Taylor & Francis
Taylor & Francis Group

Using Interest Graphs to Predict Rich-Media Diffusion in Content-Based Online Social Networks

E. Mitchell Church[1], Lakshmi S. Iyer[2], and Xia Zhao[2]

[1]*Department of Management and Decision Sciences, Coastal Carolina University, Conway, South Carolina, USA*
[2]*Department of Information Systems and Supply Chain Management, University of North Carolina at Greensboro, Greensboro, North Carolina, USA*

Rich-media, pictures, and videos, are becoming an increasingly important aspect of online social networks. Unlike social networks, where users are connected primarily because of being friends, peers, or co-workers, content-based networks build connections between individuals founded on a shared interest in rich-media content. In this study, "interest-graphs" comprised of these content-based connections were examined. As shown, interest graph analysis provides important advantages over traditional social network analysis to identify valuable network members and predicting rich-media diffusion.

Keywords   business intelligence; content-based network; interest graphs; social graphs; social network analysis; visual word of mouth

INTRODUCTION

Social networks built around rich-media digital content are becoming increasingly popular and more abundant. These content-based networks (CBN) (Church et al., 2013) contain rich-media, such as pictures and videos related to commercial products and services, and the sharing of this digital content make up the backbone of network usage. In this study, the authors show that the diffusion of rich-media information within CBN represents a form of visual word-of-mouth communication (vWOM). vWOM was coined as a special form of electronic word-of-mouth (eWOM) communication (Church et al., 2014). eWOM is traditionally focused on textual comments made online (Hennig-Thurau, Gwinner, Walsh, & Gremler, 2004). Since the content diffused in CBN is in visual format, the vWOM diffusion that can aid companies to gain a better understanding of the nature of visual content diffusion and market influence in social media environments were examined.

As people spend more time within online social network sites, it puts pressure on companies to develop marketing strategies tailored to these digital spaces (Sipior, Ward, & Volonino, 2014; Trusov, Bucklin, & Pauwels, 2009). However, despite organizational demand, winning marketing strategies in social media environments remain rare and inconsistent Susarla, Oh, and Tan (2012). This study contributes to efforts in this area by addressing the important issue of network *context*.

Every network has a context, which determines link formation within the network. Users of personal social network sites like MySpace and Facebook form links based on social relationships. Users of LinkedIn are connected based on work or business relationships. While these types of bonds may serve well when attempting to network for a job, or when looking to keep up to date with information related to friends and relatives, they may not be the most suitable for sharing information related to commercial products and services (Agarwal, Rambow, & Bhardwaj, 2009).

To maximize the potential for information diffusion, it is vital that close network nodes must share some similarity around the information that is to be diffused (Lewis, Gonzalez, & Kaufman, 2012). Recent academic research has argued that the networks of many online social network users may lack exactly this similarity. As a result, the friends and family connections present in these networks are not particularly useful for predicting commercial tastes and preferences (Agarwal et al., 2009; Lewis et al., 2012). Friends do not seem to be the deciding factor in what TV shows people watch, the things people buy, or the music people listen to (Garg, Smith & Telang 2011; Susarla, Oh, & Tan, 2012).

Achieving maximum diffusion of product information is essential for companies to realize the many desirable marketing outcomes are tied to eWOM. These outcomes include increased product sales (Chevalier & Mayzlin, 2006), greater success with long-tail market penetration (Stephen & Galak, 2012), improved customer loyalty and satisfaction (Trusov et al., 2009), among others.

In this study, business intelligence (BI) methods were applied, specifically social network analysis (SNA), to visualize

Address correspondence to Lakshmi S. Iyer, University of North Carolina at Greensboro, 1400 Spring Garden St, Greensboro, NC 27402-6170, USA. E-mail: Lsiyer@uncg.edu

and analyze two different network graphs for one company within Pinterest (www.pinterest.com), the largest and fastest-growing CBN. Pinterest has been around for about four years, but only recently has the site come to prominence. While growing quietly since its inception, the popularity of the site has since exploded, leading to Pinterest earning the distinction of becoming the fastest web site to reach 10 million unique users. A recent report found that Pinterest now ranks third among social network sites in terms of number of U.S. visitors, behind only Facebook and Twitter (Miles, & Lacey, 2012). Academic and practitioner research into Pinterest has grown substantially in recent years, with studies appearing across a number of fields, including information systems (Church et al., 2013, 2014), computer science (Sharma, Carroll, & Khune, 2013), and marketing (Belden, 2013).

As shown in this article, companies within a CBN, such as Pinterest, are not members of just one network, but instead they belong to at least two distinct network graphs. Companies participate in a "social graph," which represents the interconnections between a company and their customers. The context for this graph is often a mixture of social connections (i.e., friends) and interest-based connections (i.e., people who share interest in the company's products).

In addition, the content-based nature of Pinterest creates an underlying "interest graph." This network graph is actually created from the diffusion of rich-media content. Since connections in the interest graph are only made when content is shared, it exhibits a network context that is purely interest based. For example, even if two members are connected within Pinterest, they will only appear in the interest graph if they actively share content between one another. Conversely, individuals who share content may appear in the interest graph, but lacking any social relationship, will not appear in the social graph.

It is proposed that because it circumvents the traditional problem of social network context, the interest graph allows an organization to visualize a customer network most closely aligned with its products and services. In this way, interest graph analysis allows for BI and marketing insights not normally possible with traditional SNA. This is the central theme of the study. This question is specifically asked: "How does interest graph analysis extend and improve upon SNA when examining information diffusion in CBN?"

The remainder of the article is organized as follows. First, CBN is discussed in more detail, together with some examples. The methodology employed in both the data collection and analysis for this study is then detailed. Results of the analysis are then discussed, followed by contributions and some recommendations for future research.

CBN: BACKGROUND

As can be seen from Figure 1, pages within CBN appear quite different from those of most social network sites. CBN pages, or "boards" feature little in the way of textual information, or information pertaining to the identity or personal life of the user. In fact, even profile pictures, which have been studied so much in the social network literature (Ellison et al., 2006) are absent. This particular example comes from Pinterest.

Information presentation within CBN pages is dominated by rich-media content to such a degree that we can say these networks are literally built from this content. CBN pages, therefore, exhibit some striking differences from user pages on other social network sites like Facebook, MySpace, or Google+. First, content within CBNs is representative of users, without being directly about them. Instead, CBN content often takes the form of pictures of products and services. This distinction is fundamentally important when considering the business impact of CBN. All content within the CBN preserves a link back to the site from which the content originated. For this reason, images within CBNs can be seen as a form of advertising for content producers. As users collect and share digital content, they create digital pathways that potentially serve as powerful drivers of referral business back to the site of the content producer.

Businesses have been quick to embrace the potential of CBNs as a new form of online social media marketing (Carr, 2012). Firms within the CBN are free to add digital content from their company web sites, and interact and network with other Pinterest users, users that they hope to turn first into CBN followers or "friends," and ultimately into paying customers. "Better Homes and Gardens," an online home improvement and lifestyle magazine, manages a Pinterest network that includes more than 300,000 followers. The company freely distributes large amounts of content through the Pinterest network in an effort to drive traffic to its company-owned web sites where it can sell products and generate advertising revenues. Better Homes and Gardens recently conducted a "Pin & Win Dream Home Contest," in which users of Pinterest were encouraged to create a Better Homes and Gardens themed Pinboard on their own Pinterest page, and then fill the board with content provided by Better Homes and Gardens.

eWOM AND MARKETING OUTCOMES

The study is related to at least two important streams of research. Because we consider the diffusion of information between CBN users, the article is closely tied to research around eWOM communication. Based on Hennig-Thurau et al. (2004), Thorson and Rodgers (2006) define eWOM as "positive or negative statements made about a product, company, or media personality that are made widely available on the internet" (p. 35). eWOM has been shown at times to be more effective than traditional marketing efforts (Bickart & Schindler, 2001). Positive eWOM can increase customer opinions of products and brands (Mudambi & Schuff, 2010), customer loyalty and product sales (Chevalier & Mayzlin, 2006). Social media platforms represent natural avenues for eWOM (Trusov et al., 2009), and

FIG. 1. Pinterest's content-based social networking format.

controlling eWOM within social media has real economic value for organizations.

However, understanding eWOM is a tricky business. Misner (1994) calls word-of-mouth (WOM) the most effective, and least understood, form of marketing. For example, in online social networks, eWOM is particularly sensitive to network context which refers to the criteria used in forming links between network members. For example, past academic research has shown that within networks that have a context based around friend and family connection (e.g., Facebook, MySpace), eWOM diffusion may be limited as it pertains to commercial products and services since the family connection does not often guarantee that users share similar tastes

and preferences (Agarwal et al., 2009; Lewis et al., 2012). eWOM diffusion benefits most when close network members share commonalities on the information to be diffused (Rogers, 2010).

The current study contribute to this research by studying CBN, which as explained above, are naturally based around an "interest-based" network context. CBN graphs are, therefore, less indicative of social connections, and instead based on shared tastes and preferences. This network context makes them appropriate for the study of the diffusion of product information.

Another limitation often faced in the eWOM literature concerns the largely invisible way in which eWOM information spreads (Garg, Smith, & Telang, 2011). Since eWOM

information often exists in the form of private textual conversations between network members, the actual information being diffused can be difficult to track and analyze. This is another way that CBN offer distinct advantages in terms of analysis. CBN users do not share textual eWOM, but rather rich-media in the form of pictures of products and services. Unlike private textual conversations, this vWOM is visible to all and can easily be tracked as it travels through a network.

vWOM also offers several interesting theoretical implications that make it worthy of study. Humans can absorb a tremendous amount of visual information very quickly. This information is then also more likely to be recalled compared to textual learning (Paivio & Yarmey, 1966). Pictures are especially effective in creating paired associations (Paivio, 1990), which makes them extremely useful in product presentations and advertising Burns, Biswas, & Babin, 1993).

For these reasons, vWOM is a potentially powerful vehicle for the transmission of marketing information. However, the authors are not aware of any research that actively looks to explore the implications of vWOM, which is a significant gap in literature addressed by this study.

Strategic Network Formation

In addition to studying the spread of vWOM, its impact on network formation was also examined. The work, therefore, also contributes to a growing literature around strategic network formation. Strategic network formation seeks to understand the nature of network structures on a deep level, using principles of SNA developed in fields as diverse as sociology, mathematics, and economics (Jackson, 2010). Interest in BI and data analytics has made strategic network formation an important area of study in information systems in recent years (Jourdan, Rainer, & Marshall, 2008; Wixom, Watson, Reynolds, & Hoffer, 2008), and SNA methodologies (Sipior, Ward, Volonino, & MacGabhann, 2013) have shown the benefit of considering underlying network structure in answering questions of interest to the IS community. For example, Susarla, Oh, and Tan (2012) studies the structure of a network of YouTube videos. Hassan (2009) shows the value of network analysis in considering business process performance. Mayzlin and Yoganarasimham (2012) examine the role of network of blogs in online book sales. Oestrecher-Singer and Sundararajan (2012) consider the impact of a network of Amazon.com products on demand effects and long-tail sales performance (Oestrecher-Singer & Sundararajan, 2012). These network studies are perhaps most interesting in that they do not consider human actors, but rather networks of "things." Within this growing area of research, the current work is perhaps most closely related to research that considers networks comprised of strategically placed hyperlinks (Oestrecher-Singer & Sundararajan, 2012) The spread of vWOM examined in the present study results in lasting hyperlink connections, the placement of which offers strategic value to owning organizations.

For example, content producers rely on placement of hyperlinks to direct customers to their sites and generate revenue. As vWOM spreads through the network, it creates lasting a lasting imprint of a company's content that may provide desirable marketing benefits over time (Stephen & Galak, 2012). Strategies for how best to create such hyperlink connections between social media sources now represent an important part of the way that companies compete (Charalabidis, Loukis, & Androutsopoulou, 2014). The authors contribute to and extend the work in this area by investigating the strategic hyperlink formation and placement decisions of a number of real-world users actively competing for attention within a CBN platform.

DATA COLLECTION AND ANALYSIS

In this study, there was a specific interest in looking at two distinct types of information: (1) the spread of vWOM content and (2) the connections between network users. To begin, an illustrative example to show how the CBN platform makes it easy to uncover data useful for BI. This example focuses on a typical company operating within Pinterest. The company that was selected specializes in selling products for scrapbooking. While its small size makes it easy to illustrate, the approach can be extended to examining larger bigger firms in a CBN. Additionally, the company is desirable because it does not have the name recognition of a larger company like Better Homes and Gardens. Such name recognition could unnecessarily confound the analysis.

Data Collection: The Scrapy Web-Crawling Framework

Data for this study comes from Pinterest. In order to collect information on both the user actions and relationships around this company, the Scrapy web crawling framework was employed. Scrapy is developed in Python and extends the programming language's already excellent text-processing utilities in a number of ways that make it useful for web mining work. Scrapy makes use of Xpath selection, a modern text-selection language that uses some elements of regular expression matching to select relevant pieces of HTML code. Once selected, links and text can be extracted and written to JSON output or forwarded back to Scrapy for further crawling. All of this results in spiders that are flexible, robust, and capable of delivering quality data in a number of formats.

One particular feature of Scrapy that makes it attractive for social network data collection is its use of an index that prevents recursive scraping. Because social networks often have multiple paths leading to the same pages, preventing multiple scrapes of the same page and preserving data integrity is always a challenge. Scrapy eliminates these problems by default, by maintaining a history of scraped pages and ignoring requests that would send the spider to the same page twice. Additionally, Scrapy makes asynchronous requests, so that the spider can move through a large number of pages quickly without waiting for a web server to finish a pending request.

Data is returned by Scrapy in JSON format. JSON is a Javascript mark-up language that is being used very often today as a default output of many Application Programming Interfaces (APIs; Crockford, 2006). The data for this study is stored in a Mongo database. MongoDB is an opensource document storage solution designed for document-style storage, making it well-suited to handling a large amount of JSON text (Chodorow & Dirolf, 2010). For this study, Scrapy spiders completed two specific crawls. First, they collected information about all of the vWOM content related to the target company. Next, the spider mapped out the company's surrounding network. To do this, the spider visited the page of each user connected to the target company. Once on the user's page, the spider looked at all instances of vWOM present on the user's page, looking for matches between that user's content and the content of the target company.

Once data was obtained, it was analyzed using NodeXL, an add-on for Microsoft Excel that allows for SNA. The results of this analysis are presented in the next section.

The Social Graph

The authors begin by examining the company's social graph. This graph shows the interrelationships between the target company and its network followers. This graph is "social" in the sense that while these users are linked to the company, they are also linked to one another through the formation of following/follower relationships similar to those seen on other online social networks. Thus, while the individuals that comprise the graph likely share some interest in the products offered by the target company, their interconnections include a mixture of interest-based, friend and family connections. A visual representation of the social graph for the target company is shown in Figure 2.

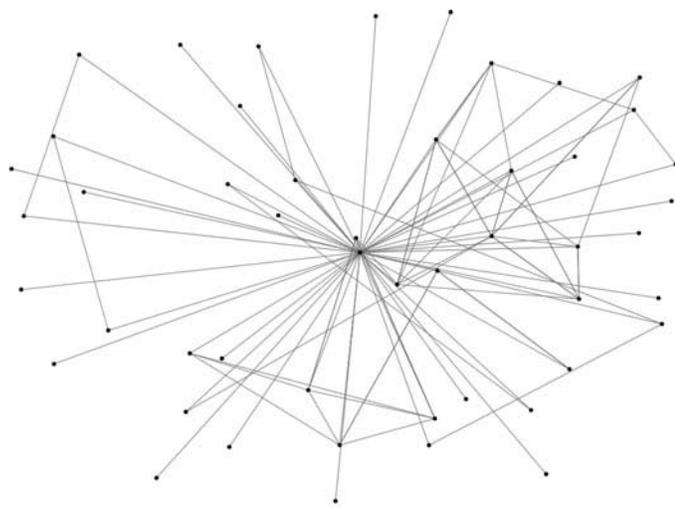Created with NodeXL (http://nodexl.codeplex.com)

FIG. 2. A social graph of the target company, showing the connections between network members.

In order to identify the most important individuals in this graph, several different dimensions of connectedness were considered. Specifically, degree was calculated, between-ness centrality, eigenvector centrality, Page Rank, and the node's clustering coefficient.

Degree is a measure of the number of connections that exist between two network nodes. A high degree means that the node is connected to many other nodes. Centrality is a measure of how important a node is in a given network. Between-ness centrality refers to how often a node is placed between two other network nodes that do not share a connection. The idea is that since these network members are disconnected, and cannot share information, the network position is afforded some advantage as a go-between for these users (Hassan, 2009). In other words, it is a measure of how often the node sits on the shortest path connecting the two users. Eigenvector is another method for measuring node influence. It is closely related to the last measure considered, Page Rank, which is based on the algorithm for evaluating webpage connectedness developed by Google. Finally, a node's clustering coefficient is a measure of its importance related to its surrounding nodes. In other words, it shows how important a node is in its local section of the network (Watts, 1999).

Table 1 shows the top 10% of the nodes based on the different measures of connectedness. These are the most important social nodes in this graph, in the sense that they occupy the most desirable network positions. As can be seen from Table 1, there is a good deal of similarity in the way that nodes perform on the measures of centrality. For this reason, only between-ness centrality was considered going forward. However, since the interest is of the diffusion of vWOM pertaining to tastes and preferences for the target company, it is important to determine whether these important social nodes are strong predictors of vWOM. Using only SNA measures of connectedness, it is difficult to know if these well-positioned individuals, are particularly powerful influences of their surrounding network nodes *on the attributes that impact our target company.*

To investigate vWOM within this network, the diffusion of vWOM content through the network was observed for a period of ten weeks. At the end of ten weeks, the amount of content that diffused from the target company to each surrounding network node was measured. Figure 3 shows the results of this analysis. Square nodes are the most important nodes in terms of social connectedness. Circle nodes are the most active network members in terms of diffusing vWOM content. Diamond nodes are those nodes that are *both* well-connected and active in vWOM diffusion. As can be seen, there is little overlap between connected nodes and vWOM diffusion, with only a single network node performing in the top 10% of nodes in both categories.

The Interest Graph

As shown in the previous section, social connections tend to create networks that are not closely aligned with desirable vWOM diffusion. In this section, an alternative network

TABLE 1
Top 10% of Network Nodes in Terms of Social Network Connectedness

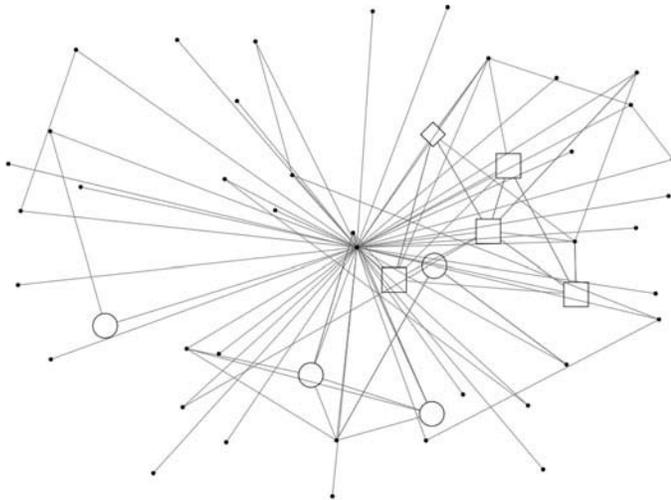| UserID	Degree	Between-ness Centrality	Eigenvector Centrality	PageRank Centrality	Clustering Coefficient
sst	8	5.750	0.038	1.907	0.500
dia	7	3.583	0.036	1.648	0.571
lin	5	0.583	0.030	1.213	0.800
lia	5	0.583	0.030	1.216	0.800
pao	5	1.000	0.030	1.220	0.700

FIG. 3. Improved social graph showing key nodes in terms of connectedness and vWOM diffusion (notes: top connected nodes denoted with squares; most important diffusion nodes denoted with circles; diamond notes represent important nodes based on both criteria).
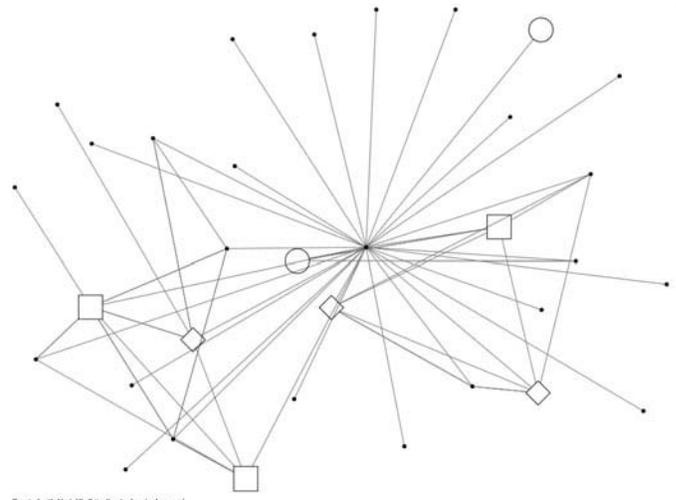

FIG. 4. Improved interest graph created from examining the diffusion of vWOM information within Pinterest (notes: top connected nodes denoted with squares; most important diffusion nodes denoted with circles; diamond notes represent important nodes based on both criteria).

structure was investigated; the interest graph. Unlike the social graph, the interest graph does not consist of voluntary follower/following relationships initiated by network members. Rather, it is built from the spread of rich-media content through the network, evolving organically as users copy content between one-another as a natural part of CBN usage. For vWOM diffusion, it may offer distinct advantages since each network node is an active participant in vWOM diffusion as a prerequisite for admission into the network.

Using interest graph data for the target company, an analysis identical to that conducted for the social graph data above was completed. The interest graph was mapped out using the Scrapy webcrawling framework. Measures of node connectedness were calculated, and vWOM diffusion within the network was observed for a period of ten weeks. Figure 4 shows the results of this analysis. Once again, squares represent the top 10% most connected network nodes. Circles are the top 10% nodes for vWOM diffusion, and diamonds are top performing nodes on both criteria.

Compared to the social graph data, it can be seen that the interest graph results in far more diamond nodes. For the target company, the interest graph results in much greater alignment between measures of social network connectedness and actual vWOM diffusion. In the next section, the implications of these results are discussed in greater detail.

Analysis Extensions

The previous discussion provides a good illustrative example of the advantages offered by interest graphs when considering vWOM. In this section, this example was extended to a much larger set of data to examine the interest and social graphs in more depth. For this work, a sample of companies taken from Pinterest were used again. Within these networks, the unit of analysis is the individual network nodes that surround these companies. These nodes can exist in the company's interest graph, social graph, or both. The variable of interest describes the instances of vWOM diffusion to a particular network node.

Through the use of webcrawling applications, we first examine the follower relationships of all nodes surrounding the target companies. Next, the interest graph for each company was created by looking at all of the company's content, and then evaluating the content of every surrounding network node looking for content in common. When content is found in common between a node and the target company, a link between these nodes in the interest graph was created. Finally, the authors then create closed triads by comparing the content of each node with every other node in the interest graph looking for shared content. When content is found, a link is made between the two nodes that share content in common. In this way, shared content is used as a measure for interest similarity, and forms the context of link formation in the interest graph.

Eight hundred sixty network nodes were examined in total. Each node was evaluated based on its between-ness centrality and clustering coefficient. Additionally, network level measures of size and density were also calculated as control variables. Once these measures were calculated, the nodes were observed over a period of six weeks. At the end of this time, a measure of content diffusion was obtained by examining the content of each network node in both social and interest graphs. Diffusion was measured as the number of new instances in which content from the target companies was now present on the content of each particular node. Table 2 shows the descriptive statistics from this analysis.

As can be seen from the descriptive statistics, interest graphs in the current sample tended to be smaller in size than social graphs. This is to be expected, since the interest graph is comprised only of individuals that share interest in content with our target company. Social graphs, by comparison, contain many individuals that have relationships to each other independent of shared interest. This characteristic of social graphs is the reason they have been the subject of debate around the diffusion of information (Lewis et al., 2012), and a major motivation for the current study.

The next step in the analysis involved the use of statistical methods to determine whether there was a significant difference between the diffusion performance of social and interest graphs. The model is described in Equation 1.

$$D_i = X_0 + X_{1i} + X_{2i} + X_{3i} + e, \tag{1}$$

where D_i represents the instances of diffusion observed for a particular node. X_0 is the intercept, X_{1i} is a Boolean variable determining whether the node belonged to a social (value 1) or interest (value 0) graph. X_{2i} is the node's centrality within its own respective network, X_{3i} is the node's clustering coefficient, and Epsilon (ε) is the standard error. Control variables were also included corresponding to network level metrics of network density and size.

This model was submitted to a Poisson regression procedure using robust standard errors within Stata. Poisson regression is suitable here given the count data format of the dependent variable. The results of the analysis are shown in Table 3.

As seen in Table 3, when considering the entire data set simultaneously, social graph nodes were observed to exhibit on average two and a half fewer instances of information diffusion

TABLE 2
Descriptive Statistics for Network Nodes Considered in Analysis

Variable	Diffusion	Centrality	Clustering Coefficient	Network Density	Network Size
Mean	0.48	7.94	0.17	0.005	620
Standard deviation	3.48	85.4	0.34	0.004	215
Minimum	0	0	0	0.0016	112
Maximum	82	2,673.2	1	0.016	761
N	2,621	2,621	2,621	2,621	2,621

TABLE 3
Results

	vWOM Coefficient		
Variable	Full Data	Social Graphs	Interest Graphs
Graph type (social versus interest)	−2.52 (0.59)**		
Node centrality	0.26 (0.017)	0.003 (0.002)	0.043 (0.19)*
Node clustering coefficient	0.45 (0.384)	0.04 (0.327)	6.95 (3.83)
N	833	659	174

Numbers in parentheses are robust standard errors.
$^*p < 0.05$; $^{**}p < 0.001$.

than interest graph nodes ($t = -4.23$; $p < 0.001$). In the combined data set, node centrality and node clustering were not shown to be significant ($t = 1.54$; $p = 0.124$; $t = 1.17$; $p = 0.242$, respectively).

Because a significant difference between social and interest graphs was observed in the previous analysis, the authors now consider the data sets separately. As seen in Table 3, column two, measures of node centrality and clustering were not significant predictors of diffusion. In other words, traditional methods of SNA do not offer reliable means of predicting information diffusion in the social graphs present in the data set.

The results for interest graphs (Table 3, column three) are much more favorable. Node centrality was observed to be a significant predictor of information diffusion ($t = 2.18$; $p = 0.031$). Interestingly, a node's clustering coefficient was not significant in any portion of the analysis. In the next section, the significance of these results will be discussed, together with their implications.

DISCUSSION

Past research (Agarwal et al., 2009; Lewis et al., 2012) has argued that traditional SNA measures may be somewhat unreliable when examining social graphs for the purpose of predicting eWOM diffusion of marketing information. This results from a lack of similarity between close network nodes in the network graph around the information to be diffused (Rogers, 2010). In this study an analysis of interest graphs have been presented, networks of users where connections are based not on social relationships but on shared interest in digital content. Compared to social graphs, interest graphs benefit from a context firmly centered around shared interest in information pertaining to commercial products and services. As such, the desired node similarity is inherent in interest graph connections, which the authors posited would make them more suitable for the diffusion of marketing information.

In the present analysis, the authors focused on using traditional measures of SNA to identify well-positioned nodes in the network graphs of one target company within Pinterest. Well-positioned nodes have numerous advantages in terms of passing on information, and ideally these nodes would be the most important for diffusion of Pinterest's rich-media content. The identification of such nodes represents an important question for companies looking to market products, not just within Pinterest but within any social network platform.

It has been shown that by considering interest graphs within CBN such as Pinterest, it is possible for companies to better predict where these key network nodes are located. Specifically, connected nodes in the interest graph are more closely aligned with information diffusion than are connected nodes in the social graph, thus providing tangible evidence of the importance of network context in eWOM diffusion.

In addition to being able to identify nodes that are both well-connected and important for diffusion, the study also offers numerous other insights for organizations. To give one example, the ability to observe information diffusion over time provides a valuable means of identifying nodes that may offer opportunities for expanding the company's existing network. For example, companies may actually want to focus on circle nodes that occupy the periphery of the interest graph. These network actors have shown an interest in the content that the central company has to offer. In fact, they are some of the most active members in terms of talking about the company, yet they are not connected to many of the company's followers. While they are not active members of any network graph related to our target company, it may be that these "long-tie" nodes are connected to users that do business with the company's competitors, or that they offer great opportunities for expanding the follower network into unexplored parts of the CBN (Centola & Macy, 2007). Once these nodes have been identified using the methods provided in this study, they can then be allocated additional marketing resources in order to improve future diffusion of marketing messages.

CONTRIBUTIONS

The first goal of this study was to show managers the importance and potential of the interest graph in terms of improving organizational performance around eWOM in social network environments. The study makes a significant contribution in this regard in that the authors lay out a method for companies to use to follow their content as it spreads from one user to another. This allows for a better prediction the direction and scale of content diffusion. Moreover, it is shown that this diffusion creates lasting network connections with significant potential for achieving marketing insights. This, in turn, will allow researchers and practitioners alike to better anticipate the outcomes of eWOM marketing initiatives CBN.

Another major contribution of the study is its development of a method for fully explicating CBN network structure. A major challenge in social network research concerns the fact that network structures are typically only viewed cross-sectionally (Jackson, 2010). As a result, only a snapshot of the network is visible, making it impossible to observe the network member actions that created the observed network structure (Jackson, 2006). This is problematic because without being able fully explain the way that the network arises as a function of observed actions, it is not possible to identify the causal effects that underlie network formation (Mayer, 2009). Put simply, if you do not know how the network was made, it is dangerous to try and study it. Because the authors are able to fully visualize the diffusion process that precludes the formation of the interest graph, they are able to speak to these causal effects. This represents a significant improvement over previous examinations of information diffusion in the presence of shared interests.

CONCLUSION

This study has shown how companies can use analytics and data mining techniques to visualize the actions and network positions of companies operating within brand outposts in the Pinterest interest-based social network. The goal of the article was to provide an overview of what can be done with some simple and widely available web-crawling tools, and discuss the potential value of the data visualization these tools can achieve. In this study, some preliminary and promising evidence has been provided for the value of the interest graph in predicting "hot spots" for eWOM diffusion. For companies interested in making the move to Pinterest or another interest-based network, this study should provide some good guidance in terms of how to go about studying their own network and the network positions of their competitors.

As with any study, there are some limitations. SNA measures considered in the study represent a fraction of what is possible with this type of analysis. Future research could examine the impact of more of these variables and methods. For example, structural holes and brokerage (Burt, 1995), two closely related concepts that deal with a nodes ability to directly capitalize on structural gaps within a surrounding network, would be a good place to begin. Additionally, with a larger sample, more advanced statistical techniques could be used to make stronger and more elaborate conclusions about the role of community response in follower growth. The authors intend to explore many of these issues, and hope that this research encourages companies and academics alike to understand and realize the potential of this type of analysis.

AUTHOR BIOS

E. Mitchell Church is an assistant professor of Information Systems in the E. Craig Wall Sr. College of Business Administration at Coastal Carolina University in Conway, SC. His research interests include online consumer behavior and the economic value of social media information. He has been published research in the *Journal of Electronic Commerce Research* and proceedings of national and international including the Americas Conference for Information Systems (AMCIS), the International Conference for Information Systems (ICIS), the Conference on Information Systems and Technology (CIST) and the Global Information Technology Management Association (GITMA). His teaching interests include the Management of Information Systems and Information Systems Project Management.

Lakshmi S. Iyer is an associate professor and director of graduate programs in the Information Systems and Supply Chain Management Department at the University of North Carolina at Greensboro. Her research interests are in the area of BI, knowledge management, emerging technologies and its organizational impact, and social inclusion in computing. Her research work has been published in *Communications of the AIS, Journal of Association for Information Systems, European Journal of Information Systems, Communications of the ACM, Decision Support Systems, eService Journal, Journal of Electronic Commerce Research, International Journal of Business Intelligence Research, Information Systems Management*, and others.

Xia Zhao is an associate professor of Information Systems at the Bryan School of Business and Economics, The University of North Carolina at Greensboro. She received her PhD degree in management science and information systems from The University of Texas at Austin. She has published papers in *Journal of Management Information Systems, Production and Operations Management, Decision Support Systems, Information Systems Frontier, IEEE Computer, International Journal of Electronic Commerce*, and many conference proceedings. She has recently won the Bryan School Junior Faculty Research Excellence Award and the Emerald Citation of Excellence Award. She is a senior editor of Decision Support Systems.

REFERENCES

Agarwal, A., Rambow, O., & Bhardwaj, N. (2009). Predicting interests of people on online social networks. In *Computational Science and Engineering, CSE'09. International Conference on, 4*, 735–740). Miami, FL: IEEE.

Belden, C. (2013). Paid, earned, and owned media: Convergence in social media. *Journal of Digital & Social Media Marketing*, *1*(3), 243–250.

Bickart, B. & Schindler, R. M. (2001). Internet forums as influential sources of consumer information. *Journal of Interactive Marketing*, *15*, 31–40. doi:10.1002/dir.1014

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *Ucinet for windows: Software for social network analysis*. In Harvard analytic technologies.

Burns, A. C., Biswas, A., & Babin, L. A. (1993). The operation of visual imagery as a mediator of advertising effects. *Journal of Advertising*, *22*(2), 71–85.

Burt, R. S. (1995). Structural holes: The social structure of competition. Cambridge, MA: Harvard University Press.

Carr, K. (2012). *Pinterest Marketing for Dummies*. Hoboken, NJ: John Wiley & Sons.

Centola, D. & Macy, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology*, *113*(3), 702–734. doi:10.1086/517892

Charalabidis, Y., Loukis, E., & Androutsopoulou, A. (2014). Fostering social innovation through multiple social media combinations. *Information Systems Management*, *31*, 225–239. Accessed Online ahead of Print. doi:10.1080/10580530.2014.923267

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*(3), 345–354.

Chodorow, K. & Dirolf, M. (2010). MongoDB: The definitive guide. Sebastopol, CA: O'Reilly Media, Inc.

Church, M., Zhao, X., & Iyer, L. S. (2013). *Pin it to win it: A study of marketing success in "curation-based" online social networks*. Proceedings of the International Conference on Information Systems Conference. Milan, Italy.

Church, M., Zhao, X., & Iyer, L. S. (2014). *Media interactions and follower growth within content-based online social networks*. Conference on Information Systems and Technology (CIST). San Francisco, CA.

Crockford, D. (2006). The application/json media type for javascript object notation (json). Internet RFC 4627.

Ellison, N., Steinfield, C., & Lampe, C. (2006). Spatially bounded online social networks and social capital. *International Communication Association*, *36*, 1–37.

Evans, D., & McKee, J. (2010). Social media marketing: The next generation of business engagement. Indianapolis IN: Wiley Publishing, Inc.

Garg, R., Smith, M. D., & Telang, R. (2011). Measuring information diffusion in an online community. *Journal of Management Information Systems*, *28*, 11–38. doi:10.2753/MIS0742-1222280202

Hassan, R. (2009). Using social network analysis to measure IT-enabled business process performance. *Information Systems Management*, *26*(1), 61–76. doi:10.1080/10580530802557762

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, *18*(1), 38–52. doi:10.1002/dir.10073

Jackson, M. O. (2006). The economics of social networks. *Econometric Society Monographs*, *41*, 1.

Jackson, M. O. (2010). Social and economic networks. Princeton, NJ: Princeton University Press.

Jourdan, Z., Rainer, R. K., & Marshall, T. E. (2008). Business intelligence: An analysis of the literature. *Information Systems Management*, *25*, 121–131. doi:10.1080/10580530801941512

Lewis, K., Gonzalez, M., & Kaufman, J. (2012). Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences*, *109*, 68–72. doi:10.1073/pnas.1109739109

Liautaud, B. & Hammond, M. (2000). e-Business intelligence: Turning information into knowledge into profit. New York, NY: McGraw-Hill, Inc.

Mayer, A. (2009). Online social networks in economics. *Decision Support Systems*, *47*, 169–184. doi:10.1016/j.dss.2009.02.009

Mayzlin, D., & Yoganarasimhan, H. (2012). Link to success: How blogs build an audience by promoting rivals. *Management Science*, *58*(9),1651–1668.

Miles, J. & Lacey, K (2012). Pinterest Power: Market Your Business, Sell Your Product, and Build Your Brand on the World's Hottest Social Network. New York, NY: McGraw-Hill Professional.

Misner, I. R. (1994). The world's best-known marketing secret: Building your business with word-of-mouth marketing. Austin, TX: *Bard & Stephen*.

Mudambi, S. M. & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly*, *34*,185–200.

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

Paivio, A. (1990). *Mental representations*. New York, NY: Oxford University Press.

Paivio, A., & Yarmey, A. D. (1966). Pictures versus words as stimuli and responses in paired-associate learning. *Psychonomic Science*, *5*(6), 235–236.

Robert, L. P. & Dennis, A. R. (2005). Paradox of richness: A cognitive model of media choice. Professional Communication. *IEEE Transactions*, *48*, 10–21.

Rogers, E. M. (2010). *Diffusion of innovations*. New York, NY: Simon and Schuster.

Sharma, V., Carroll, J., & Khune, A. (2013). *Scaling deep social feeds at Pinterest*. IEEE International Conference on Big Data, 2013 (pp. 777–783). Santa Clara, CA: IEEE.

Sipior, J., Ward, B., & Volonino, L. (2014). Benefits and risks of social business: Are companies considering e-discovery? *Information Systems Management*, *31*(4), 328–339. doi:10.1080/10580530.2014.958031

Sipior, J. C., Ward, B. T., Volonino, L., & MacGabhann, L. (2013). A framework for the e-discovery of social media content in the United States. *Information Systems Management*, *30*(4), 352–358. doi:10.1080/10580530.2013.832965

Stephen, A. T., & Galak, J. (2012). The effects of traditional and social earned media on sales: A study of a microlending marketplace. *Journal of Marketing Research*, *49*(5), 624–639.

Susarla, A., Oh, J. -H., & Tan, Y. (2012). Social networks and the diffusion of user-generated content: Evidence from YouTube. *Information Systems Research*, *23*(1), 23–41. doi:10.1287/isre.1100.0339

Thorson, K. S. & Rodgers, S. (2006). Relationships between blogs as eWOM and interactivity, perceived interactivity, and parasocial interaction. *Journal of Interactive Advertising*, *6*(2), 5–44. doi:10.1080/15252019.2006.10722117

Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing. *Findings from an Internet Social Networking Site. Journal of Marketing*, *73*, 90–102.

Watts, D. J. (1999). Small worlds: The dynamics of networks between order and randomness. Princeton, NJ: Princeton University Press.

Wixom, B. H., Watson, H. J., Reynolds, A. M., & Hoffer, J. A. (2008). Continental airlines continues to soar with business intelligence. *Information Systems Management*, *25*, 102–112. doi:10.1080/10580530801941496

Zhao, H. V., Lin, W. S., & Liu, K. J. R. (2011). Behavior dynamics in media-sharing social networks. Cambridge, MA: Cambridge University Press.