ABSTRACT


Title of Document:           **MARKETING APPLICATIONS OF**
                             **SOCIAL TAGGING NETWORKS**

                             **Hyoryung Nam, Doctor of Philosophy, 2012**

Directed By:                 **Professor P.K. Kannan and**
                             **Professor Yogesh V. Joshi,**
                             **Department of Marketing**


This dissertation focuses on marketing applications of social tagging networks. Social tagging is a new way to share and categorize content, allowing users to express their perceptions and feelings with respect to concepts such as brands and firms with their own keywords, "tags." The associative information in social tagging networks provides marketers with a rich source of information reflecting consumers' mental representations of a brand/firm/product.

The first essay presents a methodology to create "social tag maps," brand associative networks derived from social tags. The proposed approach reflects a significant improvement towards understanding brand associations compared to conventional techniques (e.g., brand concept maps and recent text mining techniques), and helps marketers to track real-time updates in a brand's associative network and dynamically visualize the relative competitive position of their brand.

The second essay investigates how information contained in social tags acts as proxy measures of brand assets that track and predict the financial valuation of firms using the data collected from a social bookmarking website, *del.icio.us,* for 61 firms across 16 industries. The results suggest that brand asset metrics based on social tags explain stock return. Specifically, an increase in social attention and connectedness to competitors is shown to be positively related to stock return for less prominent brands, while for prominent brands associative uniqueness and evaluation valence is found to be more significantly related to stock return. The findings suggest to marketing practitioners a new way to proactively improve brand assets for impacting a firm's financial performance.

The third essay investigates whether the position of products on social tagging networks can predict sales dynamics. We find that (1) books in long tail can increase sales by being strongly linked to well-known keywords with high degree centrality and (2) top sellers can be better sellers by creating dense content clusters rather than connecting them to well-known keywords with high degree centrality. Our findings suggest that marketing managers better understand a user community's perception of products and potentially influence product sales by taking into account the positioning of their products within social tagging networks.

**MARKETING APPLICATIONS OF SOCIAL TAGGING NETWORKS**


By


Hyoryung Nam




Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012




Advisory Committee:
Professor P.K. Kannan, Chair
Professor Yogesh Joshi, Co-Chair
Professor Shapour Azarm
Professor William Rand
Professor Michael Trusov

# Acknowledgements

This dissertation would not have materialized without the support of many people. First of all, I would like to thank my advisor, Dr. P.K. Kannan for his invaluable guidance and support for this dissertation. He inspired me to explore this exciting new area. His constant encouragement and caring have helped me to stay positive even during tough times in the pursuit of my Ph.D.

I would like to thank my co-advisor, Dr. Yogesh V. Joshi, who sets an example of a great researcher for his passion on research. He has graciously assisted, advised, and supported my research. I am also grateful to Dr. William Rand, who has generously given his time and expertise and helped me to gain a different perspective to the research problems through each step of the research process.

I am also obliged to Dr. Michel Wedel, who has provided a rigorous training in quantitative research and taught me the joys of research. His suggestions and insights have contributed greatly to the completion of the dissertation Special thanks go to Dr. Michael Trusov and Dr. Shapour Azarm, who graciously agreed to serve on the dissertation advisory committee and provided valuable suggestions on my research.

The support of fellow students played an instrumental role during my years as a doctoral student. I would like to thank my fellow Ph.D. students for their mentoring, friendship, and advice.

Finally, I am forever indebted to my parents for their unconditional love, endless patience and encouragement when it was most required. I am also grateful to my sister, Hyesung for her support and caring. I would like to thank my loving, supportive, encouraging, and patient husband, Ian for bearing my whining, giving me unwavering support, and standing as the best friend of mine for the past three years.

# Table of Contents

# List of Tables

# List of Figures

# Chapter I: General Introduction

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, brands and concepts. This dissertation focuses on one such form of user generated content – social tags – and develops a new approach to process this tag information for understanding how consumers think, feel, conceptualize and associate with brands, products, and firms (Chapter II) and demonstrates the informational value of social tagging networks in the context of firm valuation (Chapter III) and product sales prediction (Chapter IV).

## 1.1. Introduction to Social Tags

Social tagging is a new way online users categorize and share web content, and is fast becoming a useful alternative to a traditional search system. Under a social tagging system, users describe and categorize web content with a set of their own keywords, called "tags", and the content is searched and shared via tags. For instance, bookmarks are tagged in del.icio.us, photos are tagged and shared in Flickr and Facebook, videos are tagged in Youtube, news articles and tweets are tagged in Tweeter, products are tagged in Amazon, and scientific publications are tagged in CiteULike. Unlike traditional search mechanism, tagging systems allow people to work together to organically create a semantic structure of relevant issues as users create and update semantic associations of web content by linking them to a set of tags.

Social media platforms allow online users to produce and share rich content, news, and videos. Despite the availability of rich information, online users have a hard time "digesting" this abundant information (Wortham 2011). Social tags fill this need for organizing and discovering relevant information more efficiently by providing a system to categorize information based on topical relevance. Such systems have also been referred to as "folksonomies" in the recent literature (e.g., Pink 2005; Gruber 2007).

A key motivation for using such a tagging system is to describe and categorize content so that it can be retrieved and discovered efficiently in the future (Körner et al 2010). Users collaborate in the categorization and information discovery process by socially sharing content via these tags, enabling other users to find new content based on existing tags. Thus, tagging systems allow users to work together to organically create a semantic structure that conceptualizes a product or a brand as perceived by the network of its creators based on the online content they view and/or create. Recent research has shown that popular tags generated tend to be quite meaningful in this conceptualization process (Suchanek, Vojnović, and Gunawardena 2008).

A representative example of a social tagging system is the popular social bookmarking website *del.icio.us* which allows users to manage a collection of web links. This collection covers a number of different topics such as the latest news and trends in web technology, politics, media, business and entertainment. Figure 1.1 illustrates the kinds of data available in *del.icio.us* related to "Apple" – the focal brand chosen for illustrating our analysis. Each item is a "bookmark" generated by a specific user for a specific web link about Apple. Within each item, tags (on the right bottom) are the keywords linked to that webpage, and the number (on the right top) indicates the volume

of bookmarks created for that webpage. These bookmarks are from various sources: e.g., corporate webpages (Bookmark 1), blog posts (Bookmark 2), news articles (Bookmark 4), product reviews (Bookmark 3 and 6), tutorials (Bookmark 5), and so on.

[INSERT FIGURE 1.1 ABOUT HERE]

Users can create associations between tags and a brand by organizing brand-related content with descriptive keywords. Tags linked to each bookmark contain distinct types of mental representations of the focal brand in the minds of users based on the content they view: some tags are purely descriptive (e.g., "news"), some are identifiers for the brand and sub-brand (e.g., "mac", "iphone" and "ipod"), while others are identifiers of the category (e.g., "computers"). More importantly, some of the tags are descriptors of how consumers think and feel about the focal brand: of these, some are of a positive sentiment (e.g., "inspiration"), while others are of a negative sentiment (e.g., "defects").  Thus, while the volume of tags is a good indicator of the amount of online content focused on the brand and viewed by the consumers, the valence of these tags represent the nature of content seen and how they are perceived by the consumers creating the tags.  Therefore, tags reveal objective as well as subjective representations of a focal brand in the minds of its users; in short, a summary of consumers' perceptions of brand-related content. In this paper, we leverage the information revealed by consumers via this dual representation to understand and build brand associative maps.

Beyond specific words describing a brand, a social tagging system also provides rich information about how a brand may be related to other brands in the users' minds. Conceptually, the tripartite relationships existing within a tagging system are illustrated in Figure 1.2 (Mika 2007). There are three main entities: users, tags, and brands. Users

associate brands with concepts or sentiment, which they express via tags. The three entities are connected to each other by users' social bookmarking activity, reflecting possible interrelations between brand associations. Thus, the networks of tags connected to a brand can reveal the collective map of mental associations toward the focal, as well as related/competing brands.

[INSERT FIGURE 1.2 ABOUT HERE]

## 1.2. Overview of Three Essays

The first essay presents a methodology to create "social tag maps," brand associative networks derived from social tags. The proposed approach reflects a significant improvement towards understanding brand associations compared to conventional techniques (e.g., metaphor elicitation or brand concept maps and recent text mining techniques), and helps marketers to track real-time updates in a brand's associative network, dynamically monitor brand assets, and dynamically visualize the relative competitive position of their brand.

The second essay investigates how information contained in social tags acts as proxy measures of brand assets that track and predict the financial valuation of firms using the data collected from a social bookmarking website, *del.icio.us,* for 61 firms across 16 industries. The results suggest that brand asset metrics based on social tags explain variations in the unanticipated stock return and that the informational value of social tag metrics varies across brands. Specifically, an increase in social attention and connectedness to competitors is shown to be positively related to stock return for less prominent brands, while for prominent brands associative uniqueness and evaluation

valence is found to be more significantly related to stock return. The findings suggest to marketing practitioners a new way to proactively improve brand assets for impacting a firm's financial performance.

The third essay investigates whether the position of products on social tagging networks can predict sales dynamics. We find that (1) books in long tail can increase sales by being strongly linked to socially popular keywords and well-known keywords with high degree centrality and (2) top sellers can be better sellers by creating dense content clusters rather than connecting them to well-known keywords with high degree centrality. Our findings suggest that marketing managers understand better a user community's perception of products and potentially influence product sales by taking into account the positioning of their products within social tagging networks.

The organization of this dissertation is as follows: Chapters II, III and IV discuss the three essays in depth. Chapter V briefly summarizes each essay, points out the contributions of this dissertation, and concludes with possible future research avenues.

**Figure 1.1: Sample Bookmarks Created by a User on *del.icio.us* for Apple**
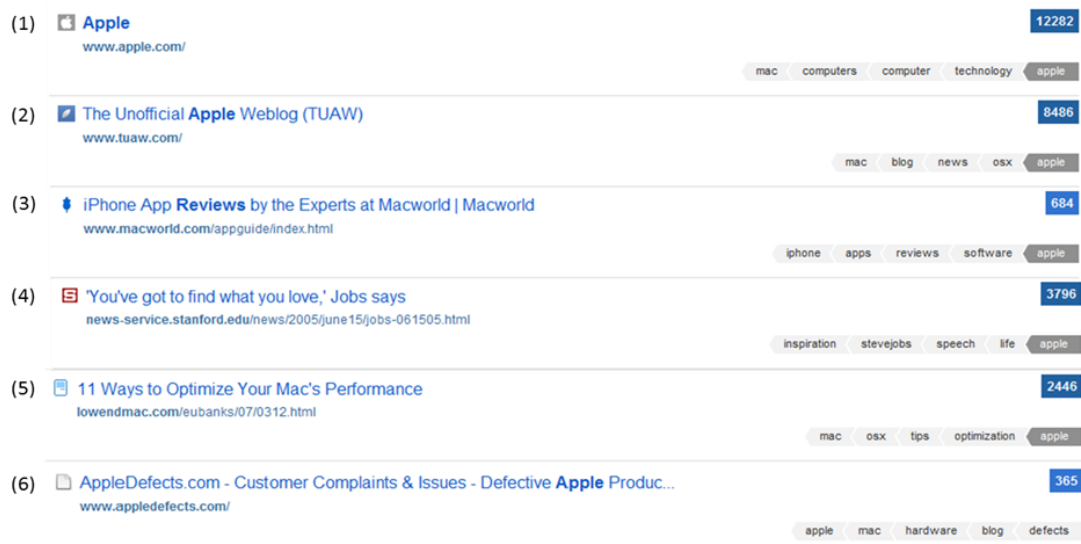


**Figure 1.2: A Tripartite Graph Representing Social Tagging Networks (Adapted from Mika 2007)**
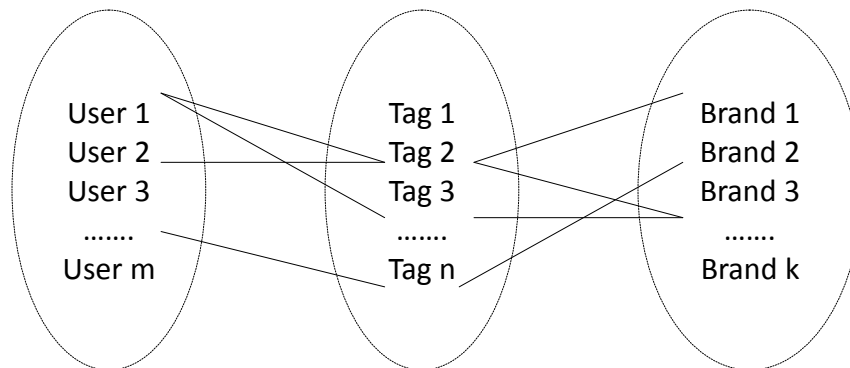
# Chapter II: Social Tag Maps: A New Approach for Understanding Brand Associations

## 2.1. Introduction

Social tags provide marketers with a significant amount of online data that reflect consumers' opinions, attitudes, and preferences for products, services, brands and concepts. This paper develops a new approach to process this tag information for understanding how consumers think, feel, conceptualize and associate with brands, products, and firms. We do so by first developing a key set of valid and reliable metrics that help us accurately extract the maximum information value contained within user-generated social tags. Using these metrics, we construct consumer associative networks, or concept maps for brands. For marketers, understanding these associative networks of social tags can provide detailed insights into the dynamics of consumer preferences, consumer demand for products, competitive market structure and firm valuation.

As demonstrated in Chapter I, social tags reveal objective as well as subjective representations of a focal brand in the minds of its users; in short, a summary of consumers' perceptions of brand-related content. Beyond specific words describing a brand, a social tagging system also provides rich information about how a brand may be related to other brands in the users' minds. The networks of tags connected to a brand can reveal the collective map of mental associations toward those focal, as well as related/competing brands.

Not only does a tagging system reveal rich associative structure, such structure is also dynamic as new tags get created in response to existing and new content, new products, and new ideas. Hence, tools that monitor social tagging systems have the potential to yield real time, updated, networked information of revealed preferences and relationships that can be of significant managerial value. Since such monitoring can be automated, it is less time consuming and less expensive than surveys and existing methods that are used to collect stated preferences and relationships. Further, it allows us to capture and analyze vast amounts of useful information that is practically infeasible to capture using existing methods – in terms of the richness of data, dynamics, and competitive positions.

Consequently, our main objective in this paper is to study social tagging systems by analyzing their informational value. We do so by developing a key set of metrics that allow us to harness the information contained within these tagging systems from the perspective of a brand. Using illustrative examples, we show how the proposed measures and metrics can enable managers to track brand associations, competitive positions, and plan future marketing actions on the basis of such measures. To achieve these objectives, our paper is structured as follows. We first review the existing literature and highlight how the use of social tags could potentially be more informative and efficient as compared to extant methods commonly used for developing brand association maps. Then, we describe the data collection and analysis procedure for creating associative tag maps that vary dynamically over time, and define a set of metrics that form the basis of our analysis. This is followed by a discussion on the reliability of social tag maps and comparison with existing methods. Subsequently, we demonstrate the informational

value of these core as well as derivative metrics for tracking and managerial actions. Finally, we conclude with a discussion of the applications of social tags in specific settings, and outlines areas for future research.

## 2.2. Comparison with Existing Approaches

Investigation of online users' motivations and incentives for social tagging in extant literature reveals motivations for tagging fall into two categories: resource organization (categorization and description) and social communication (information sharing and opinion expression). Recent work by Strohmaier, Korner, and Kern (2010) identified two primary motivations for tagging: categorizing resources according to high-level attributes for organization and navigation and describing and capturing the content in the resources for later reference. The authors suggest that tags created by describers might be more useful for understanding rich interpretations of a resource and discovering the relevant information than those by categorizers. Nevertheless, even with a categorization motivation, the social aspect of tagging behavior allows users to develop a rich semantic structure (Ames and Naaman 2007). In studying web-based photo sharing systems, Ames and Naaman (2007) identify four distinct motivations: self-oriented organization (e.g., "I like order."), self-oriented communication (e.g., "reconstruct what I was thinking"), social organization (e.g., "wanted to tell people what it was" and "tagging can connect my photos to activities"), and social communication (e.g., "I can give people basic story" and "shared social experience"). They find that self-oriented organization and social organization for the public are more commonly observed than self-oriented communication and social communication for the public, however social communication

for friends and family is more commonly observed than social organization for friends and family.

Zollers (2007) further identifies underlying social motivations for tagging on Amazon and Lastfm. The author suggests that users employ tags such as "funny" or "overrated" to reveal their thoughts and opinions, unique and witty tags such as "craptacular" or "ch-ch-check it out" to draw attention from other users, and tags such as "defectivebydesign" as a way of engaging in the protest against Microsoft Windows Vista. In addition, Thom-Santelli, Muller, and Millen (2008) identify the social roles of tagging: community-seeker, community builder, evangelist, publisher, and team-leader. It is clear from the extant literature that social tags allow users to develop rich cognitive structures through social interactions, and thus form useful input to uncovering such structures.

Extant research in cognitive psychology suggests that the knowledge structure in our minds can be represented well in the form of associative networks (e.g., Anderson and Bower 1973; Gentner and Stevens 1983). Previous literature in marketing has recognized this and proposed approaches that create brand maps to help marketers understand brand assets (Aaker 1991; Keller 1993). Conceptually, this process of constructing brand associative maps consists of three distinct stages: [i] elicitation: identify core brand associations based on customers' inputs; [ii] mapping: develop a customer mind map based on identified core brand associations; and [iii] aggregation: aggregate individual mind maps to construct a consensus brand map. Existing approaches vary in terms of the nature and the richness of information contained in the maps, as well as the resources and expertise required for successful implementation. Established

approaches include Zaltman's Metaphor Elicitation Technique – or ZMET (e.g., Zaltman and Coulter 1995; Zaltman 1997); Brand Concept Maps – or BCM (e.g., Joiner 1998; John et al. 2006); and, recently emerging text-mining approaches (e.g., Lee and Bradlow 2011; Netzer et al. 2012).[1]

The primary assumption of the ZMET approach (Zaltman and Coulter 1995; Zaltman 1997) is that a significant portion of consumers' thoughts and knowledge are stored in a nonverbal form, and cannot be fully elicited with verbal communication. Hence, ZMET employs in-depth personal interviews using qualitative techniques such as Kelly's repertoire grid, laddering exercises, and verbal as well as non-verbal cues such as images during the elicitation stage, to understand the core associations linked to a topic. This is followed by the creation of visual montages or maps by participants in a personal interview during the mapping and aggregation stage. A consensus map based on the frequency of elicited associations and association pairs is also created. While ZMET can help identify deep, unconscious thoughts and feelings related to a brand by employing multiple qualitative approaches and using both verbal and nonverbal aspects of consumer, 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., 7-10 days for subjects to collect visual images and 2-hour in-depth one-on-one interview to obtain an individual brand map). Accessibility is another issue for ZMET since it requires

---

[1] Other research on constructing brand associative networks includes the work of Henderson, Iacobucci, and Calder (1998) and Teichert and Schöntag (2010) using social network metrics such as degree centrality and betweenness to identify core perceptions or driver sub-brands of a brand; and Hui, Huang, and George (2008) using Bayesian models of graph formation to obtain parameter estimates for the likelihood of edge formation, which are further utilized to create a consensus map from multiple individual concept maps. Multidemensional scaling (MDS) has been employed to obtain a spatial representation of consumers' underlying preferences (e.g., DeSarbo et al. 1996; Ghose 1998; Shugan 1987) .We focus our discussion on ZMET, BCM, and text mining as they are the closest to the techniques proposed in this paper.

interviewers with expertise in qualitative elicitation techniques, raising administrative costs.

The BCM method (e.g., Novak and Gowin 1984; Joiner 1998; John et al. 2006), in contrast, employs more structured procedures in eliciting core associations, mapping the associations, and synthesizing individual maps into consensus maps. To elicit core associations, BCM utilizes prior consumer research and input from the brand management team. Then, individual concept maps primarily based on identified core associations are obtained though one-on-one interviews. In the BCM aggregation stage, researchers summarize each individual concept map in terms of three key aspects: (i) the presence of a set of elicited core brand associations, (ii) the associative strength (e.g., represented by single, double, triple sized links), and (iii) whether the link is directly connected to brand (level 1 link), or connected indirectly (level 2 link), etc. The final consensus map is then developed, based on the aggregated frequency of the individual maps, revealing a hierarchical associative structure with differential associative strength. Compared to ZMET, BCM is somewhat easier to administer and analyze. In addition, BCM flexibly accommodates inputs from managers. However, BCM may not be adequate for eliciting unconscious feelings and brand associations that need additional in-depth probing.

Thus, both ZMET and BCM face the following challenges: (i) they are labor-intensive as they employ qualitative analysis and one-on-one personal interviews, (ii) they often require specialized expertise, (iii) they are often implemented at specific time periods, and are a static representation of brand perception, rather than a dynamic brand map over time, (iv) they involve small sample sizes and tend to be very expensive if the

focal brand seeks to obtain a brand map from a larger sample; and (v) they are based on stated brand associations and thus bounded by the elicitation techniques.

More recent approaches based on text mining offer 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 create brand associative maps with automatically identified 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). Here, the elicitation stage consists of multiple steps: obtaining the textual data, cleaning and preparing the text, extracting appropriate information from the text, and "chunking" the information (Netzer et al. 2012). Researchers employ a rule-based approach, a machine-learning approach, or a hybrid of these two approaches to extract and chunk the information in the text (Netzer et al. 2012). In the mapping and aggregation stage, based on co-occurrence pattern of brand names, product names, or product attributes, the associative relationships are identified. As compared to ZMET and BCM, text-mining approaches involve low level of human labor, given the automation offered during the elicitation process. Nevertheless, text mining tends to be computation-intensive, as it employs multiple stages of model estimation and data training (e.g., keyword extraction, chunking, associative relationship derivation). In a text mining approach, it is often difficult to formulate network relationships from unstructured, natural text. Text mining algorithms often require human labor for better adaptation of algorithms to the domain of study. Netzer et al. (2012) have identified this as one of the key drawbacks of this approach, and suggested the need to replace this human labor aspect with tagging work

using crowd-sourced marketplaces. Our proposed social tag maps approach employs

stated keywords by online users, and hence does not require tools to infer the associations

implied by consumers, making it easier to implement and more accurate in representing

what consumers think in related to the brand. Further, in a text mining approach, salient

associations and associative structures are derived by an external algorithm, which is

based on assumptions made by the firm's market research division. This is not the case

with social tag maps, where consumers and users are the ones that personally create the

associative structure. Thus, tags reflect users' interpretations and perceptions about the

products, brands and firms they are interacting with, rather than rule-based algorithms.

Additionally, in a text mining approach, if a direct source of user-generated content is not

available, researchers need to crawl the web to collect all text related to the brands from

disparate sources (Netzer et al 2012). Finally, the cognitive costs of generating tagging

data are much lower than those required for writing a review, perhaps making it an easier

task for users to engage.

[INSERT TABLE 2.1 ABOUT HERE]

Table 2.1 summarizes discussions on the comparison of the proposed social tag

maps with extant methods. To summarize, social tag maps reduce the disadvantages of

existing methods, since: (i) social tags can be automatically acquired, thus cutting down

on labor, time, computational costs, and expertise requirements (ii) can be automatically

updated, on a real-time basis; thus providing a dynamic, rather than a static map (iii) are

constructed from a larger customer base, with minimal additional costs, (iv) can track an

extensive set of associations, and even track competitors' brand associations, and (v)

utilize unbounded, richer inputs directly stated by consumers.

**2.3. Constructing Social Tag Maps**

We now describe our method for creating associative maps from social tags generated by multiple users. It consists of three stages: (1) Data Collection, (2) Analysis of the Data / Association Elicitation, and (3) Mapping and Aggregation. For sake of illustration, we choose Apple as our focal brand.

*Stage 1: Data Collection*

This stage consists of three primary specification tasks:

[i] Specifying the set of social tags: to determine the set of social tags, researchers can use a combination of: (1) tags frequently linked to a brand, (2) tags frequently linked to competitors, (3) predefined keywords, identified from previous surveys and/or consumer interviews. Unlike existing methods, our approach does not require us to narrow down the keywords to the set of core associations given very low costs of collecting additional data. Moreover, our approach allows the data to identify what the core associations might be without a priori restricting them in any way.

[ii] Specifying the set of competing brands: the set of competitors can be defined in multiple ways. For instance, (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) can be used.

[iii] Specifying the time frame for data collection: data can be collected on any given time frame: it can be collected on a hourly, daily, weekly, monthly, quarterly, or

yearly basis. Depending on managerial objectives, an appropriate time window could be specified.

*Our data.* For the present study, we collect tags from *del.icio.us*, a top 500 global website in terms of traffic rank (Alexa.com, 2011), with the recent three-month global Alexa traffic rank being 252 (March 13-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. Since the site had over 5,500,000 unique average monthly visitors in September, 2010 (ebizmba.com 2010), the scale of system is sufficiently large to reach a collectively coherent identity (Mika 2007).

Next, we identify the set of competitors of Apple via four-digits of SIC code, Hoover's classification of competitors, Google Finance, view history provided by Yahoo Finance, and the tagging co-occurrence structure on *del.icio.us*. We include as competitors companies that appear at least three times in these five sets[2]. These include: Microsoft, Google, Blackberry, Nokia, Dell, and HP. This set can be easily redressed or expanded as desired by the focal brand manager. Then, we collect all social tags linked to the most recent 2,000 bookmarks for Apple, and its competing brands, and we construct a dictionary of tags. To exclude idiosyncratic tags from the analysis, from among 25,327 tags in the tag dictionary we select 7,019 key tags which appeared more than twice in the pool of bookmarks linked to these brands. Lastly, we obtain the historical monthly trends of the volume of bookmarks tagged with each key tag and each brand using *del.icio.us* search algorithms from 2006 to 2009. 12,353,231 brand-tag pairs are included in our data. This data can be easily aggregated to a quarterly and annual basis.

---

[2] Here we employ both external criteria such as SIC code, Hoover's and consumer driven criteria such as view history provided by Yahoo Finance and tagging co-occurrence structure on *del.icio.us*. Researchers can further consider obtaining a snowball sample of competitors based on tagging structure.

*Stage 2: Analysis of the Data / Association Elicitation*

In this stage, the goal is to identify the set of relevant core social tags for the focal brand. This goal is achieved by evaluating all tags associated with a brand on a set of defined metrics, and identifying tags that score the highest on these metrics. To do this, we propose using the following metrics[3]:

***Associative strength metrics.*** Core associations for the focal brand can be identified based on associative strength between a brand and a tag, which can be measured using the following metrics:

First metric to capture associative strength between a brand and a tag is co-occurrence volume metric, measured by how many times a brand is co-tagged with a brand via bookmark. 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 volume of bookmarks linked to both brand *i* and tag *j* during time window *t*. That is, the metric captures how many brand *i* -tag *j* pairs are created during time window *t*.

Another way to capture the associative strength is to scale the co-occurrence volume metric, $N_t(B_i, T_j)$ by the bookmark volume linked to brand and the bookmark volume linked to brand as specified in Equation (1). This metric measures the *cosine* distance between each brand and each tag, and has been proposed as being quite useful to capture the similarity between two tags in *del.icio.us* (Robu, Halpin, and Shepherd 2009).

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

---

[3] In the next section, we discuss extensively the validity and reliability of our proposed core metrics.

where, $N_t(B_i)$ is the volume of bookmarks linked to brand $i$ during time window $t$,:

$N_t(T_j)$ is the volume of bookmarks linked to tag $j$ during time window $t$.

The relevance of a tag to the competitors can also be taken into account in eliciting core associations. We employ the weighted co-occurrence volume, $WN_t(B_i, T_j)$ which is the co-occurrence volume $N_t(B_i, T_j)$, weighted by the relevance of a tag to players in the category. The relevance is calculated as sum of the co-occurrence volume of a focal tag linked to all the players in the category, scaled by sum of volume of bookmarks linked to all the players in the category:

$$(2) \qquad WN_t(B_i, T_j) = N_t(B_i, T_j) \frac{\sum_{i' \in CB_i} N_t(B_{i'}, T_j)}{\sum_{i' \in CB_i} N_t(B_{i'})}$$

where, $CB_i$ is the set of competitors for brand $i$, including brand $i$.

Table 2.2 lists the 10 strongest positive[4] brand associations linked to Apple, based on the co-occurrence volume metric, along with summary metrics that we discussed above. It is reassuring to see from a face validity viewpoint that keywords such as *technology*, *cool*, *fun*, *interesting*, and *funny* are identified as the top five strong positive associations for Apple. A similar list for Microsoft for the same period reveals *technology, excel, cool, funny* and *interesting* emerge as their top five (in order) brand associations. A similar list can be constructed using other associative strength metrics such as a scaled volume metric (Equation 1) or a weighted volume metric (Equation 2) for all keywords (also separating them out by their valence as we discuss below).

[INSERT TABLE 2.2 ABOUT HERE]

---

[4] We classified tags into three categories: positive, neutral, and negative tags. We discuss the classification process more detail in the next section.

*Eliciting core associations.* The set of relevant brand associations can now be obtained by specifying the level of explanatory power desired. Table 2.3a shows the number of tags needed to explain 95% (also 90%, 80%) 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 minimum co-occurrence volume of these tags is 31. Likewise, 90% of co-occurrence volume of tags linked to Apple is explained with 1,430 tags (24 % of all tags linked to Apple) and minimum co-occurrence volume of these tags is 65. The choice of the set of relevant associations can be further complemented by specifying a pre-requisite level of co-occurrence for each tag: e.g., co-occurrence volume greater than 1 (also 5, 10, 20). Table 2.3b 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, she 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).

[INSERT TABLE 2.3 ABOUT HERE]

### Stage 3: Mapping and Aggregation

Here, the goal is to combine the associations discovered in the elicitation stage into a holistic description of the brand. This can be achieved via the creation of various types of social tag maps and by tracking brand-level social tag metrics, to meet different managerial objectives. For instance, a brand-centric map can present the key associations for the focal brand, or for each competing brand. Such associations can also be represented in a multi-brand map to highlight inter-connected associations across the

brands. Further, distinct associative maps can be created using either all keywords, or only valenced keywords (that capture both positive and negative sentiment), or only descriptive, neutral keywords. The real power of social tag maps over previous approaches is apparent in this stage, given its ability to provide relevant managerial insight into the many distinct aspects of brand associative networks.

Figure 2.1 shows the various networks of social tags associated with the focal and competing brands 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 co-occurrence volume of the two keywords. Figure 2.1a presents various brand-centric maps for Apple. Specific associative maps are constructed based on overall strong keywords (Figure 2.1a-a1), strong category and product related keywords (Figure 2.1a-a2), strong positive keywords (Figure 2.1a-a3), and strong negative keywords (Figure 2.1a-a4). Along the same lines, Figure 2.1b (b1-b4) presents the associative networks for multiple brands, i.e., the focal brand and its top competitors: Apple, Blackberry, Dell, Google, HP, Microsoft, and Nokia. This figure shows the relative position of each brand on networks of keywords.

[INSERT FIGURE 2.1 ABOUT HERE]

Figure 2.2 shows the brand associative network for Apple with inter-tag relationships, similar to the maps constructed using ZMET or BCM, which classify tags into primary, secondary and tertiary associations[5]. Using our approach, we can impute an

---

[5] John et al. (2006) classified all associations as primary, secondary and tertiary tags on the basis of whether they have a direct, or indirect connection to the focal brand. This strict hierarchy of tags is potentially the result of the small sample size of consumers that can be employed when the brand mapping method requires intensive interviewer-interviewee discussions. If more consumers could be interviewed it might be the case that almost all tags have at least a weak relationship (i.e., a few consumers associate them) with the focal tag.

even richer interrelationship over the associations in a number of different ways: (1) Associative strength metrics such as co-occurrence volume or scaled volume can be used to classify primary, secondary and tertiary associations. For instance, primary associations can be defined as tags whose co-occurrence volume is above the top 10% quantile in all associations; secondary if it is between 10 % and 50%; and tertiary if it is below 50% quantile. (2) Inter-tag relationships can be used to determine the extent to which a tag is primary or peripheral, i.e., based on the level of connectedness (degree centrality of the tag). For instance, in Figure 2.2 tags with relatively higher degree (number of connections with other tags) and co-occurrence volume with Apple, such as "iphone" and "mac" are likely to be primary associations and positioned near the center of the graph. Similarly, tags positioned near peripheral area can be interpreted as "secondary associations."

[INSERT FIGURE 2.2 ABOUT HERE]

*Social Tag Metrics*

The aggregated social tag map can be further exploited with the following metrics, which potentially indicate diagnostic value of brand assets. These metrics can capture dynamics in social attention on a brand, richness, valence, and dispersion of brand associations, and competitiveness of a brand. Table 2.4 summarizes the social tag metrics.

[INSERT TABLE 2.4 ABOUT HERE]

*Social attention.* The volume of bookmarks linked to a brand ($BN_{it}$) can serve as a proxy measure to track dynamics in social attention/interests. This metric can not only capture as the volume of brand-related content available on the Web but also the extent to which users interpret content related to a brand. In that users provide a specific brand

21

name to associate with content in a web link, tagging process in some way can be comparable to brand recall or recognition test. Thus, marketers can use this metric as a complement to the volume of UGC on a brand or the classic brand awareness metric.

*Valence of associations.* We employ valence of tags linked to a brand as a proxy for valence of associations. Valence of tags can be determined by classifying them into subcategories such as positive descriptions, negative descriptions, and neutral descriptions. We employed three human judges to determine valence. Alternatively, automated procedures could be deployed to characterize valence of tags (e.g., Das and Chen 2007). The first judge manually classified the 7,019 tags associated with the focal brand into three subcategories - positive, negative, and neutral descriptions. We then selected a subset of keywords which includes all positive and negative keywords (318 keywords), and a random sample of 100 neutral keywords, pre-classified by the first judge. The remaining two judges were asked to classify these 418 keywords into three categories - positive, negative, and neutral descriptions. Inter-rater reliability across three judges was checked using Fleiss' Kappa index (Fleiss 1971), which for 418 subjects was .904 (z = 44.7, p < .001), indicating a reasonable level of agreement. For keywords when judges disagree, we took the majority opinion. Valence of tags linked to a brand can be measured in multiple ways: the number of distinct positive (negative) tags, the volume of positive (negative) tags, or the proportion of positive (negative) tags on all tags linked to a brand via bookmarks.

*Breadth of associations.* Breadth of brand associations can be captured by how many distinct tags are used to describe all bookmarks linked a brand ($DN_{it}$) as well as how many of tags are used per each bookmark linked to a brand ($AVGN_{it}$). These metrics

22

tells marketers the breadth of brand associations and their dynamics. Having broad range of associations can indicate richer brand associative structure, yet it can further be investigated with dispersion of tags linked to a brand. Dispersion of tags linked to a brand ($Disp_{it}$) can be measured using the entropy measure (see Table 5), which been successfully employed in the past to measure the dispersion of user-generated content in a different social media platform (e.g., Godes and Mayzlin 2004). Entropy is maximized when the volume of each tag linked to a brand is evenly distributed. One interpretation of the entropy in tag distribution is the extent to which the associations linked with a brand are diffuse or coherent. Low entropy of tag distribution indicates more coherent brand associations, since there exists a clear distinction between strong and weak associations; whereas high entropy of tag distribution indicates more diffuse brand associations, since there is less agreement on representative associations of a brand across users.

Entropy in this case is inherently relative. Thus, in order to compare dispersion measures across brands or over time, one can scale the dispersion measure by the maximum entropy that can be obtained if all tags are equally distributed, given the current number of tags and co-occurrence volume.

*Competitiveness.* We track competitiveness of a brand by computing volume share, uniqueness of associations, and the level of connectedness to competitors. One way to capture the competitiveness of a brand is to investigate the relative perceptions of a brand as compared to its competitors and identify relatively stronger set of brand associations for each brand (e.g., Keller 1993). By calculating volume share in each association of a focal brand as compared to competitors $t$ ($Share_{it}$), we can understand the extent to which a brand is more associated with the given tags in the minds of

consumers than its competitors. This metric is indicative of the extent to which a brand dominates that association over its competitors. We consider volume share in three different types of associations: positive associations, negative associations, and overall associations.

Brand uniqueness, a significant source of a competitive advantage of a brand (e.g., Keller 1993), can be obtained investigating social tagging networks. We capture uniqueness of brand associations along with two dimensions: the number of distinct unique tags not shared with competitors ($UNQ_{it}$) and proportion of unique tags on all tags linked to a brand ($SUNQ_{ijt}$). Marketers can understand the dynamics in customers' perceptions on brand uniqueness by tracking whether the proportion or distinct number of unique tags increase unique tags.

Strength of a brand's category membership or category representativeness, an important component of brand equity can be determined through the set of shared associations with competitors (e.g., Keller 1993). We capture the connectedness to competitors' associations with the average number of competitors linked to each association ($CONT_{it}$). This metric could be further enhanced by computing a weighted average, where the weight can be obtained from sources such as market share, the importance of a brand to a brand derived from the tagging structure on *del.icio.us*, etc.

In Table 2.5, we summarize the information contained in the aggregated brand map for our chosen focal brand, Apple, and its competitors with the monthly average social tag metrics for 2009. On an average, Apple is linked to 13,472 social bookmarks and 4,121 distinct tags; 62,323 Apple-tag-pairs were created each month. When it comes to the valence of tags, on an average, 105 positive tags  and 109 negative tags were linked

to Apple, and 2,043 pairs between Apple and a positive tag (3.3% of all the pairs) and 523 pairs between Apple and a negative tag (0.8% of all the pairs) were created each month. These metrics enable easy comparison of the focal brand with its competition. For instance, Apple's monthly average volume share in positive tags (18.6%) is lower than that of Google's (55.5 %), but higher than that of other competitors. For Apple, the average volume share in positive tags (18.6%) is slightly higher than that in all tags (18.3%) and the average volume share in negative tags (16.5%) is lower than that in all tags. However, for Google, HP and Microsoft, the volume share in both positive tags (55.5%, 6.0%, and 13.7%, respectively) and negative tags (54.7 %, 6.4%, and 17.0%, respectively) is higher than that in all tags (54.5%, 5.6%, and 13.5%, respectively), indicating that more negative associations were linked to them as compared to Apple. On average, Google has the most number of unique keywords (768, 14.3% of all pairs) and the least number of competitors linked to its keywords (2.61), while Blackberry has the least number of unique keywords (13, .8% of all pairs)  and the most number of competitors linked to its keywords (4.45). This result indicates that among these seven brands Google is perceived most unique. While these metrics enable ease of evaluation and comparison for the focal brand vis- à-vis the competition, such comparison becomes stronger and more valuable for the manager when we consider the dynamic evolution of competitive social tag maps as we discuss in Stage 3.

[INSERT TABLE 2.5 ABOUT HERE]

**2.4. Reliability and Comparison with Brand Concept Maps**

We now test the reliability and consistency of the co-occurrence volume metric, the core metric based upon which we construct aggregated social tag maps and derive additional social tag metrics for capturing brand-level characteristics.

*Reliability*

We test whether our approach using the co-occurrence volume metric can reliably identify the core association by comparing aggregated social tag maps from two split half sample (Churchill 1979). We selected 2,000 bookmarks linked to Apple generated in 2009 and randomly split the sample of 2,000 bookmarks into two groups. We compared the two samples in terms of the number of tags identified, the percent coverage of top 100 core associations identified in Stage 2, and the percentage of co-occurrence volume for each yearly bookmark data explained with the identified tags in each sample. 759 distinct tags are identified in sample 1, 709 tags are identified in sample 2, and 436 tags are identified in both samples. Both samples show a reliable coverage of core 100 associations identified in Stage 2: 96 core tags are identified in sample 1 and 94 core tags are identified in sample 2. It is interesting to note that with 436 tags identified both in sample 1 and sample 2, one can explain more than 50% of co-occurrence volume across the four-year data. In addition, the associative strength of each of 436 common tags and Apple, captured by the co-occurrence volume metric is highly correlated between two samples (r=.99, p<.01).

*Comparison with Brand Concept Maps*

We investigate whether the associative metrics from our approach are correlated to those from the BCM approach. To obtain BCM for Apple, we conducted a one-on-one interview with 23 subjects. Following the methodology of obtaining consensus BCM

illustrated by John et al. (2006), we (1) ask subjects what comes to mind when they think about a brand, (2) give them detailed one-on-one instructions as to how to draw a brand concept map with the example presented in John et al. (2006, Figure 2, pp.553), (3) ask them to draw their own concept map, and (4) draw the consensus brand concept map of Apple based on responses from all subjects: 47 associations, mentioned more than 25% of the subjects, are present in the consensus map. For comparison, a social tag map for Apple is constructed based on the co-occurrence relationships between each association and the brand during a corresponding time window (6-month data).

We investigate the correlation between the two metrics from the consensus brand map - the frequency of each association and the weighted frequency[6] of each association across respondents – and the co-occurrence volume of the corresponding tag with the brand from the social tag map. Overall, the frequency metric in BCM and the corresponding co-occurrence volume metric in our approach for 47 associations is significantly correlated ($r= .56$, $p <.01$) and the weighted frequency metric of each association in BCM and the corresponding co-occurrence volume metric in our approach is also significantly correlated ($r= .54$, $p <.01$). The correlation between BCM and our approach is reasonably high considering that we compared the social tag map based on more than 20,000 responses mostly generated by distinct users with brand concept map from 23 respondents.

---

[6] Similar 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 level (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.

In addition, we investigated separately the correlations for two types of associations- evaluative associations and descriptive associations[7]. The frequency of associations 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 ($r = .74$, $p < .01$). The weighted frequency of associations in the consensus BCM and the tag co-occurrence volume for evaluative keywords is also highly correlated ($r = .74$, $p < .01$) and that for descriptive keywords is highly correlated ($r = .73$, $p < .01$). We further conducted similar analysis for the other two brands -Microsoft and Google- and found a similar pattern[8].

We further investigated if the associations weakly related to the brand in BCM are also weakly related to the brand in social tag maps. We classified the associations mentioned by only one subject in the interview and thus not present on the consensus brand map as weak associations. We found that the corresponding tag co-occurrence volume of the weak associations in BCM approach is significantly lower than of the other associations ($M_{weak} = 174$, $M_{others} = 956$, $t = 2.24$, $p < .05$). For descriptive keywords, the difference is more distinct ($M_{weak} = 359$, $M_{others} = 2,517$, $t = 2.51$, $p < .05$) than for evaluative keywords ($M_{weak} = 20$, $M_{others} = 100$, $t = 1.77$, $p < .10$).

---

[7] We found that evaluative associations (e.g., "cool" and "innovative") more frequently appeared in BCM while descriptive associations (e.g., "iPod", "computer") more frequently appeared in social tags. This is because (1) the question in the survey induces subjects to generate more attitudinal and evaluative associations and (2) subjects 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.
[8] We further investigated the convergent validity of the associative map of three brands – Apple, Google, and Microsoft. Subjects were asked to draw a concept map of three brands which show the interconnections between brands and their associations. The correlation between the metrics from consensus BCM and those from social tag maps range between .49 and .57, which is reasonably high considering that this process requires significant amount of cognitive resources from subjects and subjects tend to draw less informative maps with smaller number of associations per brand for this three-brand map, than for a one-brand map.

**2.5. Value of Social Tag Maps**

    Social tag maps are valuable because they allow us to efficiently capture the dynamic evolution of brand associations, social interactions within brand association networks, as well as competitive intelligence given a focal brand of interest. We now discuss each of these aspects.

*Dynamics of Brand Associations*

    Since social tag maps can be constructed for any given time frame (e.g., weekly, monthly, etc.), tracking the dynamic evolution of tag maps can provide marketers with significant information about changes in customers' perceptions of their brand. First, marketers can track trends in the metrics of interest. To illustrate, Table 2.6 presents the monthly dynamics for the same social tag metrics from Table 2.3, but only for the focal brand Apple over the period, 2006-2009. Dynamics are summarized using velocity (the difference in value of the metric between time t and t-1) and acceleration (the difference in velocity of the metric between time t and t-1). The reported monthly mean velocity and acceleration of social tag metrics summarize the dynamics in the trends of these metrics over the four years, which are reflective of long-term growth and rates of growth. When computed for a specific month, they reflect the instantaneous changes in the related metrics in terms of short-term growth as well as the rate of growth. From Table 2.6, we see that on average, the volume of bookmarks linked to Apple has been constantly growing (velocity M = 173), but the rate of growth has been decreasing (acceleration M = -13) indicating that overall growth might be leveling off. In addition, the scaled sum of co-occurrence volume of positive tags and negative tags linked to Apple has been declining over time, implying that relatively more neutral keywords have been employed

to describe Apple's content. Overall, the volume share in all tags and the number of unique tags linked to Apple constantly declined over the four years (velocity M= -.12%, velocity M = -2.07, respectively), indicating that the perceived uniqueness of Apple may be declining over time as competitors start to steal some of the core associations that were previously uniquely linked to Apple. However, the number of competitors per association (connectedness) declined over time (velocity M= -.01), indicating the possibility that existing competitors, rather than newly connected competitors have stolen Apple's associations. It is important to realize that Table 2.6 can be readily updated at every time period, enabling marketers to proactively manage short-term and long-term brand assets by tracking instantaneous velocity and instantaneous acceleration related to various social tag metrics.

[INSERT TABLE 2.6 ABOUT HERE]

Further, the dynamics in these social tag maps allow marketers to identify managerially interesting change in brand association. Figure 2.3 shows the trend in the co-occurrence volume of selected positive tags linked to Apple. We can see the increase in the tags, *cool*, *fun*, *innovation*, and *interesting* around January 2007 when Apple unveiled its first Iphone. Likewise, the co-occurrence volume of *innovation* with Apple jumped around June 2009, when Apple introduced its iPhone 3GS on the market.

[INSERT FIGURE 2.3 ABOUT HERE]

Table 2.7 shows the five highest co-occurring valenced tags in terms of volume, volume growth, and volume decline for Apple in each quarter of 2009. We see that "technology" is the keyword the most strongly linked to Apple across all four quarters, and is one of the five fastest growing keywords in the 1st, 2nd, and 3rd quarter, but one

30

of the five fastest declining keywords in the 4th quarter. Such information enables

marketers to take steps to proactively manage their brand equity, by detecting the fastest

growing negative keywords and the fastest declining positive keywords. For instance,

"fail" and "evil" are among the fastest growing keywords in the 2nd and 3rd quarter,

indicating concerns regarding product problems and ethical issues related to Apple

products. Likewise, fastest declining positive keywords (e.g. "funny," "easy," "cool," and

"healthy" in $2^{nd}$ quarter) can imply change in customer brand perceptions. Such flags

enable marketers to take a further look at the content of bookmarks tagged with these

keywords, and see whether some keywords may chronically negatively influence overall

brand assets.

[INSERT TABLE 2.7 ABOUT HERE]

### Social Interactions within a Network

A user's bookmarking activities can be influenced by what kinds of content other

users bookmark, and what keywords other users employ to describe content. This social

aspect of tagging gives marketers the ability to find current social interests and trends,

independent of the focal brand, and determine how the focal brand may be linked to these

interests and trends. Table 2.8 presents monthly dynamics in the volume of tags from the

social bookmarking activities of all users for all content, from 2006 to 2009. The top five

most popular tags in *del.icio.us* are *blog, news, video, online,* and *web* over the four years,

and the volume velocity of all of these tags is constantly positive, reflecting a growing

interest in online content over the four years. The fastest growing tag is "ping[9]" with the

highest velocity (M=1,811, SD=3,520) and acceleration (M=87, SD=3,765). The top five

---

[9] *ping* has several different meanings, but probably the most relevant here is as a term-of-art used to describe letting various directory services know that you have created new content on your blog, website, etc. The growth of this tag coincides with the growth of the term's usage in the social media space.

fastest declining tags in the community are *reference, safari, export, programming*, and *june* over the four years. These trends can be further combined with network metrics employing network analysis such as centrality, clustering coefficient, and betweenness measures and can be tracked over time to better inform managers about the changing importance of various tags within this community (Wasserman and Faust 1994).

[INSERT TABLE 2.8 ABOUT HERE]

*Competitive Intelligence*

Social tags provide marketers with a unique ability to collect, analyze and visualize competitive information and help marketers compare brand assets of a focal brand with those of its competitors. Figure 2.4 presents an illustrative example comparing the strength of brand associations, captured by scaled volume metric, for two competing brands, Apple and Microsoft. For instance, Apple is more strongly connected to the keyword, "cool" (MAPPLE=.02, MMICROSOFT=.009, t=9.2, p<.01), while Microsoft is more strongly connected to the keyword, "bestpractices" (MAPPLE=.0038, MMICROSOFT=.01, t=-12.5, p<.01).

[INSERT FIGURE 2.4 ABOUT HERE]

The tagging networks can indicate the competitive position of a brand by dynamically revealing who is competing with whom, and in what domain. Figure 2.5 presents a perceptual map for Apple and its six competing brands on selected dimensions over time. Figure 2.5a and 2.5b show the perceptual map of seven brands on "mobile" and "computer" dimension. Standardized scaled volume metric is employed to compare the relative position of each brand on the map. Note that in 2006 Apple was more likely to be perceived as related to the "computer" market, but in 2009 Apple is perceived

related to the "mobile" market, reflecting the recent success of iPhone in the smart phones and mobile devices market. Similarly, the positioning of Google has changed, being more strongly connected to "mobile" market, given the rise of Android. The perceptions of Nokia and Blackberry as "mobile" players and HP and Dell as "computer" players are relatively unchanged from 2006 to 2009. Related to their brand image, Figure 2.5c and 2.5d show perceptual maps for seven brands on dimensions. Both in 2006 and 2009, Apple is perceived as the coolest brand while Google is perceived as the most innovative brand. Interestingly, we note that in 2006 Blackberry was perceived as being relatively innovative, which is no longer the case in 2009.

[INSERT FIGURE 2.5 ABOUT HERE]

Marketers can also track and compare the social tag metrics for the focal brand with those for competing brands. Table 2.9 presents the monthly dynamics in selected social tag metrics for Apple and its six competitors. On an average, the volume of bookmarks linked to all seven brands is increasing, but the rate of growth (acceleration) is constantly decreasing over the four years. While the volume of both positive tags and negative tags linked to Blackberry, Dell, and Google are increasing over time, the volume of both positive tags and negative tags linked to Apple, HP, and Microsoft are decreasing over time. Interestingly, the volume of negative tags linked to Nokia is increasing but that of positive tags is decreasing over time, possibly indicating lower competitiveness of the brand.

[INSERT TABLE 2.9 ABOUT HERE]

33

## 2.6. Conclusions and Discussion

In this paper, we present a novel approach for constructing brand associative networks using social tags. We demonstrate that our approach has several advantages over existing methods. First, it is less time consuming and less expensive. While existing approaches rely on elicited associations either from consumer interviews or from algorithms, our method utilizes keywords directly stated by online users to describe a brand or content related to a brand. Hence, our approach is less vulnerable to potential errors or biases involved in the elicitation stage, and is able to provide richer and unbounded associations linked to a brand. Using social tag maps, marketers can (1) have access to real-time updates of brand associative networks, and track their brand assets dynamically and (2) understand the competitive position of their brand, and track the dynamics in competitive structure.

The information contained in social tag maps 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, since both allow 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; whereas online search is more of a goal-oriented behavior. Thus, tagging data is perhaps more appropriate when marketers are interested in obtaining consumers' perceptions on a brand.

The informational value of social tag maps can be assessed by using tags as a proxy measure for intangible brand assets, which can then be used to predict and explain firm valuation, e.g., via stock market returns. Dynamics in social tag metrics can capture changes in social attention, social evaluation, and competitive advantage of a brand, and thus possibly be related to investors' expectations of a firm's prospect. In a parallel project, we investigate this issue in detail for determining the value of each social tag metric in explaining brand assets and firm value. Another potential application of social tag maps is to map an individual product onto social tagging networks at online retailers such as Amazon.com to describe the semantic position of a product with node-level network characteristics on the tagging networks. Researchers can further investigate the relationship between product demand and the semantic position of product.

There are a few limitations and possible avenues for future research that we need to highlight. First, the social tag maps we create do not take semantic distance between keywords into account; i.e., all synonyms are treated as distinct keywords in our analysis. A potential solution to this problem will be considering a lexical database of words such as *WordNet* (e.g., Miller 1995) and incorporate this information into the construction of social tag maps. Second, additional metrics that rely even more on network characteristics, betweenness and network density can be defined to provide marketers with more integrative information about the associative networks. Third, our aggregate social tag maps present the collective users' perceptions on a brand. A future research can investigate heterogeneous representation of brand maps using disaggregate level tagging data and identify dynamic customer segments based on social tags. In addition, future study can focus on modeling the evolution process of customers' brand perceptions over

time. For instance, a researcher can investigate how a brand's dominance on a specific association domain can impact customers' perceptions on competitors. Finally, future research can investigate a better representation of the growth in the dynamic network such that metrics calculated at different time points are more comparable. We hope that this work serves as a modest start towards those future directions.

**Table2.1: Comparison of Methodologies to Create Brand Maps**

| | ZMET (Zaltman and Coulter 1995; Zaltman 1997) | BCM (John et al. 2006) | Text-Mining (Lee and Bradlow 2011; Netzer et al. 2011) | Social Tag Maps |
|---|---|---|---|---|
| Association Elicitation | • In-depth personal interviews<br>- Qualitative techniques<br>- Kelly's repertory grid<br>- Laddering exercises<br>• Both verbal and nonverbal cues (e.g., photos, images) | • Prior consumer research<br>• Managers' opinions/ insights<br>• Consumer interview | • Elicited by a text-mining tool<br>- rule-based<br>- machine learning<br>- hybrid approach | • Stated by consumers/ online users |
| Mapping and Aggregation | Participants create a map or visual montage (2nd personal interview) | Respondents develop their brand maps with brand association cards (personal interview) | Maps based on elicited product attributes/ brand associations from text mining model | Maps based on tags stated by consumers/ users in the absence of a researcher, hence more informative in revealing preference |
| Richness of Information | • Deep understanding of a brand<br>• Unconscious aspects can be revealed<br>• Few Subjects | • Hierarchical associative structure<br>• Differential associative strength (single, double, triple)<br>• Few Subjects | • Dynamics<br>• Competitive intelligence (although influenced by algorithmic interpretation) | • Dynamics<br>• Competitive intelligence<br>• Social interactions/ attentions<br>• Unbounded/ undirected<br>• Revealed preferences |
| Difficulty of Administration | High (Qualitative analysis expert required) | Moderate (Standardized procedures) | Moderate (Multiple stages of text-mining processes) | Low (Public and readily accessible) |

**Table 2.2: Positive Brand Associations for Apple Sorted by Co-occurrence Volume***

| Associations | Volume of Tags $N_t(T_j)$ | Co-occurrence Volume $N_t(B_i,T_j)$ | Scaled Volume $SN_t(B_i,T_j)$ | Category relevance[a] | Weighted Volume $WN_t(B_i,T_j)$ |
|---|---|---|---|---|---|
| technology | 335,376 | 6,531 | .028 | .039 | 141.14 |
| cool | 135,016 | 2,072 | .014 | .022 | 13.21 |
| fun | 162,808 | 1,298 | .008 | .019 | 6.14 |
| interesting | 81,107 | 1,187 | .010 | .014 | 5.65 |
| funny | 165,975 | 1,134 | .007 | .018 | 4.90 |
| innovation | 78,214 | 1,040 | .009 | .006 | 4.18 |
| useful | 40,509 | 960 | .012 | .013 | 2.90 |
| new | 672,615 | 919 | .003 | .006 | 3.20 |
| best | 269,367 | 913 | .004 | .005 | 2.59 |
| awesome | 59,136 | 805 | .008 | .005 | 2.08 |

*The bookmark volume linked to Apple, $N_t(B_i)$ is 161,664 for all cells. The measures are calculated based on social tags created for Apple in 2009. The table is created within the subset of positive keywords and similar tables can be created for all the keywords.

a Category relevance is $\dfrac{\sum_{i' \in CB_i} N_t(B_{i'}, T_j)}{N_t(B_i) + \sum_{i' \in CB_i} N_t(B_{i'})}$ in Equation (2).

**Table 2.3a: Number of Tags to Explain 95% [90%, 80%] 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% |
| | Min. co-occurrence vol. | 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% |
| | Min. co-occurrence vol. | 65 | 12 | 7 | 161 | 12 | 59 | 13 |
| 80% | Number of tags | 715 | 420 | 616 | 863 | 614 | 678 | 501 |
| | % of Tags | 12% | 10% | 15% | 13% | 13% | 12% | 12% |
| | Min. co-occurrence vol. | 169 | 34 | 19 | 380 | 34 | 143 | 36 |

**Table 2.3b: The percentage of Co-occurrence Volume Explained by Decision Rules***

| Decision Rule | | Apple | Blackberry | Dell | Google | HP | Microsoft | Nokia |
|---|---|---|---|---|---|---|---|---|
| N(B,T) >1 | Number of tags | 5,613 | 3,140 | 3,014 | 6,313 | 3,872 | 4,967 | 3,275 |
| | % of Tags | 92% | 76% | 73% | 96% | 79% | 90% | 77% |
| | % Explained | 99.9% | 99.1% | 98.5% | 99.9% | 99.2% | 99.9% | 99.2% |
| 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% |
| N(B,T) >20 | Number of tags | 2,279 | 643 | 568 | 4,495 | 896 | 2,326 | 772 |
| | % of Tags | 45% | 15% | 13% | 68% | 18% | 42% | 18% |
| | % Explained | 96.7% | 85.3% | 78.7% | 99.1% | 85.4% | 96.2% | 86.1% |

* Table 2.3a and 2.3b are based on social tags created for each brand in 2009. Similar tables can be created using the scaled volume and weighted volume metrics.

**Table 2.4: Social Tag Metrics***

| Construct | Notation | Definition | Explanation |
|---|---|---|---|
| Social attention | $BN_{it}$ | $N_t(B_i)$ | Volume of bookmarks created on brand i during time t |
| Valence of associations | $PDN_{it}$ | $\sum_{j'\in POSTAG} I_t(B_i, T_{j'})$ | Number of distinct positive tags linked to brand i during time t |
| | $PSUMN_{it}$ | $\sum_{j'\in POSTAG} N_t(B_i, T_{j'})$ | Volume of positive tags linked to brand i during time t |
| | $PSN_{it}$ | $PSUMN_{it}/SUMN_{it}$ | Proportion of positive tags on all tags of brand i during time t (%) |
| | $NDN_{it}$ | $\sum_{j'\in NEGTAG} I_t(B_i, T_{j'})$ | Number of distinct negative tags linked to brand i during time t |
| | $NSUMN_{it}$ | $\sum_{j'\in NEGTAG} N_t(B_i, T_{j'})$ | Volume of negative tags linked to brand i during time t |
| | $NSN_{it}$ | $NSUMN_{it}/SUMN_{it}$ | Proportion of negative tags on all tags (%) |
| Breadth of associations | $DN_{it}$ | $\sum_{j=1}^{NT} I_t(B_i, T_j)$ | Number of distinct tags linked to brand i during time t |
| | $AVGN_{it}$ | $\sum_{j=1}^{NT} N_t(B_i, T_j)/N_t(B_i)$ | Mean number of tags per bookmark on brand i during time t |
| | $Disp_{it}$ | $-\sum_{j=1}^{NT} \frac{N_t(B_i,T_j)}{\sum_{k=1}^{NT} N_t(B_i,T_k)} log\left(\frac{N_t(B_i,T_j)}{\sum_{k=1}^{NT} N_t(B_i,T_k)}\right)$ | Dispersion of tags linked to brand i during time t |
| Competitiveness | $Share_{it}$ | $\sum_{j=1}^{NT} \frac{N_t(B_i,T_j)}{N_t(B_i,T_j)+\sum_{i'\in CB_i} N_t(B'_i,T_j)}/DN_{it}$ | Mean share of brand i on each tag compared to its competitors |
| | $UDN_{it}$ | $\sum_{j=1}^{NT} \prod_{i'\in CB_i} I_t(B_i, T_j)(1 - I_t(B_i, T_j))$ | Number of distinct unique tags not shared with competitors |
| | $USUMN_{it}$ | $\sum_{j=1}^{NT} N_t(B_i, T_j) \prod_{i'\in CB_i} I_t(B_i, T_j)(1 - I_t(B_i, T_j))$ | Volume of unique tags linked to brand i during time t |
| | $USN_{it}$ | $USUMN_{it}/SUMN_{it}$ | Proportion of unique tags on all tags of brand i during time t (%) |
| | $CONT_{it}$ | $\sum_{j=1}^{NT} \sum_{i'\in CB_i} I_t(B_{i'}, T_j) I_t(B_i, T_j)/DN_{it}$ | Mean number of competitors co-tagged with brand i during time t |

* where $N_t(B_i)$ is volume of bookmarks tagged with a brand name, for *i* is an index for brand ranging from 1 to *N*; $N_t(T_j)$ is volume of bookmarks linked to tag *j* during time window *t*, for *j* is an index for tag ranging from 1 to *NT*; $N_t(B_i, T_j)$ is the co-occurrence volume of tag *j* with brand *i* for a given time window *t*; $I_t(B_i, T_{j'})$ is an indicator, which is equal to 1 if $N_t(B_i, T_j)$ >0 else 0; *POSTAG* is the set of positive keywords classified by three judges; *NEGTAG* is the set of negative keywords classified by three judges; *CBi* is the set of competitors defined in Stage 1.

**Table 2.5: Social Tag Metrics for Selected Brands***

| | | Apple | Blackberry | Dell | Google | HP | Microsoft | Nokia |
|---|---|---|---|---|---|---|---|---|
| Social attention | Volume of bookmarks linked to a brand | 13,472 | 3,022 | 3,102 | 63,203 | 5,175 | 14,209 | 4,040 |
| Valence of associations | Number of distinct positive tags | 105 | 45 | 46 | 125 | 69 | 92 | 53 |
| | Volume of positive tags | 2,043 | 186 | 165 | 4,973 | 443 | 1,732 | 263 |
| | Proportion of positive tags (%) | 3.3% | 2.0% | 2.7% | 3.0% | 3.9% | 3.5% | 2.6% |
| | Number of distinct negative tags | 109 | 30 | 32 | 149 | 49 | 104 | 35 |
| | Volume of negative tags | 523 | 46 | 55 | 1,442 | 195 | 657 | 67 |
| | Proportion of negative tags (%) | .8% | .5% | .9% | .9% | 1.7% | 1.3% | .7% |
| Breadth of associations | Number of distinct tags linked to a brand | 4,121 | 1,651 | 1,546 | 5,369 | 2,113 | 3,541 | 1,802 |
| | Average number of tags per bookmark | 4.7 | 3.0 | 2.0 | 2.7 | 2.2 | 3.4 | 2.5 |
| | Dispersion | 6.53 | 5.30 | 6.12 | 6.88 | 6.29 | 6.60 | 5.99 |
| | Dispersion scaled by maximum entropy | .78 | .72 | .83 | .80 | .82 | .81 | .80 |
| Competitiveness | Volume share in positive tags (%) | 18.6% | 1.9% | 1.7% | 55.5% | 6.0% | 13.7% | 2.5% |
| | Volume share in negative tags (%) | 16.5% | 1.7% | 1.8% | 54.7% | 6.4% | 17.0% | 1.9% |
| | Volume share in all tags (%) | 18.3% | 2.3% | 2.7% | 54.5% | 5.6% | 13.5% | 3.1% |
| | Number of distinct unique tags | 140 | 13 | 27 | 768 | 82 | 74 | 22 |
| | Proportion of unique tags (%) | 3.4% | 0.8% | 1.8% | 14.3% | 3.9% | 2.1% | 1.2% |
| | Connectedness | 3.19 | 4.45 | 4.42 | 2.61 | 3.98 | 3.39 | 4.38 |

* The monthly average of metric is reported based on social bookmarking data generated in 2009.

**Table 2.6: Monthly Dynamics in Social Tag Metrics of Apple\***

| | Value | | Velocity | Acceleration |
|---|---|---|---|---|
| | M | SD | M | M |
| **Attention** | | | | |
| Volume of bookmarks linked to a brand | 11,213 | 2,320 | 173 | -13 |
| **Valence** | | | | |
| Number of distinct positive tags | 102 | 5 | -.02 | -.10 |
| Volume of positive tags | 2,092 | 207 | -5 | -1 |
| Proportion of positive tags (%) | 3.5% | .3% | -.017% | .0002% |
| Number of distinct negative tags | 104 | 10 | -0.12 | -.21 |
| Volume of negative tags | 523 | 92 | -1.31 | -2.50 |
| Proportion of negative tags (%) | .9% | .1% | -.004% | -.004% |
| **Breadth** | | | | |
| Number of distinct tags linked to a brand | 3,974 | 198 | -0.4 | -6 |
| Number of tags per bookmark | 5.47 | 0.99 | -.10 | -.002 |
| Dispersion | 6.51 | .06 | -.0013 | -.0013 |
| Dispersion scaled by maximum entropy | .79 | .01 | -.0001 | -.00001 |
| **Competitiveness** | | | | |
| Volume share in positive tags (%) | 20.4% | 2.1% | -.10% | -.01% |
| Volume share in negative tags (%) | 18.0% | 2.6% | -.10% | .01% |
| Volume share in all tags (%) | 20.2% | 1.7% | -.12% | -.02% |
| Number of distinct unique tags | 172 | 30 | -2.07 | -.43 |
| Proportion of unique tags (%) | 4.4% | .9% | -.1% | -.01% |
| Connectedness | 3.25 | .15 | -.01 | -.001 |

\*Measures calculated based on data collected from June 2006 to December 2009 are reported. Such a table can also be computed on a quarterly and annual basis, or any other pre-specified time period.
\*Standard deviation can also be computed for velocity and acceleration. We do not report them here to preserve space.

**Table 2.7: Quarterly Dynamics in Valenced Brand Associations Linked to Apple***

| | Highest volume | | | Fastest growing | | | Fastest declining | | |
|---|---|---|---|---|---|---|---|---|---|
| | Tag | Volume | Velocity | Tag | Volume | Velocity | Tag | Volume | Velocity |
| 1st Quarter | technology | 1623 | 202 | technology | 1623 | 202 | clean | 63 | -24 |
| | cool | 553 | 71 | cool | 553 | 71 | virus | 57 | -23 |
| | fun | 345 | 2 | creative | 144 | 52 | special | 15 | -17 |
| | funny | 330 | 39 | awesome | 210 | 47 | bailout | 2 | -12 |
| | interesting | 307 | 20 | funny | 330 | 39 | bug | 83 | -11 |
| 2nd Quarter | technology | 1654 | 31 | censorship | 88 | 70 | funny | 274 | -56 |
| | cool | 532 | -21 | innovation | 258 | 66 | hacking | 129 | -40 |
| | fun | 356 | 11 | technology | 1654 | 31 | easy | 42 | -24 |
| | interesting | 297 | -10 | fail | 63 | 21 | cool | 532 | -21 |
| | funny | 274 | -56 | clean | 81 | 18 | healthy | 33 | -20 |
| 3rd Quarter | technology | 1711 | 57 | best | 261 | 87 | creative | 105 | -42 |
| | cool | 516 | -16 | innovation | 322 | 64 | useful | 229 | -40 |
| | innovation | 322 | 64 | good | 159 | 63 | fun | 321 | -35 |
| | fun | 321 | -35 | technology | 1711 | 57 | censorship | 57 | -31 |
| | interesting | 309 | 12 | evil | 47 | 31 | star | 11 | -17 |
| 4th Quarter | technology | 1543 | -168 | healthy | 81 | 46 | technology | 1543 | -168 |
| | cool | 471 | -45 | best | 304 | 43 | funny | 231 | -68 |
| | best | 304 | 43 | lawsuit | 45 | 16 | innovation | 268 | -54 |
| | fun | 276 | -45 | awards | 23 | 12 | hacking | 104 | -52 |
| | interesting | 274 | -35 | new | 235 | 12 | good | 108 | -51 |

* The table is created based on valenced tags generated in 2009. Associations with highest volume are identified based on co-occurrence volume of a tag and a brand name and highest growing and highest declining keywords are identified based on the velocity of co-occurrence volume. Such a table can be created to include top X keywords as specified by the manager, at any time level (e.g., annual, monthly, weekly, etc.).

**Table 2.8: Monthly Dynamics in keywords at the community level***

| | | Volume | | Velocity | | Acceleration | |
|---|---|---|---|---|---|---|---|
| | | M | SD | M | SD | M | SD |
| Highest volume | blog | 94,835 | 20,381 | 1,505 | 5,590 | -99 | 9,322 |
| | news | 61,674 | 11,778 | 815 | 4,392 | -180 | 6,929 |
| | video | 55,621 | 11,138 | 904 | 3,126 | -139 | 5,090 |
| | online | 54,601 | 13,130 | 1,088 | 3,321 | -102 | 5,279 |
| | web | 54,322 | 7,390 | 478 | 3,532 | -118 | 5,655 |
| Fastest growing | ping | 16,041 | 24,213 | 1,811 | 3,520 | 87 | 3,765 |
| | blog | 94,835 | 20,381 | 1,505 | 5,590 | -99 | 9,322 |
| | youtube | 44,144 | 15,105 | 1,193 | 2,658 | -109 | 4,390 |
| | online | 54,601 | 13,130 | 1,088 | 3,321 | -102 | 5,279 |
| | free | 52,895 | 12,842 | 1,045 | 3,352 | -31 | 5,745 |
| Fastest declining | reference | 19,079 | 2,243 | -149 | 1,071 | 9 | 1,818 |
| | safari | 1,376 | 878 | -115 | 637 | 38 | 795 |
| | export | 1,009 | 617 | -112 | 471 | 31 | 378 |
| | programming | 18,880 | 1,436 | -44 | 1,185 | -8 | 1,975 |
| | june | 1,721 | 1,979 | -39 | 2,018 | -24 | 2,963 |

* The table is created based on tags generated from June 2006 to December 2009 and associations are sorted by monthly average volume and velocity

**Table 2.9: Monthly Dynamics in Competitive Relationships of Apple***

| | | Apple | Blackberry | Dell | Google | HP | Microsoft | Nokia |
|---|---|---|---|---|---|---|---|---|
| Volume | Volume | 11,213 | 1,927 | 2,455 | 44,814 | 4,738 | 12,932 | 3,438 |
| | Velocity | 173 | 57 | 50 | 1,095 | 79 | 79 | 58 |
| | Acceleration | -13 | -4 | -7 | -173 | 8 | -50 | -7 |
| Positive volume | Volume | 2,092 | 195 | 162 | 4,131 | 541 | 1,814 | 308 |
| | Velocity | -5.2 | .2 | 1.0 | 13.8 | -.9 | -13.6 | -2.5 |
| | Acceleration | -1.2 | .3 | .2 | -12.6 | .5 | -6.6 | -.4 |
| Negative volume | Volume | 523 | 39 | 59 | 1,128 | 272 | 643 | 55.3 |
| | Velocity | -1.3 | .4 | -.2 | 7.6 | -1.6 | -5.0 | .1 |
| | Acceleration | -2.5 | .03 | -.7 | -6.6 | .1 | -2.1 | -.3 |

*The results in this table are based on monthly social bookmark data linked to each brand from June 2006 to December 2009. To preserve space we report the mean of volume, velocity, acceleration. Standard deviation can be calculated for each cell.

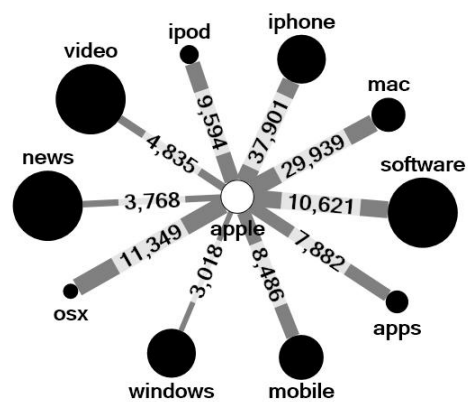**Figure 2.1a: Brand-Centric Social Tag Maps for Apple***

**(a1) overall strong tags**
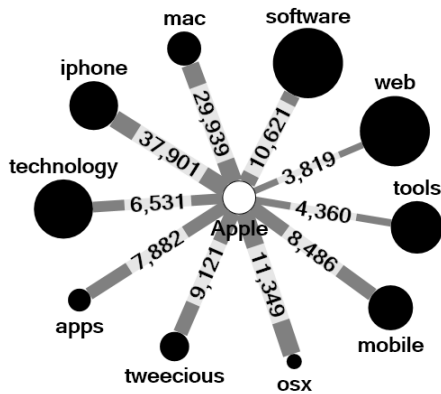


**(a2) strong category and product tags**



**(a3) Strong positive tags**
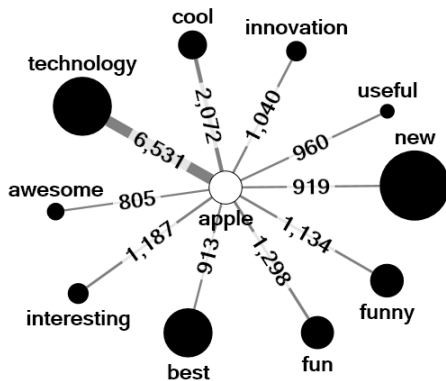


**(a4) Strong negative tags**
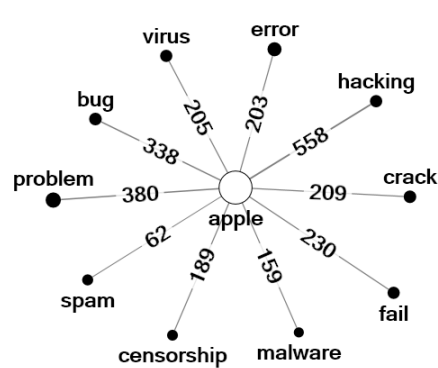


* Social tag maps are created based on delicious bookmark data generated in 2009. Size of the circle is proportional to the volume of bookmarks linked to each keyword and width of the link is proportional to the co-occurrence volume of a keyword with Apple, which is stated in the number on each link.

**Figure 2.1b: Multi-brand Social Tag Maps for Apple and Competitors***

**(b1) Overall strong tags**
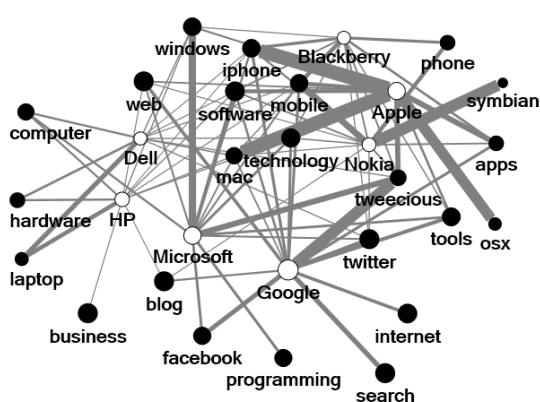


**(b2) Overall valenced tags**
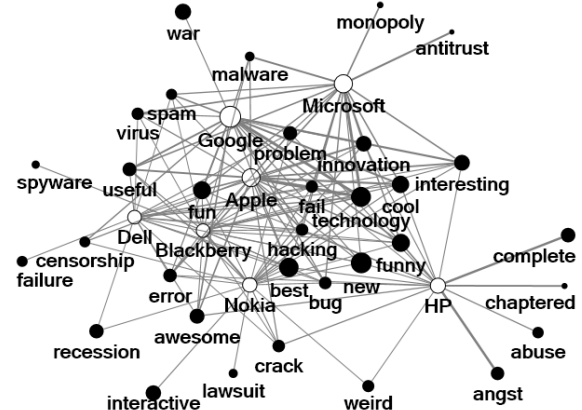


* Social tag maps are created based on delicious bookmark data generated in 2009 using Fruchterman-Reingold graph algorithm. Size of the node is proportional to the volume of bookmarks linked to each keyword and opacity of the link is proportional to the co-occurrence volume of a keyword with each brand.

45

**Figure 2.2: Brand Associative Network of Apple with Inter-tag Relationships***
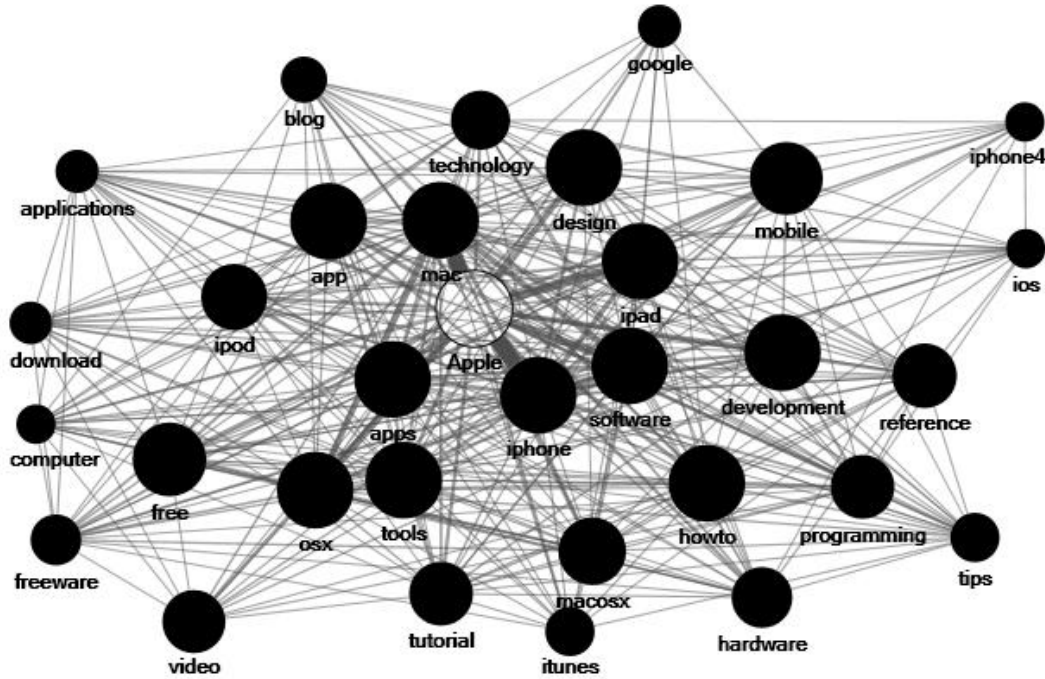


* The figure is created based on 31 primary tags whose co-occurrence volume with Apple is greater than 40, based on 2,000 bookmarks generated from 10th to 13th in June 2010. The size of the node represents the co-occurrence volume with Apple and the width and opacity of the link represents the strength of associative relationship between tags, which is proportional to the co-occurrence volume of two tags.

**Figure 2.3: Trend in Co-occurrence Volume of Selected Positive Tags Linked to Apple***
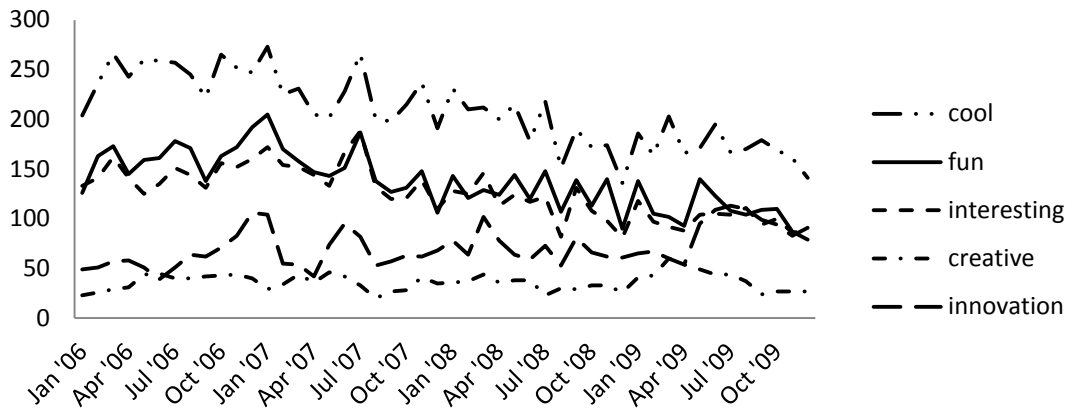


**Figure 2.4: Comparison of Selected Strong Positive Brand Associations: Apple vs. Microsoft***
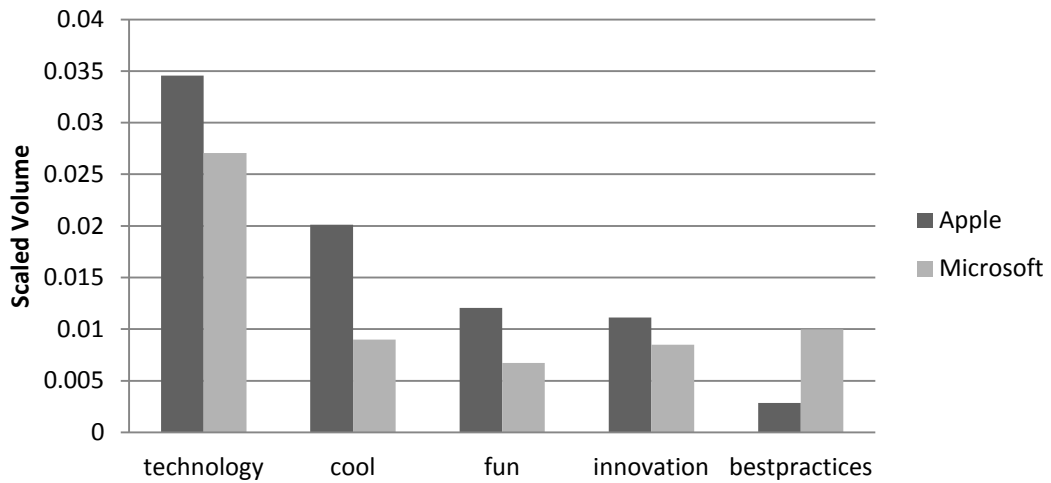


* The figure is created based on monthly mean of scaled volume of selected keywords with each brand, Apple and Microsoft, based on social bookmarks generated from 2006 to 2009. The difference of mean for each pair is significant at .01.

47

**Figure 2.5: Perceptual Maps of Competitors**

**(a) Product Category Perception (2006)**
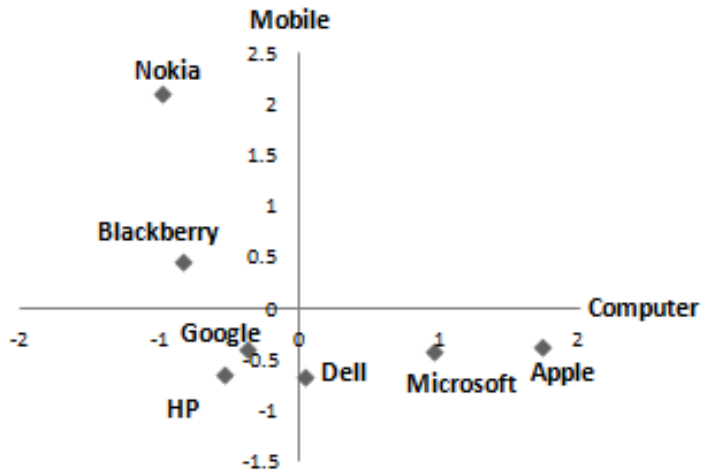


**(b) Product Category Perception (2009)**
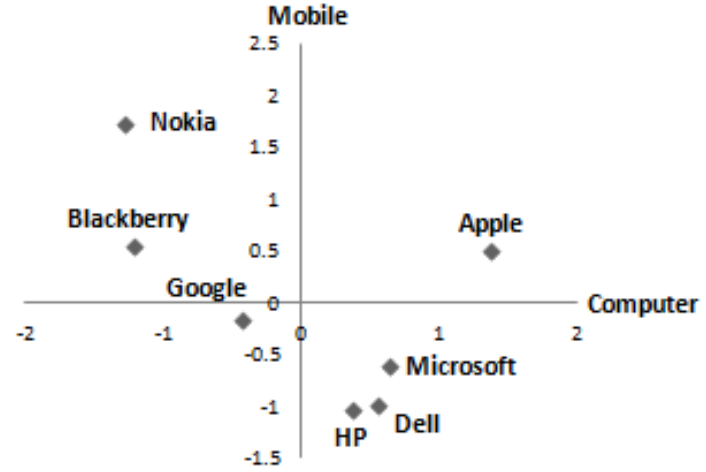


**(c) Brand Image Perception (2006)**
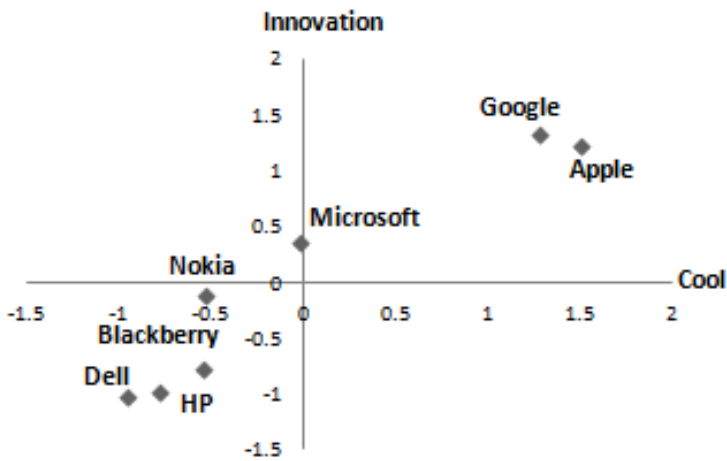


**(d) Brand Image Perception (2009)**
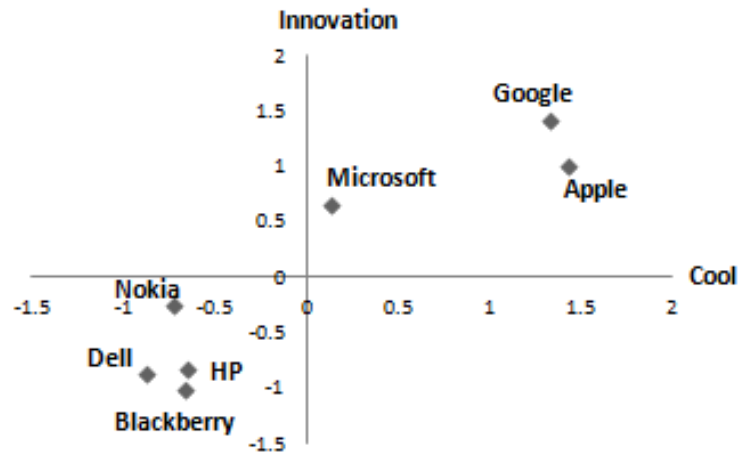
# Chapter III: Informational Value of Social Tagging Networks

## 3.1. Introduction

With the advent of social media, customers have become active content creators by expressing and sharing their feelings, thoughts, and perceptions towards brands, products, and firms. User-Generated Content (UGC) helps marketing managers to predict product sales (e.g., Liu 2006), explain firm valuation (e.g., Tirunillai and Tellis 2012), and infer competitive market structure (e.g., Netzer et al. 2012). Clearly, UGC indicates customers' brand perceptions and awareness, which constitute significant components of "customer-based brand assets" (Keller 1993). In this paper, we investigate how information contained in UGC acts as proxy measures of brand assets that track and predict the financial valuation and by doing so suggest to managers how they can proactively manage their brand assets to impact a firm's financial performance.

Prior literature suggests that information about brand assets reflected in UGC can be either directly or indirectly related to stock market performance of a firm. First of all, information in UGC can directly be related to investors' firm valuation by indicating brand equity, which investors consider as a significant signal of a brand's future performance (e.g., Tirunillai and Tellis 2012). Yet, information in UGC can also be indirectly related to firm valuation by reflecting the immediate brand performance such as sales, which can influence investors' evaluations. UGC contains abundant information regarding customers' preferences or evaluations towards a product/ brand and thus can be a significant predictor of product sales (e.g., Liu 2006). Here, we show the informational value of brand assets reflected in UGC by investigating both the direct and indirect

relationships. We further examine whether the relationships between brand assets and firm valuation vary across brands.

We infer brand associative structure from networks of a specific type of UGC, "social tags." The use of social tags to infer customer-based brand assets has two primary benefits over other forms of UGC. First, social tags provide us with brand associative structure, which contains rich semantic information compared to the volume and valence of content, the two most frequently employed metrics in extant literature. For instance, networks of social tags can reveal competitive market structure, and thus capture brand uniqueness or connectedness to competitors' associations. Second, unlike most of the UGC studied in extant literature, which is confined to a product purchase context, social tags contain a broader range of information from blog posts, news articles, and other web content about a brand and a product. While online user reviews are mostly about retrospective evaluation of brand/product performance, social tags can capture customers' perceptions or expectations of brands' future performance as well as their current performance.

We develop brand asset metrics capturing social attention, social evaluations, and competitiveness of a corporate brand by using as illustrations social tags created on a social bookmarking site, *del.icio.us*. We collect the user-generated tags of 61 firms across 16 industries (based on Standardized Industrial Classification codes) created on *del.icio.us* from January 2006 to December 2009. We find that an unexpected change in the proposed brand asset metrics derived from social tags (hereafter "social tag metrics") explains variations in unanticipated stock returns even after controlling for the firm's accounting metrics (unanticipated sales growth and unanticipated increase in return on

assets). The results suggest a differential relationship between social tag metrics and firm value across brands. Specifically, we find that an increase in social attention and connectedness to competitors is positively related to stock returns of less prominent brands while an increase in associative uniqueness and negative brand evaluations is more significantly related to stock returns of prominent brands.

Our findings contribute to extent marketing literature in three respects: First of all, we propose a conceptual model relating the social tag metrics to variations in stock returns that specifies two different routes to relate the metrics to variations in stock returns: (1) by directly reflecting the dynamics in investors' evaluations or (2) by capturing dynamics in consumer demand in the market which may influence investors' evaluations. Our research serves as a bridge between prior research showing that UGC can indicate and drive product sales (e.g., Liu, 2006) and prior research showing that the volume and valence of UGC can explain the firm valuation of stock market (e.g., Tirunillai and Tellis, 2012).

Second, we identify a moderating role of brand prominence in the relationship between social tag metrics and stock returns. Our results suggest that prominent brands benefit more from the increase in their brand uniqueness; however their losses are higher from the increase in negative associations. On the contrary, less prominent brands benefit more from being socially connected or being connected to a competitor's assets. Our results suggest that different brand asset management strategies are needed for prominent versus less prominent brands.

Lastly, we present a new way to obtain customer brand associative networks, which can be a proxy for intangible brand assets, by quantifying the information in social

tagging networks. To the best of our knowledge, our study is the first to examine the informational value of social tags in the context of firm valuation. Social tags can be a more useful resource to infer customer-based brand assets than other forms of UGC since they reveal semantic network structure of brand associations and contain broader range of information contained in blog posts, news articles, and other web content about a brand, not being confined to a retrospective product usage and purchase context .

## 3.2. Background

As demonstrated in Chapter I and II, the user-generated tags on *del.icio.us* can provide rich associative networks of users' thoughts and perceptions about various topics including brands, products, and firms. The networks of social tags have rich information about a firm's intangible brand assets.

First, social tags linked to a corporate brand can provide us with insights into the brand associations. Table 3.1 shows the classification of 7,019 sampled tags linked to 61 firms across industries and how they correspond to brand image components.[10] While most of the tags are related to some neutral descriptive information, 4.5 % of tags are related to brand attitude or brand personality and 4.9 % of tags are related to product attributes such as reliability and compatibility or non-product attributes such as price, promotion, tutorials. Social tags also contain information about how a firm is connected to product categories, sub-brands, and competitors. 6.9 % of tags are about product category associations and 12.2 % of tags are related to brand names and sub-brand names of the focal brand and other brands.

---

[10] We employed a framework from extent work on consumer knowledge typology (e.g., Brucks 1985; Cohen and Basu 1987) and extent literature on UGC (e.g. Liu 2006; Tirunillai and Tellis 2012).

[INSERT TABLE 3.1 ABOUT HERE]

Second, social tagging networks can reveal the competitive market structure. The extent to which tags are shared with other firms can reflect competitive and complementary relationships between firms. Figure 3.1 shows an example of interrelationships between three firms: Apple, Microsoft, and Google. For instance, the set of shared tags between Google and Apple - "mobile", "android", "iphone", "windows", and "apps" – reflects their rivalry in a mobile device market.

[INSERT FIGURE 3.1 ABOUT HERE]

Lastly, social tagging networks evolve dynamically, reflecting the change in social attention directed toward a firm and the change in perceptions regarding a firm over time. Figure 3.2 shows the trend in the number of bookmarks generated for six selected firms. The volume of bookmarks associated with a brand represents the extent to which users share online content related to the brand, indicating the level of social attention the brand displays in *del.icio.us*. For instance, the volume of bookmarks about Apple is smaller than that of Microsoft until 2007. Then, in 2008 and 2009 the volume of bookmarks for the two firms is almost tied, reflecting the growth in popularity and interests in Apple as compared to Microsoft, possibly due to the successful introduction of new products like iPod, iPhone, and MacBook.

[INSERT FIGURE 3.2 ABOUT HERE]


### 3.3. Conceptual Framework

In this section, we (1) present the conceptual framework of our work, (2) discuss the comparison of our paper with extant research, and (3) present the hypotheses as to how the proposed social tag metrics are related to firm value.

*Social Tags and Firm Valuation*

Previous literature shows that marketing actions such as advertising and product innovation are related to a firm's stock market performance (e.g., Joshi and Hanssens 2010; Sorescu and Spanjol 2008; Srinivasan et al. 2009). Such marketing actions can be translated to brand assets and thus stock market's firm valuation is also related to marketing metrics capturing brand perceptions, attitude, evaluations, and satisfaction. For instance, previous research finds that stock returns is related to brand attribute perceptions identified from customer survey (e.g., Mizik and Jacobson 2008), information in online user reviews (e.g., Tirunillai and Tellis 2012), information in expert reviews (e.g., Tellis and Johnson 2007), and satisfaction scores (e.g., Luo and Bhattacharya 2006). Here we propose that the associative structure of social tags provide us with a proxy measure for brand assets by reflecting the change in customers' brand schema of a focal brand as well as that of competitors.

Specifically, we propose two different paths of relationships between social tag metrics and stock returns, as presented in Figure 3.3. First, we argue that brand assets captured by social tag metrics *directly* explain variation in firm value after controlling for accounting metrics (route (1) in Figure 3.3). Prior research shows that the stock market reacts to the information about brand assets reflected in brand survey, expert review, or UGC since investors consider such information as an indicator of a brand's future performance (e.g., Mizik and Jacobson 2008; Tellis and Johnson 2007; Tirunillai and Tellis 2012). Second, we argue that social tag metrics *indirectly* explain firm value through reflecting immediate brand performance such as sales (route (2) in the Figure 6). UGC contains abundant information regarding customers' preferences or evaluations

towards a product/ a brand and thus can be a significant predictor of sales and cashflows (e.g., Godes and Mayzlin 2004; Liu 2006). Also, a firm's current-term performance such as sales or return on asset (ROA) is shown to be significant factors in firm valuation (e.g., Kothari 2001).

[INSERT FIGURE 3.3 ABOUT HERE]

### *Comparison with Extant Literature*

Exemplary prior research on the relationship between marketing metrics of brand assets and stock returns includes Tirunillai and Tellis (2012), Luo (2009), and Mizik and Jacobson (2008)[11]. Table 3.2 presents the comparison of our research with extant research:

Mizik and Jacobson (2008) investigated the relationship between brand attribute perceptions obtained from consumer survey and stock returns and found that brand relevance and energy have significant informational value in explaining firm value. The main distinction of our work from Mizik and Jacobson (2008) is that we obtain brand perceptions from an associative structure of social tags. By so doing, we obtain a more granular-level brand asset measures and richer content about brand associations not bounded to the survey questionnaires. Also, by leveraging the network structure of social tags, we can develop distinctive metrics such as brand uniqueness. Our brand uniqueness metric is measured based on the extent to which keywords (associations) of a focal brand are shared with the competing brands, while the brand "differentiation" metric in Mizik and Jacobson (2008) is obtained by directly asking subjects the extent to which a brand is differentiated.

---

[11] There is extensive literature on marketing-finance interface (e.g., Srinivasan et al. 2009; Sorescu and Spanjol 2008; Luo and Bhattacharya 2006; Tellis and Johnson 2007). Here we primarily focus on more relevant papers to our research.

Luo (2009) investigated the impact of consumer complaints filed with U.S. Department of Transportation on cash flows, stock returns, and stock volatility of firms in airline industry. He found consumer complaints have negative long-term and short-term impacts on cash flows and stock returns, and that the destructive effect of consumer complaints is stronger for the low-cost airlines than non-low-cost airlines. However, his work primarily focused on the impact by negative word of mouth and did not explore the impact of other metrics potentially derived from the rich textual content in UGC.

Tirunillai and Tellis (2012) investigated the effect of volume and valence of online product reviews on daily stock returns, trading volume, and risk to a firm. The authors found that the volume of negative reviews inversely relates to stock returns and trading volume, but is positively related to risk. However, their work did not control for the role of accounting metrics such as sales and cash flows in the relationship between UGC and stock returns. In addition, the heterogeneous relationships between UGC and stock returns across brands were not fully addressed.

In summary, this paper differs from prior research in several respects. First, we capture competitive aspects of brand using brand uniqueness and a connectedness metric inferred from the associative network structure of social tags, unlike previous studies focusing on volume and valence of UGC. Second, most of the previous literature did not note the role of sales as a mediator in the relationship between brand asset measures and firm value. We address the mediating role of sales in our model, thereby investigating the value of social tag metrics more accurately. Third, we investigate the differential relationships between social tag metrics and stock returns according to brand prominence.

[INSERT TABLE 3.2 ABOUT HERE]

*Social Tag Metrics and Hypotheses*

In this section, we present social tag metrics reflecting a firm's intangible assets and discuss how social tag metrics correspond to social attention, social evaluations, and competitiveness of a firm and how these relate to a firm's stock returns.

*Social attention.* With the advent of social media, social attention on a firm has become a significant intangible asset driving future cash flows. Marketing researchers have employed the volume of UGC as a measure of consumer interest and attention and found that dynamics in the volume of UGC can explain the future demand of products. For instance, Liu (2006) showed that the volume of online reviews of movies is positively related to the future box office revenues while the valence of the reviews does not significantly explain the revenues. Tirunillai and Tellis (2012) showed that the daily volume of user-generated reviews on a firm's products is positively related to the stock returns. We propose that the volume of social bookmarks can be a proxy for the extent of social attention on a firm, which can be a significant brand asset and thus can be related to investors' evaluation of a firm.

**H1***: The unanticipated increase (decrease) in the volume of bookmarks about a firm is positively (negatively) related to the firm's stock returns.*

Social attention on a firm can also be related to the extent to which a firm is linked to socially popular concepts or events. For instance, by being connected to fast growing keywords like "blogs" or "web2.0" a firm can indirectly obtain additional social interests. As such, we expect that an increase in the volume of fast growing keywords linked to a firm can indicate the increase in social attention on the firm, and therefore be positively related to the firm's stock returns.

57

**H2:** *The unanticipated increase (decrease) in the volume of fast growing social tags linked to a firm is positively (negatively) related to the firm's stock returns*.

*Evaluations.* Brand personalities and brand evaluations are significant brand assets to be related to investors' firm valuation. Aaker and Jacobson (2001) showed that the aggregate-level brand attitude can be a good proxy of brand assets explaining the brand's financial performance by finding that change in brand attitude is significantly related to changes in return on equity. Mizik and Jacobson (2008) showed that brand asset metrics based on five central brand attributes (differentiation, relevance, esteem, knowledge, and energy) provide incremental information content to accounting performance measures in explaining stock returns.

Social tagging networks contain rich textual information regarding brand personalities and brand evaluations, which evolve over time. Broadly, they can fall into two categories: positive associations (e.g., cool, creative, innovative, excellent, premium, etc.) and negative associations (e.g., bad, dead, dirty, hypocrisy, etc.). We posit that the dynamics in positive associations and negative associations be related to investor's expectations about firm value. Previous research also found that the valence of UGC has shown to be related to product sales and stock returns. For instance, Chevalier and Mayzlin (2006) showed that the valence of online user reviews can lead to the increase in book sales. Specifically, the authors found that the marginal (negative) impact of 1-star reviews on sales is greater than the (positive) impact of 5-star reviews. Liu (2006) investigated the impact of the textual information in online user reviews of movies by classifying online user reviews into negative vs. positive reviews yet did not find any significant effects of the valence of user reviews on the movie box office revenues. Luo

(2009) showed that negative word of mouth has negative impact on cash flows and stock returns. Tirunillai and Tellis (2012) showed that the average daily review ratings and the volume of positive messages are positively related to the stock returns. Hence, we propose that the unanticipated increase (decrease) in volume of positive tags linked to a firm be positively (negatively) related to stock returns while the unanticipated increase (decrease) volume of negative tags be negatively (positively) related to stock returns.

**H3-1:** *The unanticipated increase (decrease) in the volume of the positive social tags linked to a firm is positively (negatively) related to the firm's stock returns.*

**H3-2:** *The unanticipated increase (decrease) in the volume of the negative social tags linked to a firm is negatively (positively) related to the firm's stock returns.*

*Competitiveness.* Tagging networks can reflect the competitive market structure by revealing which firms are more strongly related to a focal firm by sharing social tags. Previous literature suggested that direct and indirect relationships between firms identified by a brand associative structure can indicate the competitive position and category representativeness of firms, and potential synergies between firms. For instance, Netzer et al. (2012) showed that the associative network structure of keywords such as brands, product attributes, or evaluations can indicate competitive market structure. Keller (1993) noted that sharing product attributes implies direct competition among brands and the extent to which brand associations are shared across firms can be indicative of market competitive structure.

First, content volume linked to a competing firm can be a proxy for social attention on competing firms which may potentially be related to the competiveness of a firm. For instance, competitors' success in marketing actions or organically arising

59

interests on competitors potentially leads to the increase in the volume of bookmarked content on competitors, which may be negatively related to investors' evaluations of a focal firm's performance. Previous research also found that the competitors' volume of UGC is negatively correlated to a focal firm's stock returns (Tirunillai and Tellis 2012). We expect that increase in social attention on competitors can be negatively related to the market evaluation of a focal firm.

**H4:** *The unanticipated increase (decrease) in competitors' volume of bookmarks is negatively (positively) related to the firm's stock returns.*

Second, brand uniqueness can be a significant source of a firm's intangible assets. Previous research shows that uniqueness of brand associations indicates sustainable competitive advantage (Keller 1993; Aaker 1982). Aaker (1982) showed that market leaders are more likely to have a "unique selling proposition" of a brand and the strong favorable associations unique to a firm are critical to a brand's success. Hence, we posit that increases in the uniqueness of associations as compared to competitors can be a firm's significant brand assets. Thus, the market evaluates the increase in uniqueness as a potential driver of future cash flows of a firm.

**H5:** *The unanticipated increase (decrease) in the uniqueness of social tags linked to a firm compared to its competitors is positively (negatively) related to the firm's stock returns.*

Third, being appropriately connected to competitor's assets can be an asset. For instance, shared associations with competitors can help to establish category membership (MacInnis and Nakamoto 1992) and define the scope of competition with other products and services (Sujan and Bettman 1989). As such, increase in connectedness with

competitors may indicate stronger category membership. In addition, a previous study showed that asset complementarities derived from the similarity based on the textual description of a firm's assets can be a measure of potential synergies between firms (Hoberg and Philips 2010). Specifically, the authors found that mergers between firms with higher asset similarity are more common and result in increased stock returns and the gains are even higher when the target is less similar to the acquirer's closest rivals. We expect that the increase in connectedness to competitors' associations can create potential synergies from the category by indicating stronger category membership, and thus is positively related to the firm's stock returns.

**H6:** *The unanticipated increase (decrease) in connectedness to competitors' social tags is positively (negatively) related to the firm's stock returns.*

However, it is possible that being strongly connected to competitors' core associations may deteriorate the focal firm's brand uniqueness and weaken the competitive position. Hence, we propose that the increase in connectedness to competitors' core associations is negatively related to the firm's stock returns.

**H7:** *The unanticipated increase (decrease) in connectedness to competitors' core social tags is negatively (positively) related to the firm's stock returns.*

*Role of brand prominence.* We allow for the possibility that the hypothesized relationships vary according to the prominence of brands. Marketing literature suggests that the reactions of prominent brands to marketing activities are systematically distinct from those of less prominent brands. For instance, researchers note double-jeopardy effects that prominent, large brands tend to gain more loyalty and attraction from customers than small brands (e.g., Ehrenberg, Goodhardt, and Berwise 1990). In addition,

prominent brands are likely to experience positive feedback on profitability of dominant market share, which signals higher product quality (e.g., Smallwood and Conlisk 1979).

First, we posit that stock market participants react more favorably to the change in social attention of less prominent brands than that of prominent brands. Less prominent brands have relatively lower level of brand awareness and "brand salience" (Ehrenberg, Barnard and Scriven 1997) while prominent brands have almost saturated level of brand awareness. We expect that the benefits of the increase in social attention can be bigger for less prominent brands since it can increase their brand familiarity and the possibility of inclusion in the consideration set by heightening the brand salience. Thus, the change in social attention can be more significantly related to investors' expectations of less prominent brands' performance.

Second, we propose that the change in valence of evaluation should be more informative for prominent brands' firm value. Stronger brands are more sensitive to the loss of established brand equity (Erdem and Swait 1998) and future sales and profits (Wernerfelt 1988). Hence, it is highly possible that stock market evaluates the change in the valence of brand evaluations of prominent brands as more pertinent indicators of their firm value than that of less prominent brands.

Third, we posit that stock market evaluates the trade-off between brand uniqueness and connectedness to competitors' associations across brands using differing criteria. For a prominent brand, the increase in brand uniqueness at the expense of connectedness to competitors can be positively related to their stock returns since their future sales can be driven by differentiation strategy (Ehrenberg, Goodhardt, and Berwise 1990). In contrast, for less prominent brands, the increase in brand uniqueness can result

in the brand positioned in a niche market, givinge rise to a smaller market share (Ehrenberg, Goodhardt, and Berwise 1990). Hence, we expect that for less prominent brands, the increase in the connectedness to competitors' associations at the expense of brand uniqueness should be positively related to their stock returns since they benefit through shared associations with prominent brands.

### 3.4. Research Design

*Data Collection*

We employ the following procedures for collecting social tags from *del.icio.us*. First, we selected firms which (1) are US-based, (2) serve in consumer goods industries (e.g., retail, consumer electronics, internet service companies), (3) had annual sales higher than $10 million in fiscal year of 2009, (4) contain at least 10 quarterly data points from the 1st quarter of 2006 to the 4th quarter of 2009, and (5) have more than 1,000 social bookmarks on del.icio.us from the 1st quarter of 2006 to the 4th quarter of 2009. As a corporate brand name may have other connotations, we investigated the content in the sample of 100 bookmarks linked to each corporate brand and excluded corporate brand names if more than 5% of bookmarks were considered unrelated. As examples, ambiguous brands such as Dominos and Blockbuster were excluded. 5,681,741 bookmarks generated for 61 firms in 16 product categories remained for the analysis. The mean volume of bookmarks generated for each firm was 113,634.8 with the standard deviation of 322,837.1. Second, we collected 2,000 recent bookmarks for each firm and obtained all of the social tags linked to the bookmarks. We constructed a dictionary of 60,377 tags and from these selected 7,019 key tags having more than five bookmarks

linked to selected firms. Lastly, we obtained the historical monthly trends of bookmarks

tagged with each key tag and each firm using *del.icio.us* search algorithms. We excluded

the bookmark data before 2006 since *del.icio.us* started in 2003 and thus the data may not

have been reliable until 2006.

Firm financial performance, including quarterly sales and stock returns were

obtained from CRSP and COMPUSTAT databases. We matched the financial data from

CRSP and COMPUSTAT to tagging data based on a calendar time window.

### *Bi-partite Networks of Firms and Social Tags*

To take advantage of rich associative relationships in tagging networks, we

employ network representation of tags[12]. Specifically, we borrow the framework of

affiliation networks (Faust 1997) to present the bi-partite networks of firms and social

tags. We denote the set of firms as $F = \{f_1, f_2, \ldots, f_{NF}\}$ and the set of tags as $K = \{k_1, k_2, \ldots, k_{NT}\}$. *NF* is the number of firms and *NT* is the number of tags in this network.

The affiliation network matrix for time window $t$, $A_t$, is consist of element, $a_{ijt}$ as

defined by Equation 1a:

(1a) $$a_{ijt} = N_t(F_i, K_j)$$

where, $N_t(F_i, K_j)$ is the number of bookmarks linked to both firm *i* and keyword *j* during

time window *t*, which captures how frequently a firm is shown together with a keyword.

Alternatively, the affiliation network can be constructed using the following

*cosine* distance measure, $a'_{ijt}$, which were shown to effectively capture the associative

relationship between tags (Robu, Halpin, and Shepherd 2009):

---

[12] Our approach is in line with Keller (1993)'s idea that brand image can be represented with an associative
network of a brand node and a variety of associations linked to a brand node.

$$(1b) \qquad a'_{ijt} = \frac{N_t(F_i, K_j)}{\sqrt{N_t(F_i)N_t(K_j)}}$$

where, $N_t(F_i)$ is the number of bookmarks linked to firm $i$ and $N_t(K_j)$ is the number of

bookmarks linked to keyword $j$ during time window $t$.

*Measures*

Table 3.3 shows the measures we employed in our analysis.

[INSERT TABLE 3.3 ABOUT HERE]

*Social tag metrics.* The level of social attentions on a firm is measured by two

metrics: (1) log of volume of bookmarks linked to a firm ($VOLBK_{it}$) and (2) log of

volume of fastest growing social tags linked to the firm ($VOLSOC_{it}$). For the second

metric, we identified as fastest growing the tags ranked in the top 10% in the bookmark

volume growth.. This metric captures the indirect social attention on a firm through being

connected to these growing tags.

Valence of brand evaluations is measured by the volume of negative tags ($POS_{it}$)

and positive tags scaled by the volume of all tags linked to a firm ($NEG_{it}$) scaled by

volume of all tags as specified by Equation 2. These metrics capture the fraction of

positive (negative) tags among all tags linked to a brand. To obtain this measure, we

manually classified 7,019 tags into three categories: positive descriptions, negative

descriptions, and neutral descriptions. Three raters participated in this classification

process. Fleiss' Kappa index (Fleiss, 1971) for the reliability across three raters was .904

($z = 44.7$, $p < .001$), indicating a reasonable level of agreement. For keywords which

raters disagreed on, we took the majority of opinion.

$$\text{(2)} \qquad POS_{it} = \frac{\sum_{j \in POSTAG} a_{ijt}}{\sum_{j=1}^{NT} a_{ijt}}, \qquad NEG_{it} = \frac{\sum_{j \in NEGTAG} a_{ijt}}{\sum_{j=1}^{NT} a_{ijt}}$$

where, *POSTAG* is the set of pre-defined positive tags and *NEGTAG* is the set of pre-defined negative tags.

Competitiveness is measured by several metrics: volume of the content connected to competitors, uniqueness of associations, the level of connectedness to competitors, and the level of connectedness to competitors' core assets. To obtain these metrics, we defined the set of competitors as the set of firms with the same four-digit SIC code[13]. For firms without the same four-digit SIC code, we used the first two digits of the SIC code. The volume of the content connected to competitors ($VOLCOM_{it}$) is defined as log of sum of the volume of bookmarks of competitors.

Uniqueness of associations of a firm as compared to its competitors is measured in two different ways. The first uniqueness metric ($UNIQUE_{it}$) is the volume of unique tags not shared with competitors scaled by the volume of all tags linked to a firm as specified in Equation 3a.

$$\text{(3a)} \qquad UNIQUE_{it} = \frac{\sum_{j=1}^{NT} a_{ijt} \prod_{i' \in CF_i} I(a_{ijt})(1 - I(a_{i'jt}))}{\sum_{j=1}^{NT} a_{ijt}}$$

where, $CF_i$ is the set of competitors for firm *i* identified by SIC code; $I(a_{ijt})$=1 if $a_{ijt}$>0, 0 otherwise.

The second metric ($SHARE_{it}$) is obtained by taking the mean of the volume share of a firm in each association as compared to competitors ($Share_{ijt}$) as specified as

---

[13] We identified the set of competitors identified with shared tags. The competitive metrics derived from this method were highly correlated to the metrics based on the competitor set identified by SIC code (Pearson's correlation coefficient ranged from .63 to .88).

Equation 3b. [14] This metric captures the extent to which a firm occupies the shared brand associations as compared to its competitors.

(3b) $$Share_{ijt} = \frac{a_{ijt}}{a_{ijt} + \sum_{i' \in CF_i} a_{i'jt}} , \quad SHARE_{it} = \frac{\sum_{j' \in STAG_{it}} Share_{ij't}}{NST_{it}}$$

where $STAG_{it}$ is the set of shared tags of firm $i$ with its competitors during time window $t$; $NST_{it}$ is the number of tags in the set, $STAG_{it}$.

We obtained connectedness of a firm with two different metrics: the first metric, connectedness to all assets of competitors ($CONNECT1_{it}$) is defined as the mean number of competitors linked to each association as presented in Equation 4a.

(4a) $$CONNECT1_{it} = \frac{\sum_{j=1}^{NT} \sum_{i' \in CF_i} I(a_{ijt}) \, I(a_{i'jt})}{\sum_{j=1}^{NT} I(a_{ijt})}$$

The second metric, connectedness to core assets of competitors ($CONNECT2_{it}$) is defined as the weighted average of number of competitors linked to each association where the weight is the associative strength between a tag and each competitor, and which is captured by the cosine distance metric, $a'_{i'jt}$. This determines the extent to which the focal brand is connected to competitors' core associations.

(4b) $$CONNECT2_{it} = \frac{\sum_{j=1}^{NT} \sum_{i' \in CF_i} I(a_{ijt}) \, a'_{i'jt}}{\sum_{j=1}^{NT} I(a_{ijt})}$$

***Stock returns.*** We employed three different approaches to measure stock returns: stock returns, Benchmark-Adjusted Buy-Hold-Abnormal-Returns, and abnormal returns based on Fama-French 3 Factor model.

The first metric, stock returns ($STKRET_{it}$) is calculated as follows:

---

[14] We considered obtaining uniqueness in subcategories of associations: positive associations, negative associations, category related associations, and brand related associations. All of these measures were highly correlated to each other (Pearson's correlation coefficient ranged from .57 to .92), We decided to calculate uniqueness across all the associations.

$$\text{(5a)} \qquad\qquad STKRET_{it} = \frac{PRICE_{it} - PRICE_{it-1} + Div_{it}}{PRICE_{it-1}}$$

where $PRICE_{it}$ is closing stock price of firm $i$ at $t$; $Div_{it}$ is dividend issued of firm $i$ at $t$.

The second metric is Benchmark-Adjusted Buy-Hold-Abnormal-Returns (BHAR). BHAR is obtained by comparing the actual return of firm over a given time window to a benchmark return (Barber and Lyon 1997). The benchmark return consists of a portfolio of stocks that belong to the same size, book-to-market, and momentum quintiles as the firm (Daniel et al. 1997; Wermers 2003).[15] Quarterly BHAR is computed as:

$$\text{(5b)} \qquad\qquad BHAR_{it} = \prod_{m=1}^{3}(I + R_{imt}) - \prod_{m=1}^{3}(I + R_{j(imt)})$$

where $t$ is the quarter of interest, $m$ is the calendar month within each quarter, $R_{imt}$ is the rate of return of firm $i$ in month $m$ of quarter $t$, and $R_{j(imt)}$ is the return of the benchmark portfolio $j$.

The third metric ($FFRET_{it}$) is abnormal returns based on Fama-French 3-Factor model. The Fama-French 3-Factor model posit that abnormal returns are a function of the overall market return, the difference between returns of small-firm and big-firm stocks and, the difference between returns of high and low book-to-market stocks (Fama and French 1993) as specified below: [16]

$$\text{(5c)} \qquad\qquad R_{it} - R_{ft} = \alpha_i + \beta_i\left(R_{mt} - R_{ft}\right) + s_i SMB_t + h_i HML_t + \varepsilon_{it}$$

where, $R_{it}$ is firm $i$'s stock market return at quarter $t$; $R_{ft}$ is the risk free rate at quarter $t$; $R_{mt}$ is market return; $SMB_t$ is the difference between returns of small-firm and big-firm

---

[15] Complete details regarding the construction of these control portfolios are available in the appendix of Daniel et al. (1997). Benchmark data is available at http://www.smith.umd.edu/faculty/rwermers/ftpsite/Dgtw/coverpage.htm.

[16] Information regarding these benchmarks as well as the actual data used for this analysis is available on Kenneth French's website at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

stocks; $HML_t$ is the difference between returns of high and low book-to-market stocks;

abnormal return for a firm is calculated as $(R_{it} - R_{ft}) - (\hat{R}_{it} - R_{ft})$ .

***Brand prominence.*** To explain the different response to social tag metrics across

brands, we classified 61 brands into two categories: prominent brands vs. less prominent

brands. First, we employed external brand equity rankings in Best Global Brands 2006-

2009 (Interbrand 2009). We classified brands that appeared in the top 100 from 2006 to

2009 as prominent brands. 14 brands in our sample are classified as prominent brands.

Second, we classified the brands based upon the average volume of bookmarks. We also

considered brands with more than 10 average daily bookmarks (upper 35%) as prominent

brands. All the prominent brands identified in the first method were classified as

prominent brands by the second method.

## 3.5. Model Formulation

***Stock response model.*** Our objective is to test whether social tag metrics defined

in the previous section have additive informational value to accounting performance

measures in explaining stock returns. Towards this objective, we employ the stock return

response modeling approach. Stock response modeling is an analytical tool to evaluate

whether information contained in a metric is associated with the changes in stock returns

(Mizik and Jacobson 2004). The underlying assumption of stock return response

modeling is the financial market efficiency that the price of a stock reflects all available

information related to the profitability of the firm (LeRoy 1989). We believe that for our

quarterly data, stock market efficiency holds[17]. Note that social tag metrics can be constructed in a more granular time window. The primary reasons why we employ quarterly data are because: (1) our focus is on more persistent dynamics in brand assets rather than transient change captured by daily or weekly trend and (2) we plan to investigate the mediating role of sales (available on a quarterly basis) on the relationship between social tag metrics and stock returns.

Following the framework proposed by Mizik and Jacobson (2004) we model the abnormal returns as a function of unanticipated change of accounting performance and unanticipated change of social tag metrics between t and t-1 as presented in Equation 6a. We include control variables such as book value to market value ratio, market value and calendar year dummies. These variables stem from extant literature and capture the impacts of economy-wide factors and firm-specific characteristics on the stock returns (Fama and French 1993; Fama 1998).

$$StkRet_{it} = Eret_{it} + Uret_{it}$$

(6a)
$$StkRet_{it} - Eret_{it} = \sum_{j=1}^{J} \gamma_j U\Delta Acc_{jit} + \sum_{k=1}^{K} \beta_k U\Delta SocTag_{kit} + \sum_{l=1}^{L} \rho_{il} Control_{ilt} + \varepsilon_{it}$$

where, $Eret_{it}$ is the expected rate of return for firm $i$ at time $t$; $Uret_{it}$ is the unexpected rate of return for firm $i$ at time $t$; $U\Delta Acc_{jit}$ is the unanticipated change in accounting variable $j$ between $t$ and $t-1$ and $U\Delta SocTag_{kit}$ is the unanticipated change in social tag metric $k$ between $t$ and $t-1$; $Control_{ilt}$ is control variable $l$ at $t$.

To test differential reaction to change in social tag metrics according to brand prominence, we modify Equation 6a as follows:

---

[17] Although there are some empirical evidences that stock market is often not efficient and needs time to incorporate the available market information (e.g., Luo 2009; Tirunillai and Tellis 2011), this market anomaly is found to be resolved in several days or weeks for most of the cases (Mizik and Jacobson 2009).

$$SRET_{it} = \sum_{j=1}^{J} \gamma_j U\Delta Acc_{jit} + \sum_{k=1}^{K} (\beta_{1k} U\Delta SocTag_{kit} + \beta_{2k} U\Delta SocTag_{kit} \times Prom_i)$$

(6b)

$$+ \sum_{l=1}^{L} \rho_{il} Control_{ilt} + \varepsilon_{it}$$

where $RET_{it}$ is stock market return; $Prom_i$ is an indicator which is equal to 1 if corporate brand $i$ is prominent, 0 otherwise.

**_Unanticipated change in metrics._** The unanticipated change of social tag metrics are obtained by regressing *t-1* levels of each social tag metric on *t* quarterly level of each variable with other control variables (see Equation 7). Dummy variables accounting for calendar year are included to control for the growth of the number of users on *del.icio.us* on the network evolution and other potential macroeconomic effects. The residual associated with each equation is the measure of unanticipated change of each social tag metric ($U\Delta SocTag_{ijt}$).

(7a) $$SocTag_{ijt} = \beta_0 + \beta_1 SocTag_{ijt-1} + \beta_{2-4} Year + \varepsilon_{it}$$

where, $SocTag_{ijt}$ is *j*th social tag metric of is firm *i* at time *t*; $Year$ is calendar year dummy.

Likewise, we obtain the measure of the unanticipated change of accounting metric (sales and ROA) by regressing t-1 levels of each social tag metric on t quarterly level of each variable.

(7b) $$ACC_{ijt} = \beta_0 + \beta_1 ACC_{ijt-1} + \beta_{2-4} Year + \varepsilon_{it}$$

where, $ACC_{ijt}$ is accounting metric of firm *i* at time *t*.

The way we obtain unanticipated measures is consistent with the methods implemented in extant marketing literature (e.g., Sorescu and Spanjol 2008; Luo 2009).

Note that endogeneity is not a concern here since we focus on the *unanticipated* change of brand assets and accounting performance (Sorescu and Spanjol 2008).

### 3.6. Results

*Descriptive Statistics*

Table 3.4 presents descriptive statistics of the measures. There is substantial variation in stock return measures, accounting metrics, as well as social tag metrics. On average, the log of volume of bookmarks published each quarter on each firm is 6.79 (SD = 1.74) and the log of the number of socially growing tags linked to each firm each quarter is.88 (SD =.16). On average, 3.3% of tags linked to a firm are positively valenced (SD =.54%), 2.1% are negatively valenced (SD = .80%). On average, the volume of unique tags takes 9.9 % of total tag volume linked to a firm (SD = 14.4%) and for shared associations each firm takes 38.9% of association share (SD= 21.7%). The average number of competitors connected to each tag linked to a firm is 2.7 (SD= 2.19) and the weighted average number of competitors connected to each tag linked to a firm is 2.98 (SD=2.42).

[INSERT TABLE 3.4 ABOUT HERE]

Table 3.5 presents the correlation of the measures. Stock return measures are significantly positively correlated to the unanticipated increase in ROA, the unanticipated sales growth, and the unanticipated increase in the connectedness to all assets of competitors, however negatively correlated to unanticipated increase in the connectedness to core assets of competitors.

[INSERT TABLE 3.5 ABOUT HERE]

*Informational Value of Social Tag Metrics*

       *Stock response model.* Table 3.6a presents the results from regression analysis of

the stock response model specified in Equation 3.6b. The results show that social tag

metrics capturing social attention, competitiveness, and valence of evaluations are

significantly related to stock returns. Since the results are consistent with three different

measures of stock returns, we focused on discussing the case of abnormal returns based

on Fama-French 3-Factor Model. As we expected, stock market response to social tag

metrics varies according to brand prominence. Specifically, the increase in social

attention captured by volume of socially popular keywords is positively related to stock

returns of less prominent brands ($\beta$=.21, p <.01). However, the added effect for those of

prominent brands is significantly negative ($\beta$=-.16, p <.05), indicating that stock market

reacts more favorably to the change in social attention of less prominent brands. In

addition, increase in uniqueness captured by association share is negatively related to

stock returns of less prominent brands ($\beta$=-3.08, p <.05), yet the added effect for those of

prominent brands is significantly positive ($\beta$=3.38, p <.05). Likewise, an increase in

connectedness to all assets of competitors is positively related to stock returns of less

prominent brands ($\beta$=3.40, p <.001), the added effect for those of prominent brands is

significantly negative ($\beta$=-5.37, p <.05). Yet, when an increase in connectedness to core

assets of competitors is negatively related to stock returns of less prominent brands ($\beta$=-

1.82, p <.001), the added effect for those of prominent brands is not significant but

negative ($\beta$=1.17, p >.10). Interestingly, an increase in negative tag volume is positively

related to stock returns of less prominent brands ($\beta$=8.86, p < .10), such that the added

effect for those of prominent brands is significantly negative ($\beta$=-15.94, p <.05).

[INSERT TABLE 3.6a ABOUT HERE]

Since the results in Table 3.6a suggest that market response to social tag metrics varies across firms, we conducted Bayesian analysis to allow for heterogeneous response to social tag metrics. We estimated the proposed hierarchical stock response model with the WinBUGS software package (Spiegelhalter, Thomas, and Best 2000). We ran 50,000 draws, thinning the chains by taking every fifth draw we took 10,000 draws after 50,000 burn-in for the estimation results.[18] Table 3.6b presents the results from Bayesian analysis of stock response model. Since it is not appropriate to discuss the significance of the Bayesian estimates, based on the proposed hypotheses, we constructed one-sided posterior probability intervals for each estimate. We find that stock response to increase in uniqueness (*U∆UNIQUE*), connectedness to competitors (*U∆CONNECTED1*) is consistently positive, while that to increase in connectedness to competitors' core assets (*U∆CONNECTED2*) is consistently negative across brands. Stock response to all other social tag metrics varies across firms.

[INSERT TABLE 3.6b ABOUT HERE]

*Mediating Role of Sales*

To test whether accounting metrics mediate the relationship between social tag metrics and stock returns, we conducted Bayesian mediation analysis (e.g., Zhang, Wedel, and Pieters 2009; Yuan and MacKinnon 2009). Employing Yuan and MacKinnon's (2009) approach, we specify the mediation model as follows:

---

[18] We tested the convergence of the chain with Geweke convergence test (Geweke 1992). We compared the means of the Markov chains obtained from the first 3,000 draws with the means of the results obtained from the last 3,000 draws. We did not find a significant difference in the equality of the means of the early chain and those of the late chain, indicating that the convergence of the chain is at an acceptable level.

$$U\Delta Sale_{it} = \alpha_{1i} + \theta_{1t} + \sum_{k=1}^{K} \beta_{1k} U\Delta SocTag_{kit} + \varepsilon_{1it},$$

(8)

$$RET_{it} = \alpha_{2i} + \theta_{2t} + \beta_M U\Delta Sale_{it} + \sum_{k=1}^{K} \beta_{2k} U\Delta SocTag_{kit} + \varepsilon_{2it},$$

where, $\varepsilon_{1it}$ and $\varepsilon_{2it}$ are residuals of the mediator, $U\Delta Sale_{it}$ and the stock returns and assumed to be independent and follow normal distributions; $\alpha_{1i}$ and $\alpha_{2i}$ are random firm-specific intercepts and $\theta_{1t}$ and $\theta_{2t}$ are random quarter-specific intercepts; $\beta_{1k}$ measures the relationship between social tag metric k and mediator and $\beta_{2k}$ measures the direct relationship between social tag metric k and stock return; $\beta_M$ measures the relationship between mediator and the stock return; $\beta = (\beta_M, \beta_{1k}, \beta_{2k})^T$ follows a multivariate normal distribution.

The average indirect effect of social tag metric k is calculated as follows:

(9) $$INDIRECT_k = E(\beta_M \beta_{1k}) = M(\beta_M)M(\beta_{1k}) + \sigma_{\beta_M \beta_{1k}},$$

where, $M(\beta_M)$ is posterior mean of $\beta_M$ and $M(\beta_{1k})$ is posterior mean of $\beta_{1k}$ and $\sigma_{\beta_M \beta_{1k}}$ is covariance between $\beta_M$ and $\beta_{1k}$.

Table 3.7 presents the results from the mediation analysis. We found that sales growth is a strong mediator of the informational value of negative evaluation (*U∆NEG*) and none of the social tag metrics consistently explain unanticipated sale growth. The mean mediating effect size relative to direct effect is 23% (95% confidence interval: [12.7%, 35.8%]). Other social tag metrics, such as uniqueness (*U∆UNIQUE*), connectedness to competitors (*U∆CONNECTED1*), and connectedness to competitors' core assets (*U∆CONNECTED2*) are directly related to stock returns. Hence, we can conclude that most of our social tag metrics have strong informational value in explaining

stock returns even after controlling for the potential mediation role of accounting metrics. The result suggests that social tag metrics possibly capture information about brand assets which may not be reflected in the current sales yet can explain investors' future-term expectations. For instance, new product preannouncements or advertising campaign may update the market expectations of a firm's future-term prospect while not being fully reflected in the current-term sales.

[INSERT TABLE 3.7 ABOUT HERE]

## 3.7. Discussion

### *Informational Value of Social Tag Metrics*

*Social attention.* We find that an increase in connectedness to socially growing keywords can be viewed as a positive signal of a firm's future cash flows while dynamics in volume of bookmarks does not contain significant information to investors. It is interesting to note that for less prominent brands, the increase in social attention is considered a more significant factor for firm valuation than for prominent brands. The result indicates that building up social attention by being connected to socially popular concepts and events can be positively related to future prospects of less prominent firms but is not the case for prominent brands. This is possibly because: (1) an increase in social tags may not fully capture changes in social attention on prominent brands or (2) an increase in social attention on them may not be considered an unanticipated shock to investors since they are already well-established brands and enjoy an almost saturated level of social attention.

***Evaluations.*** We find that the stock market reacts to the change in negative evaluation while change in positive evaluation is not a significant factor to explain stock returns. More specifically, the results indicate that investors interpret a shock in negative evaluations as incremental information for prominent brands' future-term prospects, but not so for less prominent brands. This is consistent with our expectation that for prominent brands, loss of their established brand assets may signal a more significant loss of their future cash flows than less prominent brands.

***Competitiveness.*** We find that an increase in association share (uniqueness) can be viewed as a positive signal for future cash flow for prominent brands while not the case for less prominent brands. This indicates that an increase in association share can only be related when a brand has well established assets. Another finding is that an increase in connectedness to competitors can be viewed as a positive sign for future-term prospects of less prominent brands by indicating stronger category membership while a negative signal for prominent brands by indicating lower level of brand uniqueness. However, increase in connectedness to competitors' core associations can be negatively related to stock returns but is not significantly related to investors' evaluation of prominent brands.

### *Managerial Implications*

Our findings provide marketing practitioners with insights on the kinds of marketing activities and communication strategies beneficial for developing brand assets. We suggest differential brand asset management strategy according to the brand prominence. For less prominent brands, it is more critical to focus on expanding the set of brand associations by being connected to social events or market leaders. Marketing managers of less prominent brands should invest in:  (1) boosting brand visibility by

creating content related to socially popular events, trends, or concepts and (2) creating more content which can be linked to their competitors, making the brands more assimilated and comparable to the competitors and yielding higher category membership. For instance, when promoting new products, marketers may take advantage of current fads and trends, or consider assimilation strategy to steal brand associations from the extant market leaders.

Managers of prominent brands on the contrary should be more selective in promoting their marketing activities and creating brand-related content. Since prominent brands already have a well-refined set of healthy brand associations, the manager's goal should be to create content that will expand and bolster the current assets. Managers can consider creating more content by focusing on their unique brand position and creating a distinctive image that prevent their competitors from appearing comparable. In addition, maintaining current brand assets by managing negative associations appropriately is crucial for prominent brands. For prominent brands, creating unique, new associations by developing innovative products and delivering a unique, differentiated communication message, while maintaining their current good image, can strengthen their brand assets.

We believe that the proposed social tag metrics provide marketing practitioners with a solution to track and mine a real-time measure of brand assets from UGC. The method we propose here can be easily extended to the context of online user reviews, Twitter tweets, YouTube videos, and blog posts although we conduct the analysis based on social tagging data of user-generated bookmarks. We believe that our new tools and measures can be a good complement of annual brand surveys by allowing marketers to track the dynamics of brand assets.

*Contributions, Limitations and Future Research*

Our work contributes to the extant marketing literature in three respects: First of all, we conceptualize how the social tag metrics can explain the variations in stock returns by specifying two different routes: (1) by directly being related to dynamics in investors' expectations of brand future-term performance or (2) by indirectly being related to dynamics in consumer demand in the market, which in turn can influence investors' expectations. We find that except for the negative evaluations the informative value of social tag metrics is not mediated by sales growth, indicating that social tag metrics possibly capture information about brand assets which may not be reflected in the current sales yet explain investors' future-term expectations. Our findings serve as a bridge between extent literature showing that UGC can indicate and drive product sales (e.g., Liu, 2006) and prior literature showing that the volume and valence of UGC can explain the firm valuation (e.g., Tirunillai and Tellis, 2012).

Second, our work contributes to extant research streams in brand equity by showing the moderating role of brand prominence in the relationship between social tag metrics and stock returns. The results suggest that for prominent brands, an increase in brand uniqueness at the expense of connectedness to competitors' associations is deemed as a promising signal for their future-term cash flow; however their losses are higher from the increase in negative associations. On the contrary, for less prominent brands heightened brand visibility by being socially connected and being connected to a competitor's associations are considered as a positive sign for their future cash flow. The findings suggest that different brand asset management strategies are needed for prominent versus less prominent brands.

Third, the proposed social tag metrics present a new way to track intangible brand assets using networks of user-generated keywords. To the best of our knowledge, our work is the first to quantify the information contained in social tags and investigate their informational value in the context of firm valuation. We show that social tags can be a more appropriate source to infer brand associative networks than other forms of User-Generated Content (UGC) since they provide us with the semantic network structure of keywords, which allows us to construct metrics capturing brand uniqueness and connectedness to competitors beyond the volume and valence of content. In addition, social tags are not confined to a product purchase context and thus can provide a more integrative picture of brand associations.

This paper has several limitations, which invite further research. Our analysis is based upon a quarterly time frame since (1) accounting measures are available on a quarterly time frame and (2) we judge that it is not easy to observe systematic brand associative structure change on a more granular time window (such as a weekly or monthly level). However, it will be interesting to construct metrics on a more granular time frame and investigate the explanatory power of social tag metrics in daily or weekly-level stock returns. In addition, our paper does not directly include a firm's specific marketing activities such as change in advertising expenditures and communication message, new product announcements and innovations into our model. Rather, we capture customers' integrative perceptions of those marketing activities reflected in the tagging structure of each brand as the impact of those activities are reflected in the tags. It will be interesting to investigate the chain of marketing activities, customer perceptions/ reactions captured by social tags, sales, and firm value. Lastly, the meanings of brand

associations in tagging data are often ambiguous. A firm name such as "Blockbuster" can be used as a brand name, or a general descriptor. We excluded those firms from our analysis since the results can be misleading. One possible way to resolve this problem in the future is to consider inter-tag relationships and classify only relevant tags. Future research relying on computational linguistics techniques to resolve such ambiguity in the data will be highly valuable.

**Table 3.1: Types of User-Generated Tags Linked to Firms**

| Types of Associations | Example | Volume | % |
|---|---|---|---|
| attitude/evaluation | cool, humorous, innovative, creative | 317 | 4.5% |
| product attributes | accessibility, reliability, stability | 90 | 1.2% |
| non-product attributes | price, promotion, tutorial, service | 277 | 3.9% |
| product category | mp3, television, toys, hotel | 485 | 6.9% |
| brands | Apple, Microsoft, ipod, zune, Google | 857 | 12.2% |
| unique place or name | Michigan, China, Benjamin, Ann | 1,633 | 23.2% |
| descriptive words | article, behavior, business, bus | 2,920 | 41.6% |
| other | for, abc, and | 440 | 6.2% |

**Table 3.2: Comparison with Extant Literature**

| | Our paper | Tirunillai and Tellis (2011) | Luo (2009) | Mizik and Jacobson (2008) |
|---|---|---|---|---|
| Marketing Metric | Metrics derived from social tags | Metrics derived from product reviews | Metrics derived from consumer complaints filed with U.S. Dept. of Transportation | Brand attribute perception from consumer survey |
| Dependent Metric | Stock return | Stock return Risk Trading volume | Cashflow Stock return Stock volatility | Stock return |
| Model | Stock response model | VAR | VAR | Stock response model |
| Time Window | Quarterly | Daily | Monthly | 4-14 quarters |
| Sample | 61 firms in 16 markets | 16 firms in 6 markets | 10 firms in airline industry | 275 firms |
| User-Generated Content | Yes | Yes | Yes | No |
| - Content volume | Yes | Yes | No | (five primary brand attributes: |
| - Content valence | Yes | Yes | Yes (Negativity only) | brand relevance, energy, |
| - Content uniqueness | Yes | No | No | esteem, knowledge, |
| - Content connectedness | Yes | No | No | differentiation) |
| Mediation of sales | Yes | No | Yes | No |
| Brand-level Heterogeneity | Yes | No | Yes | No |

**Table 3.3: Measures**

| Conceptual Variable | Notation | Measured Variable | Data Source | |
|---|---|---|---|---|
| Independent Variables | | | | |
| Social Attention | $VOLBK_{it}$ | Log of volume of bookmarks linked to firm $i$ at time $t$ | del.icio.us | H1 |
| | $VOLSOC_{it}$ | Log of volume of fast growing social tags linked to firm $i$ at time $t$ | del.icio.us | H2 |
| Valence of Evaluations | $POS_{it}$ | Volume of positive tags linked to firm $i$ at $t$ scaled by volume of all | del.icio.us | H3-1 |
| | $NEG_{it}$ | tags Volume of negative tags linked to firm $i$ at $t$ scaled by volume of all tags | del.icio.us | H3-2 |
| Competitiveness | $VOLCOM_{it}$ | Volume of bookmarks linked to firm $i$'s competitors at $t$ | del.icio.us, SIC | H4 |
| | $UNIQUE_{it}$ | Volume of tags unique to firm $i$ scaled by volume of all tags at $t$ | del.icio.us, SIC | H5 |
| | $SHARE_{it}$ | Mean volume share in each tag compared to competitors at $t$ | del.icio.us, SIC | H5 |
| | $CONNECT1_{it}$ | Mean number of competitors linked to tags of firm $i$ at $t$ | del.icio.us, SIC | H6 |
| | $CONNECT2_{it}$ | Weighted average number of competitors linked to each tag at $t$ (weight is associative strength between a tag and each competitor) | del.icio.us, SIC | H7 |
| Dependent Variables | | | | |
| Stock Returns | $BHAR_{it}$ | Benchmark-adjusted buy-and-hold abnormal return | CRSP, Russ Wermer's Website | |
| | $FFRET_{it}$ | Fama-French 3 Factor abnormal return | CRSP | |
| | $STKRET_{it}$ | Stock market return | CRSP | |
| Control Variables | | | | |
| Firm Characteristics | $BkMk_{it}$ | Log of book value to market value ratio of firm $i$ at time $t$ | COMPUSTAT | |
| | $MktValue_{it}$ | Log of market value of firm $i$ at time $t$ | COMPUSTAT, CRSP | |
| Firm Performance | $Sale_{it}$ | Sales of firm $i$ at time $t$ scaled by total assets of firm $i$ at time $t$ | COMPUSTAT | |
| | $ROA_{it}$ | Return on assets of firm $i$ at time $t$ | COMPUSTAT | |

**Table 3.4: Descriptive Statistics**

| Variable | Mean[a] | SD[a] | Median[b] | Minimum[b] | Maximum[b] |
|---|---|---|---|---|---|
| $VOLBK_{it}$ | 6.795 | 1.743 | 3.466 | 6.465 | 12.234 |
| $VOLSOC_{it}$ | .887 | .157 | .622 | .844 | 1.319 |
| $POS_{it}$ | .033 | .005 | .019 | .032 | .054 |
| $NEG_{it}$ | .021 | .008 | .006 | .019 | .052 |
| $VOLCOM_{it}$ | 8.361 | 2.300 | .000 | 7.826 | 12.702 |
| $UNIQUE_{it}$ | .099 | .144 | .004 | .057 | 1.000 |
| $SHARE_{it}$ | .389 | .217 | .030 | .393 | 1.000 |
| $CONNECT1_{it}$ | 2.703 | 2.189 | .000 | 1.733 | 8.040 |
| $CONNECT2_{it}$ | 2.985 | 2.419 | .000 | 2.021 | 9.255 |
| $MktValue_{it}$ | 9.558 | 1.445 | 3.272 | 9.606 | 12.713 |
| $BkMk_{it}$ | -.319 | .855 | -2.176 | -.374 | 4.112 |
| $Sale_{it}$ | .297 | .184 | .032 | .248 | 1.052 |
| $ROA_{it}$ | .000 | .016 | -.114 | .000 | .141 |
| $FFRET_{it}$ | .032 | .330 | -.938 | .019 | 5.664 |
| $BHAR_{it}$ | .022 | .256 | -.571 | .006 | 5.083 |
| $STKRET_{it}$ | .027 | .301 | -.870 | .014 | 5.205 |

a Mean and standard deviation of quarterly, firm-level values

b based on all quarterly, firm-level values

**Table 3.5: Correlation of Measures**

| | Mean | SD | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 UΔVOLSOC | -.001 | .374 | .451 | .171 | -.385 | .277 | -.056 | .202 | .132 | -.108 | .077 | -.039 | .027 | -.015 | .073 | .046 | -.022 |
| 2 UΔVOLBK | -.002 | .227 | | .116 | -.459 | .363 | -.052 | .120 | .061 | -.014 | .045 | -.064 | -.014 | .034 | -.049 | -.050 | -.047 |
| 3 UΔVOLCOM | -.002 | .180 | | | -.019 | -.321 | .000 | .282 | -.001 | -.051 | -.026 | -.004 | .008 | .027 | -.009 | -.026 | -.024 |
| 4 UΔUNIQUE | -6.43E-05 | .010 | | | | -.295 | -.123 | .010 | -.165 | .114 | .034 | -.012 | -.051 | .003 | -.029 | -.045 | -.012 |
| 5 UΔSHARE | -6.65E-05 | .017 | | | | | -.135 | -.088 | .075 | .009 | .057 | -.038 | .051 | -.011 | -.053 | -.041 | -.047 |
| 6 UΔCONNECT1 | 1.29E-04 | .014 | | | | | | -.343 | .105 | .069 | -.071 | .073 | .038 | .035 | .220 | .232 | .245 |
| 7 UΔCONNECT2 | -1.11E-04 | .036 | | | | | | | -.023 | -.089 | -.014 | -.024 | -.036 | .005 | -.205 | -.255 | -.266 |
| 8 UΔPOS | -8.80E-05 | .003 | | | | | | | | -.018 | -.010 | -.007 | .052 | .064 | .023 | .033 | .036 |
| 9 UΔNEG | -4.17E-05 | .003 | | | | | | | | | .042 | .027 | -.035 | -.009 | .028 | .023 | .054 |
| 10 MktValue | 9.558 | 1.445 | | | | | | | | | | -.288 | -.008 | .064 | -.151 | -.159 | -.144 |
| 11 BkMk | -.319 | .855 | | | | | | | | | | | -.017 | -.176 | .272 | .155 | .147 |
| 12 UΔSALE | -8.06E-05 | .033 | | | | | | | | | | | | -.234 | .096 | .092 | .094 |
| 13 UΔROA | 2.80E-04 | .016 | | | | | | | | | | | | | -.005 | .100 | .137 |
| 14 FFRET | .032 | .330 | | | | | | | | | | | | | | .787 | .762 |
| 15 BHAR | .022 | .256 | | | | | | | | | | | | | | | .883 |
| 16 STKRET | .027 | .301 | | | | | | | | | | | | | | | |

Notes: We present correlations as Pearson correlation coefficients. All the correlations greater than .10 are significant at p <.01, all the correlations greater than .07 are significant at p <.05, and all the correlations greater than .06 are significant at p <.10.

**Table 3.6a: Financial Value of Social Tag Metrics**

| | BHAR | | | FFRET | | | STKRET | | |
|---|---|---|---|---|---|---|---|---|---|
| | Estimate | SE | sig | Estimate | SE | sig | Estimate | SE | sig |
| *(Intercept)* | .28 | .07 | *** | .23 | .09 | ** | .33 | .07 | *** |
| *UΔROA* | 2.59 | .55 | *** | 1.43 | .72 | * | 3.47 | .61 | *** |
| *UΔSALE* | .87 | .26 | *** | 1.02 | .33 | ** | 1.12 | .29 | *** |
| *MktValue* | -.02 | .01 | *** | -.02 | .01 | * | -.02 | .01 | ** |
| *BkMk* | .03 | .01 | ** | .08 | .01 | *** | .03 | .01 | * |
| *UΔVOLSOC* | .14 | .03 | *** | .21 | .04 | *** | .10 | .03 | ** |
| *UΔBKVOL* | -.05 | .05 | | -.06 | .06 | | -.01 | .05 | |
| *UΔVOLCOM* | -.07 | .06 | | -.08 | .08 | | -.04 | .07 | |
| *UΔUNIQUE* | .20 | 1.22 | | .79 | 1.57 | | .88 | 1.34 | |
| *UΔSHARE* | -2.54 | .94 | ** | -3.08 | 1.22 | * | -2.37 | 1.04 | * |
| *UΔCONNECT1* | 2.64 | .75 | *** | 3.40 | .97 | *** | 3.20 | .83 | *** |
| *UΔCONNECT2* | -1.78 | .30 | *** | -1.82 | .38 | *** | -2.08 | .33 | *** |
| *UΔPOS* | -1.73 | 3.13 | | -1.94 | 4.03 | | -2.73 | 3.44 | |
| *UΔNEG* | 5.76 | 3.92 | | 8.86 | 5.05 | . | 9.97 | 4.31 | * |
| *PROM(dummy)* | .02 | .02 | | .02 | .03 | | .02 | .02 | |
| *UΔVOLSOC*PROM* | -.12 | 0.06 | * | -.16 | .08 | * | -0.17 | .07 | * |
| *UΔVOLBK*PROM* | .02 | .15 | | -.02 | .19 | | .06 | .16 | |
| *UΔVOLCOM*PROM* | .10 | .14 | | .11 | .18 | | .04 | .15 | |
| *UΔUNIQUE*PROM* | -1.01 | 2.33 | | .17 | 3.01 | | -.59 | 2.57 | |
| *UΔSHARE*PROM* | 2.83 | 1.29 | * | 3.38 | 1.67 | * | 2.61 | 1.42 | . |
| *UΔCONNECT1*PROM* | -5.37 | 1.74 | ** | -5.37 | 2.24 | * | -4.55 | 1.91 | * |
| *UΔCONNECT2*PROM* | 1.21 | .68 | . | 1.17 | .88 | | 1.51 | .75 | * |
| *UΔPOS*PROM* | 1.97 | 6.40 | | 1.89 | 8.26 | | 8.62 | 7.05 | |
| *UΔNEG*PROM* | -10.93 | 6.26 | . | -15.94 | 8.08 | * | -15.90 | 6.89 | * |
| Residual standard error | .232 | | | .299 | | | .255 | | |
| R squared | .209 | | | .208 | | | .305 | | |
| Adjusted R squared | .1833 | | | .1823 | | | .2823 | | |

\*\* : significant at p<.001

\* : significant at p<.05

. : significant at p<.10

Notes: Model includes calendar year dummies.

**Table 3.6b: Financial Value of Social Tag Metrics (Bayesian Analysis)**

| Variable | Mean | SD | 95% CI | | | | | |
|---|---|---|---|---|---|---|---|---|
| *MktValue* | .006 | .006 | [ | -.006 | , | .018 | ] | |
| *BkMk* | .112 | .024 | [ | .068 | , | .160 | ] | ** |
| *U∆SALE* | .728 | .286 | [ | .229 | , | 1.346 | ] | ** |
| *U∆ROA* | 1.560 | .360 | [ | .889 | , | 2.182 | ] | ** |
| *U∆VOLSOC* | .055 | .037 | [ | -.021 | , | .127 | ] | . |
| *U∆BKVOL* | -.011 | .050 | [ | -.111 | , | .085 | ] | |
| *U∆VOLCOM* | -.008 | .062 | [ | -.131 | , | .118 | ] | |
| *U∆UNIQUE* | 1.133 | .293 | [ | .504 | , | 1.582 | ] | ** |
| *U∆SHARE* | -.375 | .529 | [ | -1.122 | , | .590 | ] | |
| *U∆CONNECT1* | 2.656 | 1.477 | [ | -.208 | , | 5.625 | ] | * |
| *U∆CONNECT2* | -.379 | .264 | [ | -.846 | , | .218 | ] | . |
| *U∆POS* | -.526 | .641 | [ | -1.725 | , | .823 | ] | |
| *U∆NEG* | -.116 | .937 | [ | -1.996 | , | 1.678 | ] | |
| $\sigma_\theta$ | .074 | .018 | [ | .047 | , | .118 | ] | |
| $\sigma_\alpha$ | .062 | .014 | [ | .040 | , | .093 | ] | |
| $\sigma_y$ | .232 | .006 | [ | .220 | , | .245 | ] | |

**: more than 97.5% of posterior probability confidence interval (one-sided) excludes 0.

*: more than 95% of posterior probability interval (one-sided) excludes 0.

. : more than 90% of posterior probability interval (one-sided) excludes 0.

Notes: Model includes firm-specific random effects ($\alpha_i$), quarter-specific random effects ($\theta_t$) and two control variables - *log(MktCap$_{t-1}$)* and *log(Bk_Mk$_{t-1}$)*.


**Table 3.7: Bayesian Mediation Analysis**

| | Direct Effect (Return Eq.) | | Direct Effect (Sales Eq.) | | Mediated Effect | | Combined Effect | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| ***U∆VOLSOC*** | .060 | .023 ** | .003 | .004 | .004 | .004 | .064 | .023 ** |
| *U∆BKVOL* | .041 | .036 | -.006 | .007 | -.007 | .006 | .034 | .037 |
| *U∆VOLCOM* | .013 | .047 | .002 | .010 | -.008 | .009 | .005 | .049 |
| ***U∆UNIQUE*** | .644 | .139 ** | -.072 | .128 | -.157 | .101 | .487 | .131 ** |
| *U∆SHARE* | .153 | .198 | .022 | .151 | -.190 | .146 | -.037 | .261 |
| *U∆CONNECT1* | 2.042 | 1.511 | .092 | .099 | -.011 | .092 | 2.032 | 1.510 |
| ***U∆CONNECT2*** | -.013 | .006 ** | -.028 | .034 | -.030 | .030 | -.043 | .030 |
| *U∆POS* | -.002 | .002 | .434 | .262 ** | -.009 | .239 | -.011 | .239 |
| *U∆NEG* | -1.647 | .303 ** | -.413 | .143 ** | -.495 | .153 ** | -2.142 | .363 ** |
| **U∆SALE** | 0.803 | 0.073 ** | | | | | | |

**: more than 97.5% of posterior probability confidence interval (one-sided) excludes 0.

*: more than 95% of posterior probability interval (one-sided) excludes 0.

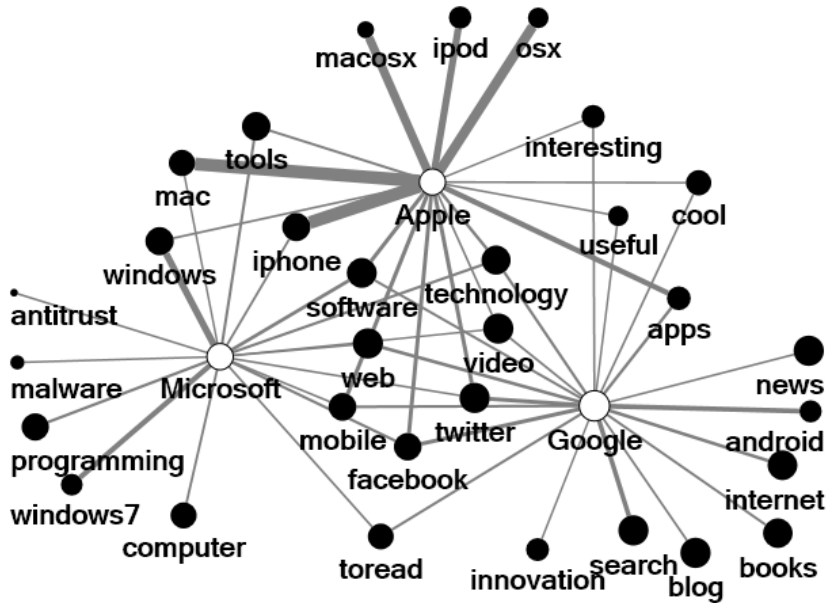**Figure 3.1: Visual Illustration of Interrelationships between Firms**



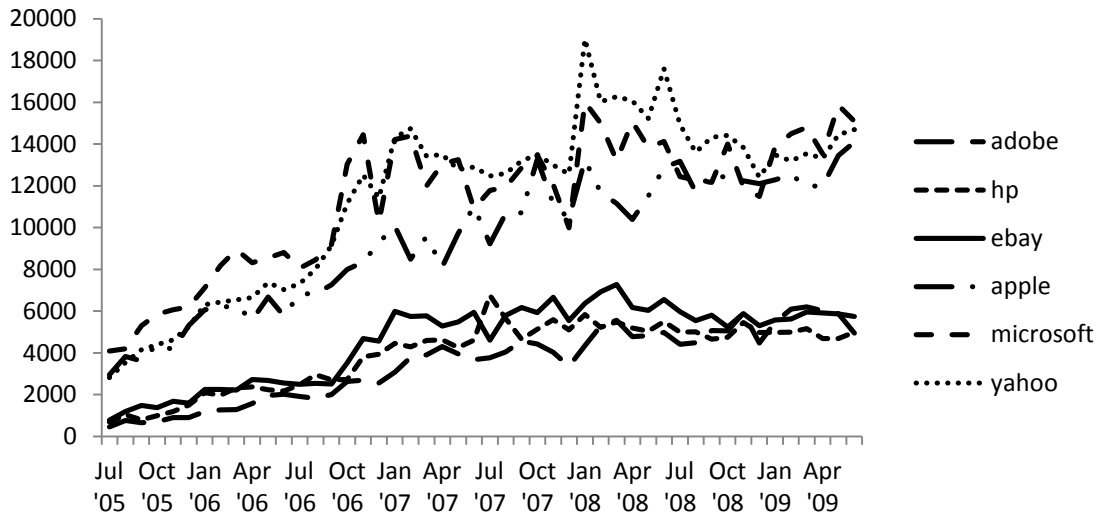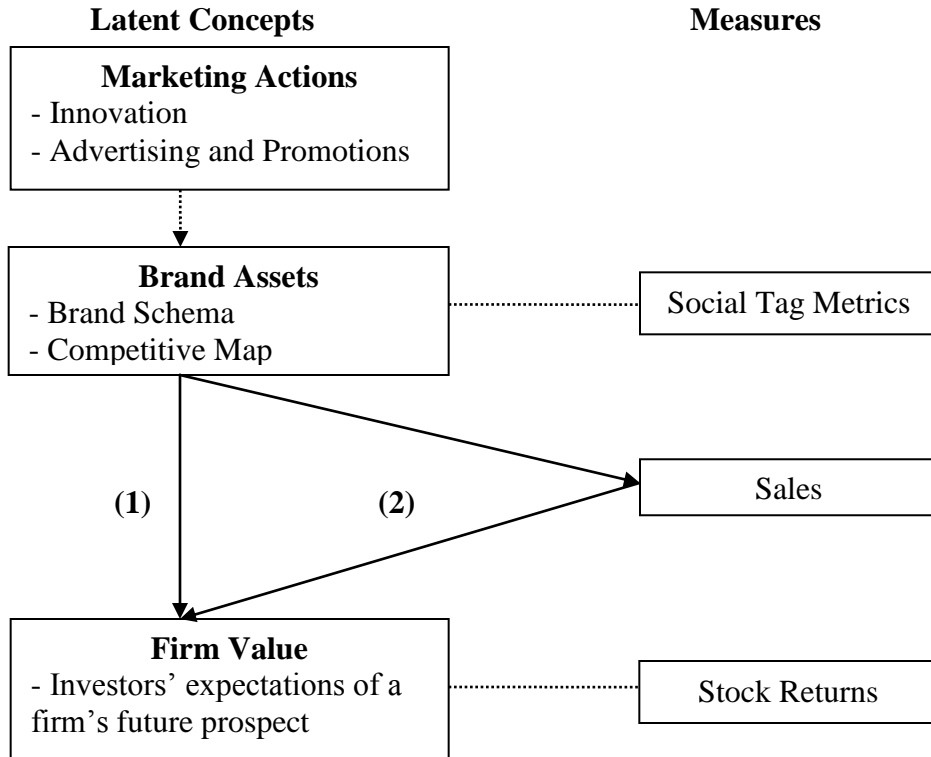**Figure 3.2: Trend of the Number of Bookmarks over Time for Six Selected Firms**

**Figure 3.3: Conceptual Framework**

**Latent Concepts**                                      **Measures**

```
┌─────────────────────────────────┐
│      Marketing Actions          │
│  - Innovation                   │
│  - Advertising and Promotions   │
└─────────────────────────────────┘
              ┊
              ▼
┌─────────────────────────────────┐            ┌──────────────────────┐
│        Brand Assets             │┈┈┈┈┈┈┈┈┈┈┈│  Social Tag Metrics  │
│  - Brand Schema                 │            └──────────────────────┘
│  - Competitive Map              │
└─────────────────────────────────┘
        │         ╲
   (1)  │    (2)   ╲                            ┌──────────────────────┐
        │           ╲─────────────────────────▶│        Sales         │
        ▼           ╱                           └──────────────────────┘
┌─────────────────────────────────┐
│        Firm Value               │┈┈┈┈┈┈┈┈┈┈┈┌──────────────────────┐
│  - Investors' expectations of a │            │    Stock Returns     │
│  firm's future prospect         │            └──────────────────────┘
└─────────────────────────────────┘
```

# Chapter IV: The Dynamics of Products on Tagging Networks: Insights for Demand Forecast and Positioning

## 4.1. Introduction

With the explosive growth of social media, networked communication has begun to play a significant role in the diffusion of products, ideas, and thoughts. A dramatic change in the social interests of a concept or content connected to a focal product can contribute to dynamics in the sales of products directly or indirectly connected to the concept or content. For instance, a recent study has shown that when a book was featured by Oprah Winfrey on her television show, not only did the appearance directly benefit the featured book, but the book's neighboring books in the co-purchase recommendation network on Amazon.com (Amazon) also benefited (Carmi, Oestreicher-Singer, and Sundararajan 2009). In order to understand and take advantage of such spillover effects, marketers track and manage the networks of content associated with their brands and products. Our primary objective is to investigate when and what types of concepts marketing managers should associate with their products to boost sales.

To answer this question, we build a semantic map of products on associative networks of concepts, perceptions, and thoughts by employing user-generated keywords, "social tags" created on an online retail Web site, Amazon. The tagging system used by Amazon allows customers to describe and classify products using descriptive keywords of their own called "tags." Users can share the tags with other users and can search products by tag. Amazon also encourages customers to interact with each other within a

tag community by participating in discussions and uploading related products and images. Under this system products are organized, searched, discovered, and shared through tags.

As demonstrated in Chapters I, II, and III, the associative structure between products and keywords in social tagging networks can provide customers' mental representation of products and thus can be a good source to estimate product demand and potential market size. To further illustrate our research question, suppose a food recipe book that has been described with several keywords like "food," "cheese," "recipe," "wine," "too expensive," "easy to follow," and "fun." We investigate the disparate impact on product sales for a book to be associated with the keyword "food" compared to the keyword "recipe." More specifically, we investigate (1) to what extent being associated with popular keywords with high degree centrality can be related to sales increase, (2) to what extent being connected to bridging keywords with high betweenness centrality can boost sales, (3) to what extent developing dense content clusters of a product can be related to sales increase, and (4) to what extent a shock on social interests on linked keywords can boost product sales.

We demonstrate that the position of a product, which can be characterized with metrics such as social activity, reach, closure, and bridging properties of tagging communities linked to the product, contains significant information related to users' perceptions and thus can facilitate an explanation of the variation in book sales. More specifically, we find that (1) books in long tail can increase sales by being strongly linked to socially popular keywords and well-known keywords with high degree centrality and (2) top sellers can make them viable by creating dense content clusters rather than connecting them to well-known keywords with high degree centrality.

This study contributes to extant literature in several respects. First, to our knowledge this is the first work developing product positioning maps utilizing products-to-tags network structure. Unlike extant methods to create product maps such as Multidimensional Scaling (e.g., DeSarbo et al. 1996; Shugan 1987) or Concept Maps (e.g., John et al. 2006), we employ user-generated keywords from multiple customers as the input for positioning maps. Our study shows that marketers can improve the accuracy of prediction of sales by taking into account the network relationships of social tags and the impact of position based on tags is more persistent than the impact of information in online ratings.

Second, we find differential mechanisms in product dynamics for top sellers versus long-tail products. Our findings suggest that for top sellers cultivating and refining the extant customer community and mental associations is more important to being a better seller while for long-tail products increasing the probability of being explored, searched, and identified is more important to boosting the sales. We believe that our findings suggest that marketing managers can better understand a user community's perception of products and potentially influence product sales by taking into account the positioning of their products within the network of content.

## 4.2. Background

A social tagging system on Amazon.com allows customers to describe and categorize products with keywords of their own, share the keywords with other customers, search products by keyword, and participate in discussions within a keyword community. Within this system, customers can associate a product with keywords, find keywords

other customers associated with the product, and further discover products linked to each keyword. The tag community facilitates customers' search and exploration within a specific keyword. As a result, products are linked to each other through common keywords and keywords are linked to each other via common products.

Figure 4.1 (a) presents how Amazon displays tags connected to a title in the *Harry Potter* series. The tags are sorted by the number of customers associating the tag with the title. "harry potter" is most strongly linked to the book and 112 customers used the tag to describe the book. "fantasy," "magic," and "jk rowling" are examples of strongly associated tags. Popular, common tags have a user community in which customers participate in listing related products (not confined to the book category), discussing tag-related topics, and posting related images. Figure 4.1 (b) presents a snapshot of one such tag community, the Fantasy Community, at Amazon. The number of products, discussions, lists and guides, images, and contributors for the tag are presented. Customers can search further information through narrowing by other tags and seeing the list of all products linked to the tag.

[INSERT FIGURE 4.1 ABOUT HERE]

Figure 4.2 (a) presents an example of a book mapped onto the networks of tag communities. Book A is strongly associated with multiple tag communities such as "mystery," "fiction," "9.99 boycott," "suspense," and "Scandinavian literature." Marketers can understand the most salient associations with the book by finding strongly linked keywords as well as customers' overall perceptions of the book. These tag communities are linked to other tag communities by being co-tagged with other products in Amazon. Figure 4.2 (b) presents an example of co-tagging networks of selected books

94

on Amazon. In this Figure, books are linked through the common tags: for instance, book A and book B are linked through tags "teen" and "evolution" and book B and book H are linked through tags "Collins," "romance," and "action-adventure."

<center>[INSERT FIGURE 4.2 ABOUT HERE]</center>

We posit that the social tagging network on Amazon is a rich source of information about customers' mental associations, perceptions, evaluations, and attitudes toward a product. All told, the social tagging network represents the customers' perception of the positioning of a product in the various communities.

First, the tags can indicate what consumers think about a product. As Figure 4.2 shows, tags contain abundant information about a book's characteristics such as author, genre, and users' evaluations and descriptions. Moreover, tags show how users perceive and interpret marketers' messages about the books. For instance, if the most strongly related tags of a book are author-related tags, we may conclude that the most salient perception of that book by customers is the author. In addition, social tags can show how the perceptions of a book change dynamically. New tags can be added and the distribution of tags can change over time.

Second, the tags can indicate social trends, fads, and how the interests of customers change. Users' activities in the tag community, summarized as the number of discussions, images, and contributors to the community can show the dynamic change of user interests as well as the extent of user interests in the tag-related issues. Marketers can track the change of user interests and activities in tag communities and thus can identify the dynamics of trends and fads.

Third, the tagging network can show the position of a product on the tagging

network. As Figure 4.2 (a) shows, we can map a product onto the content space

represented by the tagging network. In Figure 4.2 (a) book A is strongly linked to the

"fiction" tag which has a relatively large number of neighboring tag communities and

plays a role as a bridge between several distinct communities. We believe that by digging

into the characteristics of tags we can find the demand drivers of a product. For instance,

tags with a relatively large number of neighboring communities added to a book may

suggest that the reach of the book to potential customers is increasing while tags with a

relatively high bridging property may suggest that the appeal of the book to various

diverse customer communities is increasing.


## 4.3. Conceptual Development

### *Building Perceptual Maps from Social Tags*

Figure 4.3 depicts the underlying mechanism in the relationships among social

tags, customers' perceptions, and product demand. We argue that customers' underlying

thoughts, perceptions, and mental associations can be revealed through various forms of

User-Generated Content (UGC) such as user reviews, blog posts, Tweets, social tags, and

so forth. Numerous existing marketing literatures have also noted the informative value

of UGC and employed various methods to capture customers' thoughts, sentiments,

opinions, and awareness about a product/brand through UGC toward a brand/product.

The predictive power of volume and valence of UGC in sales dynamics is well

demonstrated in several existing studies (e.g., Godes and Mayzlin 2004; Liu 2006).

Recent marketing studies further investigated the value of associative structure in textual

information in UGC (e.g., Lee and Bradlow 2011; Netzer et al. 2012): Lee and Bradlow (2011) showed automatically identified keywords from online user reviews can provide insights on customers' perceptions on product positioning. Netzer et al. (2012) demonstrated how the competitive market structure can be derived from posts on online user forums.

In line with Lee and Bradlow (2011) and Netzer et al. (2012), we posit that the informative value of associative structure in UGC can provide marketers with additional information regarding customers' perceptions. To investigate the associative structure of products, rather than mining full online reviews we employ social tags to extract information about consumer evaluations, attitudes, and social trends. We believe that without complicated text mining analysis we can obtain rich semantic information from social tags which can be simple and easily mineable indicators of customers' activities such as online user reviews, Tweets, or blog posts.

[INSERT FIGURE 4.3 ABOUT HERE]

Although recent studies investigate the value of associative structure in textual information in UGC, a standard methodology of quantifying textual information is not yet well established, and the predictive value of associative networks built upon UGC has not been empirically investigated. In this paper, we propose a method to create product positioning maps and empirically investigate the value of product position on the map in predicting the sales.

*Networks of Products*

Researchers have investigated the value of networks to explaining the diffusion process of ideas, messages, and products. For instance, Granovetter (1973) suggested that

97

the weak ties in the network play a role in bridging the two clusters of close friends and individuals with few weak ties will be deprived of information and confined to the provincial news and views of close friends. Moreover, the author suggests that social systems lacking in weak ties will be fragmented and incoherent and in such systems new ideas spread slowly. Burt (1997) demonstrated how the value of social capital to an individual is contingent on the number of people doing the same work. The author argues that people with high social capital stand at the crossroads of a large social organization and therefore are more likely to be a candidate for inclusion in new opportunities since their contacts are more diverse.

Connectedness in the network can influence the information dissemination process (e.g., Schott 1987; Weimann 1994; Goldenberg et al. 2009; Stonedahl, Rand, and Wilensky 2010; Watts and Dodds 2007). Schott (1987) while examining interpersonal influence in science, suggested that a national community's influence is enhanced by its expertise and that the influence of one community on another is prompted by collegial and educational ties between them (indicated by co-authorships and student exchanges). Similarly, Weimann (1994) suggested that centrally positioned scholars, i.e., scientific opinion leaders, determine the direction of scientific progress because innovations adopted by central figures are more widely accepted by other members of the profession.

Extant marketing literature investigates the impact of influencer characteristics on the adoption process of a product/content. Goldenberg et al. (2009) investigated the role of social hubs (people with an exceptionally large number of social ties) in the diffusion process of content. The authors found innovative hubs have more impact on the speed of the adoption process, while follower hubs have more impact on market size (the total

number of adoptions). In addition, the authors showed that a small sample of hubs can predict success versus failure of content early in the diffusion process. Katona, Zubcsec, and Sarvary (2011) investigated the diffusion process in online social networks. The authors found that the more an individual is connected to many adaptors and the higher the density of connections in the adopter group is, the higher the probability an individual adopts a product.

Networks of products based on browsing history or purchase history have been employed to predict customer preference and the product diffusion process. The user-product networks based upon purchase history has been shown to help design the recommendation system (Huang, Zeng, and Chen 2007). The product co-purchase network is shown to accelerate the media feature effects on product demand (Carmi, Oestreicher-Singer, and Sundararajan 2009).

## 4.4. Social Tag Metrics and Hypotheses

### Bi-partite Networks of Products-to-Tags

To capture the network positions and characteristics of a product, we first define a time-varying bi-partite network of products-to-tags, $G_t(B, N_t, L_t)$ with the edges $L_t$ between the pairs of book $i$ in $B$ and tag $j$ in $N_t$. For each book $i$, we define the set of connected tags of book $i$ at time $t$, represented by the set $Tag_{it} = \{n_j : l_{ij} \in L_t\}$ for a tag $n_j \in N_t$. Edges $(l_{ij})$ are defined when a tag $j$ is added to a book $i$ at time $t$. We define the associative strength of edge $l_{ij}$ as the level of attention tag $j$ gains from book $i$ as compared to other tags linked to book $i$, as specified in Equation (1) by employing the method proposed by De Choudhury et al. (2010).

$$(1) \qquad AscStr_t(l_{ij}) = \frac{N_t(b_i, n_j)}{\sum_{n_j \in N_t} N_t(b_i, n_j)}$$

, where $AscStr_t(l_{ij})$ is the strength of an association between book $i$ and tag $j$ at time $t$;

$N_t(b_i, n_j)$ is the number of customers who have associated tag $j$ to book $i$ at time $t$.[19] In

this network, dynamics in network positions of a book come from the set of tags linked to

book $i$, $Tag_{it}$ and associative strength between book $i$ and tag $j$, $AscStr_t(l_{ij})$.

The time-varying network of tags-to-products, $G_t(B, N_t, L_t)$, enables us to define

a time-varying network of products (Faust 1997). The networks of products, $g_t(B, E_t)$, is

represented with the edges $E_t$ between the pairs of nodes $b_i$ and $b_{i'}$ in $B$. Edges $(e_{ii't})$ are

defined when book $i$ and $i'$ have been co-tagged at time $t$ based on network of tags-to-

products $G_t(B, N_t, L_t)$. We employ the similarity of tags between two books, as measured

by the geometric mean of associative strength of tags, as the edge weight in this network

(e.g., De Choudhury et al. 2010).

$$(2) \qquad AscStr_t(e_{ii'}) = \sum_{n_j \in N_t} N_t(b_i, n_j) N_t(b_{i'}, n_j)$$

We characterize network properties of each tag community with three different

measures: degree, clustering coefficient, and betweenness centrality of the tag community.

These three measures are most widely employed to capture the node-level features in the

network in extant literatures on network analysis (e.g., De Choudhury et al. 2010;

Newman 2001).

---

[19] The tagging system on Amazon allows tags to have a negative number of customers when a tag is added
and then deleted by a customer and there are no other customers using that tag. We ignore negative links
for the analysis.

*Social Tag Metrics*

**Degree or reach.** We capture the extent to which a book can reach its potential

customers with degree centrality of a book on the network, $g_t(B, E_t)$. The degree of

book $i$ ($Degree_{it}$) is proportional to the number of books sharing tags with the focal

book $i$. A book with high degree centrality is likely to be connected to more tag

communities with a large number of neighbors. Tag communities with a large number of

neighbor tags can be viewed as "hub" in the network (Goldenberg et al. 2009). Thus, we

posit that a book with high degree centrality tends to have higher reach to mass customers,

and is more likely to be discovered and searched by customers.

**Clustering coefficient.** We capture the extent to which a book tends to closely

cluster with other books via shared tags. The clustering coefficient measures how well the

neighbors in the network are connected. If neighbors of a book are fully connected via

social tags, the clustering coefficient is 1, while if none of the neighbors are linked then

the clustering coefficient is 0. The clustering coefficient of book $i$ on the network,

$g_t(B, E_t)$ is defined as the average probability that two neighbor books of book $i$ are

neighbors of each other, as presented by Equation (3). A book with a high clustering

coefficient can be a highly isolated book in the networks, or a highly unique book with its

own content clusters. Demand of a book with high clustering coefficient tends to be less

affected by changes in tagging networks.

(3) $$Cluster_{it} = \frac{2|e_{ii't}|}{Degree_{it}(Degree_{it} - 1)}$$

, where $e_{ii't}$ is the edge weight between a book $b_i$ and $b_{i'}$ when $b_i$ , $b_{i'} \in Neighbor_{it}$.

**Betweenness centrality or bridging.** We capture the extent to which a book can

reach various groups of customers with the betweenness centrality book $i$ on the network

$g_t(B, E_t)$, as presented in Equation (4). Betweenness centrality represents the extent to which a book connects between distinct clusters of tags. A book of high betweenness centrality can be viewed as a bridging book connecting various tag communities and books. Bridging ties have been known to play a significant role in the information dissemination process (Grenovetter 1973; Burt 1997).

$$(4) \qquad Between_{it} = \sum_{i \neq i' \neq i''} \frac{SP_{i'i''}(i)}{NSP_{i'i''}}$$

, where $NSP_{i'i''}$ is the number of shortest paths between book $b_{i'}$ and $b_{i''}$; $SP_{i'i''}(i)$ is the number of shortest paths between book $b_{i'}$ and $b_{i''}$ that pass through book $b_i$ on $g_t(B, E_t)$.

*Social activeness.* We capture the social activeness of a tag community with multiple measures: number of tags linked to the community, number of products tagged in the community, number of contributors in the community, and number of discussions in the community. Since all these measures are highly correlated (correlation coefficient $> 0.8$), we decided to employ the number of tags linked to the tag community as a measure of social activeness of a tag community. We summarize social activeness of each tag community linked to a product with the weighted average of social activeness of each tag community, where the weight is the associative strength between product $i$ and tag $j$ in Equation (1).

$$(5) \qquad Active_{it} = \sum_{n_{jt} \in Tag_{it}} AscStr_t(l_{ij}) \, Active_{jt}$$

, where $AscStr_t(l_{ij})$ is the strength of an association between book $i$ and tag $j$ from time at $t$, defined in Equation (1); $Tag_{it}$ is the set of connected tags of book $i$ at $t$.

*Hypotheses*

Regarding product positioning, marketing managers are faced with a choice to refine and cultivate current associations/concepts (i.e., "Exploitation") or to expand the horizon of associations/concepts (i.e., "Exploration") (e.g., March 1991). Exploitation strategy is closely related to increasing the clustering coefficient of a book in the networks by cultivating and refining the extant set of associations. Exploration strategy corresponds to increasing the degree centrality of a book by connecting to a new customer community, concept, and association.

We expect that a distinct mechanism in product dynamics exists for top sellers versus books in long tail. For long-tail products, being identified and discovered through various methods (e.g., search technologies and recommendation systems) is critical for viability (e.g., Hinz, Eckert, and Skiera 2011; Brynjolfsson, Hu, and Smith 2006). Hence, books in long tail will be more benefitted from exploration strategy than exploitation strategy. That is, having higher reach to mass customers by being linked to popular, common keywords in the community can boost sales of long-tail products while having a high clustering coefficient can make long-tail books isolated in the networks, less likely to be discovered by customers, and thus is negatively related to sales.

**H1-1**: *The increase in degree centrality of long-tail products is positively related to product sales.*

**H1-2**: *The increase in clustering coefficient of long-tail products is negatively related to product sales.*

We expect that top sellers can benefit more from exploitation strategy than exploration strategy. That is, cultivating extant content/associations by creating dense,

unique content clusters is more helpful than creating new associations. Since for them

being discovered by customers is not an issue, influencing and attracting non-buyers is

more critical. Having high/increasing clustering coefficients can further facilitate product

sales by influencing extant non-buyers (Katona, Zubcsec, and Sarvary 2011).

**H2-1**: *The increase in degree centrality of top seller books is negatively related to*

*product sales.*

**H2-2**: *The increase in clustering coefficient of top sellers is positively related to product*

*sales.*

 We also expect the impact of betweenness centrality on sales. High betweenness

centrality indicates that the book is linked to distinct clusters of tags and thus can be

viewed as a bridge across other books. This indicates more possibilities to attract

customers with distinct interests and thus a bigger target customer community. However,

it may indicate diluted perception of the product, which interferes with the product being

included into the consideration set (MacInnis and Nakamoto 1992). We expect mixed

relationships in betweenness centrality and sales.

**H3-1**: *The increase in betweenness centrality of a book is positively related to product*

*sales if it attracts distinct customer communities.*

**H3-2**: *The increase in betweenness centrality of a book is negatively related to product*

*sales if it indicates association dilution.*

 In line with previous research showing that the media feature of a book can

indirectly impact demand for books (Carmi, Oestreicher-Singer, and Sundararajan 2009),

we posit that the increase in social activities of the tag communities linked to a product

can indicate increased interests in the book and thus be positively related to the increase

of book sales.

**H4**: *The increase in social activeness of tag communities of a book is positively related to*

*product sales.*

## 4.5. Research Design

*Data*

To test whether the information contained in the social tagging networks can

indicate the dynamics of product sales, we collect product-level information and tag

community information for a sample of four hundred books in the hardcover book

category on Amazon.com every week. To obtain a representative sample covering books

in various sales ranks, we sample books in four different strata defined by sales rank. We

select all the book titles ranked in the top 100, randomly select 100 titles from books

ranked between $101^{st}$ and $1,000^{th}$ in sales, randomly select 100 titles from books between

$1,001^{st}$ and $2,000^{th}$ in sales, and randomly select 100 titles from books between $2,001^{st}$

and $3,000^{th}$ in sales.

For this set of 400 books, each week we collect product information including

sales rank, pricing information (list price and actual price), online review volume and

ratings, all the associated tags and the strength of the associations, and product-level

information (author, reading level, publisher, publication date, and genre). Then, for each

tag associated with any of the products in our dataset we collect tagging community

information such as the number of contributors, the number of associated products, the

number of lists and guides, the number of discussions, the number of images, and all the

associated tags and the strength of the associations. We track weekly product-level data

of 400 books and weekly tag-level data of 11,294 tag communities for 31 weeks (August

17, 2010-March 15, 2011). For our analysis, we exclude textbooks and reference books

(107 books), books with missing information (four books), and books with no user

activity during 30 weeks (77 books). The final set of data contains weekly dynamics in

tags and online reviews of 212 books.

*Measures*

Table 4.1 describes the set of variables we use for the analysis. In addition to the

proposed social tag metrics, we capture dynamics in the volume and valence of online

user reviews which can indicate the level of customer interest and evaluations of a

product. In addition, as control variables we consider product-level characteristics such as

genre, book publication date, and price.

For the analysis we converted Amazon sales rank to sales quantities employing

the method suggested by Chevalier and Goolsbee (2003) who assume that sales rank data

follows a power law distribution. Following their approach, we obtain sales quantity

using Equation (6).

(6) $$ln( Sales_{it}) = c - \frac{ln\,( Salesrank_{it})}{\theta}$$

, where $Salesrank_{it}$ is Amazon sales rank for product $i$ at week $t$; $\theta$ and $c$ are constants

and we set $\theta$ =1.2 and $c$ = 10 based on Chevalier and Goolsbee (2003)'s findings.

[INSERT TABLE 4.1 ABOUT HERE]

Table 4.2 shows descriptive statistics of measures. The mean sales rank of the

sample is 16,065 (SD=46,715) and the mean converted sales quantity is 33 (SD=124).

The mean price of the sample is $16.70 (SD=$14.20) and the mean list price of the

sample is $26.80 (SD=$15.20); the average discount rate is 38%. On average, books have 244 reviews (SD=501) and the average review rating is 4.15 (SD=0.85). The mean degree centrality is 776 (SD=807), clustering centrality is 0.0040 (SD=0.0058), betweenness centrality is 0.0082 (SD=0.0042). On average, books 1,018 days old (SD=2,006) since they are published. About 30% of books are categorized as fiction, 26% are health-related books, 21% are business and technology, 19% are children books, 17% are science fiction, 12% books are history books, 12% are religion, and 3% are arts and crafts.

[INSERT TABLE 4.2 ABOUT HERE]

*Model Specification*

We employ Panel Vector Autoregressive (PVAR) to test whether the semantic position of a product, i.e., the position of the product relative to the tags describing the product, can explain dynamics in book demand (e.g., Sismeiro, Mizik, and Bucklin 2012). To capture the concurrent relationship among sales, online reviews, and social tag metrics capturing semantic position of a product, we allow the error terms to be correlated. In addition, our model controls for the effects of potential exogenous covariates including genres, tenure of a product, and price. The model determines the long-term and the short-term dynamics of the online reviews and social tagging networks on book demand.

$$
\begin{bmatrix}
Sales_{it} \\
\Delta Nreviews_{it} \\
Ratings_{it} \\
\Delta Degree_{it} \\
Clustering_{it} \\
Between_{it} \\
\Delta Active_{it}
\end{bmatrix}
=
\begin{bmatrix}
\alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7
\end{bmatrix}
+ \sum_{j=1}^{k}
\begin{bmatrix}
\beta_{11j} & \beta_{12j} & \beta_{13j} & \beta_{14j} & \beta_{15j} & \beta_{16j} & \beta_{17j} \\
\beta_{21j} & \beta_{22j} & \beta_{23j} & \beta_{24j} & \beta_{25j} & \beta_{26j} & \beta_{27j} \\
\beta_{31j} & \beta_{32j} & \beta_{33j} & \beta_{34j} & \beta_{35j} & \beta_{36j} & \beta_{37j} \\
\beta_{41j} & \beta_{42j} & \beta_{43j} & \beta_{44j} & \beta_{45j} & \beta_{46j} & \beta_{47j} \\
\beta_{51j} & \beta_{52j} & \beta_{53j} & \beta_{54j} & \beta_{55j} & \beta_{56j} & \beta_{57j} \\
\beta_{61j} & \beta_{62j} & \beta_{63j} & \beta_{64j} & \beta_{65j} & \beta_{66j} & \beta_{67j} \\
\beta_{71j} & \beta_{72j} & \beta_{73j} & \beta_{74j} & \beta_{75j} & \beta_{76j} & \beta_{77j}
\end{bmatrix}
\begin{bmatrix}
Sales_{it-j} \\
\Delta Nreviews_{it-j} \\
Ratings_{it-j} \\
\Delta Degree_{it-j} \\
Clustering_{it-j} \\
Between_{it-j} \\
\Delta Active_{it-j}
\end{bmatrix}
$$

(7)

$$
+
\begin{bmatrix}
\gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \\ \gamma_6 \\ \gamma_7
\end{bmatrix}
Z_{it}
+
\begin{bmatrix}
\epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7
\end{bmatrix}
$$

, where $Z_{it}$ is the set of exogenous control variables including genre indicators, number of days since the book was introduced, list price, and actual price of book.[20] Error terms for each book follow multivariate normal distribution.

The presented model in Equation (7) captures the following dynamic effects.

*Direct effects of social tag metrics and online reviews on sales.* $\beta_{12j}$ - $\beta_{17j}$ capture direct effects of social tag metrics on the product demand. We expect that for both long-tail and top sellers, an increase in social attention and evaluation is related to a sales increase. We expect that product positioning can contribute to sales. We expect that for top sellers developing densely-developed clusters helps them to be viable while for long-tail books it is important to be noticed and thus increasing degree centrality by being connected to popular concepts and social activeness by associating with socially popular concepts or keywords is important.

*Cross relationships between social tag metrics and online reviews.* $\beta_{22j}$ - $\beta_{77j}$ capture cross relationships between social tag metrics and online reviews. We expect that

---

[20] It is possible that Amazon adjusts the retail price of a book as sales change overtime and price dynamics can possibly be indirectly influenced by dynamics in social tag metrics and user reviews. However, for our sample the results of Granger's Causality Test indicate that price variable is exogenous to this system.

social attention and product evaluations captured by online user reviews are highly correlated to social tag metrics. For instance, an increase in degree centrality may lead to a higher level of social attention for a product and thus encourage more customers to write reviews on the product and vice-versa.

***Feedback effects of sales on social tag metrics and online reviews.*** Customers' tagging and reviewing activities are highly dependent on product sales performance at previous periods. We capture such feedback effects of book sales on social tag metrics with $\beta_{21j}$ - $\beta_{71j}$.

***Reinforcement effects.*** $\beta_{11j}$ - $\beta_{77j}$ capture the reflexive effects of sales, online reviews, and social tag metrics.

## 4.6. Results

We first investigate whether the variables are evolving over time based on augmented Dickey-Fuller test and Panel unit root test. We find volume of user reviews, degree centrality, and social activeness to be evolving and take the difference for the analysis. The appropriate number of lags in the VAR model is selected based on Schwarz BIC. Overall, the VAR model shows a good fit (R Square values for pooled model, top sellers, long tail group: .73, .68, and .82, respectively).

*Contemporaneous Correlation*

Table 4.3 shows contemporaneous correlation of residuals of the VAR model. For top seller books, sales is highly positively related to review ratings, clustering coefficient, and social activeness of tagging community, while negatively related to volume of reviews and betweenness centrality. For long-tail books, sales is highly positively related

to review ratings, clustering coefficient, and social activeness of tagging community, while negatively related to the volume of reviews and betweenness centrality.

[INSERT TABLE 4.3 ABOUT HERE]

*Impulse Response Function*

We derive impulse response functions (IRFs) from the model. The IRFs track the dynamics in the impact of an innovation to an endogenous variable on the other endogenous variables in the system. We employ generalized IRFs (Dekimpe and Hanssens 1999; Villanueva, Yoo, and Hanssens 2008) which allow us to account for contemporaneous correlation between endogenous variables. Table 4.4 shows accumulated elasticities of one (1) standard deviation innovation for social tag metrics and online reviews on sales.

For top sellers, the clustering coefficient has strong positive persistent effects on sales user review ratings while degree centrality has strong negative persistent impacts on sales. As we expected, building up densely connected clusters linked to a product helps top sellers to boost the sales rather than connecting to popular keywords. The impact of social activeness of tag communities and betweenness centrality on sales is not significant. Consistent with extant literature, we find user review ratings have strong positive persistent impact on sales. The volume of user reviews has a negative, persistent impact on sales.

For long-tail books, we find that degree centrality and social activeness have a positive impact on sales. Interestingly, the impact of degree centrality is not significant after one week, yet has strong lagged impact on sales for nine weeks. Likewise, the impact of social activeness is not significant after one week, yet has strong lagged impact

110

after two to three weeks. The results suggest that the spillover effects of being connected to popular keywords (with high degree centrality) and socially-popular keywords (with active tag communities) on sales might take a couple of weeks for long-tail books. Consistent with extant literature, user review ratings have a strong positive persistent impact on sales. We do not find a significant impact of betweenness centrality and volume of user reviews on sales.

[INSERT TABLE 4.4 ABOUT HERE]

**4.7. Discussion**

*Contributions*

The primary contributions of our paper are two-fold. First, we show how semantic position of products can be automatically acquired and quantified with the information contained in associative networks of user-generated tags. To our best knowledge this is the first work developing product positioning maps utilizing products-to-tags network structure.

Second, the semantic position of products on associative maps is found to have distinctive value in predicting sales. More specifically, we find that (1) books in long tail can increase sales by being strongly linked to socially-popular keywords and well-known keywords with high degree centrality (exploration strategy) and (2) top sellers can become even better sellers by creating dense content clusters rather than connecting the top sellers to well-known keywords with high degree centrality (exploitation strategy). We believe that our findings suggest that marketing managers can better understand a

111

user community's perception of products and potentially influence product sales by taking into account the positioning of their products within the network of content.

*Managerial Implications*

With the growth of the electronic book market, the number of publishing start-ups is gradually increasing. One of the main problems that publishing start-ups or small publishers face is how to position their books in the appropriate segment and how to take advantage of the so-called network effects and buzz effects to sell their books (Lacy 2008; Tonkery 2010). In the present study, we present a way to deal with publishers' questions by investigating the impacts of network positions of a book mapped onto a social tagging network on product sales dynamics. Our findings indicate that publishers can increase product sales by appropriately positioning a product in the network-based content map. We suggest three strategies: targeting a specific content cluster, connecting a product to socially-popular keywords, and connecting a product to keywords with high degree. In addition, we believe that the proposed methods can also provide added explanatory power in demand forecasting of new products, multi-functional products, and informative products.

*Limitations and Future Research*

Our data describe weekly dynamics of tagging networks. We believe that there might be more information at a more granular level such as daily or even hourly since Amazon updates its sales ranks every hour for bestsellers. Although we do not observe significant dynamics in users' reviewing and tagging activity at the hourly level in our sample, future work should be done to investigate if it is possible to predict demand for a product with data collected at a more granular level.

Our classification between long tail and top sellers is based on post-hoc observation. Future work could find a better method to capture long-tail vs. top seller dynamics. In addition, since book sales evolve over time, it is possible that a book in long tail become a top seller. Our VAR model does not capture such systematic evolution. It will be interesting to investigate the evolution process of books.

**Table 4.1: Measures**

| Conceptual Variable | Notation | Measured Variable |
|---|---|---|
| Social Tag Metrics | | |
| Reach | | |
| Bridging | $Degree_{it}$ | degree centrality of book $i$ on product networks at time $t$ |
| Clustering | $Between_{it}$ | betweenness centrality of book $i$ on product networks at time $t$ |
| Coefficient | $Cluster_{it}$ | clustering coefficient of book $i$ on product networks at time $t$ |
| Social Activeness | $Active_{it}$ | weighted average of social activeness of tag communities linked to a book $i$ at time $t$ |
| Online Reviews | | |
| Volume of reviews | $NReviews_{it}$ | The number of user reviews on a book $i$ at time $t$ |
| Valence of reviews | $Rating_{it}$ | Average rating of user reviews on a book $i$ at time $t$ |
| Control Variables | | |
| Genre | $GR_{ik}$ | Indicator for genre $k$ for book $i$ |
| Tenure | $Tenure_{it}$ | Log of the number of days since book $i$ was introduced |
| Price | $Price_{it}$ | Actual price of book $i$ at time $t$ |
| List price | $Listprice_i$ | Actual price of book $i$ |

**Table 4.2: Descriptive Statistics**

|  | Pooled | | Top sellers | | Long-tail | |
|---|---|---|---|---|---|---|
|  | Mean | SD | Mean | SD | Mean | SD |
| Number of books | 212 |  | 50 |  | 162 |  |
| Salesrank | 16,065 | 46,715 | 386 | 449 | 20,905 | 52,504 |
| Sales (Converted) | 33 | 124 | 115 | 214 | 8 | 59 |
| *Price* | 16.7 | 14.2 | 13.5 | 3.0 | 17.7 | 16.0 |
| *Listprice* | 26.8 | 15.2 | 24.1 | 4.7 | 27.6 | 17.1 |
| *NReviews* | 244.23 | 501.18 | 509.62 | 621.77 | 162.32 | 425.44 |
| *Ratings* | 4.15 | 0.85 | 3.88 | 1.20 | 4.23 | 0.68 |
| *Degree* | 776 | 807 | 1,332 | 1,023 | 604 | 636 |
| *Cluster* | 0.0040 | 0.0058 | 0.0085 | 0.0085 | 0.0026 | 0.0038 |
| *Between* | 0.0082 | 0.0042 | 0.0077 | 0.0043 | 0.0084 | 0.0042 |
| *Active* | 253 | 555 | 206 | 374 | 268 | 599 |
| *Tenure* | 1,018 | 2,006 | 596 | 881 | 1,149 | 2,226 |
| Genre |  |  |  |  |  |  |
| Art | 3% |  | 0% |  | 4% |  |
| Religion | 12% |  | 4% |  | 14% |  |
| Fiction | 30% |  | 30% |  | 30% |  |
| Business and Technology | 21% |  | 28% |  | 19% |  |
| History | 12% |  | 14% |  | 11% |  |
| Health | 26% |  | 30% |  | 25% |  |
| Children | 19% |  | 24% |  | 17% |  |
| Science Fiction | 17% |  | 20% |  | 17% |  |

**Table 4.3: Contemporaneous Correlation of Residuals**

a. Top seller group

|  | Sales | NReviews | Rating | Degree | Cluster | Between | Active |
|---|---|---|---|---|---|---|---|
| *Sales* | 1.000 | -0.047 | 0.376 | 0.023 | 0.114 | -0.009 | 0.023 |
| *NReviews* | -0.047 | 1.000 | 0.003 | 0.009 | -0.003 | 0.033 | -0.006 |
| *Ratings* | 0.376 | 0.003 | 1.000 | 0.116 | -0.115 | 0.041 | 0.008 |
| *Degree* | 0.023 | 0.009 | 0.116 | 1.000 | -0.255 | 0.178 | 0.054 |
| *Cluster* | 0.114 | -0.003 | -0.115 | -0.255 | 1.000 | -0.250 | 0.004 |
| *Between* | -0.009 | 0.033 | 0.041 | 0.178 | -0.250 | 1.000 | 0.055 |
| *Active* | 0.023 | -0.006 | 0.008 | 0.054 | 0.004 | 0.055 | 1.000 |

b. Long-tail group

|  | Sales | NReviews | Rating | Degree | Cluster | Between | Active |
|---|---|---|---|---|---|---|---|
| Sales | 1.000 | -0.027 | 0.072 | 0.011 | -0.004 | 0.011 | -0.011 |
| Nreviews | -0.027 | 1.000 | 0.009 | 0.009 | 0.000 | 0.005 | -0.009 |
| Ratings | 0.072 | 0.009 | 1.000 | 0.101 | 0.090 | 0.023 | 0.079 |
| Degree | 0.011 | 0.009 | 0.101 | 1.000 | -0.037 | 0.107 | 0.012 |
| Cluster | -0.004 | 0.000 | 0.090 | -0.037 | 1.000 | -0.222 | 0.014 |
| Between | 0.011 | 0.005 | 0.023 | 0.107 | -0.222 | 1.000 | -0.008 |
| Active | -0.011 | -0.009 | 0.079 | 0.012 | 0.014 | -0.008 | 1.000 |

c. All groups (pooled model)

|  | Sales | NReviews | Rating | Degree | Cluster | Between | Active |
|---|---|---|---|---|---|---|---|
| Sales | 1.000 | -0.051 | 0.247 | 0.036 | 0.068 | -0.014 | -0.003 |
| Nreviews | -0.051 | 1.000 | 0.001 | 0.012 | -0.006 | 0.024 | -0.008 |
| Ratings | 0.247 | 0.001 | 1.000 | 0.109 | -0.003 | 0.033 | 0.057 |
| Degree | 0.036 | 0.012 | 0.109 | 1.000 | -0.101 | 0.136 | 0.012 |
| Cluster | 0.068 | -0.006 | -0.003 | -0.101 | 1.000 | -0.240 | 0.015 |
| Between | -0.014 | 0.024 | 0.033 | 0.136 | -0.240 | 1.000 | -0.003 |
| Active | -0.003 | -0.008 | 0.057 | 0.012 | 0.015 | -0.003 | 1.000 |

**Table 4.4: Long-term and Short-term Elasticities of Social Tag Metrics and Reviews on Sales ***

a. Top seller group

|  | NReviews | Rating | Degree | Cluster | Between | Active |
|---|---|---|---|---|---|---|
| 1 week | -6.06 | 48.38 | 2.90 | 14.64 | -1.16 | 2.95 |
|  | (4.14) | (3.99) | (4.14) | (4.12) | (4.14) | (4.14) |
| 2 week | -12.74 | 64.53 | -9.03 | 36.30 | -5.32 | 6.55 |
|  | (7.50) | (7.15) | (7.29) | (7.44) | (7.49) | (6.81) |
| 3 week | -15.12 | 73.82 | -21.63 | 53.19 | -8.28 | 10.02 |
|  | (11.48) | (10.31) | (10.69) | (10.72) | (10.60) | (9.70) |
| 6 week | -24.13 | 57.13 | -62.10 | 86.74 | -16.85 | 16.33 |
|  | (18.36) | (19.51) | (21.23) | (18.86) | (18.44) | (16.97) |
| 9 week | -29.62 | 11.69 | -92.56 | 99.50 | -22.33 | 18.83 |
|  | (22.66) | (28.86) | (29.28) | (25.69) | (25.09) | (21.79) |

b. Long-tail group

|  | NReviews | Rating | Degree | Cluster | Between | Active |
|---|---|---|---|---|---|---|
| 1 week | -0.56 | 1.50 | 0.24 | -0.07 | 0.23 | -0.24 |
|  | (0.38) | (0.38) | (0.38) | (0.38) | (0.38) | (0.38) |
| 2 week | -0.41 | 3.05 | 1.50 | -0.81 | -0.15 | 0.94 |
|  | (0.72) | (0.73) | (0.72) | (0.73) | (0.71) | (0.73) |
| 3 week | 0.11 | 10.71 | 3.06 | 0.08 | -0.19 | 1.74 |
|  | (1.16) | (1.26) | (1.12) | (1.16) | (1.13) | (1.18) |
| 6 week | 2.17 | 17.58 | 5.82 | 1.01 | -0.91 | 0.74 |
|  | (2.13) | (2.27) | (1.99) | (1.99) | (1.88) | (2.26) |
| 9 week | 2.48 | 11.95 | 5.49 | 0.43 | -0.55 | -0.68 |
|  | (2.17) | (2.57) | (2.29) | (2.14) | (2.14) | (2.59) |

c. All groups (pooled model)

|  | NReviews | Rating | Degree | Cluster | Between | Active |
|---|---|---|---|---|---|---|
| 1 week | -3.53 | 16.99 | 2.45 | 4.70 | -0.99 | -0.18 |
|  | (1.10) | (1.09) | (1.10) | (1.10) | (1.10) | (1.10) |
| 2 week | -7.24 | 21.26 | 2.71 | 11.65 | -3.28 | 1.31 |
|  | (2.06) | (2.01) | (2.03) | (2.05) | (2.02) | (2.07) |
| 3 week | -9.23 | 20.94 | 1.10 | 19.62 | -6.75 | 1.51 |
|  | (3.21) | (3.19) | (3.00) | (3.14) | (3.06) | (3.14) |
| 6 week | -14.87 | 6.40 | -7.89 | 35.14 | -12.31 | -0.40 |
|  | (5.98) | (6.17) | (5.86) | (5.95) | (5.62) | (6.37) |
| 9 week | -18.33 | -15.83 | -15.93 | 42.47 | -13.84 | -3.30 |
|  | (7.75) | (8.66) | (8.36) | (8.05) | (7.61) | (8.86) |

* Mean accumulated elasticities are reported and the values in the parentheses are standard deviation of the accumulated elasticities.

**Figure 4.1: Snapshot of Social Tagging Networks in Amazon.com**

**(a) Snapshot of Presentation of Tags Connected to one of Harry Potter series**



**(b) Snapshot of Community of tag, "fantasy"**

**Figure 4.2: Graph of Product Network Mapped onto the Tag Network**

**(a) Book-Centric Network Mapped onto Tag Communities**



**(b) Co-tagging Networks of Products**

**Figure 4.3: Conceptual Framework**

```
┌─────────────────────────┐              ┌──────────────────────┐
│ Customers'              │              │ UGC                  │
│ Mental Associations,    │ ───────────► │ - Reviews            │
│ Perceptions, and        │              │ - Tweets             │
│ Evaluations             │              │ - Blog Posts         │
│                         │              │ - Social Tags        │
└─────────────────────────┘              └──────────────────────┘
            │                                       │
            ▼                                       │
┌─────────────────────────┐                         │
│ Product Sales           │ ◄───────────────────────┘
└─────────────────────────┘
```
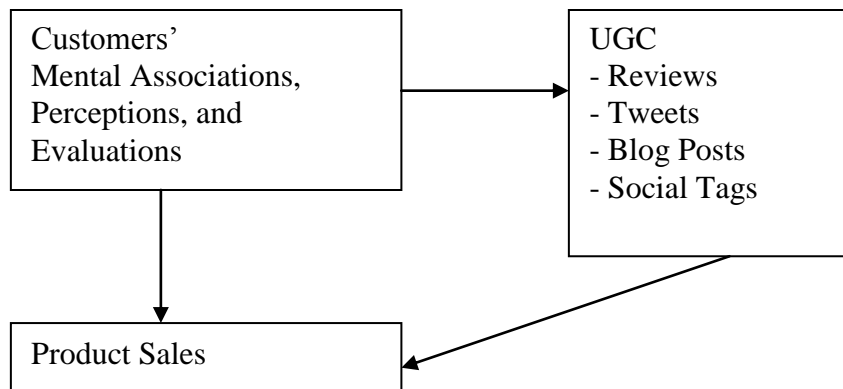
# Chapter V: General Conclusion

This dissertation proposes a method to create associative networks using social tags, a short-form of UGC and empirically investigates the informative value of social tagging networks in firm valuation and product positioning. In this chapter, we will summarize the findings from Chapter II, III, and IV, discuss the contributions of this dissertation, and conclude with an exploration of possible avenues of future research.

## 5.1. Summary of Three Essays

The first essay presents a new methodology to create brand associative networks using social tags. We demonstrate that our approach has several advantages over existing methods (e.g., metaphor elicitation or brand concept maps and recent text mining techniques). Proposed approach (1) is less time consuming and less expensive, (2) less vulnerable to potential errors or biases involved in the elicitation stage since it does not rely on elicited associations either from consumer interviews or from algorithms, and (3) is able to provide richer and unbounded associations linked to a brand by utilizing keywords directly stated by online users to describe a brand or content related to a brand. Using social tag maps, marketers can (1) have access to real-time updates of brand associative networks, and track their brand assets dynamically and (2) understand the competitive position of their brand, and track the dynamics in competitive structure.

The second essay investigates how information contained in social tags acts as proxy measures of brand assets that track and predict the financial valuation of firms using the data collected from a social bookmarking website, *del.icio.us,* for 61 firms

across 16 industries. The results suggest that brand asset metrics based on social tags explain variations in the unanticipated stock return and that the informational value of social tag metrics varies across brands. Specifically, an increase in social attention and connectedness to competitors is shown to be positively related to stock return for less prominent brands, while for prominent brands associative uniqueness and evaluation valence is found to be more significantly related to stock return. The findings suggest to marketing practitioners a new way to proactively improve brand assets for impacting a firm's financial performance.

The third essay investigates whether the position of products on social tagging networks can predict sales dynamics. We find that (1) books in long tail can increase sales by being strongly linked to socially popular keywords and well-known keywords with high degree centrality and (2) top sellers can be better sellers by creating dense content clusters rather than connecting them to well-known keywords with high degree centrality. Our findings suggest that marketing managers better understand a user community's perception of products and potentially influence product sales by taking into account the positioning of their products within social tagging networks.

## 5.2. Contributions and Managerial Implications

The information contained in social tag maps 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, since both allow 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; whereas online search is more of a goal-oriented behavior. Thus, tagging data is perhaps more appropriate when marketers are interested in obtaining consumers' perceptions on a brand.

In the first essay we demonstrate that the proposed social tag map has several advantages over existing methods. First, it is less time consuming and less expensive. While existing approaches rely on elicited associations either from consumer interviews or from algorithms, our method utilizes keywords directly stated by online users to describe a brand or content related to a brand. Hence, our approach is less vulnerable to potential errors or biases involved in the elicitation stage, and is able to provide richer and unbounded associations linked to a brand. Using social tag maps, marketers can (1) have access to real-time updates of brand associative networks, and track their brand assets dynamically, (2) understand changes and trends in social interests related to their brand and incorporate these into their communication plans, and (3) understand the competitive position of their brand, and track the dynamics in the competitive structure.

The informational value of social tag maps can be further assessed by using tags as a proxy measure for intangible brand assets, which can then be used to predict and explain firm valuation, e.g., via stock market returns. Dynamics in social tag metrics can capture changes in social attention, social evaluation, and competitive advantage of a brand, and thus possibly be related to investors' expectations of a firm's prospect. Another potential application of social tag maps is to map an individual product onto

social tagging networks to describe the semantic position of a product with node-level network characteristics on the tagging networks.

The second essay shows how the social tag metrics can explain the variations in stock returns by investigating both direct and indirect relationships between social tag metrics and firm valuation. The findings serve as a bridge between extant literature showing that UGC can indicate and drive product sales (e.g., Liu, 2006) and prior literature showing that the volume and valence of UGC can explain the firm valuation (e.g., Tirunillai and Tellis, 2012). In addition, this paper contributes to extant research streams in brand equity by showing the moderating role of brand prominence in the relationship between social tag metrics and stock returns. The findings suggest that different brand asset management strategies are needed for prominent versus less prominent brands. For less prominent brands, it is more critical to focus on expanding the set of brand associations by being connected to social events or market leaders. Marketing managers of less prominent brands should invest in: (1) boosting brand visibility by creating content related to socially popular events, trends, or concepts and (2) creating more content which can be linked to their competitors, making the brands more assimilated and comparable to the competitors and yielding higher category membership. Managers of prominent brands on the contrary should be more selective in promoting their marketing activities and creating brand-related content. Since prominent brands already have a well-refined set of healthy brand associations, the manager's goal should be to create content that will expand and bolster the current assets.

The third essay shows that the semantic position of products on associative maps of user-generated tags has distinctive value in predicting sales. Specifically, our results

suggest that (1) books in long tail can increase sales by being strongly linked to socially popular keywords and well-known keywords with high degree centrality (exploration strategy) and (2) top sellers can become even better sellers by creating dense content clusters rather than connecting the top sellers to well-known keywords with high degree centrality (exploitation strategy). We believe that our findings provide publishing start-ups or small publishers with insights as to how to position their books in the appropriate segment and how to take advantage of the so-called network effects and buzz effects to sell their books.

## 5.3. Future Research

We hope this dissertation can serve as a modest start towards those future directions and stimulate more research on social tagging networks and networked content in marketing. This dissertation mostly investigates the aggregate picture of customers' associative networks towards a brand, a product, and a firm. An interesting extension of this dissertation will be investigating heterogeneous representation of brand maps using disaggregate level tagging data and identify dynamic customer segments based on social tags. Identifying heterogeneous semantic maps across customers can help marketers to implement a better segmentation/targeting strategy.

Second, investigating dynamics of brand associative network formation will be a promising venue. Future research can explore how consumers' brand perceptions evolve over time and how competing brands/products interact with each other in the process of evolution. For instance, a researcher can investigate how a brand's dominance on a specific association domain can impact customers' perceptions on competitors.

Third, the social tag maps we create do not take semantic distance between keywords into account; i.e., all synonyms are treated as distinct keywords in our analysis. A potential solution to this problem will be considering a lexical database of words such as *WordNet* (e.g., Miller 1995) and incorporate this information into the construction of social tag maps.

Lastly, the meanings of brand associations in tagging data are often ambiguous. A firm name such as "Blockbuster" can be used as a brand name, or a general descriptor. We excluded those firms from our analysis since the results can be misleading. One possible way to resolve this problem in the future is to consider inter-tag relationships and classify only relevant tags. Future research relying on computational linguistics techniques to resolve such ambiguity in the data will be highly valuable.

The second essay has several limitations, which invite further research. Our analysis is based upon a quarterly time frame since (1) accounting measures are available on a quarterly time frame and (2) we judge that it is not easy to observe systematic brand associative structure change on a more granular time window (such as a weekly or monthly level). However, it will be interesting to construct metrics on a more granular time frame and investigate the explanatory power of social tag metrics in daily or weekly-level stock returns. In addition, our paper does not directly include a firm's specific marketing activities such as change in advertising expenditures and communication message, new product announcements and innovations into our model. Rather, we capture customers' integrative perceptions of those marketing activities reflected in the tagging structure of each brand as the impact of those activities are reflected in the tags. It

will be interesting to investigate the chain of marketing activities, customer perceptions/ reactions captured by social tags, sales, and firm value.

We acknowledge a few limitations of the third essay. First, our data describe weekly dynamics of tagging networks. We believe that there might be more information at a more granular level such as daily or even timely since Amazon updates its sales ranks every hour for bestsellers. Although we do not observe significant dynamics in user's reviewing and tagging activity at hourly level in our sample, future work should be done to investigate if it is possible to predict demand for a product with data collected at a more granular level. Our classification between long tail and top sellers is based on post-hoc observation. Future work could find a better method to capture long-tail vs. top seller dynamics. In addition, since book sales evolve over time, it is possible that a book in long tail will become a top seller. Our VAR model does not capture such systematic evolution. It will be interesting to investigate the evolution process of books.

# Bibliography

Aaker, David A. (1982), "Positioning Your Product," *Business Horizons*, 25 (May/June), 56-62.

___ (1991), Managing Brand Equity: Capitalizing on the Value of a Brand Name, New York: Free Press.

___ and Robert Jacobson (2001), "The Value Relevance of Brand Attitude in High-Technology Markets," *Journal of Marketing Research*, 38 (November), 485-93.

Alexa.com (2011), "Statistics Summary for icio.us," http://www.alexa.com/siteinfo/icio.us# , Retrieved on June 12, 2011.

Ames, Morgan and Mor Naaman (2007), "Why We Tag: Motivations for Annotation in Mobile and Online Media," CHI 2007 Proceedings on Tagging and Notetaking, April 28-May 3, San Jose, California, USA, pp.971-980.

Anderson, J.R. and G.H. Bower (1973), Human Associative Memory: A Brief Edition, Hillsdale, NJ: Erlbaum.

Brynjolfsson, Erik, Yu Jeffrey Hu, Michael D. Smith (2006), "From Niches to Riches: the Anatomy of the Long Tail," *Sloan Management Review*, 47(4), 67-71.

Brucks, Merrie (1985), "A Typology of Consumer Knowledge Content," *Advances in Consumer Research*, 58-63.

Burt, Ronald S. (1997), "The Contingent Value of Social Capital," *Administrative Science Quarterly*, 42 (2), 339-58.

Carmi, Eyal, Gal Oestreicher-Singer, Arun Sundararajan (2009), "Spreading the Oprah Effect: The Diffusion of Demand Shocks in a Recommendation Network," ICIS 2009 Proceedings, Paper 78.

Chevalier, Judith A. and Goolsbee, Austan (2003) "Measuring prices and price competition online: Amazon.com and BN.comandNoble.com", *Quantitative Marketing and Economics I* (2).

___ and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345-354.
Churchill, Gilbert A. Jr. (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research*, 16 (Feb), pp. 64-73.

Cohen, Joel B. and Kanul Basu (1987), "Alternative Models of Categorization: Towards a Contingent Processing Framework," *Journal of Consumer Research*, 13 (March), 455-72.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers (1997), "Measuring Mutual Fund Performance with Characteristic-Based Benchmarks," *Journal of Finance*, 52 (3), 1035–1058.

Das, Sanjiv R. and Mike Y. Chen (2007), "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web," *Management Science*, 53 (9), 1375-88.

De Choudhury, Munmun, Winter A. Mason, Jake M. Hofman, and Duncan J. Watts (2010), "Inferring Relevant Social Networks from Interpersonal Communication," *WWW 2010: Proceedings of the 19th international conference on World Wide Web*.

Dekimpe, Marnik G. and Dominique M. Hanssens (1999), "Sustained Spending and Persistent Response: A New Look at Long-Term Marketing Profitability," *Journal of Marketing Research*, 36 (November), pp. 397-412.

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.

Ebiz MBA (2010), "Top 15 Most Popular Social Bookmarking Websites," http://www.ebizmba.com/articles/social-bookmarking-websites, Retrieved on December 1, 2010.

Ehrenberg, Andrew S.C., Gerald J. Goodhardt, and T. Patrick Barwise (1990), "Double Jeopardy Revisited," *Journal of Marketing*, 54 (3), 82-91.

___, Neil Barnard, and John Scriven (1997), "Differentiation and Salience," Journal of Advertising Research, 37 (November/December), 82-91.

Erdem, Tulin and Joffre Swait (1998), "Brand Equity as a Signaling Phenomenon," *Journal of Consumer Psychology*, 7(2), 131-157.

Fama, Eugene and Kenneth French (1993), "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33(1), 3-56.

___ (1998), "Market Efficiency, Long-Term Returns and Behavioral Finance," *Journal of Financial Economics*, 49 (3), 283-306.
Faust, Katherine (1997), "Centrality in Affiliation Networks," *Social Networks*, 19, 157-191.

Fleiss, J.L. (1971), "Measuring Nominal Scale Agreement Among Many Raters," *Psychological Bulletin*, 76, 378-82.

Gentner, D. and A. L. Stevens (1983), Mental Models, Hillsdale, NJ: Erlbaum.

Geweke, John. (1992), "Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior Moments, 11 in Bayesian Statistics 4, J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds.), Oxford University Press, Oxford.

Ghose, Sanjoy (1998), "Distance Representations of Consumer Perceptions: Evaluating Appropriateness by Using Diagnostics," *Journal of Marketing Research*, 35 (May), 137-53.

Godes, David and Dina Mayzlin (2004), "Using Online Conversation to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545-560.

Goldenberg, Jacob, Sangman Han, Donald R. Lehmann, and Jae Weon Hong (2009), "The Role of Hubs in the Adoption Process," *Journal of Marketing*, 73 (2), pp. 1-13.

Granovetter, Mark (1973), "The Strength of Weak Ties," *American Journal of Sociology*, 78 (6), 1360-1380.

Gruber, Thomas (2007), "Collective Knowledge Systems: Where the Social Web Meets the Semantic Web," *International Journal on Semantic Web Information Systems*, 3(2), 1-11.

Henderson, Geraldine R., Dawn Iacobucci, and Bobby J. Calder (1998), "Brand Diagnostics: Mapping Branding Effect Using Consumer Networks," *European Journal of Operational Research*, 111 (December), 306-327.

Hinz, Oliver, Jochen Eckert, and Bernd Skiera (2011), "Drivers of the Long Tail Phenomenon: An Empirical Analysis," *Journal of Management Information Systems*, 27 (4), 43-69

Hoberg, Gerard and Gordon Philips (2010), "Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis," *Review of Financial Studies*, 23 (10), 3773-3811.

Huang, Zeng, and Chen (2007), "Analyzing Consumer-Product Graphs: Empirical Findings and Applications in Recommender Systems," *Management Science*, 53 (7), pp. 1146–1164

Hui, Sam K., Yanliu Huang, and Edward I. George (2008), "Model-based Analysis of Concept Maps," *Bayesian Analysis*, 3, 479-512.

Interbrand (2009), "Best Global Brands 2009," (accessed April 13, 2011), [available at http://www.interbrand.com/best_global_brands.aspx]

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.

Joshi, Amit M. and Dominique M. Hanssens (2010), "The Direct and Indirect Effects of Advertising Spending on Firm Value," *Journal of Marketing*, 74 (January), 20-33.

Katona, Zsolt, Peter Zubcsek and Miklos Sarvary (2011), "Network Effects and Personal Influences: Diffusion of an Online Social Network," *Journal of Marketing Research*, 48(3), 425-443.

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

Körner, Christian, Dominik Benz, Andreas Hotho, Markus Strohmaier, and Gerd Stumme (2010), "Stop Thinking, Start Tagging: Tag Semantics Emerge From Collaborative Verbosity," In *Proceedings of the 19th International Conference on World Wide Web* ,WWW '10, ACM, New York, NY, 521-530.

Kothari, S.P. (2001), "Capital Markets Research in Accounting," *Journal of Accounting and Economics*, 31 (September), 105-231.

Lacy, Sarah (2008), "Book Publishers: Learn from Digg, Yelp – Even Gawker", Business Week, August 21, 2008, available at http://www.businessweek.com/technology/content/aug2008/tc20080820_194967.htm.

Lee, Thomas Y. and Eric T. Bradlow (2011), "Automated Marketing Research Using Online Customer Reviews," *Journal of Marketing Research*, 48 (5), 881-894.

LeRoy, Stephen (1989), "Efficient Capital Markets and Martingales," *Journal of Economic Literature*, 27 (December), 1583-1621.

Liu, Yong (2006), "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing*, 70 (3), 74-89.

Luo, Xueming and C.B. Bhattacharya (2006), "Corporate Social Responsibility, Customer Satisfaction, and Market Value," *Journal of Marketing*, 70 (July), 1-18.

Luo, Xueming (2009), "Quantifying the Long-term Impact of Negative Word of Mouth on Cash Flows and Stock Prices," *Marketing Science*, 28 (1), 148-165.

MacInnis, Deborah J. and Kent Nakamoto (1992), "Factors that Influence Consumers' Evaluations of Brand Extensions," working paper, Karl Eller Scholl of Management, University of Arizona.

March, James G. (1991), "Exploration and Exploitation in Organizational Learning," *Organizational Science*, 2, 71-87

Mika, Peter (2007), "Ontologies are on Us: A Unified Model of Social Networks and Semantics," in Web Semantics: Science, Services and Agents on the World Wide Web, International Semantic Web Conference, International Semantic Web Conference (ISWC2005), 5-15.

Miller, George A. (1995), "WordNet: A Lexical Database for English," *Communications of ACM*, November, 38 (11), 39-41.

Mizik, Natalie and Robert Jacobson (2004), "Stock Return Response Modeling," in Assessing Marketing Strategy Performance, Christine Moorman and Donald R. Lehmann, eds. Boston: Marketing Science Institute, 29-46.

___ and ___ (2008), "The Financial Value Impact of Perceptual Brand Attributes", *Journal of Marketing Research*, 47 (1), 15-32.

___ and ___ (2009), "Financial Markets Research in Marketing," *Journal of Marketing*, 46 (3), 320-324.

Netzer Oded, Ronen Feldman, Jacob Goldenberg, and Fresco Moshe (2012), "Mine Your Own Business: Market Structure Surveillance Through Text Mining," *Marketing Science*, Forthcoming.

Newman, Mark. E. J. (2001), "Scientific Collaboration Networks. II. Shorted Paths, Weighted Networks, and Centrality," *Physical review E*, 64 (1), pp. 016132 (7). Novak, Joseph D. and D. Bob Gowin (1984), Lerning How to Learn. New York: Cambridge University Press.

Pink, Daniel H. (2005), "Folksonomy," New York Times, http://www.nytimes.com/2005/12/11/magazine/11ideas1-21.html, Retrieved 14 July 2009.

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), 14:1-14:34.

Schott, Thomas (1987), "International Influence in Science: Mathematicians in Denmark and Israel," Social Network, 9, pp. 351-75.

Shugan, Steven M. (1987), "Estimating Brand Positioning Maps Using Supermarket Scanning Data," *Journal of Marketing Research*, 24 (February), 1-18.

Sismeiro, Catarina, Natalie Mizik, and Randolph E. Bucklin (2012), "Modeling Coexisting Business Scenarios with Time-series Panel Data: A Dynamics-based Segmentation Approach," *International Journal of Research in Marketing*, 29.

Stonedahl, Forrest, William Rand, and Uri Wilensky (2010), "Evolving Viral Marketing Strategies," Genetic and Evolutionary Computation Conference (GECCO), July 7-11, Portland, OR, USA.

Smallwood, Denise E. and John Conlisk (1979), "Product Quality in Markets Where Consumers are Imperfectly Informed," *Quarterly Journal of Economics*, 93 (1), 1-23.

Sorescu, Alina B. and Jelena Spanjol (2008), "Innovation's Effect on Firm Value and Risk: Insights from Consumer Packaged Goods," *Journal of Marketing*, 72 (March), 114-132.

Spiegelhalter, David, Andrew Thomas, and Nicky Best (2000), "WinBUGS Version 1.3 User Manual," http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml.

Srinivasan, Shuba, Koen H. Pauwels, Jorge M. Silva-Risso, Dominique M. Hanssens (2009), "Product Innovations, Advertising, and Stock Returns," *Journal of Marketing*, 73 (January), 24-43.

Strohmaier, Markus, Christian Körner, Roman Kern (2010), "Why Do Users Tag? Detecting Users' Motivation for Tagging in Social Tagging Systems," Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, pp. 339-342.

Suchanek, Fabian M., Milan Vojnović, and Dian Gunawardena (2008), "Social Tags: Meaning and Suggestions," In *Proceedings of the 17th ACM conference on Information and Knowledge Management*, CIKM '08 ACM, New York, NY, 223-232.

Sujan, Mita and James R. Bettman (1989), "The Effects of Brand Positioning Strategies on Consumers' Brand and Category Perceptions: Some Insights from Schema Research," *Journal of Marketing Research*, 26 (November), 454-67.

Teichert, Thorsten A. and Katja Schöntag (2010), "Exploring Consumer Knowledge Structures Using Associative Network Analysis," *Psychology and Marketing*, 27 (4), 369-98.

Tellis, Gerard J. and Joseph Johnson (2007), "The Value of Quality," *Marketing Science*, 26 (6), 758-773.

Thom-Santelli, Jennifer, Michael J. Muller, and David R. Millen (2008), "Social Tagging Roles: Publishers, Evangelists, Leaders," *CHI 2008 Proceedings on Online Social Networks*, April 5-10, Florence, Italy, pp. 1041-1044.

Tirunillai, Seshadri and Gerard J. Tellis (2012), "Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance," *Marketing Science*, 31 (2), 198-215.

Tonkery, D. (2010). "Amazon Versus Ebook Publishers," *Searcher,* 18(4), pp. 41-43.

Villanueva, Julian, Shijin Yoo, and Dominique M. Hanssens (2008), "The Impact of Marketing-Induced Versus Word-of-Mouth Customer Acquisition on Customer Equity Growth," *Journal of Marketing Research*, 45 (February), 48-59.

Wasserman, S. and Faust, K. (1994), Social Network Analysis: Methods and Applications, New York: Cambridge University Press.

Watts, D. J. and P. S. Dodds (2007), "Influentials, Networks, and Public Opinion Formation," *Journal of Consumer Research*, 34, 441-458.

Weimann, Gabriel (1994), "The Influentials: People Who Influence People," Albany, NY: State University of New York Press.

Wermers, Russ (2003), "Is Money Really 'Smart'? New Evidence on the Relation Between Mutual Fund Flows, Manager Behavior and Performance Persistence," working paper, Robert H. Smith School of Business, University of Maryland, (May), (accessed June 1, 2009), [available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=414420#]

Wortham, Jenna (2011), "YouTube Founders Revamping a Site for Link Sharing," published September 11 2011, available at http://www.nytimes.com/2011/09/12/technology/youtube-founders-aim-to-revamp-delicious.html?_r=1&smid=tw-nytimesbusiness&seid=auto
Yuan, Ying and David P. MacKinnon (2009), "Bayesian Mediation Analysis," *Psychological Methods*, Vol. 14, No.4, 301-322.

Zaltman, Gerald and Robin Higie Coulter (1995), "Seeing the Voice of the Customer: Metaphor-Based Advertising Research," *Journal of Advertising Research*, 35 (July-August), 35-51.

___ (1997), "Rethinking Market Research: Putting People Back In," *Journal of Marketing Research*, 35 (November), 424-37.

Zhang, Jie, Michel Wedel, and Rik Pieters (2009), "Sales Effects of Attention to Feature Advertisements: A Bayesian Mediation Analysis," *Journal of Marketing Research*, 669-681.