Identifying Key Users in Online Social Networks: A PageRank Based Approach

Completed Research Paper

By Julia Heidemann, Mathias Klier, Florian Probst

Abstract

Online social networks evolved into a global mainstream medium that generates an increasing social and economic impact. However, many online social networks face the question how to leverage on their fast growing popularity to achieve sustainable revenues. In that context, particularly more effective advertising strategies and sophisticated customer loyalty programs to foster users' retention are needed. Thereby, key users in terms of users' connectivity and communication activity play a decisive role. However, quantitative approaches for the identification of key users in online social networks merging concepts and findings from research on users' connectivity and communication activity are missing. Based on the design science research paradigm, we therefore propose a novel PageRank based approach bringing together both research streams. To demonstrate its practical applicability, we use a publicly available dataset of Facebook.com. Finally, we evaluate our novel PageRank based approach in comparison to existing approaches, which could alternatively be used.

Keywords: Online Social Network, PageRank, Weighted Activity Graph, Centrality Measures, Social Network Analysis, Design Science Research

Introduction

Since the first recognizable online social network (OSN) SixDegrees.com launched in 1997 (Boyd and Ellison 2007), numerous OSN such as Facebook.com, MySpace.com, and LinkedIn.com became popular Internet platforms, which connect people around the globe. The active use of OSN enjoys great popularity both in private and corporate context. While in 2008 41% of the US Internet user population visited OSN at least once per month, an estimated 52% of all US Internet users will be regular OSN visitors by 2013 (Williamson 2009b). Worldwide the fast growing number of OSN users reached its latest peak on February 4, 2010, when Facebook.com celebrated six years in business and its number of active users exceeded 400 million (Facebook 2010). A couple of weeks later, Facebook.com even surpassed Google.com to become the most visited website of the week in the US (Dougherty 2010). Thus, this technical and social phenomenon evolved into a global mainstream medium that generates an increasing social and economic impact. Therefore, media and IT companies have been acquiring OSN for considerable amounts. In 2005, for example, the media company News Corporation acquired the OSN MySpace.com for US$ 580 million (BBC 2005), and two years later, Microsoft paid US$ 240 million for a 1.6% minority interest in the OSN Facebook.com (MSNBC 2007).

Despite the rising number of users, the purchase prices for OSN are also being considered critically. For instance, Martin Sorrell, CEO of the WWP Group, seriously questioned the valuation of Facebook.com at US$ 15 billion (Andrews 2009). In fact, OSN face the question how to leverage on their fast growing popularity to achieve sustainable revenues. For example, many OSN are not sure how to generate adequate revenues through advertising and membership fees (Clemons 2009; Lu and Hsiao 2010). This is critical, since nowadays the majority of OSN relies on the advertisement based and/or the two-tiered business model, the latter meaning that basic services are offered for free and premium services are provided for a fee (Riggins 2003). Particularly these business models pose major challenges to OSN providers: On the one hand, more effective advertising strategies are needed in order to remain financially viable (Wen et al. 2009). Even though worldwide advertisement spending on OSN are expected to grow from US$ 2.0 billion in 2008 to US$ 3.5 billion in 2013 (Williamson 2009a), OSN often do not know how to unleash this potential. Consequently, there are already indicators for unexpected low advertising sales (Delany et al. 2008). MySpace.com for instance, recently “has fallen ‘significantly’ short of expectations and is jeopardising a critical US$ 900 million [...] agreement with Google” (Edgecliffe-Johnson and Li 2009). On the other hand, OSN need to foster users’ retention, i.e. they need to ensure that users don’t leave the OSN or become inactive, since “retention of users and virality are crucial to growth and survival of large online social networks” (Nazir et al. 2009). Especially for OSN operating under the two-tiered business model, acquiring and retaining users that are willing to pay fees for premiums services is essential.

To overcome these challenges and to tap the enormous potential originated by the dramatic increase in the popularity of OSN, key users play a decisive role (Bampo et al. 2008; Xu et al. 2008; Xu et al. 2009). In our context, a key user is characterized by one or more of the following aspects: (1) He or she can affect a large number of his or her friends, acquaintances, or other users in an OSN. Such a user can for instance be addressed in marketing campaigns to achieve a high awareness of a product or service (Zahng et al. 2010). This strategy is very promising, since Ray et al. (2010) found “that people in the US generate more than 500 billion online impressions *on each other* regarding products and services” and that only “16% of online consumers generate 80% of these impressions”. (2) He or she is very unlikely to leave the OSN or to become inactive. Such a loyal user can also be helpful to increase stickiness, i.e. the ability to attract and hold users’ interest (Bhat et al. 2002), which is for instance an important success factor for web-based advertisement (Wang and Fesenmaier 2006). (3) He or she is more likely to be willing to pay for premium services in an OSN, which are provided for a fee. Such a user is particularly interesting for OSN operating under the two-tiered business model. To enable more effective and user centric advertising strategies as well as sophisticated customer loyalty programs by addressing users deliberately, approaches for the identification of such key users in OSN are needed. For the identification of key users, users’ connectivity and communication activity are particularly important regarding advertisement in OSN (Cheung and Lee 2010; Ganley and Lampe 2009; Staab et al. 2005; Wen et al. 2009; Xu et al. 2008), users’ loyalty (Algesheimer and Von Wangenheim 2006; Xu et al. 2009), and users’ willingness to pay for services in OSN (Oestreicher-Singer and Zalmanson 2009). However, even though studies emphasize the importance of both a user’s connectivity *and* communication activity (Ganley and Lampe 2009; Staab et al. 2005; De Valck et al. 2009; Willinger et al. 2009), quantitative approaches for the identification of key users in OSN merging both aspects are missing. Therefore, we propose a novel PageRank based approach for identifying key users in OSN bringing together concepts and findings from both research streams. In addition, we demonstrate the practical applicability by using a publicly available dataset of Facebook.com and evaluate our novel PageRank based approach in comparison to existing approaches, which could also be used to identify key users in OSN.

The paper is based on the design science research paradigm and in particular on the guidelines for conducting design science research by Hevner et al. (2004). Since Hevner et al. (2004) do not propose an approach for structuring and organizing design science research contributions, we follow Peffers et al. (2008) and their "nominal process model for the conduct of design science research", which is based on the guidelines by Hevner et al. (2004) and contains six activities. Hence, after the discussion of the general relevance of the problem and its motivation within this introduction (activity 1: *"problem identification and motivation"*), we specify the problem context for which the novel approach is relevant and review prior research on users' connectivity and communication activity in OSN. Afterwards, we identify the research gap (activity 2: *"define the objectives for a solution"*). In the third section, we develop our artifact as a novel PageRank based approach for the identification of key users in OSN (activity 3: *"design and development"*). The penultimate section illustrates the applicability of the artifact (activity 4: *"demonstration"*) by using a publicly available dataset of Facebook.com. Furthermore, the artifact's utility (activity 5: *"evaluation"*) is extensively assessed in comparison to *"competing artifacts"*. Finally, the last section summarizes our results and provides an outlook on future steps (activity 6: *"communication"*).

Problem Context and Related Work

After the identification of the problem and its motivation in the previous section, we specify the problem context. Subsequently, we focus on relevant literature regarding the identification of key users in OSN. Thus, we review prior research on users' connectivity and communication activity in OSN. Drawing on these two research streams, we finally identify the research gap.

Problem Context

Boyd and Ellison (2007) define OSN as "web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system". Aroused by the web 2.0 boom, OSN have evolved into a mass medium, where users present themselves to a broad public and establish or maintain connections to other users. Hence, OSN provide a basis for "maintaining social relationships, for finding users with similar interests, and for locating content and knowledge that has been contributed or endorsed by other users" (Mislove et al. 2007). Particularly the aspect of networking, i.e. establishing and maintaining connections between users, plays a decisive role. Thereby, the visibility and searchability of the users' social network of friends, or at least acquaintances, is a distinctive feature of OSN. Thus, OSN can "create substantial value for the individuals who participate in them, the organizations that sponsor them, and the larger society in multiple ways" (Agarwal et al. 2008). However, the majority of OSN relying on advertisement based and/or two-tiered business models face the challenge to tap the enormous potential originated by the dramatic increase in the popularity of OSN in order to generate sustainable revenues (Clemons 2009; Lu and Hsiao 2010). Therefore, approaches for the identification of key users in OSN are needed to enable for instance more effective advertising strategies (e.g. viral marketing campaigns, targeted marketing) and sophisticated customer loyalty programs by addressing users deliberately (Bampo et al. 2008; Xu et al. 2008; Xu et al. 2009). In this context, literature indicates that particularly users' connectivity based on social structures in the network and users' communication activity are essential (Ganley and Lampe 2009; Kiss and Bichler 2008; Staab et al. 2005; De Valck et al. 2009; Willinger et al. 2009).

Users' Connectivity in Online Social Networks

Users' connectivity in OSN is primarily based on the structural characteristics of the network, i.e. patterns of connections among users (cf. Oinas-Kukkonen et al. 2010). Prior research suggests that a user's connectivity plays a decisive role for the identification of key users in OSN. Wen et al. (2009) for instance point out that a user's connectivity in the whole network could be a significant factor that may impact advertising effectiveness in OSN. This is underpinned by further studies, which illustrate that well-connected users, i.e. users with many direct and indirect connections to other users, are particularly important for OSN, as they can be highly relevant for the promotion of brands, products, and viral marketing campaigns (Domingos and Richardson 2001; Kiss and Bichler 2008; Staab et al. 2005; De Valck et al. 2009). Moreover, well-connected users tend to be more loyal, as for example every additional direct or indirect connection raises a user's barrier to leave the network (Algesheimer and Von

Wangenheim 2006; Xu et al. 2009). Thus, a user's connectivity based on the structural characteristics of the network needs to be considered when identifying key users in OSN.

In general, structural characteristics have been extensively studied for instance to understand and explain human behavior in multiple social networks (Monge and Contractor 2003; Nohria and Eccles 1992; Shapiro and Varian 1999). Thereby, particularly interesting elements in the context of OSN include social capital (Burt 1992; Granovetter 1974) and embeddedness (Saxenian 1994; Uzzi 1997). The structure invoked by the binary connections among users in OSN is mostly perceived as a set of nodes (users), and a set of undirected edges (ties or in the following social links) connecting pairs of nodes (Adamic and Adar 2003; Bampo et al. 2008). These nodes and undirected edges determining the network structure can be represented by a graph (Wasserman and Faust 1994), as shown in Figure 1.

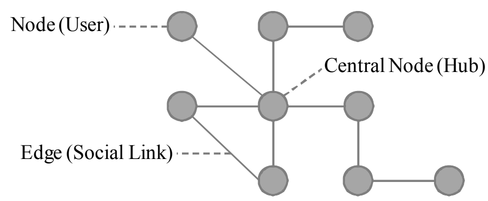

Figure 1. Visualization of the Network Structure (Social Graph)

Since this graph is based on binary social links among users irrespective of their actual interactions, it is usually called social graph (Benevenuto et al. 2009; Wilson et al. 2009). Its visualization especially highlights so-called hubs (Bampo et al. 2008), i.e. users who have an exceedingly large number of social links to other users. Users who are in such a hub position (Constant et al. 1996) are characterized by a great potential for communication and interaction within networks. Hence, OSN allow users to draw on resources from others in the network and to leverage connections from multiple social and geographically dispersed contexts (Haythornthwaite 2002). Thereby, the whole network structure, i.e. direct and indirect connections, plays a decisive role when identifying key users in OSN. Kiss and Bichler (2008) for example emphasize that a connection to a user with many connections is more valuable than to a user with only one or no further connection. Therefore, direct and indirect connections need to be considered when identifying key users in OSN.

Approaches for the identification of important nodes that consider direct and/or indirect connections in networks can be found not only in social network analysis, but also in many other fields for instance in biology for the identification of genes (e.g. Özgür et al. 2008) or in scientometrics for the ranking of scientific journals (e.g. Bollen et al. 2006). These approaches' interpretations highly depend on the particular context (Borgatti 2005; Borgatti and Everett 2006; Freeman et al. 1980). For the specific context of social networks, several measures have been suggested to identify influential and prestigious nodes (Bonacich 1972; Bonacich 1987; Scott 2000; Wasserman and Faust 1994). Additional measures indicate the social influence of nodes on other nodes in a network (Friedkin 1991) or assess a node's integration into a network (Valente and Foreman 1998). The three most common centrality measures to quantify the centrality of a certain node in social networks are presented in Freeman's article "Centrality in Social Networks: Conceptual Clarification" (Freeman 1979): Degree centrality, closeness centrality, and betweenness centrality. The first centrality measure called degree centrality represents the simplest instantiation of centrality, assuming that a node with many direct connections to other nodes is central to the network. The second measure named closeness centrality expands the definition of degree centrality by focusing on how close a node is to all other nodes in the network. The idea behind the third measure referred to as betweenness centrality is that if a node is more often on the shortest paths between other nodes, it is more central to the network. A fourth popular centrality measure, namely eigenvector centrality, is proposed by Bonacich (1972). Eigenvector centrality extends the logic of degree and closeness centrality, since a node's connectivity in the whole network is incorporated (Bolland 1988). Thus, eigenvector centrality tries to quantify the centrality of a node in terms of the global or overall structure of the network, and pays less attention to local patterns (Hanneman and Riddle 2005). To calculate the centralities of the nodes in the network, eigenvector centrality uses the primary eigenvector of a graph's adjacency matrix (Rodriguez 2008). Thereby, the adjacency matrix represents, which nodes of the graph are adjacent, i.e. connected by an edge (the formal representation of a graph's adjacency matrix can be found in the third section). For a detailed description of how to calculate eigenvector centrality and the primary eigenvector see for instance Kiss and Bichler (2008) or Newman (2003b). The primary eigenvector has been applied extensively to rank nodes in all types of networks. It has been used for instance for the ranking of web pages (Brin and Page 1998; Kleinberg 1998; Xing and Ghorbani 2004) and to evaluate the influence of scientific journals (Bollen et al. 2006; Pinski and Narin 1976), articles, and authors (Ding et al. 2009; Liu et al. 2008). These approaches acknowledge explicitly that not all connections are equal, as connections to nodes that are themselves influential are assumed to lend a node more influence than connections to less influential nodes (Newman 2003b). Therefore, the concept underlying eigenvector centrality qualifies particularly for the quantification of a user's connectivity in OSN. Thus, approaches based on the primary eigenvector can be conducive to the identification of key users in OSN.

Users' Communication Activity in Online Social Networks

Latest studies show that not only the structural characteristics underlying a user's connectivity, but also the user's communication activity, i.e. the exchange of information for instance via messages or wall posts, is highly relevant for advertising effectiveness, a user's loyalty, and a user's willingness to pay for services in OSN (Cheung and Lee 2010; Ganley and Lampe 2009; Oestreicher-Singer and Zalmanson 2009). Hence, users' communication activity among each other plays an important role for the identification of key users in OSN. Prior research emphasizes the importance of users' communication activity: "No matter what resources are available within a structure, without communication activity those resources will remain dormant, and no benefits will be provided for individuals" (Butler 2001). Ridings and Wasko (2010) further illustrate, how users' retention in online discussion groups increases as communication activity rises. Moreover, recent work in the context of OSN indicates that the value of OSN lies in the communication activity between users (Krasnova et al. 2009; Willinger et al. 2009). Xu et al. (2008) for instance emphasize "that interaction information is invaluable to marketers, more important than the static links". Thus, a user's communication activity should be considered when identifying key users in OSN.

However, high levels of communication activity cannot be taken for granted (Cummings et al. 2002). Thus, prior studies focus on the network that is based on users who actually interact rather than on users connected by mere social links. This network is usually called activity network (Viswanath et al. 2009) and the resulting graph is referred to as activity graph (Nazir et al. 2008). Thereby, nodes represent users and usually directed edges (activity links) represent communication activity between pairs of users. Here, an edge from node A to B exists if and only if the nodes A and B interacted directly with each other in a way that communication activity was initiated by node A and received by node B. Thus, the activity graph is a visual representation of communication activity among nodes in the network irrespective of their social relations. While previous studies on activity networks examined instant messengers or telecommunication networks (Leskovec and Horvitz 2008; Onnela et al. 2007), initial studies in the context of OSN indicate that the activity graph can provide a sound basis for the identification of key users in OSN (Chun et al. 2008; Wilson et al. 2009).

In the activity graph of an OSN all edges between nodes are the same, regardless whether the corresponding users have a strong connection (i.e. interact frequently) or a weak connection (i.e. interact infrequently). However, literature states that there may be stronger and weaker connections between users in social networks (Newman 2004) and in OSN particularly (Gilbert and Karahalios 2009; Kahanda and Neville 2009; Wen et al. 2009; Xiang et al. 2010). In general, strong connections between users are for instance more likely to be activated for information flow and more influential (Brown and Reingen 1987). In contrast, weak connections provide people with access to information and resources beyond those available in their social circle (Granovetter 1973; Granovetter 1983) and bridge cliques of strong connections (Constant et al. 1996). Further studies emphasize that the strength of connections facilitates awareness in the context of electronic referrals (De Bruyn and Lilien 2008) and that the influence of a reference group and word of mouth recommendations strongly depends on the strength of connections (De Valck et al. 2009). In the context of OSN, for instance Wen et al. (2009) conclude that the strength of connections "denotes an irresistible element for [...] advertising". Nevertheless, previous work on activity graphs in OSN does often not distinguish between strong and weak connections and leaves exploration of this facet to future work (Nazir et al. 2008; Viswanath et al. 2009; Wilson et al. 2009). Only a few authors consider the strength of connections based on users' activity when identifying important nodes in customer networks (Kiss and Bichler 2008) or when comparing structural characteristics of social graphs and weighted activity graphs in OSN (Chun et al. 2008). In order to distinguish between strong and weak connections, these studies started to examine each connection's communication activity level. In this context, communication activity can be any sort of interaction among users facilitated by methods provided by OSN, for example messages or wall posts (cf. Schneider et al. 2009). Since almost every OSN provides such infrastructure for communication and transfer of information, the record of communication activities between users can be used to identify which activity link can be considered as strong and weak, respectively (Xiang et al. 2010). Thus, the strength of a user's activity link can be a measure of intensity, duration, intimacy, or exchange of information between users (Barrat et al. 2004; Granovetter 1973). Furthermore, in accordance to the above mentioned findings from research on users' connectivity, Benevenuto et al. (2009) discovered that users do not only interact with directly connected users, but also have significant exposure to users "that are 2 or more hops away". Therefore, not only a user's communication activity represented by the activity graph but also the strength of a user's direct and indirect activity links based on each activity link's communication activity level should be incorporated when identifying key users in OSN.

Research Gap

Multiple authors emphasize the importance of both a user's connectivity *and* activity in OSN (Ganley and Lampe 2009; Staab et al. 2005; De Valck et al. 2009; Willinger et al. 2009). However, to the best of our knowledge, quantitative approaches for the identification of key users in OSN bringing together concepts and findings from both research streams are missing. Therefore, we merge concepts from research on users' connectivity *and* on users' communication activity in order to identify key users in OSN. Figure 2 summarizes the previously introduced concepts and findings from research on users' connectivity and users' communication activity and highlights which aspect of the novel PageRank based approach that is developed in the following section is informed by which research stream.

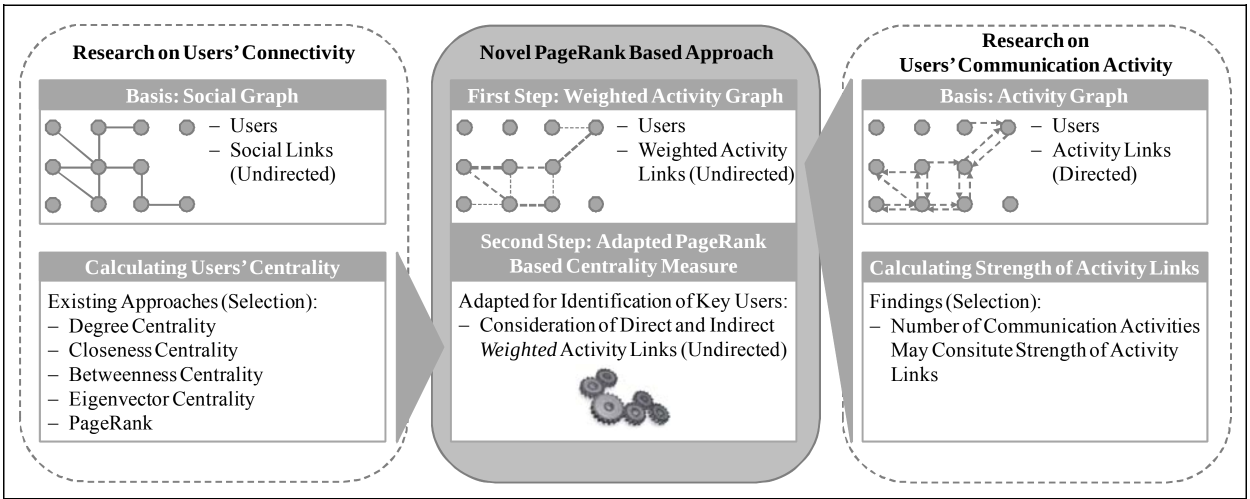

Figure 2. Novel PageRank Based Approach

Novel PageRank Based Approach

For the identification of key users in OSN, we develop a novel PageRank based approach, which is composed of two steps. First, we derive a weighted activity graph. Thus, we incorporate users' communication activity and the strength of users' connections. The weighted activity graph provides the basis for our second step towards the identification of key users in OSN. Therefore, we design a PageRank based centrality measure to determine users' centrality scores in terms of their connectivity in the weighted activity graph. Hence, we consider the structural characteristics of the network based on users' communication activity and direct as well as indirect connections among users. In combination, the weighted activity graph and the PageRank based centrality measure add up to our novel PageRank based approach for the identification of key users in OSN, which merges concepts from research on users' connectivity and communication activity in OSN (cf. Figure 2).

First Step: Deriving the Weighted Activity Graph

The weighted activity graph constitutes the basis of our novel PageRank based approach. First, we define the basic concept of activity graphs. Afterwards, we adapt the activity graph for the identification of key users in OSN and extent the basic concept to account for the strength of users' connections. Thereby, we finally derive the weighted activity graph.

First of all, we define the activity graph as a graph that is based on users who actually communicate with each other instead on users who are connected by a static social link (cf. Chun et al. 2008; Nazir et al. 2008; Wilson et al. 2009). In the activity graph, a node represents a user and an edge (activity link) represents communication activity (e.g. a wall post, a message) between a pair of users. Thus, the activity graph differs from the social graph, as inactive social links are not considered in the activity graph. However, users who are not connected by a social link in the social graph can be connected by an activity link in the activity graph, if there has been communication activity between this pair of users. An example in Figure 3 highlights the possible differences between a social graph and an activity graph.

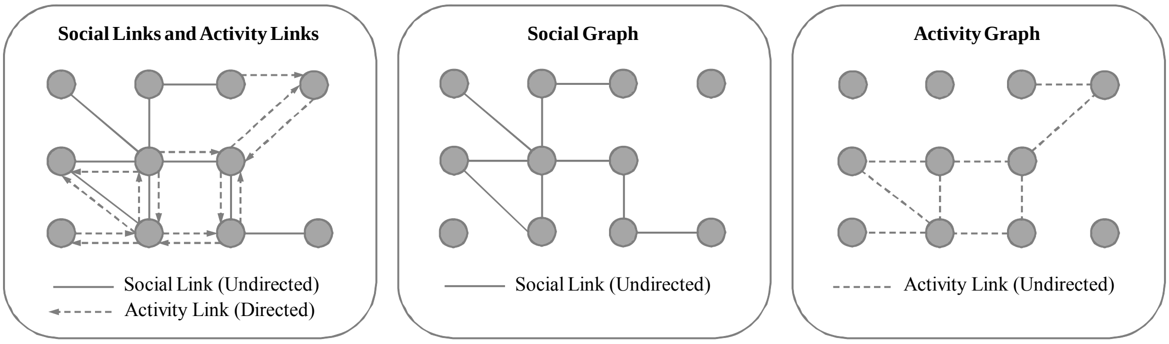

Figure 3. Example: Social Graph vs. Activity Graph

For the identification of key users in OSN, we need to adapt the basic concept of activity graphs. As illustrated in the left picture of Figure 3, activity links are usually assumed to be directed, since communication activity needs an initiator and a receiver. However, the direction of influence (e.g. word of mouth or peer pressure) through communication activity, which can lead to higher advertising effectiveness, users' loyalty, and users' willingness to pay for services in OSN, can be bidirectional. Theoretically, this influence can be classified according to social influence literature as informational social influence and normative social influence (Deutsch and Gerard 1955). While informational social influence means that users rely on information provided by others, normative social influence describes the pressure or assumed need to align the own attitude with that of some other valued users (Bass 1969; Kraut et al. 1998; Wen et al. 2009). In the special case of OSN however, it is hard to tell if the initiator or the receiver of communication activity is more likely to be affected by each type of social influence. For instance, a user who writes a message on another user's wall can either point attention to a brand, product, or service himself or he or she can be influenced by an advertisement placed on the other user's profile (e.g. the user is member of a brand community, i.e. he or she declares himself as a fan of a certain brand). Or a user who receives a lot of messages can be more loyal and likely to stay in a network in almost the same manner than a user who sends a lot of messages. Thus, we model communication activity as undirected activity links to cover bidirectional social influence. Moreover, since pairs of users usually perform reciprocal communication activity, modeling undirected activity links represents to a great extent users' communication behavior in OSN (Chun et al. 2008; Wilson et al. 2009). For instance, for 65% of the users in the largest Facebook.com regional networks all interactions via wall posts were reciprocated (Wilson et al. 2009). Therefore, the loss of information by modeling undirected activity links is limited and the advantages of a bidirectional interpretability of social influence prevail. Hence, we model the activity graph for the identification of key users in OSN by using undirected activity links.

Formally we define the activity graph according to graph theory as $G = (V, E)$, where V denotes a set of nodes (users) and E a set of undirected edges (activity links) (cf. Albert and Barabási 2002; Wassermann and Faust 1994). Thereby, $|V| = n$ represents the number of users in the OSN and $|E| = m$ the number of undirected activity links between them. Two nodes i and j are called adjacent, if and only if they are connected by an activity link $\{i, j\} \in E$. Thus, the activity graph can be represented by its symmetric adjacency matrix $A = (a_{ij}) \in \{0; 1\}^{n \times n}$, whose elements take the value 1 if an undirected activity link connects the nodes i and j, and 0 otherwise. Furthermore, we let t (with $t = 1, 2, ...$) determine a window of time, during which at least once communication activity between two nodes i and j must have occurred in order to create an activity link between them. Thereby, t denotes the number of periods (e.g. days) counted backwards from the point in time when the activity graph is constructed. To account for stronger and weaker activity links (cf. the second section), we further extend our activity graph to include weights of the undirected activity links. Thereby, c_{ij} (with $c_{ij} = 0, 1, ...$) denotes the number of communication activities initiated by node i and received by node j during the time interval stipulated by t (Chun et al. 2008; Onnela et al. 2007). Respectively, c_{ji} (with $c_{ji} = 0, 1, ...$) constitutes the number of communication activities initiated by node j and received by node i. Thus, we define the weight w_{ij} of an undirected activity link between two users i and j as the number of communication activities between that pair of users:

$$w_{ij} = c_{ij} + c_{ji} . \tag{1}$$

Our weighted activity graph $G' = (V', E')$ can again be represented by a symmetric adjacency matrix, where $A' = (a'_{ij})^{n \times n}$, with $a'_{ij} = \begin{cases} w_{ij} & \text{if } \{i, j\} \in E \\ 0 & \text{otherwise} \end{cases}$.

Thus, in contrast to the activity graph, our weighted activity graph does not only contain binary information about whether communication activity occurred at least once between two users i and j during the time interval stipulated by t (existence of an activity link), but also indicates the strength of activity links between users (cf. w_{ij} in formula (1)). Based on this definition, the weighted activity graph derived in the first step provides the basis for the second step of our novel PageRank based approach towards the identification of key users in OSN.

Second Step: Determining Users' Centrality Scores

In the second step, we develop a PageRank based centrality measure to determine each user's centrality score in terms of his or her connectivity in the weighted activity graph. Finally, sorting users by their centrality scores in descending order allows us to define a ranking of key users in OSN.

For the determination of users' centrality scores based on users' connectivity in the weighted activity graph, we consider particularly approaches based on the primary eigenvector of a graph's adjacency matrix. These approaches acknowledge explicitly that not all connections are equal (cf. the second section). Connections to nodes that are themselves influential are rather assumed to lend a node more influence than connections to less influential nodes (Newman 2003b). Since the nodes' connectivity in the whole network is incorporated (Bolland 1988), approaches based on the primary eigenvector try to find well-connected nodes in terms of the global or overall structure of the network, and pay less attention to local patterns (Hanneman and Riddle 2005). Thus, these approaches qualify particularly to rank nodes in a network. Consequently, approaches based on the primary eigenvector of a graph's adjacency matrix have been applied extensively to calculate centrality scores in all types of networks. In single-relational networks, i.e. networks with a data structure that can only represent a single type of relationship, such as social links or undirected activity links, the primary eigenvector can be computed using the power method. Thereby, the power method simulates the behavior of random walkers traversing the network. Hence, the nodes that have a higher probability of being traversed are the most central or important nodes in the network and gain consequently a higher centrality score (Brandes and Erlebach 2005). Single-relational networks can result in different types of graphs. First, there can be strongly connected, aperiodic graphs, i.e. graphs that contain paths from all nodes to all other nodes, whose lengths are sufficiently long (Kemeny and Snell 1960). In this type of graphs, for instance eigenvector centrality can be used to rank nodes (Bonacich 1987). However, graphs as our weighted activity graph G' do not certainly fulfill these properties, as they are not always strongly connected or are even periodic, i.e. there exist isolated nodes (cf. Figure 3). For this second type of graphs, the network's topology can be altered, such that a "teleportation network" is overlaid with the graph G' to construct an irreducible and aperiodic network (Rodriguez 2008). This "teleportation network" introduces an artificial activity link with equal weights between all possible pairs of nodes, even if they are not connected according to our weighted activity graph G'. Thus, when there exists a non-zero probability of "teleportation" to every node in V', the network becomes strongly connected (cf. Rodriguez 2008). This idea was introduced by Brin and Page (1998) who developed the well-known random web surfer model of the PageRank algorithm to rank web pages in the World Wide Web (WWW) (Brin and Page 1998; Page et al. 1999). PageRank interprets the web pages as nodes and directed edges represent the links between them. Thus, PageRank uses the link structure of the WWW as an indicator of an individual web page's importance relative to other web pages by interpreting a link from web page A to web page B as a vote by web page A for web page B.

Following Langville and Meyer (2004), the PageRank $PR(i)$ for a web page i can be defined as:

$$PR(i) = \frac{(1-d)}{N} + d \cdot \sum_{j \in B_i} \frac{PR(j)}{O_j}, \tag{2}$$

such that $||PR||_1$=1 ($||PR||_1$ denotes the L_1 norm of PR). In formula (2), N is the total number of web pages in the network and O_j is the number of outgoing links from page j. B_i denotes the set of web pages pointing to web page i, and d (with $0 \leq d \leq 1$) is a dampening factor that is usually set to 0.85 (cf. e.g. Langville and Meyer (2004) for a detailed derivation of the formula and the optimal dampening factor). As discussed before, methodically PageRank is based on the primary eigenvector of the underlying graph's adjacency matrix. Therefore, in the second part of formula (2) web page i inherits a proportion of centrality from all web pages pointing to it, i.e. all web pages connected to i by ingoing links. To calculate the proportion, which web page i inherits from each web page j in B_i, web page j's rank $PR(j)$ is divided by the number O_j of j's outgoing links. Hence, web page j contributes equally to the centrality of all web pages it points to. Consequently, $PR(i)$ not only depends on the quantity of links, but also on their qualities. Thus, PageRank deviates from degree, closeness, and betweenness centrality by modeling inherited or transferred status (Liu et al. 2008).

Due to its characteristics, the general concept of PageRank seems to be appropriate regarding the identification of key users in OSN. However, for our context we need to adapt the PageRank formula by two modifications. First, a general difference between the WWW and our weighted activity graph in OSN exists. While links in the WWW are directed, the activity links in our weighted activity graph are considered to be undirected. To account for this distinction when identifying key users in OSN, we have to adapt the original PageRank formula accordingly by substituting the set B_i (set of web pages connected to i by ingoing, i.e. directed links) by a set F_i, which represents a set of users connected to i by undirected activity links. The second modification concerns the activity links' weights. A reduction of the activity links' weights to binary values as in the original PageRank formula would entail a severe loss of information (Newman 2004). We therefore have to define a modification of PageRank, which considers the undirected activity links' weights. Our second modification is based on an adaption of the original PageRank's assumption, that a node transfers its centrality evenly to all the web pages it connects to (cf. Xing and Ghorbani 2004). However, in our weighed activity graph the distribution should be determined by the level of communication activity between user i and the users it connects to (cf. the second section). Therefore, we need to consider the weights w_{ij} of each undirected activity link as defined in formula (1). Thus, we remove the dominator O_j and the undirected activity link's weight w_{ij} is added to account for strong and weak connections among users. Finally, we define the formula of our adapted PageRank based centrality measure $S(i)$ for a user i as:

$$S(i) = \frac{(1-d)}{N} + d \cdot \sum_{j \in F_i} S(j) \cdot w_{ij}, \tag{3}$$

such that $||S||_1$=1. We apply the PageRank based centrality measure to determine the centrality score $S(i)$ for each user i based on his or her connectivity in the weighted activity graph. Thereby, we calculate the PageRank based centrality measure recursively. This procedure entails that a user ceteris paribus inherits a higher centrality score from a well-connected user than from a sparsely connected one. Consequently, the network structure and direct as well as indirect connections are considered. Moreover, a user j connected to i by an undirected activity link with a higher weight w_{ij} contributes more to i's centrality score than a user connected by an undirected activity link with a lower weight. Hence, the PageRank based centrality measure accounts for the strength of connections based on each undirected activity link's communication activity level. As the computation of the PageRank based centrality measure can be traced back to the problem of finding an eigenvector (cf. e.g. Brin and Page 1998) the computational complexity can be reduced to $O(n^2)$. Therefore, its computational complexity is manageable with today's computing power. Thus, we developed a PageRank based centrality measure to calculate users' centrality scores in terms of their connectivity in the weighted activity graph. Taken together, the weighted activity graph and the PageRank based centrality measure allow us to identify key users in OSN by sorting users in terms of their centrality scores in descending order.

Demonstration and Evaluation

To demonstrate and evaluate our novel PageRank based approach for the identification of key users in OSN, we use a publicly available dataset of the Facebook.com New Orleans Network. First, we introduce Facebook.com and the dataset. After validating that the dataset exhibits the OSN specific characteristics, we demonstrate the applicability of our novel PageRank based approach and evaluate it in comparison to existing approaches, which could be used to identify key users in OSN. Finally, we highlight and critically discuss limitations of our novel PageRank based approach.

Facebook.com New Orleans Network Dataset

Facebook.com is the largest OSN in the world with over 400 million active users, as of February 2010 (Facebook 2010). As many other OSN, Facebook.com allows users to set up personal profiles. These can include various information, for instance on users' background (e.g. university, hometown), demographics (e.g. date of birth, gender), or personal interests (e.g. favorite music, sports). Furthermore, users are able to establish undirected social links by entering virtual "friendship relationships". One of the most popular mechanisms for communication activity in many OSN in general and in Facebook.com in particular is a message board called "wall" that is included in every profile (Benevenuto et al. 2009; Wilson et al. 2009). Unlike personal messaging or email, wall posts are by default public, meaning that anyone with a Facebook.com account can initiate and receive wall posts. Furthermore, users' history of wall posts can be accessed. However, users can set their wall to be private, so that for instance only users connected by a direct social link are able to access their wall. A special characteristic of Facebook.com is that users can join networks that represent schools, institutions, and geographic regions. Thereby, membership in regional networks is unauthenticated and open to all users. Since the majority of Facebook.com users belong to a regional network, and most users do not modify their default privacy settings, crawling regional networks allows researchers to cover a large fraction of a regional network's users and social links among them (Wilson et al. 2009).

For the demonstration and evaluation of our novel PageRank based approach, we use a dataset provided by Viswanath et al. (2009). This dataset focuses on the New Orleans Network in Facebook.com and consists of two parts. The first part includes a snapshot of the social network structure, i.e. a set of users and social links, which represent "friendship relationships" among these users. The second part of the dataset contains communication activity in terms of wall posts exchanged among the users covered in the first part of the dataset. To gather the social network structure, a crawler started from single users in the New Orleans Network and visited all connected users of these users and their connected users in a breadth first search (BFS) fashion during December 29, 2008 and January 3, 2009. This procedure is consistent with crawls in OSN conducted in prior studies (cf. e.g. Mislove et al. 2007). Earlier research on OSN further indicates that the majority of users in the social graph are part of a single, large, weakly connected component (WCC) (Mislove et al. 2007). Since social links on Facebook.com are undirected, BFS crawling of social links is able to generate complete coverage of the WCC, assuming that at least one of the initial seeds of the crawl is linked to the WCC (Wilson et al. 2009). Prior research verifies that the only inaccessible users could be ones that lie outside the regional network of the crawl, ones who have changed their default privacy settings, or ones that are not connected to the WCC (Mislove et al. 2007; Wilson et al. 2009). Hence, 52% of the users in the New Orleans Network at the time of the crawl could be covered based on the statistics provided by Facebook.com (Viswanath et al. 2009). This corresponds to 90,269 users connected by 1,823,331 undirected social links. However, not all of these users made their wall public. Thus, the entire history of wall posts of a subset of 63,731 (70.6%) of the previously crawled users could be accessed. The first part of the dataset was therefore aligned and represents finally a subset of the Facebook.com New Orleans Network including these 63,731 users connected by 817,090 undirected social links. The second part of the dataset contains 876,687 wall posts initiated and received by these users. Wall posts initiated or received by users who are not included in the subset of 63,731 users are not covered. Each wall post in the second part of the dataset contains information about the initiator of the wall post, the receiver of the wall post, and the time at which the wall post was made. Overall, the wall posts span from September 14, 2004 to January 22, 2009. Taken together, the first and the second part of the New Orleans Network dataset represent the network structure and communication activity of a subnetwork of the Facebook.com New Orleans Network. Therefore, we are able to derive the social graph and the activity graph of this subnetwork.

Characteristics of the Facebook.com New Orleans Network Dataset

To validate that the New Orleans Network dataset exhibits the OSN specific characteristics, we examine the social graph as well as the activity graph and compare them to graphs used in prior research on OSN. For that purpose, we draw on the social graph derived from the first and the activity graph derived from the second part of the dataset. As described in the previous section, the social graph consists of 63,731 users connected by 817,090 undirected social links. To analyze whether the social graph is characteristic of an OSN, we determine the average path length, the average clustering coefficient, and the assortativity coefficient. Table 1 provides an overview of the social graph's statistics compared to social graphs from prior research on OSN.

Table 1. Comparison of Social Graphs’ Statistics					
Network	Users	Undirected Social Links	Path Length	Clustering Coefficient	Assortativity Coefficient
10 Largest Regional Networks of Facebook.com (Wilson et al. 2009)	10,697K	408,265K	4.89	0.164	0.166
Orkut.com (Mislove et al. 2007)	3,072K	223,534K	4.25	0.171	0.072
Facebook.com New Orleans Network Dataset (Social Graph)	63,731	817,090	4.32	0.221	0.177

The average path length of 4.32, which is the average of all pairs’ shortest paths in the social graph, lends credence to the six degrees of separation hypothesis, i.e. that everyone is just a few steps apart in the global social network (Milgram 1967). This so-called “small world” effect is typical for modern networks such as OSN (cf. Schnettler 2009). Furthermore, the New Orleans Network dataset’s social graph has an average clustering coefficient of 0.221. This compares favorably with the average clustering coefficient of 0.164 in the ten largest regional networks in Facebook.com and 0.171 for Orkut.com. Since the average clustering coefficient is higher than those in either similarly sized random graphs or random power law graphs, our average clustering coefficient indicates a tightly clustered fringe that is characteristic of OSN (Mislove et al. 2007). Combined with the relatively low average path length, the average clustering coefficient suggests that our network fulfills the properties of a small world network (Watts and Strogatz 1998; Wilson 2009). The assortativity coefficient indicates the probability for users in a graph to link to other users with a similar number of direct connections. Thereby, an assortativity coefficient greater than zero indicates that users tend to connect with similar users in terms of their number of direct connections, while an assortativity coefficient less than zero denotes that users connect to dissimilar ones (Newman 2002). The assortativity coefficient value of 0.177 closely resembles those for other large OSN (Newman 2003a; Wilson 2009). Thus, connections between users with many direct connections in the social graph are numerous. This core of well-connected users forms the backbone of small world networks, which enables the highly clustered users at the edge of the network to achieve low average path lengths to all other users. To sum it up, the social graph derived from our New Orleans Network dataset is consistent with other social graphs used in prior research on OSN and exhibits the OSN specific characteristics.

To derive the corresponding activity graph, we use the wall posts contained in the second part of the dataset, which represent the most popular form of communication activity between users in OSN (Benevenuto et al. 2009). As described before, the social graph was crawled during December 29, 2008 and January 3, 2009. For our activity graph, we use a fraction of 832,277 wall posts spanning from September 14, 2004 to January 3, 2009. Thus, the end of the considered period of communication activity equals the date when the crawl of the underlying network structure ended. The remainder of 44,410 wall posts spanning from January 4, 2009 to January 22, 2009 were written and received after the social structure was crawled. In the subsection after next, we evaluate our novel PageRank based approach in comparison to alternative approaches for the identification of key users in OSN, which are based on the social graph. Since we do not want to discriminate these approaches, we do not consider the remainder of wall posts for our activity graph. The activity graph $G = (V, E)$ contains the same set of users V as the social graph, with $|V| = 63{,}731$ (cf. Figure 3 for an example). These users are connected by a set of undirected activity links E, with $|E| = 171{,}711$. Thereby, an undirected activity link between a user A and a user B exists if and only if the users A and B interacted during September 14, 2004 and January 3, 2009 at least one time directly with each other, in a way that a wall post was initiated by user A and received by user B, or vice versa. 6,392 (3.7%) of these undirected activity links in our activity graph do not have a corresponding social link in the social graph. This equals to 191,980 (23.1%) wall posts exchanged via these activity links. This finding is in line with prior research on users communication activity in OSN. Benevenuto et al. (2009) for instance discovered that 22.0% of users’ wall posts in their Orkut.com dataset were exchanged between users, which were not connected by a social link in the social graph. To further examine the activity graph, we determine again the average path length (5.39), the average clustering coefficient (0.109), and the assortativity coefficient (0.220). The activity graph’s statistics are in line with the little prior research on activity graphs in OSN. Wilson et al. (2009) for instance display average path lengths in the range of 5.00 to 7.00, average clustering coefficients between 0.030 and 0.080, and assortativity coefficients around 0.200. Chun et al. (2008) show similar properties and comparable correlations between their social graph’s and activity graph’s measurements. To sum it up, the activity graph derived from our New Orleans Network dataset is in line with prior studies on activity graphs in OSN. Since both the social and the activity graph exhibit the OSN specific characteristics, the New Orleans Network dataset provides a sound basis for the demonstration and evaluation of our novel PageRank based approach for the identification of key users in OSN.

Demonstration of the Novel PageRank Based Approach

We demonstrate the applicability of our novel PageRank based approach developed in the third section by using the New Orleans Network dataset. Thereby, we conduct the two major steps of the approach. In the first step, we derive the weighted activity graph as a basis for the identification of key users in the network. In the second step, we determine each user's centrality score in terms of his or her connectivity in the weighted activity graph. Hence, we apply the PageRank based centrality measure developed in the previous section. Sorting users by their centrality scores in descending order allows us to define a ranking of key users based on the New Orleans Network dataset.

First, we build the weighted activity graph on the basis of the collected dataset of the Facebook.com New Orleans Network. Therefore, we use the activity graph derived in the previous subsection and extend it to include weights for the undirected activity links. Consequently, the weighted activity graph contains 63,731 users, which are connected by 171,711 undirected activity links. Since the activity graph is based on wall posts spanning from September 14, 2004 to January 3, 2009, we set the parameter t of our weighted activity graph to t = 1,573 days. Thus, all wall posts initiated and received by the 63,731 users in the activity graph during that period of time are covered. To calculate each undirected activity link's weight w_{ij}, we apply formula (1). Thereby, c_{ij} (respectively c_{ji}) denotes the number of wall posts initiated by user i and received by user j (respectively initiated by user j and received by user i) between September 14, 2004 and January 3, 2009. We represent our weighted activity graph as a symmetric adjacency matrix, where $A' = (a'_{ij})^{63,731 \times 63,731}$, with $a'_{ij} = \begin{cases} w_{ij} & \text{if } \{i, j\} \in E \\ 0 & \text{otherwise} \end{cases}$. Based on the weighted activity graph, we calculate the centrality score $S(i)$ of each user i in the second step.

Therefore, we apply the PageRank based centrality measure defined in formula (3). For that purpose, we need to choose the dampening factor d first. When d takes a value close to 1, the measure places greater emphasis on the structure of the weighted activity graph and less on the teleportation network modeled in the first part of formula (3). However, higher values of d slow down the convergence of the power method (Langville and Meyer 2004). Moreover, Boldi et al. (2005) provided a mathematical analysis of different values for d, finding that values close to 1 do not give a more meaningful ranking than other high damping factors. Pretto (2002) further found that when d changes, the top section of the ranking changes only slightly. As we are especially interested in users with high centrality scores, i.e. the top section, the impact of the dampening factor's choice is limited. Thus, we set the dampening factor to $d = 0.85$. This value is favorable in terms of computational performance and is also often considered as the default value for PageRank calculations in literature (cf. Langville and Meyer 2004). Finally, we calculate the centrality scores $S(i)$ applying the PageRank based centrality measure. For that purpose, we use the software package "NetworkX" for the exploration and analysis of networks and network algorithms (cf. Hagberg et al. 2008). In conclusion, we derive a centrality score $S(i)$ for every user i included in our weighted activity graph. By sorting these centrality scores in descending order, we receive a ranking of users. Based on this ranking of all users included in the New Orleans Network dataset, the key users in the network can be identified by choosing a designated top segment of the ranking.

Evaluation of the Novel PageRank Based Approach

Building on the ranking of identified key users in the network, we evaluate our novel PageRank based approach. As we highlighted in the introduction, in our context the term key user stands for users who can affect a large number of other users in terms of marketing, users who are unlikely to leave an OSN or to become inactive, and/or users who are more likely to be willing to pay for premium services in an OSN. Here, we use users' retention as evaluation criterion, since in particular retention of users is "crucial to growth and survival of large online social networks" (Nazir et al. 2009). Thereby, we define that a user is retained, if he or she stays active in the network. A user's retention strongly affects the retention of other users in the network, since every additional connection raises users' barrier to leave the network (Algesheimer and Von Wangenheim, 2006). In addition, retained users are particularly valuable, as they support a sense of familiarity and community (Figallo 1998; Hagel and Armstrong 1997; Wellman and Gulia 1999). Gan et al. (2009) further illustrate, that as individuals become more involved in online communities, their "habit effect strengthens". Thus, users who are continuously retained, have a higher probability to remain involved “as participation becomes more automatic” (Gan et al. 2009). Finally, retained users are particularly important for OSN providers, since they can only leverage users, for instance for targeted marketing or premium services, if they stay active in the network.

Based on users’ retention, we compare our novel PageRank based approach to existing approaches, which could also be used to identify key users in OSN. This comparison to alternative approaches, so-called *“competing artifacts”*, is integral to design science research (Hevner et al. 2004). For our context, we consider the common centrality measures degree centrality, closeness centrality, and betweenness centrality. We do not employ eigenvector centrality, since graphs as social graphs and activity graphs are usually not connected and aperiodic graphs, as required for the calculation of eigenvector centrality (cf. the third section). Even though the same holds true for closeness centrality, we computed closeness centrality for each connected part of the graphs separately for comparison reasons. However, the results indicate the bias when identifying key users based on closeness centrality in not connected and aperiodic graphs. Applying the common centrality measures to the social graph derived from the New Orleans Network dataset allows us to identify key users based on their connectivity in the network as it is common practice in social network analysis. Hence, we first evaluate our novel PageRank based approach in comparison to the application of common centrality measures to the social graph (evaluation step 1). However, the application of common centrality measures to the social graph derived from the New Orleans Network dataset focuses solely on users’ connectivity, but does not incorporate users’ communication activity. Second, we evaluate our novel PageRank based approach in comparison to an approach, which is solely based on users’ prior communication activity in the network, but does not incorporate users’ connectivity (evaluation step 2). So far, we consider existing approaches taking either users’ connectivity or users’ communication activity into account. However, in contrast to the common centrality measures applied to the social graph and users’ prior communication activity, our novel PageRank based approach merges concepts from research on users’ connectivity and users’ communication activity. Even though existing approaches do not incorporate both aspects, we finally compare our novel PageRank based approach to the common centrality measures applied to the activity graph of the New Orleans Network dataset (evaluation step 3). Thus, we extend our evaluation by approaches, which are also based on both users’ connectivity and users’ communication activity.

As basis for our three evaluation steps, we use the fraction of wall posts in the New Orleans Network dataset spanning from January 4, 2009 to January 22, 2009 to determine users’ retention. Thereby, following Java et al. (2007) and Kolari et al. (2007), we consider a user retained, if he or she wrote at least one wall post during this period. For evaluation step 1, we calculate the common centrality measures degree centrality, closeness centrality, and betweenness centrality for every user in the social graph of the New Orleans Network dataset by using the software package “NetworkX”. For each common centrality measure, we are therefore able to rank users based on each user’s corresponding centrality score, which represents his or her connectivity in the social graph. For evaluation step 2, we determine users’ prior communication activity. Therefore, we draw on users’ wall posts between September 14, 2004 and January 3, 2009. Thus, we rank users solely based on the number of wall posts, i.e. their prior communication activity. Gan et al. (2009) refer to this cumulative number as “rank” in order to determine status in the context of online communities. To compare our novel PageRank based approach to the alternative approaches for the identification of key users in OSN (evaluation steps 1-3), we use a method, which has been similarly applied in biology to evaluate competing approaches for the identification of genes (Özgür et al. 2008). Thereby, we create top segments of u identified key users in every ranking, which has been either derived by applying our novel PageRank based approach, by the common centrality measures applied to the social graph (evaluation step 1), by users’ prior communication activity (evaluation step 2), and by the common centrality measures applied to the activity graph (evaluation step 3). Afterwards, we compare the percentages of retained users in these segments to evaluate how many identified key users were actually retained. In Table 2, we display segments of top u identified key users and the corresponding percentages of actually retained users for the common centrality measures applied to the social graph (evaluation step 1) and for the ranking by users’ prior communication activity (evaluation step 2). Table 2 highlights that by applying our novel PageRank based approach, 92% of the top 100 identified key users were actually retained. However, the application of the common centrality measures applied to the social graph leads to 48% retained top 100 identified key users for degree centrality, 43% for closeness centrality, and 54% for betweenness centrality. Ranking users solely based on users’ prior communication activity resulted in 90% retained top 100 identified key users. Thus, Table 2 illustrates that our PageRank based approach leads to better results for all top segments of identified key users than the common centrality measures applied to the social graph (evaluation step 1). Furthermore, also the percentages of retained identified key users compared to the ranking solely based on users’ prior communication activity are higher for every top segment (evaluation step 2).

Table 2. Percentages of the Actually Retained Top *u* Identified Key Users (Common Centrality Measures Applied to Social Graph)					
Top *u* Identified Key Users	PageRank Based Approach	Degree Centrality	Closeness Centrality	Betweenness Centrality	Prior Communication Activity
100	92%	48%	43%	54%	90%
500	87%	61%	55%	60%	84%
637 (1%)	82%	62%	55%	58%	80%
1000	86%	62%	55%	58%	83%
6373 (10%)	65%	51%	48%	50%	61%

As the results presented in Table 2 indicate, our novel PageRank based approach, which merges concepts from research on users' connectivity and communication activity, identifies more users that are retained than approaches based on solely users' connectivity (evaluation step 1) or users' prior communication activity (evaluation step 2). However, these existing approaches do not incorporate both users' connectivity and users' communication activity. Thus, we finally compare our novel PageRank based approach to the common centrality measures applied to the activity graph derived from the New Orleans Network dataset (evaluation step 3). Thereby, we extend our evaluation by approaches, which also merge concepts from research on users' connectivity and users' communication activity. In Table 3, we display segments of top *u* identified key users and the corresponding percentages of actually retained users. Thereby, we applied the common centrality measures degree centrality, closeness centrality, and betweenness centrality to the activity graph (evaluation step 3). To improve the clarity and comparability of Table 3, we once more display the results of our novel PageRank based approach and of the solely prior communication activity based approach (evaluation step 2).

Table 3. Percentages of the Actually Retained Top *u* Identified Key Users (Common Centrality Measures Applied to Activity Graph)					
Top *u* Identified Key Users	PageRank Based Approach	Degree Centrality	Closeness Centrality	Betweenness Centrality	Prior Communication Activity
100	92%	89%	80%	83%	90%
500	87%	80%	72%	83%	84%
637 (1%)	82%	76%	68%	78%	80%
1000	86%	78%	71%	81%	83%
6373 (10%)	65%	59%	52%	62%	61%

Table 3 illustrates that the common centrality measures identify more key users that are retained when they are applied to the activity graph. Nevertheless, our novel PageRank based approach still leads to better results than the common centrality measures applied to the activity graph (evaluation step 3). In order to test whether our results are significant, we ran a paired t-test. Thus, we came to the result that the novel PageRank based approach is significantly better than each of the other approaches in comparison (e.g. for the top 10% identified key users and $\alpha = 0.05$). In addition, we evaluated our novel PageRank based approach in comparison to the common centrality measures applied to the weighted activity graph derived from the New Orleans Network dataset. Therefore, we applied adapted common centrality measures, which have been extended to account for the activity links' weights (cf. Barrat et al. 2004). However, the application of weighted degree centrality, closeness centrality, and betweenness centrality to the weighted activity graph did not lead to better results than the ones of our novel PageRank based approach.

Taken together, we evaluated our novel PageRank based approach regarding users' retention in comparison to existing approaches, which are based on either users' connectivity or users' communication activity. Furthermore, we compared the novel PageRank based approach to approaches, which incorporate both users' connectivity and users' communication activity. Thereby, we illustrated that the proposed approach leads to significantly better results regarding the retained users for the New Orleans Network dataset than all approaches in comparison. Based on the evaluation using the New Orleans Network dataset, we believe that our novel PageRank based approach is better suited to identify key users in OSN than existing approaches, which could alternatively be used.

Discussion and Limitations of the Novel PageRank Based Approach

Besides the previously highlighted benefits, the underlying assumptions, the evaluation criterion and the real-world applicability of our novel PageRank based approach offer scope for discussion and implicate limitations.

Due to its formal representation and the underlying assumptions, the approach does not entirely consider and formalize all aspects of social connections and communication activities. Users have for instance a broad variety of different purposes, motivations, and ways regarding their usage of OSN. While some focus on making new connections, many users try to find out more about offline contacts (cf. e.g. Lampe et al. 2006). Thereby, communication with offline contacts might also occur through other media or face to face. However, in our paper we focus on OSN and consider communication activity within an OSN but not interactions between users occurring beyond that network. In addition, our approach incorporates the number of communication activities (cf. weights w_{ij} in formula (2)) to quantify the strength of connections between users but not the quality and the direction of the posts, messages etc. This fact might also be critical, since not only the number but also the quality and the direction of communication activities may influence the impact of a connection, for instance in terms of marketing. Moreover, the implicit assumption that users without communication activity have no influence on advertisement effectiveness, users' loyalty, and users' willingness to pay for services in OSN can be regarded as worth discussing. Hence, even though the number of users' communication activities allows a first indication of the strength of connections, formalizing social phenomena such as social connections needs to be critically discussed. However, prior research and the evaluation of our approach indicate the exceptionally high importance of users' communication activity in the context of OSN. Finally, we neglected any possible counterproductive and negative effects of high levels of users' connectivity and communication activity.

Taking users' retention as evaluation criterion indicates that our novel PageRank based approach allows to identify key users who are likely to be retained. Based on literature we argued that these users are particularly important and valuable for OSN, since only retained users can be leveraged, for instance for targeted marketing or premium services. However, taking users' retention as evaluation criterion is only one possibility towards evaluating our approach. According to the definition of key users stated above, other evaluation criteria – for instance users' willingness to pay for premium services in an OSN – are also reasonable. Future work is encouraged to address this issue, for instance by surveying users for their willingness to pay for premium services and analyzing the results of all approaches in comparison using this evaluation criterion. Currently, we are cooperating with a German OSN provider, which allows us to further evaluate our novel PageRank based approach using advertisement revenues and users' e-commerce revenues as evaluation criteria. In this context, we also analyze the costs and benefits when applying our novel PageRank based approach in practice and aim at conducting business cases with is an important future step to underline the practical benefit of the approach.

Finally, besides the discussion on how OSN can create value, there is an ongoing debate about the privacy risks they involve (cf. e.g. Gross and Acquisti 2005; Krasnova et al. 2009), which might influence the real-world applicability of our novel PageRank based approach. On the one hand, as users are becoming more and more aware and sensitive regarding privacy issues, they might change their behavior in OSN. Therefore, it is important to keep in mind that not only users' connectivity and communication activity but also exogenous factors might have a strong impact on advertisement effectiveness, users' loyalty, and users' willingness to pay for services in OSN. On the other hand, new privacy practices and novel privacy protection directives might come up and reduce the available amount of data to conduct analyses etc. Against this background, the data requirements of approaches for the identification of key users in OSN have to be critically discussed. As the weighted activity graph constitutes the basis of our approach, data about the number of communication activities is required for each pair of users in the OSN. However, besides that, no personal data of the users (content of messages etc.) is needed, which is very important to preserve the applicability of the approach. Nevertheless, future changes regarding privacy control in OSN might pose new challenges here.

Conclusion

OSN face the challenge to tap the enormous potential originated by the dramatic increase in the popularity of OSN in order to generate sustainable revenues. In that context, particularly more effective advertising strategies and sophisticated customer loyalty programs to foster users' retention are necessary. Therefore, quantitative approaches for the identification of key users in OSN are needed to address users deliberately and to enable for instance more effective and user centric marketing campaigns. In this paper, we propose a novel PageRank based approach bringing together concepts and findings from research on users' connectivity and users' communication activity in OSN. Related to the seven guidelines for conducting design science research articulated by Hevner et al. (2004), we can summarize as follows: We propose an *"artifact"* (cf. guideline 1) that is a method in terms of a PageRank based approach, which is composed of two steps. In the first step, a weighted activity graph is derived as basis for the identification of key users in OSN. In the second step, users' centrality scores are determined by using a novel PageRank based centrality measure. For the design of our artifact, we specified our *"problem context"* and focused on relevant literature regarding the identification of key users in OSN. Thereby, statements in literature support that the identification of key users in OSN is an *"important and relevant business problem"* (cf. guideline 2). Moreover, we reviewed prior research on users' connectivity and communication activity in OSN. Drawing on these two research streams, we identified the research gap: Quantitative approaches for the identification of key users in OSN bringing together concepts and findings from research on users' connectivity and users' communication activity were missing. Thus, we developed a novel PageRank based approach to "address an important organizational problem" (Hevner et al. 2004). We believe that our artifact contributes as a first, but essential step to overcome the challenges faced by the majority of OSN. We *"evaluated"* our novel PageRank based approach (cf. guideline 3) regarding its applicability and its practical utility by using a publicly available dataset of Facebook.com. For the evaluation, we chose users' retention as evaluation criterion and compared our novel PageRank based approach with *"competing artifacts"*, which could also be used to identify key users in OSN. Thus, we illustrated the advantages of our *"research contribution"* (cf. guideline 4), i.e. of our novel PageRank based approach. According to literature, we highlighted the importance of both users' connectivity and users' communication activity when identifying key users in OSN. We incorporated users' communication activity and the strength of users' connections in the first step of our approach by deriving a weighted activity graph. For the second step of our approach, we designed a PageRank based centrality measure to determine users' centrality scores in terms of their connectivity in the weighted activity graph. Hence, we developed a first quantitative approach for the identification of key users in OSN bringing together concepts and findings from research on users' connectivity *and* users' communication activity in OSN and addressed the research gap stated above. The evaluation based on the Facebook.com New Orleans Network dataset illustrates that the novel PageRank based approach leads to (significantly) better results regarding the retained users than all other approaches in comparison. Therefore, the proposed approach, which allows to identify key users in OSN, seems to be quite promising and may contribute to overcome current challenges of OSN (e.g. regarding their monetization by enabling more effective advertising strategies etc.). Nevertheless, future work is needed and intended to further evaluate the approach.

To support a *"rigorous"* definition and presentation of our artifact (cf. guideline 5), we denoted it formally. Thereby, we drew on Hevner et al. (2004), who state: "to be mathematically rigorous, important parts of the problem may be abstracted". This implicates assumptions and limitations of the approach, which were critically discussed. Future work should address these issues either by confirming our assumptions or by relaxing the assumptions when developing further approaches for the identification of key users in OSN. Furthermore, upcoming challenges, for instance due to changing privacy practices, need to be carefully observed and considered. Thus, the *"search process"* (cf. guideline 6) can be distinguished in present and future steps. In this paper we presented the initial design of a PageRank based approach for the identification of key users in OSN, which may represent a starting point for OSN to overcome the described challenges. Thereby, the design process was guided by existing literature and the identified main factors of influence, namely users' connectivity and communication activity in OSN. Certainly, we abstracted quite strongly when initially designing our novel PageRank based approach. Future iterations need to relax assumptions and particularize and enhance the artifact accordingly. We are currently collaborating with a German OSN provider to additionally analyze our approach "in depth in business" (Hevner et al. 2004) and to extend our basic approach for the identification of key users in OSN. Regarding the *"communication"* of our results (cf. guideline 7), we chose a formal, mathematical presentation in order to be able to demonstrate and evaluate our artifact in a rigorous and unambiguous way. However, we also tried to attract a managerial audience by means of the extensive explanations of the used concepts and formulas as well as detailed description of the application and the practical utility of our novel PageRank based approach.

References

Adamic, L. A., and Adar, E. 2003. “Friends and Neighbors on the Web,” *Social Networks* (25:3), pp. 211-230.

Agarwal, R., Gupta, A. K., and Kraut, R. 2008. “The Interplay Between Digital and Social Networks,” *Information Systems Research* (19:3), pp. 243-252.

Albert, R., and Barabási, A.-L. 2002. “Statistical Mechanics of Complex Networks,” *Reviews of Modern Physics* (74:1), pp. 47-97.

Algesheimer, R., and Von Wangenheim, F. 2006. “A Network Based Approach to Customer Equity Management,” *Journal of Relationship Marketing* (5:1), pp. 39-57.

Andrews, A. 2009. “Facebook and Other Social Networking Sites Face a Critical Year”, *Telegraph*. http://www.telegraph.co.uk/technology/facebook/4436417/Facebook-and-other-social-networking-sites-face-a-critical-year.html. Accessed April 12, 2010.

Bampo, M., Ewing, M. T., Mather, D. R., Stewart, D., and Wallace, M. 2008. “The Effect of the Social Structure of Digital Networks on Viral Marketing Performance,” *Information Systems Research* (19:3), pp. 273-290.

Barrat, A., Barthélemy, M., Pastor-Satorras, R., and Vespignani, A. 2004. “The Architecture of Complex Weighted Networks,” *Proceedings of the National Academy of Sciences of the United States of America* (101:11), pp. 3747-3752.

Bass, F. 1969. “A New Product Growth Model for Consumer Durables,” *Management Science* (15:5), pp. 215-227.

BBC 2005. “News Corp in $580m Internet Buy”, *BBC*. http://news.bbc.co.uk/2/hi/business/4695495.stm. Accessed April 12, 2010.

Benevenuto, F., Rodrigues, T., Cha, M., and Almeida, V. 2009. “Characterizing User Behavior in Online Social Networks”, in *Proceedings of the Ninth ACM SIGCOMM Conference on Internet Measurement*, Chicago, IL, November 2009, pp. 49-62.

Bhat, S., Bevans, M., and Sengup, S. 2002. “Measuring Users' Web Activity to Evaluate and Enhance Advertising Effectiveness,” *Journal of Advertising* (31:3), pp. 97-106.

Boldi, P., Santini, M., and Vigna, S. 2005. “PageRank as a Function of the Damping Factor”, in *Proceedings of Fourteenth International Conference on World Wide Web*, Chiba, Japan, May 2005, pp. 557-566.

Bolland, J. M. 1988. “Sorting Out Centrality: An Analysis of the Performance of Four Centrality Models in Real and Simulated Networks,” *Social Networks* (10:3), pp. 233-253.

Bollen, J., Rodriguez, M. A., and Van De Sompel, H. 2006. “Journal Status,” *Scientometrics* (69:3), pp. 669-687.

Bonacich, P. 1972. “Factoring and Weighting Approaches to Status Scores and Clique Identification,” *Journal of Mathematical Sociology* (2:1), pp. 113-120.

Bonacich, P. 1987. “Power and Centrality: A Family of Measures,” *American Journal of Sociology* (92:5), pp. 1170-1182.

Borgatti, S. P. 2005. “Centrality and Network Flow,” *Social Networks* (27:1), pp. 55-71.

Borgatti, S. P., and Everett, M. G. 2006. “A Graph-Theoretic Perspective on Centrality,” *Social Networks* (28:4), pp. 466-484.

Boyd, D. M., and Ellison, N. B. 2007. “Social Network Sites: Definition, History, and Scholarship,” *Journal of Computer-Mediated Communication* (13:1), Article 11.

Brandes, U., and Erlebach, T. 2005. *Network Analysis: Methodological Foundations*, Berlin, Germany: Springer.

Brin, S., and Page, L. 1998. “The Anatomy of a Large-Scale Hypertextual Web Search Engine,” *Computer Networks and ISDN Systems* (30:1-7), pp. 107-117.

Brown, J. J., and Reingen, P. H. 1987. “Social Ties and Word-of-Mouth Referral Behavior,” *Journal of Consumer Research* (14:3), pp. 350-362.

Burt, R. S. 1992. *Structural Holes: The Social Structure of Competition*, Cambridge, MA: Harvard University Press.

Butler, B. S. 2001. “Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures,” *Information Systems Research* (12:4), pp. 346-362.

Cheung, C. M. K., and Lee, M. K. O. 2010. “A Theoretical Model of Intentional Social Action in Online Social Networks,” *Decision Support Systems* (49:1), pp. 24-30.

Chun, H., Kwak, H., Eom, Y., Ahn, Y., Moon, S., and Jeong, H. 2008. “Comparison of Online Social Relations in Volume vs. Interaction: A Case Study of Cyworld”, in *Proceedings of the Eighth ACM SIGCOMM Conference on Internet Measurement*, Vouliagmeni, Greece, October 2008, pp. 57-70.

Clemons, E. K. 2009. “The Complex Problem of Monetizing Virtual Electronic Social Networks,” *Decision Support Systems* (48:1), pp. 46-56.

Constant, D., Sproull, L., and Kiesler, S. 1996. “The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice,” *Organization Science* (7:2), pp. 119-135.

Cummings, J. N., Butler, B., and Kraut, R. 2002. "The Quality of Online Social Relationships," *Communications of the ACM* (45:7), pp. 103-108.
De Bruyn, A., and Lilien, G. L. 2008. "A Multi-Stage Model of Word-of-Mouth Influence Through Viral Marketing," *International Journal of Research in Marketing* (25:3), pp. 151-163.
De Valck, K., Van Bruggen, G. H., and Wierenga, B. 2009. "Virtual Communities: A Marketing Perspective," *Decision Support Systems* (47:3), pp. 185-203.
Delaney, K. J., Steel, E., Vara, V., and Vascellaro, J. E. 2008. "Social Sites Don't Deliver Big Ad Gains," *Wall Street Journal-Eastern Edition*, February 5.
Deutsch, M., and Gerard, H. B. 1955. "A Study of Normative and Informational Social Influences Upon Individual Judgment," *Journal of Abnormal and Social Psychology* (51:3), pp. 629-636.
Ding, Y., Yan, E., Frazho, A., and Caverlee, J. 2009. "PageRank for Ranking Authors in Co-Citation Networks," *Journal of the American Society for Information Science and Technology* (60:11), pp. 2229-2243.
Domingos, P., and Richardson, M. 2001. "Mining the Network Value of Customers", in *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, August 2001, pp. 57-66.
Dougherty, H. 2010. "Facebook Reaches Top Ranking in US", *Hitwise*. http://weblogs.hitwise.com/heather-dougherty/2010/03/facebook_reaches_top_ranking_i.html. Accessed April 12, 2010.
Edgecliffe-Johnson, A., and Li, K. 2009. "News Corp Says Sluggish MySpace Jeopardises $900m Google Deal," *Financial Times*, November 5.
Facebook 2010. "Six Years of Making Connections", *Facebook*. http://blog.facebook.com/blog.php?post=287542162130. Accessed April 12, 2010.
Figallo, C. 1998. *Hosting Web Communities: Building Relationships, Increasing Customer Loyalty, and Maintaining a Competitive Edge*, New York, NY: John Wiley & Sons, Inc.
Freeman, L. C. 1979. "Centrality in Social Networks: Conceptual Clarification," *Social Networks* (1:3), pp. 215-239.
Freeman, L. C., Roeder, D., and Mulholland, R. R. 1980. "Centrality in Social Networks: Experimental Results," *Social Networks* (2:2), pp. 119-141.
Friedkin, N. E. 1991. "Theoretical Foundations for Centrality Measures," *American Journal of Sociology* (96:6), pp. 1478-1504.
Gan, L., Gu, B., Jarvenpaa, S., and Yang, G. 2009. "Habit Formation in Online Communities", in *Proceedings of the Thirtieth International Conference on Information Systems*, Phoenix, AZ, December 2009, Paper 189.
Ganley, D., and Lampe, C. 2009. "The Ties That Bind: Social Network Principles in Online Communities," *Decision Support Systems* (47:3), pp. 266-274.
Gilbert, E., and Karahalios, K. 2009. "Predicting Tie Strength with Social Media", in *Proceedings of the Twenty-Seventh International Conference on Human Factors in Computing Systems*, Boston, MA, April 2009, pp. 211-220.
Granovetter, M. S. 1973. "The Strength of Weak Ties," *American Journal of Sociology* (78:6), pp. 1360-1380.
Granovetter, M. S. 1974. *Getting a Job: A Study of Contacts and Careers*, Chicago, IL: University Of Chicago Press.
Granovetter, M. S. 1983. "The Strength of Weak Ties: A Network Theory Revisited," *Sociological Theory* (1:1), pp. 201-233.
Gross, R. and Acquisti, A. 2005. "Information Revelation and Privacy in Online Social Networks", in *Proceedings of the ACM Workshop on Privacy in the Electronic Society*, Alexandria, VA, November 2005, pp. 71-80.
Hagberg, A. A., Schult, D. A., and Swart, P. J. 2008. "Exploring Network Structure, Dynamics, and Function Using NetworkX", in *Proceedings of the Seventh Python in Science Conference*, Pasadena, CA, August 2008, pp. 11-15.
Hagel, J., and Armstrong, A. G. 1997. *Net Gain: Expanding Markets Through Virtual Communities*, Boston, MA: Harvard Business School Press.
Hanneman, R. A., and Riddle, M. 2005. *Introduction to Social Network Methods*, Riverside, CA: University of California. http://www.faculty.ucr.edu/~hanneman/nettext/, Accessed August 10, 2010.
Haythornthwaite, C. 2002. "Building Social Networks via Computer Networks: Creating and Sustaining Distributed Learning Communities", in *Building Virtual Communities: Learning and Change in Cyberspace*, Renninger, K. A., and Shumar, W. (eds.), Cambridge, United Kingdom: Cambridge University Press, pp. 159-190.
Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75-105.

Java, A., Song, S., Finin, T., and Tseng, B. 2007. "Why We Twitter: Understanding Microblogging Usage and Communities", in *Proceedings of the Ninth WebKDD and First SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, San Jose, CA, August 2007, pp. 56-65.

Kahanda, I., and Neville, J. 2009. "Using Transactional Information to Predict Link Strength in Online Social Networks", in *Proceedings of the Third International AAAI Conference on Weblogs and Social Media*, San Jose, CA, May 2009, Paper 106.

Kemeny, J. G., and Snell, J. L. 1960. *Finite Markov Chains*, Princeton, NJ: Van Nostrand Co., Inc.

Kiss, C., and Bichler, M. 2008. "Identification of Influencers - Measuring Influence in Customer Networks," *Decision Support Systems* (46:1), pp. 233-253.

Kleinberg, J. M. 1998. "Authoritative Sources in a Hyperlinked Environment", in *Proceedings of the Ninth ACM-SIAM Symposium on Discrete Algorithms*, San Francisco, CA, January 1998, pp. 668-677.

Kolari, P., Finin, T., Yesha, Y., Yesha, Y., Lyons, K., Perelgut, S., and Hawkins, J. 2007. "On the Structure, Properties and Utility of Internal Corporate Blogs", in *Proceedings of the International Conference on Weblogs and Social Media*, Boulder, CO, March 2007.

Krasnova, H., Hildebrand, T., and Günther, O. 2009. "Investigating the Value of Privacy in Online Social Networks: Conjoint Analysis", in *Proceedings of the Thirtieth International Conference on Information Systems*, Phoenix, AZ, December 2009, Paper 173.

Kraut, R. E., Rice, R., Cool, C., and Fish, R. 1998. "Varieties of Social Influence: The Role of Utility and Norms in the Success of a Communication Medium," *Organizational Science* (9:4), pp. 437-453.

Lampe, C., Ellison, N., and Steinfield, C. 2006. "A Face(book) in the Crowd: Social Searching vs. Social Browsing", in *Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work*, Banff, Alberta, Canada, November 2006, pp. 167-170.

Langville, A. N., and Meyer, C. D. 2004. "Deeper Inside PageRank," *Internet Mathematics* (1:3), pp. 335-380.

Leskovec, J., and Horvitz, E. 2008. "Planetary-Scale Views on an Instant-Messaging Network", in *Proceedings of the Seventeenth International Conference on World Wide Web*, Beijing, China, April 2008, pp. 915-924.

Liu, X., Bollen, J., Nelson, M. L., and Van De Sompel, H. 2005. "Co-Authorship Networks in the Digital Library Research Community," *Information Processing & Management* (41:6), pp. 1462-1480.

Lu, H.-P., and Hsiao, K.-L. 2010. "The Influence of Extro/Introversion on the Intention to Pay for Social Networking Sites," *Information & Management* (47:3), pp. 150-157.

Milgram, S. 1967. "The Small World Problem," *Psychology Today* (1:1), pp. 61-67.

Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., and Bhattacharjee, B. 2007. "Measurement and Analysis of Online Social Networks", in *Proceedings of the Seventh ACM SIGCOMM Conference on Internet Measurement*, San Diego, CA, October 2007, pp. 29-42.

Monge, P. R., and Contractor, N. S. 2003. *Theories of Communication Networks*, New York, NY: Oxford University Press.

MSNBC 2007. "Microsoft Invests $240 Million in Facebook", *MSNBC*. http://www.msnbc.msn.com/id/21458486. Accessed April 12, 2010.

Nazir, A., Raza, S., and Chuah, C.-N. 2008. "Unveiling Facebook: A Measurement Study of Social Network Based Applications", in *Proceedings of the Eighth ACM SIGCOMM Conference on Internet Measurement*, Vouliagmeni, Greece, October 2008, pp. 43-56.

Nazir, A., Raza, S., Gupta, D., Chuah, C.-N., and Krishnamurthy, B. 2009. "Network Level Footprints of Facebook Applications", in *Proceedings of the Ninth ACM SIGCOMM Conference on Internet Measurement*, Chicago, IL, November 2009, pp. 63-75.

Newman, M. E. J. 2002. "Assortative Mixing in Networks," *Physical Review Letters* (89:20), Article 208701.

Newman, M. E. J. 2003a. "Mixing Patterns in Networks," *Physical Review E* (67:2), Article 026126.

Newman, M. E. J. 2003b. "The Structure and Function of Complex Networks," *SIAM Review* (45:2), pp. 167-256.

Newman, M. E. J. 2004. "Analysis of Weighted Networks," *Physical Review E* (70:5), Article 056131.

Nohria, N., and Eccles R. G. 1992. *Networks and Organizations*, Boston, MA: Harvard Business School Press.

Oestreicher-Singer, G., and Zalmanson L. 2009. "'Paying for Content or Paying for Community?' The Effect of Social Involvement on Subscribing to Media Web Sites", in *Proceedings of the Thirtieth International Conference on Information Systems*, Phoenix, AZ, December 2009, Paper 9.

Oinas-Kukkonen, H., Lyytinen, K., and Yoo, Y. 2010. "Social Networks and Information Systems: Ongoing and Future Research Streams," *Journal of the Association for Information Systems* (11:2), pp. 61-68.

Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., De Menezes, M. A., Kaski, K., Barabási, A.-L., and Kertész, J. 2007. "Analysis of a Large-Scale Weighted Network of One-to-One Human Communication," *New Journal of Physics* (9:6), Article 179.

Özgür, A., Vu, T., Erkan, G., and Radev, D. R. 2008. "Identifying Gene-Disease Associations Using Centrality on a Literature Mined Gene-Interaction Network," *Bioinformatics* (24:13), pp. 277-285.

Page, L., Brin, S., Motwani, R., and Winograd T. 1999. *The PageRank Citation Ranking: Bringing Order to the Web*, Stanford, CA: Technical Report, Stanford InfoLab. http://ilpubs.stanford.edu:8090/422/. Accessed March 12, 2010.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2008. "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems* (24:3), pp. 45-77.

Pinski, G., and Narin, F. 1976. "Citation Influence for Journal Aggregates of Scientific Publications: Theory, with Application to the Literature of Physics," *Information Processing and Management* (12:5), pp. 297-312.

Pretto, L. 2002. "A Theoretical Analysis of Google's PageRank", in *Proceedings of the Ninth International Symposium on String Processing and Information Retrieval*, Laender, A. H., and Oliveira, A. L., (eds.), Lecture Notes In Computer Science, Vol. 2476, London, United Kingdome: Springer, pp. 131-144.

Ray, A., Bernoff, J, and Wise, J. 2010. "Peer Influence Analysis - Mass Influencers Are The Key To Achieving Scale In Social Media Marketing", *Forrester*. http://www.forrester.com/rb/Research/peer_influence_analysis/q/id/56766/t/2. Accessed August 29, 2010.

Ridings, C., and Wasko, M. 2010. "Online Discussion Group Sustainability: Investigating the Interplay Between Structural Dynamics and Social Dynamics over Time," *Journal of the Association for Information Systems* (11:2), pp. 95-121.

Riggins, F. J. 2003. "Market Segmentation and Information Development Costs in a Two-Tiered Fee-Based and Sponsorship-Based Web Site," *Journal of Management Information Systems* (19:3), pp. 69-81.

Rodriguez, M. A. 2008. "Grammar-Based Random Walkers in Semantic Networks," *Knowledge-Based Systems* (21:7), pp. 727-739.

Saxenian, A. L. 1994. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*, Cambridge, MA: Harvard University Press.

Schneider, F., Feldmann, A., Krishnamurthy, B., and Willinger, W. 2009. "Understanding Online Social Network Usage from a Network Perspective", in *Proceedings of the Ninth ACM SIGCOMM Conference on Internet Measurement*, Chicago, IL, November 2009, pp. 35-48.

Schnettler, S. 2009. "A Structured Overview of 50 Years of Small-World Research," *Social Networks* (31:3), pp. 165-178.

Scott, J. 2000. *Social Network Analysis: A Handbook*, Thousand Oaks, CA: Sage.

Shapiro, C., and Varian, H. R. 1999. *Information Rules: A Strategic Guide to the Network Economy*, Boston, MA: Harvard Business School Press.

Staab, S., Domingos, P., Mika, P., Golbeck, J., Ding, L., Finin, T., Joshi, A., Nowak, A., and Vallacher, R. R. 2005. "Social Networks Applied," *IEEE Intelligent Systems* (20:1), pp. 80-93.

Uzzi, B. 1997. "Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness," *Administrative Science Quarterly* (42:1), pp. 35-67.

Valente, T. W., and Foreman, R. K. 1998. "Integration and Radiality: Measuring the Extent of an Individual's Connectedness and Reachability in a Network," *Social Networks* (20:1), pp. 89-105.

Viswanath, B., Mislove, A., Cha, M., and Gummadi, K. P. 2009. "On the Evolution of User Interaction in Facebook", in *Proceedings of the Second ACM Workshop on Online Social Networks*, Barcelona, Spain, August 2009, pp. 37-42.

Wang, Y. and Fesenmaier, D. R. 2006. "Identifying the Success Factors of Web-Based Marketing Strategy: An Investigation of Convention and Visitors Bureaus in the United States," *Journal of Travel Research* (44:3), pp. 239-249.

Wassermann, S., and Faust, K. 1994. *Social Network Analysis: Methods and Applications*, Cambridge, United Kingdom: Cambridge University Press.

Watts, D. J., and Strogatz, S. H. 1998. "Collective Dynamics of 'Small-World' Networks," *Nature* (393), pp. 440-442.

Wellman, B., and Gulia, M. 1999. "Virtual Communities as Communities", in *Communities in Cyberspace*, Smith, M. A., and Kollocks, P. (eds.), New York, NY: Routledge, pp. 167-194.

Wen, C., Tan, B. C. Y., and Chang, K. T.-T. 2009. "Advertising Effectiveness on Social Network Sites: An Investigation of Tie Strength, Endorser Expertise and Product Type on Consumer Purchase Intention", in *Proceedings of the Thirtieth International Conference on Information Systems*, Phoenix, AZ, December 2009, Paper 151.

Williamson, D. A. 2009a. "Social Networking Worldwide: Ad Spending and Usage", *eMarketer*. http://www.emarketer.com/Reports/All/Emarketer_2000567.aspx. Accessed April 12, 2010.

Williamson, D. A. 2009b. "Social Networks: Five Consumer Trends for 2009", *eMarketer*. http://www.emarketer.com/Report.aspx?code=emarketer_2000566. Accessed April 12, 2010.

Willinger, W., Rejaie, R., Torkjazi, M., Valafar, M., and Maggioni, M. 2010. "Research on Online Social Networks: Time to Face the Real Challenges," *ACM SIGMETRICS Performance Evaluation Review* (37:3), pp. 49-54.

Wilson, C., Boe, B., Sala, A., Puttaswamy, K. P. N., and Zhao B. Y. 2009. "User Interactions in Social Networks and Their Implications", in *Proceedings of the Fourth ACM European Conference on Computer Systems*, Nuremberg, Germany, April 2009, pp. 205-218.

Xiang, R., Neville, J., and Rogati, M. 2010. "Modeling Relationship Strength in Online Social Networks", in *Proceedings of the Ninth International Conference on World Wide Web*, Raleigh, NC, April 2010, pp. 981-990.

Xing, W., and Ghorbani, A. 2004. "Weighted PageRank Algorithm", in *Proceedings of the Second Annual Conference on Communication Networks and Services Research*, Washington, District of Columbia, May 2004, pp. 305-314.

Xu, Y., Lu, X., Goh, K. Y., Jiang, Z., and Zhu, X. 2009. "The Impact of Online Social Network on Consumer Loyalty: An Empirical Study of an Online Dining Community", in *Proceedings of the Thirtieth International Conference on Information Systems*, Phoenix, AZ, December 2009, Paper 17.

Xu, Y., Zhang, C., Xue, L., and Yeo, L. L. 2008. "Product Adoption in Online Social Network", in *Proceedings of the Twenty Ninth International Conference on Information Systems*, Paris, France, December 2008, Paper 200.