can guide this training process by providing the set of keywords retrieved by consumers/readers/users and complement the data-training process in keyword extraction text mining. For instance, in the previous example of text mining, researchers can employ the start list of keywords identified in social tags to mine the text. Some of the keywords that were not included in the top ten keywords using text mining's keyword extraction process can be discovered when researchers incorporate the input from social tags.

In summary, social tags are an inexpensive source of brand associations derived from large-scale content preprocessed by customers and online users. Thus, we recommend the use of social tags as a complementary tool to the keyword extraction of a text mining approach when a marketing manager needs to obtain a succinct summary directly generated by engaged online users on large-scale data.

Analysis of Disaggregate Information in Social Tags

In this section, we show the value of disaggregate level data in social tags. In addition, we illustrate how the challenges of understanding disaggregate-level information in social tags can be resolved by employing existing language processing and data-reduction techniques.

Value of User-Level Disaggregate Information in Social Tags

A reasonable assumption for tagging behavior is that each individual user interprets a single content/object differently. Such a difference mainly arises from (1) heterogeneous knowledge structure characterizing individual mental schema and (2) heterogeneous motivations for tagging, which may be time and context dependent within each person.⁵ As Figure 2

⁵Consumers' perceptual maps can vary across situation and time within each consumer (DeSarbo et al. 2008). The heterogeneity we discuss in this section captures both individual-specific heterogeneity and context-dependent heterogeneity but does not differentiate between the two because of the nature of the data, such that only a single tagging behavior on a single piece of content is observed for each person.

illustrates, each user associates different keywords with an object/content depending on his or her interest, motivation, and knowledge structure, which may be reflected in previous tagging activities. Thus, understanding the disaggregate-level associative structure of keywords is critical for harnessing rich information contained within social tags.

To further investigate heterogeneous perceptions of the same content, we collected 57 users' social tags on Kawasaki's blog post and conducted a clustering analysis on the similarity of tagging patterns across users. The hierarchical clustering using a parameterized Gaussian finite mixture model (Fraley and Raftery 2007) finds four segments as optimal (log-likelihood = 760.25; n = 54; d.f. = 247; Bayesian information criterion = 535.22). The model allocates 21 users in segment 1, 24 users in segment 2, 4 users in segment 3, and 5 users in segment 4. Users in segments 1 and 2 associated with fewer social tags ($M = 4.19$ and $M = 4.45$, respectively) than users in segments 3 and 4 ($M = 8.25$ and $M = 7.4$, respectively). Figure 3, Panel A, presents the aggregate perceptual map based on all social tags associated with this blog, and Figure 3, Panel B, presents the perceptual map for each segment. Segments 1 and 2's perceptions are similar (e.g., Apple, Stevejobs, Kawasaki, inspiration, lesson), yet segment 2's focus is more on Steve Jobs than on Apple. Segments 3 and 4 are using relatively fewer tags: segment 3's focus is similar to that of segments 1 and 2; yet this group tends to be more specific about other reference sites (e.g., Lifehacker, Twitter). Segment 4's primary focus is on Steve Jobs and does not link the article to Apple.

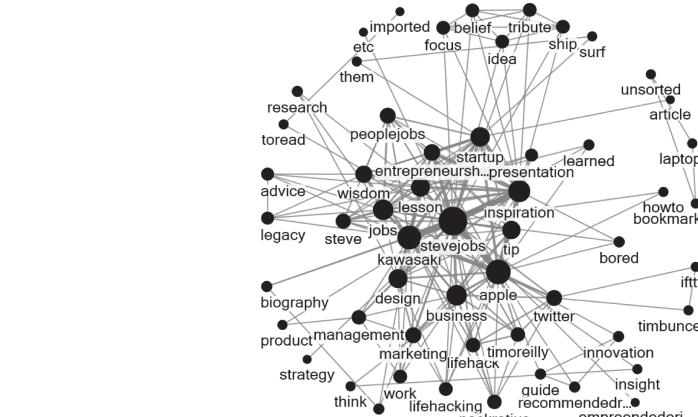
Thus, by employing heterogeneous representations of brand maps using disaggregate-level tagging data, marketers can understand and visualize heterogeneity in brand perceptions and gain insights for segmentation and targeting. For example, marketers can analyze popular online content such as a news article on a brand or a product recall using tags to understand how consumers code it—what are the variations in how consumers react to and tag the content? What are the relative volumes of tags? What are relative sizes of the segments? This analysis can provide useful insights into the impact of such online content on brand perceptions as well as how the firm should react, if need be.

Discovering Representative Topics for Customer Segments

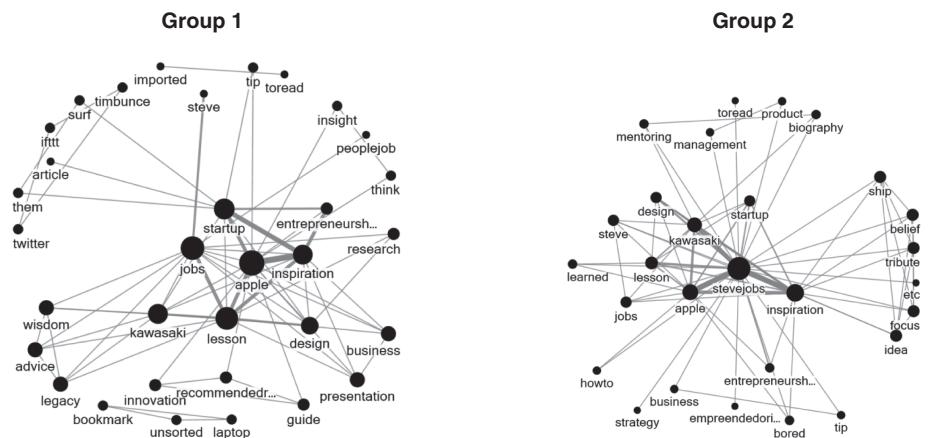
The collection of disaggregate social tagging information provides a large volume of semantic information because many consumers associate content with their own keywords. It is common to observe more than 1,000 bookmarks created for a brand each day on social tagging platforms. The volume of the social tags associated with these bookmarks is also large (over several thousand) and the distribution of keywords is sparse and has a long tail. Thus, to interpret such sparse, high-dimensional, qualitative information generated by heterogeneous customers, dimensionality reduction is critical. We illustrate how marketing managers can extract useful insights from such a large volume of semantic information in disaggregate social tags by identifying latent topics with LDA topic models (e.g., Blei 2012; Blei, Ng, and Jordan 2003; Griffiths and Steyvers 2004; Tirunillai and Tellis 2014).

FIGURE 3
Social Tag Maps of Article A

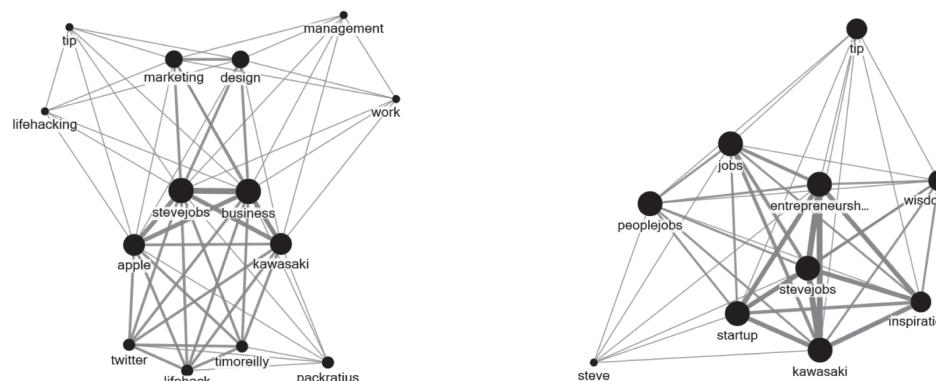
A: Aggregated



B: Cluster-Level



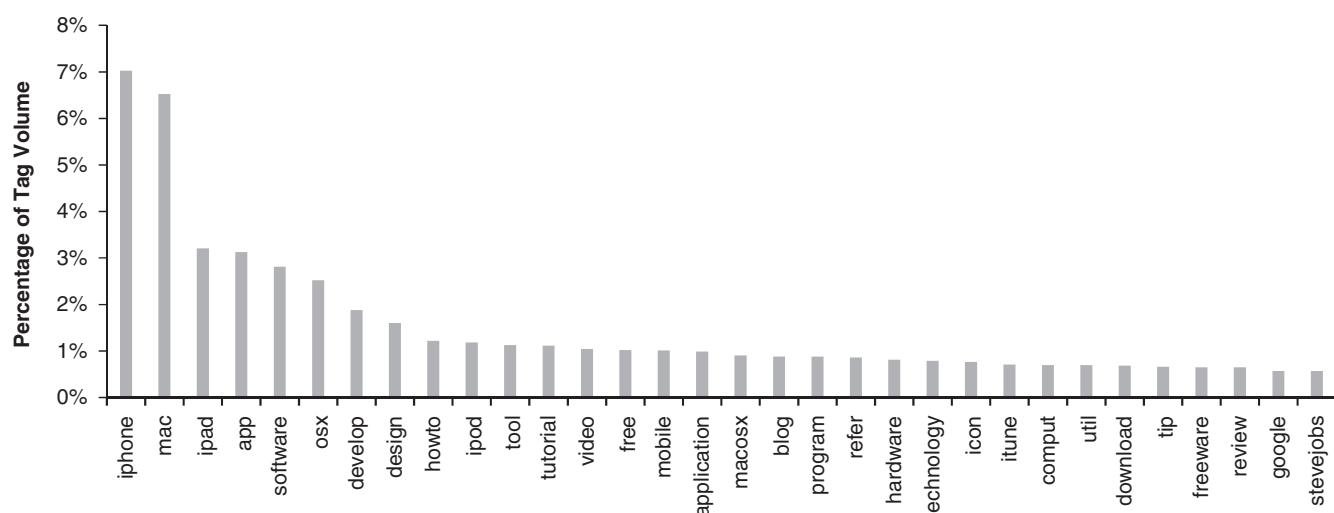
Group 3



Model. The goal of topic models is to explain the collection of documents (corpora) with a mixture distribution of probabilistic topics. Topic models assume that a word in a document is generated by sampling a topic from the topic distribution and

then sampling a word from the word distribution given the selected topic. As such, the probability of a user associating a tag with content can be modeled with a similar rationale: a user first selects a topic that best describes the content to be tagged and

FIGURE 4
Distribution of Social Tags



Notes: The percentage of tag volume represents the percentage of each tag in the entire set of tags associated in our sample of bookmarks (1,869 bookmarks associated with 8,610 tags; 763 distinct tags).

then selects the tags on the basis of the word distribution given the latent topic. More specifically, the topic model is specified as follows:

$$(2) \quad P(\text{tag}_{bi}) = \sum_{k=1}^K P(\text{tag}_{bi} | z_{bi} = k)P(z_{ib} = k),$$

where $P(\text{tag}_{bi})$ is the probability of observing tag i in bookmark b ; b represents a bookmark where $b = 1, \dots, B$; i represents a tag where $i = 1, \dots, N$; N is the number of distinct tags in all the bookmarks; k represents a latent topic where $k = 1, \dots, K$; $P(z_{ib} = k)$ is the probability that the k th topic was sampled for the i th tag in bookmark b ; and $P(\text{tag}_{bi} | z_{bi} = k)$ is the probability of tag i in bookmark b given topic k . We estimated the model using LDA (e.g., Blei, Ng, and Jordan 2003; Griffiths and Steyvers 2004; Tirumillai and Tellis 2014) employing Markov chain Monte Carlo Gibbs sampling with a conjugate Dirichlet prior distribution (for the likelihood functions and sampling procedures, see Appendix D).

Data and results. We employed an LDA topic model to uncover the representative topics in social tags in the sample of 2,000 bookmarks associated with Apple. The 2,000 bookmarks contain 11,851 social tags (1,982 distinct tags) after we (1) deleted non-English words and numbers, (2) removed stop words, and (3) stemmed the tags in line with Porter (1997)'s methods. For the analysis, we removed bookmarks that only had the "Apple" tag as well as tags that appeared only once in the collection of bookmarks. As a result, we had 1,869 bookmarks associated with 8,610 tags (763 distinct tags). The distribution of 8,610 keywords is sparse and has a long tail. Figure 4 shows the distribution of the most frequently mentioned tags, defined as tags with more than .5% of the entire volume of tags. Apple's products (iPhone, Mac, and iPad) are the three most frequently mentioned tags.

We started the estimation of LDA topic model with two latent topics and gradually increased the number of latent topics to find the best model. We determined the optimal number of latent topics using the posterior log-marginal likelihood following the method used in prior studies (e.g., Griffiths and Steyvers 2004). We selected the model with 15 latent topics (marginal log-likelihood = -35,468.52) as the best model.

Table 7 summarizes the most important tags in each latent topic based on the highest distribution probability in each latent topic. We interpreted these topics as "iPhone video photography," "Mac software products," "iPhone computer software," "Mobile technology rivalry," "Mac operations system and music interface," "Safari and JavaScript for Apple products," "iPod and technology review blog," "Web design and graphic," "Comic and humor," "Free tools and software," "Hardware repair for Apple products," "Apple developer," "New products and Steve Jobs," "Apps and download interface," and "Tutorials and tips." All the latent topics are strongly related to Apple's products (e.g., "iPhone," "mac," "osx," "iPod," "iPad," "safari," "iTunes"). These latent topics, which represent the entire set of social tags, reflect characteristics of the original content (e.g., "Apple developer" and "iPod and technology review blog"), users' interpretations of the content associated with Apple (e.g., "Mobile technology rivalry" and "New products and Steve Jobs"), and users' motivations to tag the content ("Tutorials and tips" and "Free tools and software"). That is, these tags provide a summary of users' interpretation of the brand-related content through the lens of the users' own knowledge, experience, and schema. For instance, even though an article about Google's new mobile phone does not discuss Apple or Apple's products, some consumers may interpret the article related to Apple and tag the article with Apple and/or Apple's products. Thus, understanding these latent topics in the entire set of social tags helps marketers grasp consumers' deeper interpretations of brand-related content.

TABLE 7
Characteristics of Representative Topics

Topics	% of Bookmarks	Top Five Keywords
iPhone video photography	12.4%	iphone, video, photography, refer, tech
Mac software products	11.6%	mac, software, web, product, audio
iPhone computer software	9.7%	iphone, computer, software, accessory, business
Mobile technology rivalry	8.6%	mobile, google, html, technology, iphone
Mac operation system and music interface	6.7%	mac, osx, macintosh, music, itunes
Safari and JavaScript for Apple products	6.5%	iphone, safari, javascript, macosx, ipad
iPod and technology review blog	6.2%	blog, ipod, review, technology, game
Web design and graphic	6.3%	design, icon, refer, webdesign, graphic
Comic and humor	5.5%	ipad, comic, funny, flash, humor
Free tools and software	5.7%	tool, free, util, software, osx
Hardware repair for Apple products	4.0%	mac, hardware, macbook, ipod, repair
Apple developer	5.6%	develop, program, video, cocoa, wwdc
New products and Steve Jobs	3.4%	mac, stevejobs, wallpaper, new, ipad
Apps and download interface	4.0%	app, application, itunes, interface, download
Tutorials and tips	3.7%	howto, tutorial, macosx, tip, hack

Our data contain different associative keywords for Apple generated by different users. Thus, the latent topics listed in Table 7 can also be interpreted as the topics representing overlapping user segments⁶ with different perceptions and interests related to the brand. To understand which topic is the most prominent in our data, we calculated the posterior probability of each topic for each of 1,896 bookmarks and classified the bookmarks into the topic with the highest posterior probability. The most prominent user segments were users engaging on the topics of “iPhone video photography” (12.4% of bookmarks), “Mac software products” (11.6% of bookmarks), “iPhone computer software” (9.7% of bookmarks), and “Mobile technology rivalry between Google and Apple’s iPhone” (8.6% of bookmarks). Such information helps marketing managers identify distinct interests and brand perceptions within customer segments and gain insights for segmentation and targeting.

Furthermore, the identified important keywords in each latent topic in Table 7 help us understand the associative structure between brand associations by revealing the underlying interrelationship between tags based on user-level tagging activity. For instance, the topics strongly associated with iPhone are “video and photography,” “computer products,” “mobile market rivalry,” and “Safari JavaScript,” whereas the topics strongly associated with iPad are “Safari JavaScript,” “comic and humor,” and “Steve Jobs and new products.” Such interconnectedness between keywords provides a better understanding of the associative structure of brand associations across subbrands and product attributes.

In summary, social tag–based topic models help managers understand the representative topics capturing distinct user segments’ brand perceptions and interests. The results provide insights on what the most prominent topics associated with a brand and its products are, and how users interpret and categorize content related to the brand and its products. This creates an incisive snapshot of how the brand is perceived through

online user-generated content. Such topics can be useful for a brand to understand the impact of its short-term tactics (e.g., reaction to a new product launch or a public relations campaign, a new advertisement campaign), gauge users’ perceptions through the online buzz surrounding these events (e.g., Hewett et al. 2016), and refine its tactics accordingly. Furthermore, marketing managers can employ distinct topics for better segmentation and targeting strategies and understand the characteristics of distinct interest segments (e.g., size of the segments and types of brand associations). If the brand perceptions of a segment are not quite aligned with the core brand perceptions the firm desires, the firm can design and target communication campaigns and/or other marketing campaigns at the appropriate segment to set the perceptions right. Developing topics by applying LDA on social tags generated on either side of a major event such as a brand recall (before and after the event) can also provide how brand perceptions, their importance, and associated segment sizes change as a result of the event.

Understanding the Evolution of Top-of-Mind Brand Associations

Analyzing trends and the temporal dynamics of information at the disaggregate level in social tags helps marketers identify managerially interesting changes within top-of-mind brand associations. Such information enables marketers to take steps to proactively manage their brand equity by detecting trending keywords (for see further discussion, see Appendix C). Tracking trends in social tags requires an understanding of how keywords associated with the brand evolve over time and which keywords move together. Shedding light on the common trends hidden behind trending social tags provides deeper insights into the dynamics of brand associations.

Model. To understand how consumers’ mental associations connected with a brand evolve over time, we employed a DFA (e.g., Du and Kamakura 2012; Zuur et al. 2003), an analytical tool that can uncover common trends in multivariate time series. Here, our objective was to find latent trends that

⁶Because our LDA model is at a document level, the user segments allow for multiple memberships for each user.

explain the dynamics in social tags associated with a brand. We specified the DFA model as follows:

$$(3) \quad y_t = Zf_t + u_t, \quad u_t \sim MVN(0, I), \text{ and}$$

$$(4) \quad f_t = f_{t-1} + v_t, \quad v_t \sim MVN(0, V),$$

where y_t ($N \times 1$ vector) is a vector of standardized co-occurrence volume of N keywords associated with a brand, and f_t ($m \times 1$ vector) is a vector of latent dynamic factors, with $m < N$. Equation 3 specifies how the comovements of the N -dimensional vector y_t can be explained by the m -dimensional latent dynamic factors f_t . The factor-loading matrix Z ($N \times m$ matrix) determines the correlation between observations y_t and latent factors f_t . The dynamics of latent factors f_t are assumed to be governed by the state equation specified in Equation 4. Following Du and Kamakura (2012) and Zuur et al. (2003), we calibrated the model with the expectation–maximization algorithm.

Data and results. We employed the DFA model to analyze the 36-month time series of co-occurrence volume of 25 keywords with the brand name Apple in the Delicious platform from January 2007 to December 2009. (These keywords were a subset of the most frequently used tags in our LDA analysis.) The DFA results suggest that this 36-month time series of co-occurrence volume of 25 keywords can be represented by 9 dynamic latent factors (for the factor loadings for 25 associations and the prediction performance, see Appendix E). Drawing on the factor loadings following a varimax rotation, we interpreted each latent factor as follows:

- Factor 1: “Mac and Howto,” with high factor loading scores on “tips,” “howto,” “osx,” “mac,” and “tools,” indicating that user interest on content about tips and tools moves with the interest on content about Macs and OS X.
- Factor 2: “Mobile and Technology,” with high factor loading scores on “technology,” “mobile,” “iPhone,” “iPod,” and “design.”
- Factor 3: “(Hardware and Computer),” with high negative factor loading scores on “hardware,” “technology,” and “computer.”
- Factor 4: “Windows,” with highest factor loading score on “windows.”
- Factor 5: “iTunes and Music,” with high factor loading scores on “iTunes,” “music,” and “iPod.”
- Factor 6: “Design, Fun, and Cool,” with high factor loading scores on “design,” “fun,” and “cool,” indicating that positive perceptions about Apple (“cool” and “fun”) move with consumer interest on content about Apple’s design.
- Factor 7: “Growth Trend,” representing the growth of the usage of 25 keywords with Apple over three years and indicating that keywords with relatively lower factor loading score (e.g., “Internet”) show relatively slower growth patterns and keywords with a relatively higher factor loading score (e.g., “iPad”) show relatively higher growth compared with the dynamics of this growth trend factor.
- Factor 8: “(Design and iPod),” with high negative factor loading scores on “design” and “iPod.”
- Factor 9: “(Internet and Inspiration),” with high negative factor loading scores on “Internet,” “design,” and “inspiration.”

Figure 5 shows the trends of these nine latent factors plotted on the basis of the factor scores. We find that consumer interest in

the content about “Mac and Howto” peaked at October 2007, when a new version of Mac OS X was introduced, and declined gradually. Consumer interest in “Mobile and Technology” associated with Apple increased gradually and peaked at August and September in 2008 and 2009, when Apple launched a new iPhone, while consumer interest in “Hardware and Computer” and “Windows” in relation to Apple gradually declined over the three years. Such trends indicate the change in positioning for Apple over the time frame: consumers became less likely to associate content in the computer/hardware category and the competing operating system (Windows) with Apple but more likely to associate content related to mobile and technology with Apple. Consumer interest in “iTunes and Music” in relation to Apple was relatively constant over the three years, with peaks during the summer season. Consumer interest in “Design, Fun, and Cool” in relation to Apple peaked in June 2007, when the first generation of the iPhone was launched, and declined gradually. Consumer interest in “Design and iPod” peaked in September 2007, when the first generation of iPod touch and other new iPod products were launched, and declined gradually. Consumer interest in “Internet and Inspiration” associated with Apple increased gradually with a peak in March 2009.

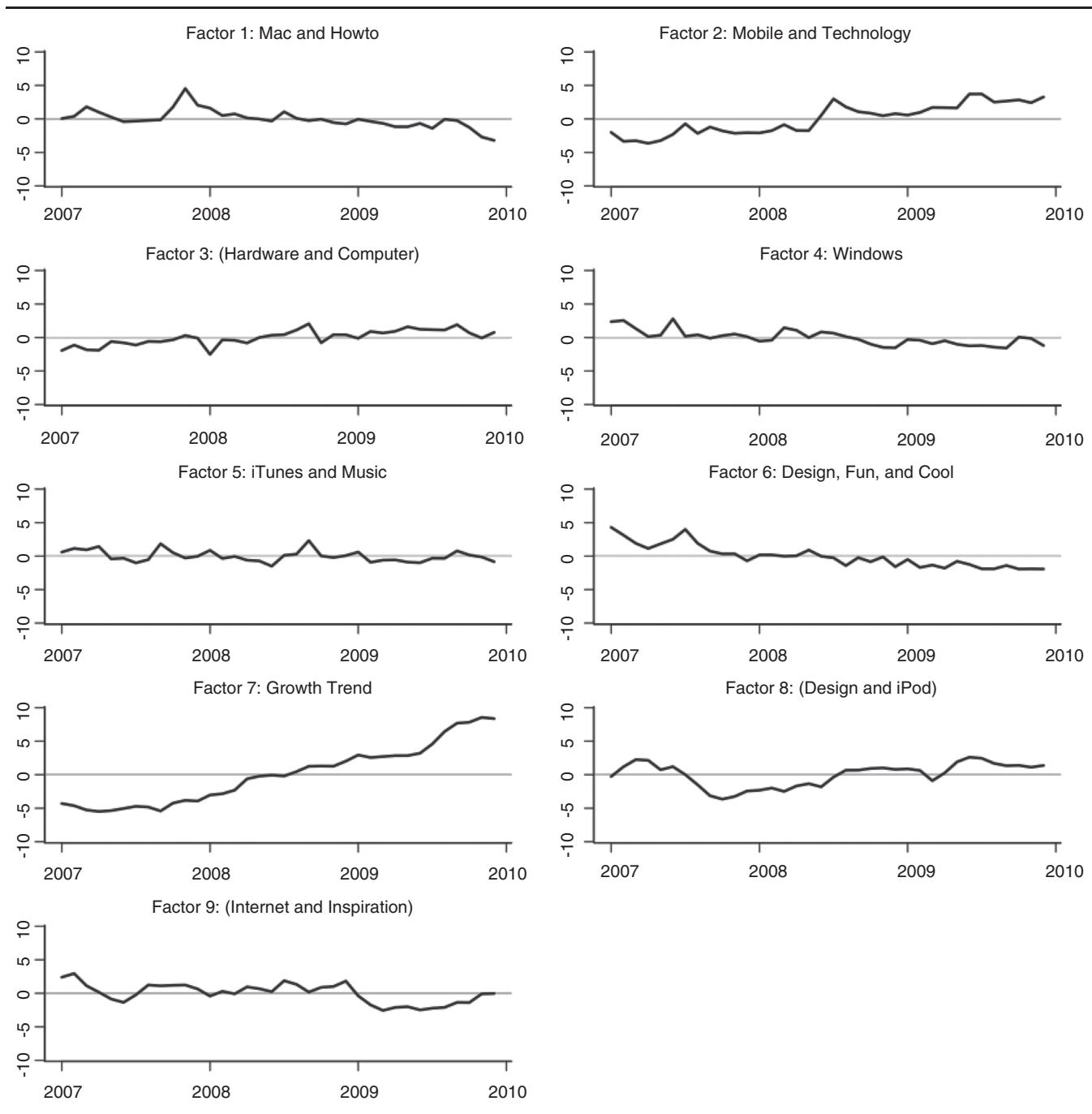
Thus, we demonstrate that rich dynamic information contained in social tags associated with a brand can be summarized with latent dynamic factors without significant loss of information. Monitoring such latent trends provides marketing managers with a better understanding of the trends in top-of-mind associations for a brand and the evolving brand image over time. Although LDA is useful in getting instantaneous feedback on how users perceive the brand, latent dynamic factors are useful to track the specific combination of highly correlated tags over time and for monitoring brand perceptions over time. By comparing such trends with those for a competing brand, a firm can also monitor the relative positioning of its brand relative to the competition in users’ minds through the way people interpret online content.

Discussion and Conclusions

Managerial Implications

The power of social tags resides in the interpretations that users generate for any type of content they observe, in the form of unconstrained and open-ended keywords. Social tags, compared with keywords identified from text mining, provide powerful insights into how users view the content or items filtered through their own knowledge structures and social influences, using their own words and phrases. Thus, social tags complement the current automatic keyword identification in text mining by providing the set of start list of keywords to mine. Our comparison analysis suggests that keywords underrepresented or not discovered by automatic keyword identification in text mining were frequently used as social tags—taggers deemed these keywords important for describing and categorizing the content. By balancing the start list of keywords (social tags) and the list of frequently mentioned keywords discovered by automatic keyword identification in text mining, marketing

FIGURE 5
Trends of the Latent Dynamic Factors



researchers can obtain better insights on consumer-generated textual data.

Social tags generated by individual consumers provide marketing researchers with a unique opportunity to observe consumers' heterogeneous interpretations of content about brands and products. As online content such as newspaper articles, blogs, commentaries, and reviews emerges for a brand in response to an event (e.g., Chipotle's food safety incidents, an exposé on Amazon's corporate culture), an analysis of the associated tags using clustering techniques

can reveal how consumers interpret such content conditional on their mental schema. An LDA analysis on disaggregate social tagging data can show topics representing distinct consumer groups with different perceptions and interpretations of content about the brand. Such use of social tags helps marketing researchers discover answers to the following questions: Are there heterogeneous groups of users who think differently about the event—some who view it much more seriously than others, and vice versa? How different are the perceptions before and after an event of consequence (e.g.,

Puranam, Narayan, and Kadiyali 2017)? Are some users willing to forgive the brand, whereas others vow never to patronize it again? The relative sizes of these segments can also indicate how widespread the change in consumers' perceptions about the brand is. An instantaneous snapshot of the higher-level topics under which consumers cluster helps marketing managers discover distinct interests and perceptions in customer segments and gain insights for segmentation and targeting activities related to their brand using different marketing campaigns.

An investigation into the dynamic relationships between user-generated social tags can reveal how—as more content emerges on the event, its aftermath, and firm actions—consumers' perceptions of the brand change on major factors comprising the tags they use. As we observed in the example of Apple, the factor “Design, Fun, and Cool” in relation to Apple peaked in June 2007, when the first generation of iPhone was launched, and declined gradually. Such information can be immensely useful for the firm to understand how customer perceptions are changing as a function of its actions and competitor actions over the longer term. While similar information can be obtained using other methods such as text mining and primary research methods, it is the ubiquitous nature and type of tagging data that leads to its several advantages over the other methods. It is an inexpensive way to obtain brand associations derived from large-scale content preprocessed by customers and online users.

Given these advantages of analyzing individual-level social tags, they form an inexpensive and continuous data input to implement a brand monitoring dashboard (1) to understand how brand associations vary across segments, and to estimate the size of such segments; (2) to monitor the changes in the topics associated with a brand over time, and specifically in response to important events related to the brand; and (3) to understand changing perceptions of the brand over time. While some of these capture short-term changes, others (using DFA) reflect longer-term changes in brand perceptions. Such a dashboard could be very useful to managers because it would enable them to identify different perception-driven segments for targeting, positioning, and other marketing efforts.

Usage Situations, Limitations, and Further Research

The information contained in social tags is distinct from that in other forms of user-generated content. A unique characteristic of tagging data is that it reflects the associative structure that forms the basis for developing rich semantic networks between keywords and brands. Social tagging data could be perceived as similar to online search data because both enable researchers to obtain the trend of co-occurrence between two or multiple keywords. However, social tagging activity is distinct in that it is more reflective of user perceptions or interpretations about an event, content, or news related to a brand; in contrast, online search is more of a goal-oriented behavior. Thus, tagging data are perhaps more appropriate when marketers are interested in obtaining consumers' perceptions on a brand. Such situations are common, and there is significant interest on the part of brand managers and industry advisors in providing meaningful

solutions that can help capture and represent such perceptions. For instance, recent buzz-tracking solutions including YouGov BrandIndex⁷ and McKinsey's Brand Navigator⁸ are aimed at brand managers, helping them find answers to questions such as how their brand performs with respect to competitors, how the brand performance varies across markets and segments, and how brand managers can improve brand positioning. Social tags can address these questions by providing brand managers with data that represent consumer perceptions regarding both their focal brand and competing brands in a dynamic, real-time setting. In addition, while most of the current solutions focus on the valence and volume of buzz, social tags reveal how individual consumers categorize and describe brands, thus helping brand managers understand associative relationships between brand attributes and identify heterogeneous perceptions regarding a brand.

While the use of social tags is less vulnerable to potential errors involved in the elicitation stage (e.g., algorithm choice in text-mining approaches and/or assumptions by researchers and managers), its interpretation is bounded by each consumer's own point of view and potential social influences (s)he faces. Thus, when such social, individual-level bias is not desired (e.g., when the goal is to obtain an objective interpretation of the content), other approaches such as text mining could perhaps be more effective. However, when a researcher wants to obtain customers' subjective interpretation of content, in which customers' biases are incorporated, a social tag-based approach may be better. Thus, we recommend the use of social tags over a text-mining approach when marketing managers need to obtain a succinct subjective summary directly generated from online users on large-scale data.

We must acknowledge several caveats in using the proposed social tag-based approach. First, like other types of user-generated data, one must question how representative the selected sample of data is. Social tagging data highly rely on users' input, and thus users' self-selection plays a role in their participation decision on tagging platforms and their choice of content to tag. For instance, those who tag a brand name or subbrand name might be more knowledgeable about the brand and more engaged with the news and content related to the brand. Furthermore, the characteristics of taggers can depend on the characteristics of a particular social tagging platform. For instance, on Delicious, users tend to be more tech savvy; on Pinterest, the majority of users are female. To tackle this self-selection bias and obtain more representative data, it is critical for marketing managers to understand whether characteristics of their aggregate customer base are in line with the characteristics of taggers.

Second, as indicated in comparisons with the primary data-based approach and the text-mining approach, tags only indicate the gist of information generated by consumers. Given that most users associate five to ten keywords with a brand name in our data, it is possible that social tags may not be exhaustive in terms of the associations individual users generate for a brand. For instance, in the event that a user associates “cool” and “innovation” with the brand, it clearly tells us that these two

⁷<https://today.yougov.com/find-solutions/brandindex>.

⁸<http://solutions.mckinsey.com/brandnavigator>.

keywords represent the strongest top-of-the-mind brand associations for this user. However, if a user did not associate “inspirational” with the brand, it does not necessarily mean that (s)he does not think the brand is inspirational. It is just that the other terms could have been more salient under the norms or other situational constraints. In addition, it is possible that the user may rate the brand as being very inspirational either on a survey questionnaire or during in-depth discussions. Thus, it would be beneficial for marketing managers occasionally to complement the findings from a social tag-based approach with a primary data-based approach such as sorting, personal interview, and surveys, in which marketing managers directly ask participants to recall, think about, and evaluate all the important aspects of brand associations.

Third, the interpretation of social tags can depend on the characteristics of the social tagging platform. For instance, social tags generated on content management platforms provide insights on perceptions and categorizations related to the brand, whereas social tags generated on microblogs and social network platforms provide insights on user engagement and the context of brand usage experiences. Thus, researchers need to take into account the platform’s characteristics when interpreting the information contained within social tags.

Fourth, selecting the appropriate semantic unit for analysis (i.e., tokenization) can be a challenging process in the social tag-based approach. The social tag-based approach relies on users’ input and thus allows for different levels of tokenization. Compared with automatic keyword extraction in text mining (e.g., 1-gram, 2-gram tokenization), the keywords elicited in a social tag-based approach are more flexible for incorporating users’ interpretations. However, it is questionable if a 1-gram tokenization in social tags is more appropriate than the natural

language susceptible to individual biases and habits. Thus, we recommend that the findings from social tags be complemented with the findings from primary data such as surveys, sorting, and interviews. In addition, social tags may not always be available for all brands. When a brand is unable to engage a sufficient number of users in tagging activities, there may not be enough data to extract brand-related information and insights.

Finally, there are many possible avenues for further research in this area. First, although we did not take the semantic distance between keywords into account while in our analysis (i.e., all synonyms are treated as distinct keywords), a potential future direction could be to consider a lexical database of words such as WordNet (e.g., Miller 1995) and incorporate this information into the brand association elicitation process. Second, additional metrics that rely even more on network characteristics, such as centrality metrics and network density, can be used to provide marketers with more integrative information about their brands’ associative networks. Third, tags can be especially useful to understand perceptions about nontextual content (e.g., pictures, music, video). Future studies can model the categorization process of nontextual content related to a brand and show how such tags can be used for brand equity management. Finally, further research can investigate a better representation of the growth in the dynamic network, such that metrics calculated at different points in time are more comparable. We hope that this work serves as a modest start toward these future directions.⁹

⁹We provide all materials, including the data, data manipulations, and codes, in one package at the following link for researchers interested in more details on our methodology and to encourage future work in this area: <https://www.dropbox.com/sh/wg1c3z4mbgys3hz/AABWWo1h9VNGaGpLc81lmHVwa?dl=0>.

REFERENCES

- Alexa.com (2011), “Statistics Summary for iclo.us,” (accessed June 12, 2011), <http://www.alexa.com/siteinfo/icio.us#>.
- Ames, Morgan, and Mor Naaman (2007), “Why We Tag: Motivations for Annotation in Mobile and Online Media,” in *Proceedings of ACM CHI Conference on Human Factors in Computing Systems*. New York: Association for Computing Machinery, 971–80.
- Bijmolt, Tammo H.A., and Michel Wedel (1995), “The Effects of Alternative Methods of Collecting Similarity Data for Multi-dimensional Scaling,” *International Journal of Research in Marketing*, 12 (4), 363–71.
- Blanchard, Simon J., Daniel Aloise, and Wayne S. DeSarbo (2016), “Extracting Summary Piles from Sorting Task Data,” *Journal of Marketing Research*, 54 (June), 398–414.
- Blanchard, Simon J., and Wayne S. DeSarbo (2013), “A New Zero-Inflated Negative Binomial Methodology for Latent Category Identification,” *Psychometrika*, 78 (2), 1–19.
- Blei, David M. (2012), “Probabilistic Topic Models,” *Communications of the ACM*, 55 (4), 77–84.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003), “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, 3, 993–1022.
- Buchanan, Lauranne, Carolyn J. Simmons, and Barbara A. Bickart (1999), “Brand Equity Dilution: Retailer Display and Context Brand Effects,” *Journal of Marketing Research*, 36 (August), 34–55.
- Carroll, J. Douglas, Paul E. Green, and Catherine M. Schaffer (1986), “Interpoint Distance Comparisons in Correspondence Analysis,” *Journal of Marketing Research*, 23 (August), 271–80.
- Culotta, Aron, and Jennifer Cutler (2016), “Mining Brand Perceptions from Twitter Social Network,” *Marketing Science*, 35 (3), 343–62.
- DeSarbo, Wayne, A. Selin Atalay, David LeBaron, and Simon J. Blanchard (2008), “Estimating Multiple Consumer Segment Ideal Points from Context-dependent Survey Data,” *Journal of Consumer Research*, 35 (1), 142–53.
- DeSarbo, Wayne, Donald Lehmann, Gregory Carpenter, and Indrajit Sinha (1996), “A Stochastic Multidimensional Unfolding Approach for Representing Phased Decision Outcomes,” *Psychometrika*, 61 (3), 485–508.
- Du, Rex, and Wagner Kamakura (2012), “Quantitative Trend-spotting,” *Journal of Marketing Research*, 49 (August), 514–36.
- Fraley, Chris, and Adrian E. Raftery (2007), “Bayesian Regularization for Normal Mixture Estimation and Model-Based Clustering,” *Journal of Classification*, 24, 155–81.
- Gilbert, Eric, Saeideh Bakhshi, Shuo Chang, and Loren Terveen (2013), “I Need to Try This?: A Statistical Overview of Pinterest,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: Association for Computing Machinery, 2427–36.

- Godes, David, and Dina Mayzlin (2004), "Using Online Conversation to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545–60.
- Griffiths, Thomas L., and Mark Steyvers (2004), "Finding Scientific Topics," *Proceedings of the National Academy of Sciences of the United States of America*, 101 (Supplement 1), 5228–35.
- Griffiths, Thomas L., Mark Steyvers, and J.B. Tenenbaum (2007), "Topics in Semantic Representation," *Psychological Review*, 114 (2), 211–44.
- Hamilton, Rebecca W., Debora V. Thompson, Zachary G. Arens, Simon J. Blanchard, Gerald Haubl, P.K. Kannan, et al. (2014), "Consumer Substitution Decisions: An Integrative Framework," *Marketing Letters*, 25 (3), 305–17.
- Hewett, Kelly, William Rand, Roland T. Rust, and Harald J. van Heerde (2016), "Brand Buzz in the Echoverse," *Journal of Marketing*, 80 (May), 1–24.
- Huang, Jeff, Katherine M. Thornton, and Efthimis N. Efthimiadis (2010), "Conversational Tagging in Twitter," in *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*. New York: Association for Computing Machinery, 173–78.
- Hui, Sam K., Yanliu Huang, and Edward I. George (2008), "Model-Based Analysis of Concept Maps," *Bayesian Analysis*, 3, 479–512.
- John, Deborah Roedder, Barbara Loken, Kyeongheui Kim, and Alokparna Basu Monga (2006), "Brand Concept Maps: A Methodology for Identifying Brand Association Networks," *Journal of Marketing Research*, 43 (November), 549–63.
- Joiner, Christopher (1998), "Concept Mapping in Marketing: A Research Tool for Uncovering Consumers' Knowledge Structure Associations," in *Advances in Consumer Research*, Vol. 25, Joseph W. Alba and Welsey Hutchinson, eds. Ann Arbor, MI: Association for Consumer Research, 311–17.
- Lee, Thomas Y., and Eric T. Bradlow (2011), "Automated Marketing Research Using Online Customer Reviews," *Journal of Marketing Research*, 48 (October), 881–94.
- Miller, George A. (1995), "WordNet: A Lexical Database for English," *Communications of ACM*, 38 (11), 39–41.
- Nam, Hyoryung (2012), "Marketing Applications of Social Tagging Networks," doctoral dissertation, University of Maryland.
- Nam, Hyoryung, and P.K. Kannan (2014), "Informational Value of Social Tagging Networks," *Journal of Marketing*, 78 (July), 21–40.
- Netzer, Oded, Ronen Feldman, Moshe Fresko, and Jacob Goldenberg (2012), "Mine Your Own Business: Market-Structure Surveillance Through Text Mining," *Marketing Science*, 31 (3), 521–43.
- Novak, Joseph D., and D. Bob Gowin (1984), *Learning How to Learn*. New York: Cambridge University Press.
- Porter, M.F. (1997), "An Algorithm for Suffix Stripping," in *Readings in Information Retrieval*, Karen Sparck Jones and Peter Willett, eds. San Francisco: Morgan Kaufmann Publishers, 313–16.
- Puranam, Dinesh, Vishal Narayan, and Vrinda Kadiyali (2017), "The Effect of Calorie Posting Regulation on Consumer Opinion: A Flexible Latent Dirichlet Allocation Model with Informative Priors," working paper.
- Ratneshwar, Srinivasan, and Allan D. Shocker (1991), "Substitution in Use and the Role of Usage Context in Product Category Structures," *Journal of Marketing Research*, 28 (August), 281–95.
- Ringel, Daniel M., and Bernd Skiera (2016), "Visualizing Asymmetric Competition Among More Than 1,000 Products Using Big Search Data," *Marketing Science*, 35 (3), 511–34.
- Robu, Valentin, Harry Halpin, and Hana Shepherd (2009), "Emergence of Consensus and Shared Vocabularies in Collaborative Tagging Systems," *ACM Transactions on the Web*, 3 (4), Article 14.
- Shugan, Steven M. (1987), "Estimating Brand Positioning Maps Using Supermarket Scanning Data," *Journal of Marketing Research*, 24 (February), 1–18.
- Strohmaier, Markus, Christian Körner, and Roman Kern (2010), "Why Do Users Tag? Detecting Users' Motivation for Tagging in Social Tagging Systems," in *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. Menlo Park, CA: Association for the Advancement of Artificial Intelligence, 339–42.
- Teevan, Jaime, Daniel Ramage, and Meredith Ringel Morris (2011), "#TwitterSearch: A Comparison of Microblog Search and Web Search," in *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*. New York: Association for Computing Machinery, 35–44.
- Tirunillai, Seshadri, and Gerard J. Tellis (2014), "Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation," *Journal of Marketing Research*, 51 (August), 463–79.
- Zaltman, Gerald (1997), "Rethinking Market Research: Putting People Back In," *Journal of Marketing Research*, 35 (November), 424–37.
- Zaltman, Gerald, and Robin Higie Coulter (1995), "Seeing the Voice of the Customer: Metaphor-Based Advertising Research," *Journal of Advertising Research*, 35 (4), 35–51.
- Zuur, A.F., R.J. Fryer, I.T. Jolliffe, R. Dekker, and J.J. Beukema (2003), "Estimating Common Trends in Multivariate Time Series Using Dynamic Factor Analysis," *Environmetrics*, 14 (7), 665–85.

Copyright of Journal of Marketing is the property of American Marketing Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.