Research Center
Finance & Information Management

Fraunhofer
FIT

Project Group
Business & Information
Systems Engineering

Who will lead and who will follow: Identifying Influential Users in Online Social Networks - A Critical Review and Future Research Directions

by

Florian Probst, Laura Grosswiele, Regina Pfleger

Europäische Union
„Investition in Ihre Zukunft"
Europäischer Fonds für
regionale Entwicklung

in: Business & Information Systems Engineering 5 (2013) 3, p. 179-193

University of Augsburg, D-86135 Augsburg
Visitors: Universitätsstr. 12, 86159 Augsburg
Phone: +49 821 598-4801 (Fax: -4899)

University of Bayreuth, D-95440 Bayreuth
Visitors: F.-v.-Schiller-Str. 2a, 95444 Bayreuth
Phone: +49 921 55-4710 (Fax: -844710)

www.fim-rc.de

WI-401

Who will lead and who will follow:[1]
Identifying Influential Users in Online Social Networks

A Critical Review and Future Research Directions

Die Autoren

Dr. Florian Probst

Dipl.-Kffr. Laura Grosswiele

Dipl.-Kffr. Regina Pfleger

Autorenadresse zur Veröffentlichung

FIM Kernkompetenzzentrum Finanz- & Informationsmanagement

Universität Augsburg

Universitätsstr. 12

86159 Augsburg

Deutschland

florian.probst@wiwi.uni-augsburg.de

http://www.fim-online.eu

http://www.fim-online.eu/fb-crm

[1] Adopted from Katz (1957, p. 73).

Who will lead and who will follow:[1]
Identifying Influential Users in Online Social Networks
A Critical Review and Future Research Directions

Abstract

Along with the explosive growth of the phenomenon Online Social Networks (OSN), identifying influential users in OSN has received a great deal of attention in recent years. However, the development of practical approaches for the identification of influential users is still in its infancy and researchers face numerous challenges. By means of a structured literature review, we analyze and synthesize the growing number of publications particularly from two perspectives. From a research perspective, we find that existing approaches mostly build on users' connectivity and activity but hardly consider further characteristics of influential users. Moreover, we outline two major research streams. It becomes apparent that most marketing-oriented articles draw on real-world data of OSN, while more technology-oriented papers rather have a theoretical approach and mostly evaluate their artifacts by means of formal proofs. We find that an even stronger collaboration between the scientific Business & Information Systems Engineering (BISE) and Marketing community than observed today could be mutually beneficial. With respect to a practitioner's perspective, we compile advice on the practical application of approaches for the identification of influential users. It is hoped that the results can stimulate and guide future research.

Keywords viral marketing, information diffusion, word-of-mouth, influence, contagion, influentials, literature review, online social networks

Outline

One of the most important questions at the heart of viral marketing is how companies can identify and target the "right" initial set of influential users in Online Social Networks (OSN). Even though we find that both the scientific Business & Information Systems Engineering (BISE) and Marketing community engage in research on the identification of influential users in OSN, the development of practical approaches is still in its infancy. Therefore, we analyze and synthesize the growing number of scientific publications and hope that the results can stimulate and guide future research.

[1] Adopted from Katz (1957, p. 73).

1 Introduction

For decades, marketers have been intensively investigating the effects driving the diffusion and adoption of new products and services. In this context, major developments could be observed over the last couple of years: First, the impact of traditional marketing techniques has been constantly decreasing (Clemons 2009, p. 48 f.; Hinz et al. 2011, p. 55; Trusov et al. 2009, p. 90). Second, consumers increasingly trust the recommendations of other consumers, acquaintances, and friends (Chen and Xie 2008; Iyengar et al. 2011b; Narayan et al. 2011; Schmitt et al. 2011). Third, it recently has become widely accepted that social influence actually affects the diffusion process and that there are influential people who have disproportionate influence on others (Godes and Mayzlin 2009; Goldenberg et al. 2009; Hinz et al. 2013; Iyengar et al. 2011a). Such social influence can be defined as "[…] change in the belief, attitude, or behavior of a person […], which results from the action, or presence, of another person […]" (Erchul and Raven 1997, p. 138), usually denoted as influencer. To respond to these developments and to leverage the effect of social influence on product adoption, companies increasingly try to actively initiate and control the diffusion process by targeting the most influential people in a social network (Bonchi et al. 2011, p. 21; Hinz et al. 2011, p. 55; Libai et al. 2010, p. 271). Thus, with small marketing costs a considerable part of the network may be reached. However, among others, one key prerequisite needs to be fulfilled: Companies need to be able to identify and target the "right" initial set of influential people (Iyengar et al. 2011b, p. 195; Hinz et al. 2011, p. 55 f.).

Traditionally, self-designation, that is, people report on their own influence in surveys (cf. Rogers and Cartano 1962), has been popular to identify influential people. More sophisticated sociometric techniques, that is, using network data on social connections, could only scarcely be used at a larger scale, as data sets have often been too small (Corey 1971, p. 52; Watts 2004, p. 5). However, due to the rise of modern communication networks and the Internet, the usage of network data for the identification of influential people has gained increasing popularity in research and practice (cf. e.g., Bampo et al. 2008; Hill et al. 2006; Hinz et al. 2011; Nitzan and Libai 2011). Especially along with the explosive growth of the phenomenon of Online Social Networks (OSN) to currently more than one billion active users and 140 billion friendship connections as of October 2012 solely on Facebook (Facebook 2012), identifying influential users in OSN is receiving a great deal of attention in recent years (Bonchi et al. 2011, p. 21; Hinz et al. 2013; Katona et al. 2011, p. 426). Besides mere social connections, which for instance could be observed in telecommunication networks as well, OSN allow for analyzing the diffusion process taking into account additional information such as detailed demographic data, personal interests, the level of activity with respect to different technical features of OSN (e.g., comments, likes), and partly even the content and sentiment of communication (e.g., in public wallposts). Moreover, users thereby usually reveal more information than in an offline context, as online communications tend to be more uninhibited, creative, and blunt (Wellman et al. 1996, p. 213).Thus, OSN provide a unique and vast amount of user data (also referred to as “digital trace data”, cf. Howison et al. 2011) that was not available before and can now be leveraged for marketing purposes[2] (Bonchi et al. 2011, p. 2; Katona et al. 2011, p. 425 f.; Subramani and Rajagopalan 2003, p. 301).

However, the development of practical approaches for the identification of influential users in OSN is still in its infancy (Richter et al. 2011, p. 98) and researchers face numerous challenges: First, the processing of previously unknown large amounts of (digital trace) data and the consequently required scalability of existing approaches for the identification of influential people are not trivial (cf. e.g., Watts 2004). Second, research based on such data faces numerous validity issues (cf. Howison et al. 2011) and several sources of bias might confound the identification of influential users in OSN (cf. section 2.1). Third, findings from research on viral marketing and the identification of influential people in an offline environment or from the “old Internet” may not be transferred to the context of OSN without critical reflection (cf. e.g., Brown et al. 2007; Eccleston and Griseri 2008, p. 608; Howison et al. 2011, p. 768; Susarla et al. 2012). Therefore, further research is needed in order to overcome these challenges and to achieve a better understanding in research and practice.

What can a critical literature review contribute? We believe that the growing number of publications on the identification of influential users in OSN needs to be analyzed and synthesized to assess the applied methods, knowledge, and theories (Scandura and Williams 2000) as well as to identify research gaps that can be addressed in future research (Webster and Watson 2002). For our following analysis, we 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” (Boyd and Ellison 2007, p. 211) but focus on user-oriented sites (Pallis et al. 2011, p. 220), “[…] where, to a certain extent, networking is the main preoccupation” (Beer 2008, p. 518). In contrast, content-oriented sites such as Twitter, YouTube, or Flickr exhibit some features of OSN but are rather microblogging sites or content communities with different characteristics than OSN (Heidemann et al. 2012, p. 3867; Pallis et al. 2011, p. 220; Richter et al. 2011, p. 90; Smith et al. 2012, p. 103). For instance, Wu et al. (2011, p. 707) found that Twitter “[…] does not

[2] For a critical discussion of related fundamental problems such as the access to data from OSN, privacy issues, and validity concerns see for instance Howison et al. (2011), Lazer et al. (2009) and with respect to the identification of influential users in OSN section 5.

conform to the usual characteristics of social networks, which exhibit much higher reciprocity […] [Kossinets and Watts 2006]". Prior research also emphasizes that on content-oriented sites "[…] the primary motivation and goal of the majority of users is the content instead of socialization" (Laine et al. 2011, p. 2). Some content-oriented sites are therefore even perceived as a "[…] mixture of one-way mass communications and reciprocated interpersonal communications" (Wu et al. 2011, p. 707). Consequently, (partly) different data can be collected in OSN and content-oriented sites (e.g., friendship connections in Facebook versus followers in Twitter). Treating them interchangeably might raise several validity issues along the chain of reasoning when drawing conclusions on a construct under consideration (e.g., social influence) based on data from these information systems (i.e., a content-oriented site or an OSN) (cf. Howison et al. 2011, p. 772). For instance, theoretical cohesion might not be given when operationalizing constructs deduced from theories on (offline) *social networks* with data from content-oriented sites. Before further research can focus on the identification of influential users in content-oriented sites and commonalities and differences to their identification in OSN, this paper aims at laying the foundations by concentrating on OSN as the currently predominant phenomenon. Thereby, two particular perspectives are taken into account (cf. Poeppelbuss et al. 2011, p. 506): A research perspective that relates to the theoretical and methodological aspects and a practitioner's perspective that covers issues relevant to users of approaches for the identification of influential users in OSN.

The remainder of this paper is organized as follows: In the next section, we provide an overview on important foundations from the context of social influence as well as from the identification of influential people in social networks and delineate three research questions (Q): (Q.1) How are influential users characterized in the context of OSN? (Q.2) Which approaches have been developed and applied for the identification of influential users in OSN? (Q.3) How have these approaches been evaluated and which implications have been derived? In section 3, we outline the procedure of our structured literature search. In the subsequent section 4, we present our findings regarding the three research questions and critically discuss the identified articles from a research perspective. By highlighting nine implications of our literature review, we point out future research directions in section 5. These may also be beneficial for an audience from practice who adopt approaches for the identification of influential users. Finally, in section 6 we draw an overall conclusion and explicate limitations.

2 Foundations and Research Questions

As previously mentioned, marketers aim at targeting the most influential people in social networks in order to initiate a diffusion process that makes it possible to reach a large part of a network with small marketing cost (Bonchi et al. 2011, p. 21). To do so, three key assumptions need to be fulfilled (Iyengar et al. 2011b, p. 195): (1) social influence ought to be at work, (2) there actually have to be influential people in the social network who have disproportionate influence on others, and (3) companies need to be able to identify and target these influential people. With respect to these three assumptions, we briefly review relevant literature from economics, marketing, and sociology beyond the context of OSN that constitutes the foundation for research on the identification of influential users in OSN. In doing so, we also derive our research questions that are addressed in the subsequent structured literature review.

2.1 Social Influence in the diffusion process

After Moreno (1934) coined the term "sociometry" when formalizing social relationships, Rapoport (cf. e.g., Rapoport 1952; 1953; Rapoport and Rebhun 1952) was one of the first who applied "[…] sociometric ideas to large-scale social systems […]" and "[…] elaborated on the formal implications […]" in the context of predictive epidemiological models of contagion (Scott 2000, p. 15 f.). Similar ideas have been used to understand the diffusion of innovations (cf. e.g., Rogers 1962), such as technical innovations in an agricultural context (Beal and Bohlen 1955; 1957; Ryan and Gross 1943), or new drugs in physicians' networks (Coleman et al. 1966). While these studies implied that diffusion was driven by communication (cf. also Valente 1995; Valente and Rogers 1995), others found contradicting results showing that diffusion was rather a result of imitation (Mansfield 1961) or comparison (Burt 1987). Strang and Tuma (1993) even found traces of both, communication and comparison effects. In the field of marketing, Arndt (1967) studied product-related word-of-mouth with respect to the diffusion of information, which led to groundbreaking product growth models (cf. e.g., Bass 1969; Mahajan and Muller 1979). Hereby, diffusion has traditionally been perceived again only as theory of interpersonal communication (Peres et al. 2010, p. 92). Besides this interpersonal communication, some more recent studies suggest incorporating additional potential sources of influence on the diffusion process (e.g., Goldenberg et al. 2010; Van den Bulte and Lilien 2001). Peres et al. (2010, p. 92) consequently state that influence should "[…] include all of the interdependencies among consumers that affect various market players with or without their explicit knowledge". In this context, we generally need to distinguish between social influence and heterogeneity as driving forces of diffusion (Peres et al. 2010, p. 92 f.; Van den Bulte and Stremersch 2004).

In line with French and Raven (1959), who developed one of the most recognized frameworks in the area of social and interpersonal power (Mintzberg 1983), social influence can be defined as "[…] change in the belief, attitude, or behavior of a person […], which results from the action, or presence, of another person […]" (Erchul and Raven 1997, p. 138). Such social influence can be induced by all kinds of consumer interactions such as traditional one-to-one word-of-mouth, the observation of others, or one-to-many communication as in the case of OSN (Godes et al. 2005, p. 416; Nitzan and Libai 2011, p. 25). In literature, the process of social influence is also often referred to as social contagion (e.g., Hinz et al. 2013; Iyengar et al. 2011b; Van den Bulte and Stremersch 2004). Van den Bulte and Wuyts (2007) distinguish five reasons for social contagion (cf. also Van den Bulte and Lilien 2001), with the first two being especially relevant for viral marketing (Hinz et al. 2011, p. 59). First, awareness and interest for a product or innovation might be induced by information transferred for instance by word-of-mouth (cf. e.g., Katz and Lazarsfeld 1955). Second, social learning about benefits, costs, and risks of products, services, or innovations might allow to reduce search efforts and uncertainty (cf. e.g.,Iyengar et al. 2011a). Third, normative pressures might lead to discomfort if a new product or innovation is not adopted, that is, people feel the need to conform to the expectations of their peer group as they wish to fit in (cf. e.g., Asch 1951; Deutsch and Gerard 1955). Fourth, not adopting a product or innovation might even lead to status or competitive disadvantages. In literature, the first three reasons are also referred to as cohesion and the fourth as structural equivalence (Burt 1987). In this context, a recent study by Hinz et al. (2013) indicates that structural equivalence drives adoption more than cohesion. Fifth, network externalities might drive social contagion due to an increasing utility that originates from the consumption of a good when the number of other people consuming this good grows (cf. e.g., Granovetter 1978; Katz and Shapiro 1994).

In contrast, research under the heterogeneity hypothesis claims that diffusion rather depends on heterogeneous consumer characteristics such as innovativeness, price sensitivity, or needs that influence the probability and time of adoption (Peres et al. 2010, p. 92). Since common diffusion models (e.g., Bass 1969) often assume a fully connected and homogenous social network or omit marketing efforts (e.g., Coleman et al. 1966), doubts have been raised whether social influence has been overestimated (Van den Bulte and Lilien 2001; Van den Bulte and Stremersch 2004). Further studies show that the role of social influence may also have been confounded due to several potential sources of bias (cf. e.g., Aral and Walker 2012; Garg et al. 2011; Hartmann et al. 2008), such as simultaneity (i.e., the tendency for connected users to be exposed to the same external stimuli) (Godes and Mayzlin 2004), homophily and endogenous group formation (i.e., the tendency to choose friends and to form social groups with similar tastes and preferences) (Aral et al. 2009; Hartmann 2008; McPherson et al. 2001; Nair et al. 2010), or other contextual and correlated effects (Manski 1993; 2000; Moffitt 2001). Therefore, recent studies attempt to control for heterogeneity and other potential sources of bias (cf. e.g., Garg et al. 2011; Hinz et al. 2013; Nair et al. 2010; Susarla et al. 2012), for instance by conducting large-scale randomized experiments in real-world settings (cf. e.g., Aral and Walker 2012). Other studies have been decomposing the adoption process in its different phases (e.g., awareness and evaluation phase, adoption phase) while incorporating marketing efforts (Manchanda et al. 2008; Van den Bulte and Lilien 2003). All in all, even though also heterogeneity and several other factors play an important role in the diffusion process, the presence of social influence could thus be confirmed and is generally acknowledged today (Iyengar et al. 2011a).

2.2 Characterization of Influential People in Social Networks

Already since Katz and Lazarsfeld (1955) started the discussion about the "flow of mass communications", it is agreed that some people are more influential than others (cf. e.g., Godes and Mayzlin 2009; Goldenberg et. al. 2009; Iyengar et al. 2011a). Their original definition of influential people as "[…] individuals who were likely to influence other persons in their immediate environment" (Katz and Lazarsfeld 1955, p. 3) with respect to their opinions and decisions has remained more or less unchanged until today (Watts and Dodds 2007, p. 442). A central question in this context is how these influential people can be characterized. Katz (1957) states that the ability to influence is related to three (personal and social) factors (cf. Weimann 1991, p. 2): (1) the personification of certain values ("who one is"), (2) the competence ("what one knows"), and (3) the strategic social location ("whom one knows"). This categorization finds also affirmation in the works of Gladwell (2000) and Watts and Dodds (2007). The first factor alludes to distinct characteristics, that is, abilities which make a person persuasive. For instance, usually salesmen have these charismatic traits and communication abilities to successfully convince people (Gladwell 2000, p. 70; Eccleston and Griseri 2008, p. 595). Watts and Dodds (2007, p. 442) characterize such people to be respected by others. The second factor relates to mavens, that is, highly informed individuals (Watts and Dodds 2007, p. 442) or even experts in distinct fields of knowledge (Gladwell 2000; Eccleston and Griseri 2008). Mavens might be especially influential in the case of cohesion driven by information transfer and social learning (cf. e.g., Iyengar et al. 2011a), whereby it is important to bear in mind that peoples' influence might be contextually sensitive. The last factor describes the position of an individual within a society. It specifically refers to connectors, characterized as "[…] people with a special gift for bringing the world together" (Gladwell 2000, p. 38). Such people are usually well-connected (Watts and Dodds 2007, p. 442) and enjoy meeting new people as well as introducing them to others they know (Eccleston and Griseri 2008, p. 594). Thus, people with a high degree of connectedness have the opportunity to influence the behavior of others (Barabási 2003; Van den Bulte and Wuyts 2007). Van den Bulte and Stremersch (2004) point out that such well-connected people might be particularly influential when cohesion (cf. section 2.1.) is at work. In case of competition for status, however, this might not be the case (Burt 1987). Furthermore, tie strength, that is, the intensity of the connections, moderates the impact of social influence (cf. e.g., Brown and Reingen 1987; Burt 1992; Granovetter 1973).

By means of these three – not mutually exclusive – factors, Katz (1957) provided a classification scheme for characterizing influential people in general. With the provided context at hand, we first examine how influential people are characterized in literature on the identification of influential users in OSN:

Q.1 How are influential users characterized in the context of OSN?

2.3 **Identification of Influential People in Social Networks**

Multiple studies investigating the question whether and to what extent people might be influential focused primarily on the strategic location within a social network by taking into account its structural characteristics (cf. e.g., Borgatti 2006, p. 21; Bampo et al. 2008; Kiss and Bichler 2008) (cf. third factor that characterizes influential people, section 2.2). Structural characteristics are here defined as patterns of connections among actors in a social network (cf. Oinas-Kukkonen et al. 2010). The structure resulting from connections among people is mostly described as a set of nodes and directed or undirected edges that connect pairs of nodes. These nodes and edges determining the network structure can be represented by a graph (Watts 2004; Wasserman and Faust 1994).

Several approaches for the identification of important nodes in such a graph can be found in social network analysis (SNA) (for an overview of SNA in the context of marketing cf. e.g., Iacobucci 1996). For instance, numerous measures exist that indicate the social influence of nodes on other nodes in a network (Friedkin 1991). The three most common 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 (for a critical review with respect to a marketing context cf. e.g., Kiss and Bichler 2008; Landherr et al. 2010). 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. Such well-connected nodes are often called "hubs" (Bampo et al. 2008). As Hinz et al. (2011, p. 57 ff.) point out, some studies suggest that these hubs should be considered as influential people (cf. e.g., Iyengar et al. 2011b; Kiss and Bichler 2008; Van den Bulte and Joshi 2007). However, other studies found that "fringes", that is, poorly connected nodes characterized by low degree centrality might be particularly influential (cf. e.g., Galeotti and Goyal 2009; Sundararajan 2006). 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. Prior work also indicates that such "bridges" connecting otherwise unconnected parts of a network should be considered as influential people (cf. e.g., Rayport 1996; Hinz and Spann 2008). A further popular centrality measure, namely eigenvector centrality, is proposed by Bonacich (1972). Since a node's connectivity in the whole network is incorporated (Bolland 1988), approaches based on the 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). Connections to nodes that are themselves influential are therefore assumed to add more to a nodes' influence than connections to less influential nodes (Newman 2003). Thus, eigenvector centrality and related measures such as PageRank deviate from degree, closeness, and betweenness centrality by modeling inherited or transferred status (Liu et al. 2005) that also allows for modeling network effects in the context of viral marketing (cf. e.g., Richardson and Domingos 2002). In conclusion, it can be stated that despite the extensive usage of these well-established centrality measures, "[…] little consensus exists regarding recommendations for optimal seeding strategies" (Hinz et al. 2011, p. 58).

In addition to SNA, the second research stream regarding the identification of influential people goes back to Domingos and Richardson (2001), who studied the so-called "*influence maximization problem*". This refers to the combinatorial optimization problem of identifying the target set of influential people (also often referred to as "top-*k* nodes") that allows for maximizing the information cascade in the context of viral marketing (cf. also Richardson and Domingos 2002). By applying three approximation algorithms to their NP-hard problem, Domingos and Richardson (2001) were able to prove that the selection of the "right" target set can make a substantial difference for a marketing campaign. Based on these works, Kempe et al. (2003) investigated two of the "[…] most basic and widely-studied diffusion models" (Kempe et al. 2003, p. 138), that is, the linear threshold (LN) and the independent cascade (IC) model. Both models are so-called susceptible/infectious/recovered (SIR) models that do not allow for multiple activations of the same node: The IC model is usually considered as a push model, since nodes (information sender) independently try to propagate information to connected nodes in the network. In contrast, the LN model can be considered as a pull model, where nodes (information receiver) accept information if many connected nodes have already accepted it. In this case, acceptance of propagated information is determined by a random threshold. Even though Kempe et al. (2003, p. 138) found that also under the IC and LN model it is NP-hard to determine the target set of influential people, they were able to derive the first approximation guarantee for the proposed greedy algorithm by arguing that their objective function is monotone and submodular (for a more general model and further approximation algorithms cf. e.g., Chen et al. 2009;

Leskovec et al. 2007). Moreover, the proposed approximation algorithm significantly out-performed heuristics based on centrality measures (Kempe et al. 2003). Even-Dar and Shapira (2011) apply another approach to solve the influence maximization problem, namely the so-called voter model. While the IC and LN model consider only the status of the network in the case of convergence to the steady state (Bonchi et al. 2011, p. 24), the voter model can be applied with different target times. Furthermore, it also overcomes a major limitation of the approach by Kempe et al. (2003), that is, the assumption that only one player introduces a product into the market. Besides Even-Dar and Shapira (2011), also Bharathi et al. (2007) and Carnes et al. (2007) suggested approaches for solving the influence maximization problem in a competitive environment.

In summary, the first major research stream concerning the identification of influential people in social networks focuses on the strategic location while the second solves the influence maximization problem by applying diffusion models and (greedy) algorithms. However, as discussed within the introduction, these findings may not be transferable to OSN without further reflection. Therefore, we investigate which of the above mentioned and which further approaches are applied in the context of OSN in order to identify influential users. Furthermore, the specific evaluation of these approaches and implications for theory and practice shall be outlined. Hence, we address two further questions in the following:

Q.2 Which approaches have been developed and applied for the identification of influential users in OSN?

Q.3 How have these approaches been evaluated and which implications can be derived for theory and practice?

3 Literature Search

A systematic, comprehensive as well as replicable literature search strategy is regarded essential for a profound literature analysis on a certain topic of interest (vom Brocke et al. 2009). Bandara et al. (2011, p. 4) delineate two important cornerstones for the literature review process: First, one has to define which *sources* shall be searched (Webster and Watson 2002). Second, the precise *search strategy* needs to be defined, that is, relevant search terms, search fields, and an appropriate period of time (Cooper 1998; Levy and Ellis 2006). Finally, we outline the (number of) included and excluded articles and the selection procedure to allow for intersubjective comprehensibility (vom Brocke et al. 2009).

3.1 **Sources**

In order to identify relevant publication organs, some authors suggest focusing on leading journals of the research discipline under investigation (Webster and Watson 2002, p. 16). However, as this restricts the search results beforehand, this approach should only be applied if the topic of interest can be narrowed down to specific journals. Elsewise, a broad database search is advised (Bandara et al. 2011, p. 4). As research on OSN is quite broad and wide-spread over diverse disciplines such as Management Science, Marketing, Information Systems Research, or Computer Science, we conducted an extensive query in quality scholarly literature databases (cf. Table 1) (Levy and Ellis 2006, p. 189; vom Brocke et al. 2009, p. 8). We purposely accept duplicates instead of being limited to journals or conferences provided by a certain vendor (Levy and Ellis 2006, p. 189).

3.2 **Search Strategy**

For querying the scholarly databases, we derived the following search terms from literature, and applied them by string concatenations: As several synonyms for the definition of OSN can be found in literature, we searched for *"social network"* as an umbrella term to cover different term variations, such as Online Social Network or Social Network(ing) Site (cf. Richter et al. 2011). Additionally, we applied the search terms *"influential"* (covering also influential user), *"influencer"*, *"key user", "hub"*, and *"opinion leader"* (cf. Goldenberg et al. 2009, p. 1; Libai et al. 2010, p. 271). We searched the databases with these terms per title, abstract, and keywords. As the first recognizable OSN SixDegrees launched in 1997 (Boyd and Ellison 2007), we chose a six-teen year period for our search reaching from 1997 to 2012. Table 1 summarizes the search strategy.

Table 1 Summary of the Search Strategy

Databases	AIS eLibrary, EBSCOhost, EmeraldInsight, IEEEXplore, INFORMS, ProQuest, ScienceDirect ,SpringerLink, Wiley InterScience
Search Terms	("social network") AND ("influential" OR "influencer" OR "key user" OR "hub" OR "opinion leader")
Search Fields	Title, Abstract, Keywords
Time Period	1997 – 2012

3.3 **Search Results**

In order to identify the relevant articles with respect to our research questions (cf. section 2), at least two authors screened all search results. Only articles were selected which in essence provide a clear proposition on how influential users can be identified. At the same time, at least one of the following criteria had to be fulfilled: (1) The article explicitly focuses on OSN, either as defined within the introduction or on OSN in general without further definition. (2) The article explicitly states that the derived results are applicable for OSN or the applicability is actually demonstrated by means of using an OSN data set.

The initial database queries resulted in 1,912 articles. In a first step, we analyzed each article regarding its title, abstract, and publication organ in order to exclude all articles, which obviously did not match our research focus. This reduced the set of articles to 180. In a second step, we examined these articles by means of a full-text review to verify whether an article corresponds to our research question and to assess the quality of the article's publication organ. In this process, we excluded articles that were obviously not subject to some kind of formalized peer-review or quality verification (Levy and Ellis 2006, p. 185). Apart from journals, also conferences[3] were considered (Webster and Watson 2002, p. 16) as they offer valuable contributions in the exchange of ideas and promote the development of new research agendas (Levy and Ellis 2006, p. 185). Articles that were too short for a thorough content analysis (e.g., contributions for a poster session) (Poeppelbuss et al. 2011, p. 509), professional magazines, newspapers, or patents were excluded (Levy and Ellis 2006, p. 185). As on the one hand the field of research on OSN is relatively young (Richter et al. 2011, p. 89) and on the other hand methods as well as theories need some time to be verified and generally accepted, we also excluded books.. By these means, we obtained 12 articles proposing actual approaches for the identification of influential users in OSN. By backward search, that is, by studying each article's references (Levy and Ellis 2006, p. 191), we located another four relevant articles. In summary, a set of 16 articles serves as the basis for our subsequent analysis.

4 Findings and Critical Discussion

In the following, we analyze the relevant articles with respect to the three research questions. As all these articles deal with the identification of influential *people* in the context of OSN, we hereafter refer to them as influential *users*.

Q.1 How are influential users characterized in the context of OSN?

The broadly accepted fact that some people are more influential than others (Katz and Lazarsfeld 1955) seems to hold true also in OSN (Libai et al. 2010). As outlined in section 2.2, Katz (1957) observed in an offline context that personal influence is related to three (personal and social) factors, namely: "who one is", "what one knows", and "whom one knows" (Katz 1957, p. 73). These categories have been confirmed to be also applicable for a Web 2.0 context by Eccleston and Griseri (2008). To determine the influence of users in OSN, Eirinaki et al. (2012) deduced two properties, namely popularity and activity, together with several parameters for their measurement in OSN. Looking closely at the parameters of popularity suggested by Eirinaki et al. (2012), the factors "who one is" and "whom one knows" by Katz (1957) can be found to be covered. However, the original three (personal and social) factors need to be complemented by users' activity for the analysis of influence in the context of OSN: First, influential people in general tend to be more involved in personal

[3] If workshop or conference papers were identified that have been published also in a journal, only the journal article were considered when in essence the key findings remained the same.

communication than others (Weimann et al. 2007, p. 175). Second, users in OSN like Facebook have up to several hundred friends whereof only a very small portion actually interacts (Heidemann et al. 2010) and some users are actually totally inactive (Cha et al. 2010). Consequently, pure connectedness of users does not necessarily guarantee influence (Goldenberg et al. 2009; Trusov et al. 2010, p. 646). Additionally, implicit connections that cannot be gathered via explicit friendship connections between users, for instance, explicated via voting, sharing, or bookmarking, can be captured by accounting for users' activity (Bonchi et al. 2011, p. 6). Third, new possibilities induced by the previously unknown amount of data on users' activity allows for incorporating users' activity as further factor. Accordingly, we analyzed the relevant articles by means of the four (not mutually exclusive) factors "who one is", "what one knows", "whom one knows", and "how active one is". Table 2 illustrates the findings.

Table 2 Overview of the Characteristics Considered by the Relevant Articles

References	"Who one is"	"What one knows"	"Whom one knows"	"How active one is"
Aral and Walker (2012)	●	○	●	○
Canali and Lancellotti (2012)	○	○	●	●
Eirinaki et al. (2012)	●	○	●	●
Goldenberg et al. (2009)	○	○	○	●
Heidemann et al. (2010)	○	○	○	●
Hinz et al. (2011)[4]	○	○	●	◑
Ilyas and Radha (2011)	○	○	●	●
Kim and Han (2009)	○	○	●	●
Kimura et al. (2007)	○	○	◑	◑
Lerman and Ghosh (2010)	○	○	●	●
Ma et al. (2008)	○	○	◑	◑
Narayanam and Narahari (2011)	○	○	◑	◑
Saito et al. (2012)	○	○	◑	◑
Trusov et al. (2010)	○	○	○	●
Zhang et al. (2010)	●	○	◑	◑
Zhang et al. (2011)	○	●	◑	◑

Not Considered ○ ***Considered*** ● ***Not further explicated*** ◑

[4] Hinz et al. (2011) identify influential users in OSN by means of users' social position ("whom one knows") and thereby also reveal a significant correlation between users' social position and activity in OSN ("how active one is"). Based on a real-life viral marketing campaign of a mobile phone provider, the authors confirm the influence of these two factors on viral marketing success and further reveal a significant influence of customer characteristics ("who one is"). As we restricted our focus to OSN, however, these findings based on a telecommunication network have not been incorporated in Table 2.

Overall, the majority of the relevant articles relies on broad definitions of influential users or remains imprecise about which characteristics are taken into account. Surprisingly, two factors ("who one is" and "what one knows") are hardly considered, although for instance Zhang et al. (2011, p. 1512) find that different topics ("what one knows") lead to different results regarding the set of users that should be selected in order to influence most users in an OSN. In summary, we observe that current approaches barely consider user specific attributes or users' knowledge on certain topics.

After the synthesis of how influential users are characterized within our set of articles, in the following we examine the articles with respect to the proposed methods along with their evaluation and implications.

Q.2 Which approaches have been developed and applied for the identification of influential users in OSN?

Q.3 How have these approaches been evaluated and which implications have been derived?

With respect to the two outlined major research streams (cf. section 2.3), six of the relevant articles apply approaches that are generally based on the strategic location of nodes in a graph (cf. Table 3). Since a static and potentially inactive social link (often so-called "friendship relationship") in OSN does not guarantee an exchange of information and thus influence, Goldenberg et al. (2009) and Heidemann et al. (2010) define activity graphs where links between users do not represent friendship connections but the activity between nodes (e.g., messages, visits). Based on a directed activity graph, Goldenberg et al. (2009, p. 5) identify influential users by looking for hubs "[…] with in- and out-degrees larger than three standard deviations above the mean". By analyzing Cyworld, the authors find that users with high degree centrality scores adopt products earlier due to their large number of active connections to other users. Furthermore, a users' innovativeness was estimated in terms of adoption timing across multiple products. The authors differentiate innovators (who adopt before anyone else in the neighborhood) and followers and thereby reveal that the former mainly influence the speed of adoption and the latter market size. Thus, Goldenberg et al. (2009, p. 10) conclude that hubs "[…] could be an efficient target for word-of-mouth campaigns, leading to both faster growth and increased market size". Heidemann et al. (2010) define an undirected activity graph with weighted activity links representing the number of exchanged communication activities among users. By adapting the PageRank algorithm to account for the weighted graph, influential users are identified by means of high rankings among all users' PageRank scores. The authors apply their approach to a Facebook data set and show that their algorithm allows to identify more users that can be retained as active users in the future than when drawing on other centrality measures or users' prior communication activity.

Besides these two articles focusing on the activity graph, the remaining four articles model a social graph consisting of social links, that is, friendship connections among users in OSN. Lerman and Ghosh (2010) argue that in general, dynamic social processes (e.g., information diffusion) as well as centrality measures to identify influential users can either be conservative (random walk-based) or non-conservative (broadcast-based). Since the diffusion of information is a non-conservative process, they hypothesize that accordingly non-conservative centrality measures (e.g., degree centrality, (normalized) α-centrality) perform better than conservative ones (e.g., PageRank, betweenness centrality). By analyzing a Digg data set, Lerman and Ghosh (2010) confirm this hypothesis and find that in their case (normalized) α-centrality performs best. Hinz et al. (2011), however, find that targeting users in OSN with both high degree (non-conservative) and betweenness centrality scores (conservative) is particularly beneficial as well-connected users are more likely to participate in viral marketing campaigns. The authors further observed that hubs do not have more influence on other users per se, they only use their greater reach more actively. In contrast to the so far discussed articles, Ilyas and Radha (2011) rather aim at identifying influential neighborhoods than single influential users. Therefore, they apply principal component centrality (PCC) in an undirected (weighted) social graph. Using the example of an Orkut and a Facebook data set (in order to incorporate also user activity, the authors weight the social links by the number of users' interactions in the latter case), they show that in comparison to the application of eigenvector centrality the number of identified influential neighborhoods and users can be increased by applying PCC. The authors further find that the tendency of eigenvector centrality to identify a set of influential users within the same region of a massive graph of an OSN can be overcome by their proposed approach (Ilyas and Radha 2011). Finally, Kim and Han (2009) propose to first rank users by their corresponding degree centrality scores in an undirected social graph. Second, the authors suggest identifying influential users by selecting the users with the highest centrality scores and the highest activity indices calculated as weighted sum of selected activity indicators (e.g., number of groups, updated content per day). By analyzing the diffusion of a Facebook game, the authors find that targeting their identified influential users achieves increasing growth rates and a higher number of new adopters than when addressing mediocrities (Kim and Han 2009). Table 3 summarizes the approaches and findings.

Table 3 Articles Focusing on the Strategic Location of Users in OSN

Articles	Approaches and Findings
Goldenberg et al. (2009)	Propose to identify influential users by looking for **hubs** in a **directed graph** based on **activity links**. Define hubs as users "[…] with both **in- and out-degrees larger than three standard deviations above the mean**". Analyze **Cyworld** and suggest targeting hubs, who lead to both **faster growth** and **increased market size**.
Heidemann et al. (2010)	Propose an **adapted PageRank** to identify influential users in an **undirected** and **weighted graph** based on **activity links**. Evaluate the approach by means of a **Facebook** data set and find that **more users** that are **retained** can be identified than when users' prior communication activity (second best) or applying other centrality measures such as degree centrality (third best).
Hinz et al. (2011)	Propose **degree** and **betweenness centrality** to identify influential users in graphs based on **social links**. Apply different seeding strategies in **anonymous OSN** and a telecommunication network. Find that **hubs** and **bridges** are **more likely to participate in viral marketing campaigns** and hubs use their greater reach more actively.
Ilyas and Radha (2011)	Propose **principal component centrality** (PPC) to identify influential users at the center of **influential neighborhoods** in an **undirected (weighted) graph** based on **social links**. Apply their approach to **Orkut** and **Facebook** and find that in comparison to the application of eigenvector centrality the **number of identified influential neighborhoods** and **users** can be **increased**.
Kim and Han (2009)	Propose to identify influential users by first computing **degree centrality** in an **undirected graph** based on **social links** and then estimating an **activity index**. Evaluate their approach by means of the diffusion of a **Facebook** game. Find that targeting their identified influential users **increases growth rates** and leads to **higher numbers of new adopters**.
Lerman and Ghosh (2010)	Propose **(normalized) α-centrality** to identify influential users in **non-conservative diffusion processes** in a **directed (weighted) graph** based on **active social links**. Evaluate the approach by means of a **Digg** data set and find that the **non-conservative** model of **(normalized) α-centrality performs better** than conservative models of influence when identifying influential users in non-conservative processes such as **information propagation**.

Besides the six articles that apply approaches based on the strategic location of users in OSN (cf. Table 3), another six of all relevant articles focus on solving the influence maximization problem (top-*k* nodes problem) by different approximation algorithms (cf. Table 4). In contrast to the former articles, it becomes apparent that none of the latter ones, which will be discussed in the following, specifies whether the underlying directed or undirected graph is based on social or activity links. Four of the articles use SIR models (cf. section 2.3) to model the diffusion process. While Kimura et al. (2007) mainly focus on the design of an efficient approximation algorithm for the solution of the influence maximization problem based on bond percolation, Zhang et al. (2010) and Zhang et al. (2011) aim at incorporating more personal and social factors of influential users (cf. section 2.2) instead of solely their connectivity. Therefore, Zhang et al. (2010) incorporate similarity between users and Zhang et al. (2011) account for users' preferences for specific topics by weighting the graphs' links. Contrary to Kempe et al. (2003), Zhang et al. (2010) were able to show that due to richer information incorporated in the social graph, a degree centrality-based algorithm often performs even better than the general and hill-climbing greedy algorithm. Narayanam and Narahari (2011) select a fundamentally different approach and suggest a Shaply value-based influential nodes (SPIN) algorithm based on an appropriately defined cooperative game. The authors show that their algorithm cannot only solve the top-*k* nodes problem investigated in all articles displayed in Table 4, but also the λ-coverage problem, that is, finding a minimum set of influential nodes that influences a given percentage λ of nodes in the network. Furthermore, the authors show that their algorithm is more computationally efficient and yields a higher performance in terms of quality than the algorithms proposed by Kempe et al. (2003), Leskovec et al. (2007), and Chen et al. (2009). The article of Ma et al. (2008) also differs from the previously discussed approaches. Instead of using a SIR model, the authors model diffusion by a heat diffusion process. Thus, the approach cannot only capture users that diffuse positive information but also those who exert negative influence on other users (even if these users already adopted, e.g., a product). Moreover, their approach allows for planning marketing strategies sequentially in time, as a time factor is included. Apart from Ma et al. (2008), also Saito et al. (2012) take into account the time factor. Thus, the authors apply a susceptible/infected/susceptible (SIS) model and define a final-time and an integral-time maximization problem. While the first problem solely concentrates on how many nodes are influenced at a point in time, the second problem focuses on the question of how many nodes have been influenced throughout a period of time. By solving the two problems with a greedy algorithm, Saito et al. (2012) find that more influential users can be discovered than by applying approaches based on centrality measures. Furthermore, the identified influential users differ remarkably depending on the chosen influence maximization problem. Therefore, the authors conclude that "[…] it is crucial to choose the right objective function that meets the need for the task" (Saito et al. 2012, p. 632). Table 4 summarizes the approaches and findings.

Table 4 Articles Focusing on the Solution of the Influence Maximization Problem

Articles	**Approaches and Findings**
Kimura et al. (2007)	Examine the influence maximization problem (**top-*k* nodes problem**) using **SIR** models (namely the **IC** and **LT** model) in a **directed graph**. Solve the problem by means of the greedy hill climbing algorithm based on **bond percolation** and demonstrate a **higher performance** and a large **reduction in computational cost** in comparison to the conventional method that simulates the random process many times.
Ma et al. (2008)	Examine the influence maximization problem (**top-*k* nodes problem**) using a **heat diffusion process** in a **directed** and an **undirected graph**. Solve the problem under a **top-*k***, ***k*-step greedy**, and **enhanced *k*-step greedy algorithm**. Apply their approach to an **Epinion** data set and show that not only the diffusion of positive but also of **negative information** can be modeled. Furthermore, the included **time factor** allows for planning viral marketing campaigns sequentially in time.
Narayanam and Narahari (2011)	Examine the influence maximization problem (**top-*k* nodes problem**) and the **λ-coverage problem** (finding a minimum set of influential nodes that influences a given percentage λ of nodes in the network) using a **SIR** model (namely **LT**) in a **directed graph**. Solve both problems by means of the **Shaply value-based influential nodes (SPIN)** algorithm based on a **cooperative game**. Show that the SPIN algorithm is **more powerful** and **computationally efficient** than existing algorithms.
Saito et al. (2012)	Examine the influence maximization problem (**top-*k* nodes problem**) using **SIS** models as **final-time** and **integral-time maximization problem** in a **directed graph**. Solve the problems under the greedy algorithm on the basis of **bond percolation**, **pruning**, and **burnout**. Find that **more influential users** can be discovered than by approaches based on **centrality measures** and that the **identified influential users differ remarkably** depending on the chosen influence maximization problem.
Zhang et al. (2010)	Examine the influence maximization problem (**top-*k* nodes problem**) using a **SIR** model (namely **LT**) in a **directed graph**. Adapt the LT model by weighting edges that **account for similarity between users**. Solve the problem by applying **centrality**, **greedy**, and **combined algorithms.** Apply their approach to an **Epinion** data set and show that the graph built by "trust" and "review-rate" includes more information on the social network. Thus, a **degree centrality-based algorithm** often **performs** even **better** than the general and hill-climbing greedy algorithm.
Zhang et al. (2011)	Examine the influence maximization problem (**top-*k* nodes problem**) using a **SIR** model (namely **IC**) in an **undirected graph**. Adapt the IC model by weighting edges that **account users' preferences for specific topics**. Solve the problem under an optimized **greedy algorithm** including Monte Carlo simulation. Experimental results show that the approach significantly **outperforms the traditional greedy algorithm** in terms of information diffusion on **specific topics**.

Finally, four of the identified articles apply approaches for the selection of influential users in OSN which cannot be attributed to one of the two above mentioned research streams. The first article by Aral and Walker (2012) proposes hazard models to measure the moderating effect of individual level attributes (e.g., gender, age) on influence, susceptibility, and dyadic peer-to-peer influence. By conducting a large-scale in vivo randomized experiment in Facebook, bias by confounding effects, homophily, unobserved heterogeneity etc. could be eliminated (Aral and Walker 2012). The results indicate that there are remarkable differences between the individual level attributes characterizing influencers and susceptibles. For instance, susceptibility decreases with age and women are less susceptible than men. Influence is also exerted mostly to users of the same age, men are more influential than women, and influential users cluster in the network. In conclusion, Aral and Walker (2012, p. 340) highlight that (1) influential users need to be targeted, since they are unlikely to adopt due to influence by other users, (2) "[…] being influential is not simply a consequence of having susceptible peers […]", as diffusion depends on both influence and susceptibility, and that (3) "[…] targeting should focus on the attributes of current adopters […] rather than attributes of their peers […]", since there are more users with high influence scores than with high susceptibility scores. Canali and Lancellotti (2012) also differentiate and analyze "sources", that is, users that propagate information that receives the most attention of other users, and "targets", that is, users that access most information. The authors propose a principal component analysis (PCA) to select and combine relevant user attributes (e.g., number of friends, number of comments). By applying their approach to a YouTube and Flickr data set, they show that the approach is robust and effective, as it identifies more targets and sources than by applying in-degree centrality. Eirinaki et al. (2012) apply a similar approach and suggest selecting and combining a set of profile-based characteristics representing popularity (e.g., number of friends, received comments) and activity (e.g., number of updates, last log-in time). By applying their approach to a synthetic and a MySpace data set, the authors find that influential users that might not have been included by betweenness centrality or PageRank can be identified as not only users' connectedness but also activity is taken into account. To account for the importance of users' activity, Trusov et al. (2010) suggest a nonstandard form of Bayesian shrinkage implemented in a Poisson regression, which is based on users' daily log-ins. The authors apply their approach to an anonymous OSN and find that only few social links of a user actually have influence on his or her behavior. They further show that their approach identifies more users that influence others' activity than simpler alternatives such as degree centrality or an approximation by the number of a user's profile views. Table 5 summarizes the approaches and findings.

Table 5 Articles Focusing on Further Approaches

Articles	Approaches and Findings
Aral and Walker (2012)	Propose to identify influential users by applying **hazard models** to measure the moderating effect of **individual level attributes** on influence, susceptibility, and dyadic peer-to-peer influence. By conducting a large scale in vivo randomized experiment in **Facebook** it is shown that **susceptibility decreases with age**, but **increases** with increasing **relationship commitment** until marriage, that **men are more influential than women**, users exert most influence on other users of the **same age**, and influential users **cluster in the network**.
Canali and Lancellotti (2012)	Propose to apply **principal component analysis** (PCA) to **select and combine user attributes** that allow for identifying influential users. Differentiate between "**sources**" and "**targets**". Apply their approach to a **YouTube** and **Flickr** data set to show that it is robust and effective. Find that their approach makes it possible to **identify more targets** and **sources** than when applying in-degree centrality.
Eirinaki et al. (2012)	Propose to identify influential users by **selecting and combining** a set of **profile-based characteristics** representing popularity and activity. Apply their approach to a **synthetic** and **MySpace** data set. Find that their approach allows for identifying influential users that might have not been found by betweenness centrality or PageRank as not only users' **connectedness** but also **activity** is **taken into account**.
Trusov et al. (2010)	Propose to identify influential users by a **nonstandard form of Bayesian shrinkage implemented in a Poisson regression**. Apply their approach to an **anonymous** OSN and find that only few social links of a user actually have influence on his or her behavior. Also their approach **identifies more users** that **influence others' activity** than simpler alternatives such as degree centrality or an approximation by the number of a user's profile views.

5 Future Research Directions

Online and offline social influence might not be the same.

Even though there have been first studies comparing offline and online social network constructs, such as tie strength (cf. e.g., Brown et al. 2007), many articles on the identification of influential users in OSN draw on theories and previous findings that have been originally derived in an offline context without critical reflection (cf. section 2.1). For instance, the visibility of social actions in OSN might lead to new forms of social influence, "[…] which rather than flowing from the actor to the observer, flows from the observer to the actor" (Sundararajan et al. 2012, p. 8). Thus, companies might be able to develop marketing strategies that "[…] incorporate targeting advisees, not just advisers", as suggested by Hinz et al. (2013, p. 8). Future research should therefore especially focus on differences and commonalities of offline and online networks (Howison et al. 2011, p. 773). Are there differences between online and offline social systems, and if yes, what are these differences? Are online influencers also influential offline and vice versa? Are online traces reliable mirrors of offline social influence and contagion and does social influence invoked in online settings further spread into the offline world? More work regarding such questions should be encouraged and practitioners need to be aware that concepts developed offline might not work alike in online settings such as OSN.

BISE and Marketing could mutually benefit from more collaboration.

We find that most articles on the identification of influential users in OSN originate either from the scientific Business & Information Systems Engineering (BISE) or the Marketing community. Viewed jointly with our findings presented in section 4, it becomes apparent that marketing-oriented articles extensively draw on rich real-world data of OSN and even collaborate with OSN providers (cf., e.g., Trusov et al. 2010). In contrast, technology-oriented papers from the field of Computer Science have a more theoretical approach and evaluate their artifacts in most cases by use of formal proofs, for instance regarding efficiency and run-time, or in a few cases apply synthetical or other networks' data (e.g., authorship networks) (cf. e.g., Narayanam and Narahari 2011). This may account for the fact that some of the central findings of these rather design-oriented articles are contrary to empirical findings from the Marketing community (e.g., regarding the applicability of degree centrality for the identification of influential users in OSN). Therefore, we believe that an even stronger collaboration between the scientific BISE and Marketing community than we find today could be mutually beneficial by exchanging data on OSN, knowledge about efficient and automated algorithms that actually can handle the vast amount of data in OSN, or contacts to OSN providers. Furthermore, the actual design and implementation of algorithms in cooperation with companies or OSN providers, for instance by conducting Action Design Research (cf. Sein et al. 2011), could be facilitated in future research. To do so, however, access and privacy challenges need to be overcome in order to acquire reliable data (Howison et al. 2011, p. 775; Lazer et al. 2009, p. 722). Therefore, "[r]obust models of collaboration and data sharing between industry and academia are needed" and "[r]esearchers themselves must develop technologies that protect privacy while preserving data essential for research" (Lazer et al. 2009, p. 722).

A human being and his or her behavior are not just nodes and links in a graph.

The majority of the articles do neither incorporate personal information on users that allows for assessing "who one is" or "what one knows" (cf. Table 2). However, Trusov et al. (2010, p. 645) and Hinz et al. (2011, p. 68), for instance, find that having many friends (i.e., social links) does not make users influential per se. Thus, focusing solely on "whom one knows" (cf. Table 2) might not be sufficient to identify influential users in OSN. Instead, there is remarkable heterogeneity among users in OSN, that is, the average user is influenced by relatively few other users and in turn, influences few other users (Trusov et al. 2010, p. 645). Prior research states that "[…] influence […] cannot be simply traced back to the graph properties […] but also depends on the personality and emotions of the human being behind it" (Quercia et al. 2011, p. 1). Furthermore, it has been emphasized that social influence in OSN is not a "[…] unidimensional measure, but a combination of personal traits with social network positioning […]" (Weimann 1991, p. 276). However, empirical studies of how individual attributes of users moderate social influence can hardly be found. A first study by Aral and Walker (2012) finds that social influence and susceptibility of users heavily depends on the individual level attributes of users (e.g., age, gender). This is also confirmed by Katona et al. (2011), who find that some demographic variables are good predictors of adoption. On the other hand, social influence is often overestimated, as homophily actually accounts for a large share of social contagion (cf. section 2.3). Zhang et al. (2011) emphasize that the identification of influential users also depends on users' preferences for specific topics as the diffusion of information differs among topics (cf. e.g., Saito et al. 2009; Saito et al. 2010). Thus, practitioners targeting influential users in OSN should take into account not only the specific characteristics of the users but also of their advertised products and services. We consequently believe that more research is needed to investigate the relationships between the personal and social factors of influential users, the distribution of these factors across users, and the homophily in the formation of social and activity links in OSN. With respect to these links, also questions regarding the selection and combination of different link types (e.g., social and activity links), their intensity (e.g., denoted by weights based on the number of communication activities, cf. Heidemann et al. 2010), and the role of missing links (e.g., does the absence of traces for a link in the data set under consideration provide evidence for the absence of social influence?) should be addressed in more detail in future research (Howison et al. 2011).

Not just positive information might be propagated.

Except for the article by Ma et al. (2008) (cf. Table 4), none of the analyzed articles explicitly models the diffusion of positive *and* negative information in OSN. However, prior research on word-of-mouth in general found that negative word-of-mouth is more likely and stronger than positive word-of-mouth (Anderson 1998; Bone 1995): While on average dissatisfied customers can be expected to tell eleven persons, satisfied ones only tell about five persons about their experiences (Heskett et al. 1997). Thus, negative word-of-mouth is about twice as likely as positive word-of-mouth (Mangold et al. 1999). Also in an online context, Chevalier and Mayzlin (2006) found that the impact of a negative review on sales was greater than the impact of a positive one, and Berger and Milkman (2012) showed that content provoking negative emotions such as anger or anxiety tended to be exceptionally viral. Therefore, practitioners need to be aware that targeting influential users in OSN can also incorporate a certain risk of negative information diffusion. In order to better understand the role of influential users propagating negative information in OSN, future research should also develop diffusion models that incorporate a certain degree of (influential) users that do not solely spread positive information.

The one who leads might not follow.

Most of the discussed approaches (cf. section 4) try to identify the most influential users that should be targeted in order to maximize the impact of a marketing campaign. However, as Watts and Dodds (2007, p. 442) state, "[…] it is generally the case that most social change is driven not by influentials but by easily influenced individuals influencing other easily influenced individuals". Aral and Walker (2012) point out that the susceptibility hypothesis is for instance well represented in theoretical threshold-based models (cf. section 2.3), which are also used by some of the approaches discussed in section 4 (cf. Table 4). However, besides Aral and Walker (2012) and partly Canali and Lancellotti (2012) as well as Trusov et al. (2010), none of the discussed articles analyzes the role of susceptibility in depth. Particularly behind the backdrop of the findings of Aral and Walker (2012) outlined in section 4, it still seems to be promising for practitioners to address influential users in OSN. Nevertheless, further research is needed to enrich our understanding of the role of susceptible users and their individual characteristics as well as their interplay with influential users in OSN (cf. e.g., Hinz et al. 2013).

You are not alone.

None of the discussed articles considers optimal seeding strategies in a competitive environment (cf. section 2.3). However, due to the sheer size and the high number of connections to other users in OSN, isolated diffusion processes may not be representative for reality. Furthermore, users in OSN are exposed to a tremendous amount of information (Canali and Lancelotti 2012, p. 29). This information overload may cause users in OSN to be less easily influenced as they simply cannot process all the information that they are exposed to (Hinz et al. 2011, p. 58). Therefore, practitioners need to be aware that competing marketing campaigns or information overload may diminish the effects of viral marketing campaigns. We believe that further research is needed to better understand the consequences of parallel (competing) viral marketing campaigns, for example regarding different products of one company or simultaneous marketing campaigns of different companies, and the impact of information overload.

Degree centrality is not that bad.

Our analysis shows that most articles focusing on the solution of the influence maximization state that their approaches outperform simpler approximations such as degree centrality (cf.

Table 4). However, this is in contrast to a number of articles which find that particularly users with high degree centrality scores (i.e., hubs) are in fact the influential users in OSN (cf. Table 3). This finding is also verified by Zhang et al. (2010), who show that degree centrality-based algorithms often perform even better than greedy algorithms when approximating the optimal solution of the influence maximization problem. This might be due to richer information, which is incorporated in social graphs of OSN (Zhang et al. 2010). In a similar context also Tang and Yang (2010) find that a simple degree centrality-based algorithm performs almost as well as a complex PageRank based approach. One explanation for these deviating results could be the different evaluation methods as outlined above. In line with related studies (e.g., Kiss and Bichler 2008) we find that degree centrality can be a reasonable measure for the identification of influential users in OSN. However, practitioners targeting users with high degree centrality scores need to be aware of further findings, which indicate that the influential power of users and susceptibility decrease with a rising number of contacts (e.g., Katona et al. 2011; Narayan et al. 2011). Moreover, some articles show that users with high degree centrality scores do not have higher conversion rates due to a higher persuasiveness but are rather more active (e.g., Hinz et al. 2011; Iyengar et al. 2011b). Thus, further research on the optimal centrality of influential users, the actual role of social influence in OSN, and further validations using large-scale data from actual OSN should be encouraged.

Methods, diffusion processes, and network properties need to be aligned.

As Lerman and Ghosh (2010) point out, the diffusion of information is a non-conservative process. However, not only the diffusion process but also centrality measures make implicit assumptions about the nature of the diffusion process (Borgatti 2006). Therefore, the actual underlying diffusion process affects the applied approaches (Ghosh et al. 2011), which hence need to be aligned accordingly. However, for instance Hinz et al. (2011, p. 69) find that it is beneficial to target users with high betweenness centrality scores. This is a conservative centrality measure (Lerman and Ghosh 2010) applied in the context of viral marketing campaigns, whereby diffusion is usually considered as a non-conservative process (Ghosh et al. 2011). Furthermore, Narayanam and Narahari (2011, p. 145) find that "[t]he presence of communities strongly affects the process of identifying influential nodes". This is in line with findings by Kimura et al. (2008), who found that certain community structures are strongly correlated with the greedy solution of their influence maximization problem under the IC model. Ilyas and Radha (2011) go one step further and identify users that form centrality maxima within influential neighborhoods. This is a promising approach for future research, as hardly only one single influential neighborhood exists in OSN with millions of users. Consequently, several users might have relatively low influence scores compared to the whole OSN, but relatively high influence scores within their relevant neighborhoods. Therefore, practitioners and researchers should carefully consider and align their applied methods and approaches to the underlying diffusion processes and network properties when identifying influential users in OSN (cf. Howison et al. 2011, p. 790 f.). However, since not all studies confirm the propositions of Lerman and Ghosh (2010), further research should be encouraged to achieve a deeper understanding about the interplay of centrality measures and diffusion processes.

Efficiency and validity are crucial.

When taking a look at the articles focusing on the solution of the influence maximization problem by using diffusion models and solving them by (greedy) algorithms (cf. Table 4), it becomes apparent that the efficiency of the applied algorithms is a crucial success factor for their applicability in a real-world context (Saito et al. 2012). Therefore, as discussed above, solutions based on well-established centrality measures from SNA are often preferable, even though more sophisticated algorithms might be more accurate (cf. e.g., Zhang et al. 2011). However, the application of SNA in new contexts such as OSN raises several challenges and corresponding validity issues (cf. Howison et al. 2011 for an overview). For instance, building an activity graph requires the aggregation of activity links over time (cf., e.g., Heidemann et al. 2010). This might lead to "[…] networks with different structural properties than the network experienced by participants" (Howison et al. 2011, p. 784), which offers starting points for future research. All in all, practitioners and researchers need to be aware of the trade-off between high accuracy as well as validity and sufficient efficiency for large-scale data sets of OSN. Further research could thus also address questions of optimal levels of accuracy and efficiency from an economical perspective when identifying influential users for marketing purposes in OSN.

6 Conclusion

Who will lead and who will follow? The question of identifying those people who mobilize and propagate influence in networks and society in the most effective way has been intensively analyzed in different research streams over the last decades. Along with the explosive growth of OSN, related changes regarding access and availability of user data, a decreasing impact of traditional marketing techniques, and changes in customer behavior, a great deal of attention was paid to identifying influential users in OSN in recent years. With this context at hand, we focused on identifying relevant publications by means of a structured literature search in order to analyze, synthesize, and assess applied characteristics of and methods for identifying influential users in OSN. It is hoped that the results can stimulate and guide future research in the field.

However, our findings are also subject to limitations: First, although we conducted a broad and structured database search, there is still a certain chance that not all relevant articles have been identified. Furthermore, we selected appropriate search terms derived from literature, but nevertheless additional phrases might have also uncovered a few more relevant papers. Second, due to our focus on OSN we excluded articles that analyze content-oriented sites such as Twitter or YouTube. Thus, our perspective is narrowed and certain approaches and findings that have only been researched on such sites are not considered. Future research could build upon the presented findings, first extending the analysis to also content-oriented sites and second investigating commonalities and differences regarding the identification of influential users in content-oriented sites and OSN. Additionally, the focus on influential users in OSN could in the future be broadened in order to discuss also commonalities and differences of social influence in online and offline settings. Further research might therefore apply a broader definition of OSN and also incorporate studies on offline networks. Besides these limitations, we hope that our findings help interested parties from BISE, Marketing, and beyond to obtain a first overview and better understanding of the body of knowledge regarding the identification of influential users in OSN. Additionally, we hope to have provided directions for future research in this field.

Literature

Anderson EW (1998) Customer satisfaction and word of mouth. Journal of Service Research 1(1):5-17

Aral S, Muchnika L, Sundararajana A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Proceedings of the National Academy of Sciences of the United States of America 106(51):21544-21549

Aral S, Walker D (2012) Identifying influential and susceptible members of social networks. Science 337(6092):337-341

Arndt J (1967) Role of product-related conversations in the diffusion of a new product. Journal of Marketing Research 4(3):291-295

Asch SE (1951) Effects of group pressure upon the modification and distortion of judgment. In: Guetzkow H (ed) Groups, leadership and men. Carnegie Press, Pittsburgh, pp 177-190

Bampo M, Ewing MT, Mather DR, Stewart D, Wallace M (2008) The effect of the social structure of digital networks on viral marketing performance. Information Systems Research 19(3):273-290

Bandara W, Miskon S, Fielt E (2011) A systematic, tool-supported method for conducting literature reviews in information systems. In: Proc 19th European conference on information systems, Helsinki, paper 221

Barabási A-L (2003) Linked: How everything is connected to everything else and what it means. Plume, New York

Bass FM (1969) A new product growth model for consumer durables. Management Science 15(5):215-227

Beal GM, Bohlen JM (1955) How farm people accept new ideas. Cooperative extension service report 15. Ames, Iowa

Beal GM, Bohlen JM (1957) The diffusion process. Cooperative extension service report 18. Ames, Iowa

Beer D (2008) Social network(ing) sites ... revisiting the story so far: a response to Danah Boyd & Nicole Ellison. Journal of Computer-Mediated Communication 13(2):516-529

Berger J, Milkman KL (2012) What makes online content viral? Journal of Marketing Research 49(2):192-205

Bharathi S, Kempe D, Salek M (2007) Competitive influence maximization in social networks. In: Proc 3rd international workshop on Internet and network economics, San Diego, pp 306-311

Bolland JM (1988) Sorting out centrality: An analysis of the performance of four centrality models in real and simulated networks. Social Networks 10(3):233-253

Bonacich P (1972) Factoring and weighting approaches to status scores and clique identification. Journal of Mathematical Sociology 2(1):113-120

Bonchi F, Castillo C, Gionis A, Jaimes A (2011) Social network analysis and mining for business applications. ACM Transactions on Intelligent Systems and Technology 2(3):1-37

Bone PF (1995) Word-of-mouth effects on short-term and long-term product judgments. Journal of Business Research 32(3):213-223

Borgatti SP (2006) Identifying sets of key players in a social network. Computational & Mathematical Organization Theory 12(1):21-34

Boyd DM, Ellison NB (2007) Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication 13(1):210-230

Brown J, Broderick AJ, Lee N (2007) Word of mouth communication within online communities: Conceptualizing the online social network. Journal of Interactive Marketing 21(3):2-20

Brown J, Reingen P (1987) Social ties and word-of-mouth referral behavior. Journal of Consumer Research 14(3):350-362

Burt RS (1987) Social contagion and innovation: Cohesion versus structural equivalence. American Journal of Sociology 92(6):1287-1335

Burt RS (1992) Structural holes. Harvard University Press, Cambridge

Canali C, Lancellotti R (2012) A quantitative methodology based on component analysis to identify key users in social networks. International Journal of Social Network Mining 1(1): 27-50

Carnes T, Nagarajan C, Wild SM, Van Zuylen A (2007) Maximizing influence in a competitive social network: A follower's perspective. In: Proc 9th international conference on electronic commerce, Minneapolis, pp 351-360

Cha M, Haddadi H, Benevenuto F, Gummadi KP (2010) Measuring user influence in Twitter: The million follower fallacy. In: Proc 4th annual conference on weblogs and social media, Barcelona, pp 10-17

Chen W, Wang Y, Yang S (2009) Efficient influence maximization in social networks. In: Proc 15th ACM SIGKDD international conference on knowledge discovery and data mining, New York, pp 199-208

Chen Y, Xie J (2008) Online consumer review: Word-of-mouth as a new element of marketing communication mix. Management Science 54(3):477-491

Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. Journal of Marketing Research 43(3):345-354

Clemons EK (2009) The complex problem of monetizing virtual electronic social networks. Decision Support Systems 48(1):46-56

Coleman JS, Katz E, Menzel H (1966) Medical innovation, a diffusion study. Bobbs-Merrill, New York

Cooper HM (1998) Synthesizing research: A guide for literature review. Sage Publications, Thousand Oak

Corey LG (1971) People who claim to be opinion leaders: Identifying their characteristics by self-report. Journal of Marketing 35(4):48-63

Deutsch M, Gerard HB (1955) A study of normative and informational social influences upon individual judgment. The Journal of Abnormal and Social Psychology 51(3):629-636

Domingos P, Richardson M (2001) Mining the network value of customers. In: Proc 7th ACM SIGKDD international conference on knowledge discovery and data mining, San Francisco, pp 57-66

Eccleston D, Griseri L (2008) How does Web 2.0 stretch traditional influencing patterns? International Journal of Market Research 50(5):591-161

Eirinaki M, Monga SPS, Sundaram S (2012) Identification of influential social networkers. International Journal of Web Based Communities 8(2):136-158

Erchul WP, Raven BH (1997) Social power in school consultation: A contemporary view of French and Raven's bases of power model. Journal of School Psychology 35(2):137-171

Even-Dar E, Shapira A (2011) A note on maximizing the spread of influence in social networks. Information Processing Letters 111(4):184-187

Facebook (2012) One billion fact sheet. http://newsroom.fb.com/imagelibrary/download media.ashx?MediaDetailsID=4227&SizeId=-1, accessed 2012-10-27

Freeman LC (1979) Centrality in social networks: Conceptual clarification. Social Networks 1(3):215-239

French JRP, Raven BH (1959) The bases of social power. In: Cartwright D (ed) Studies in social power. Institute for Social Research, Ann Arbor, pp 150-167

Friedkin NE (1991) Theoretical foundations for centrality measures. American Journal of Sociology 96(6):1478-1504

Galeotti A, Goyal S (2009) Influencing the influencers: A theory of strategic diffusion. RAND Journal of Economics 40(3):509-532

Garg R, Smith MD, Telang R (2011) Measuring information diffusion in an online community. Journal of Management Information Systems 28(2):11-38

Ghosh R, Lerman K, Surachawala T, Voevodski K, Teng SH (2011) Non-conservative diffusion and its application to social network analysis. http://arxiv.org/abs/1102.4639v3, accessed 2012-10-20

Gladwell M (2000) The tipping point: How little things can make a big difference. Abacus Books, London

Godes D, Mayzlin D (2004) Using online conversations to study word-of-mouth communication. Marketing Science 23(4):545-560

Godes D, Mayzlin D (2009) Firm-created word-of-mouth communication: Evidence from a field test. Marketing Science 28(4):721-739

Godes D, Mayzlin D, Chen Y, Das S, Dellarocas C, Pfeiffer B, Libai B, Sen S, Shi M, Verlegh P (2005) The firm's management of social interactions. Marketing Letters 16(3/4):415-428

Goldenberg J, Han S, Lehmann D, Hong, J (2009) The role of hubs in the adoption process. Journal of Marketing 73(2):1-13

Goldenberg J, Libai B, Muller E (2010). The chilling effect of network externalities. International Journal of Research in Marketing 27(1):4-15

Granovetter MS (1973) The strength of weak ties. American Journal of Sociology 78(6): 1360-1380

Granovetter MS (1978) Threshold models of collective behavior. American Journal of Sociology 83(6):1420-1443

Hanneman RA, Riddle M (2005) Introduction to social network methods. University of California, Riverside. http://www.faculty.ucr.edu/~hanneman/nettext/, accessed 2012-06-25

Hartmann WR, Manchanda P, Nair H, Bothner M, Doods P, Godes D, Hosanagar K, Tucker C (2008) Modeling social interactions: Identification, empirical methods and policy implications. Marketing Letters 19(3):287-304

Heidemann J, Klier M, Probst F (2010) Identifying key users in online social networks: A PageRank based approach. In: Proc international conference of information systems, St. Louis, paper 79

Heidemann J, Klier M, Probst F (2012) Online social networks: A survey of a global phenomenon. Computer Networks 56(18):3866-3878

Heskett JL, Sasser WE, Schlesinger LA (1997) The service profit chain. The Free Press, New York

Howison J, Wiggins A, Crowston K (2011) Validity issues in the user of social network analysis with digital trace data. Journal of the Association for Information Systems 12(12):767-797

Hill S, Provost F, Volinsky C (2006) Network-based marketing: Identifying likely adopters via consumer networks. Statistical Science 21(2):256-276

Hinz O, Schulze C, Takac C (2013) New product adoption in social networks: Why direction matters. Journal of Business Research (in press)

Hinz O, Skiera B, Barrot C, Becker JU (2011) Seeding strategies for viral marketing: An empirical comparison. Journal of Marketing 75(6):55-71

Hinz O, Spann M (2008) The impact of information diffusion on bidding behavior in secret reserve price auctions. Information Systems Research 19(3):351-368

Iacobucci D (1996) Networks in marketing. Sage Publications, Thousand Oaks

Ilyas MU, Radha H (2011) Identifying influential nodes in online social networks using principal component centrality. In: Proc international conference on communications, Budapest, pp 1-5

Iyengar R, Van den Bulte C, Choi J (2011a) Distinguishing between drivers of social contagion: Insights from combining social network and co-location data. Working Paper at the Wharton School of the University of Pennsylvania, Philadelphia

Iyengar R, Van den Bulte C, Valente TW (2011b) Opinion leadership and social contagion in new product diffusion. Marketing Science 30(2):195-212

Katona Z, Zubcsek PP, Sarvary M (2011) Network effects and personal influences: The diffusion of an online social network. Journal of Marketing Research 48(3):425-443

Katz E (1957) The two-step flow of communication: An up-to-date report on an hypothesis. Public Opinion Quarterly 21(1):61-78

Katz E, Lazarsfeld PF (1955) Personal influence: The part played by people in the flow of mass communications. Transaction Publishers, New Brunswick

Katz ML, Shapiro C (1994) Systems competition and network effects. The Journal of Economic Perspectives 8(2):93-115

Kempe D, Kleinberg J, Tardos E (2003) Maximizing the spread of influence through a social network. In: Proc 9th ACM SIGKDD international conference on knowledge discovery and data mining, New York, pp 137-146

Kim ES, Han SS (2009) An analytical way to find influencers on social networks and validate their effects in disseminating social games. In: Proc international conference on advances in social network analysis and mining, Athens, pp 41-46

Kimura M, Saito K, Nakano R (2007) Extracting influential nodes for information diffusion on a social network. In: Proc 22nd national conference on artificial intelligence, Vancouver, pp 1371-1376

Kimura M, Yamakawa K, Saito K, Motoda H (2008) Community analysis of influential nodes for information diffusion on a social network. In: IEEE international joint conference on neural networks, Hong Kong, pp 1358-1363

Kiss C, Bichler M (2008) Identification of influencers - Measuring influence in customer networks. Decision Support Systems 46(1):233-253

Kossinets G, Watts DJ (2006) Empirical analysis of an evolving social network. Science 311(5757):88-90

Landherr A, Friedl B, Heidemann J (2010) A critical review of centrality measures in social networks. Business & Information Systems Engineering 2(6):371-385

Laine MSS, Ercal G, Bo L (2011) User groups in social networks: An experimental study on YouTube. In: Proc 44th Hawaii international conference on system sciences, Kauai, pp 1-10

Lazer D, Pentland A, Adamic L, Aral S, Barabási A-L, Brewer D, Christakis N, Contractor N, Fowler J, Guttmann M, Jebara T, King G, Macy M, Roy D, Alstyne MV (2009) Computational social science. Science 323(5915):721-723

Lerman K, Ghosh R (2010) Information contagion: An empirical study of the spread of news on Digg and Twitter social networks. In: Proc 4th international AAAI conference on weblogs and social media, Washington D.C., pp 90-97

Leskovec J, Adamic LA, Hubermann BA (2007) The dynamics of viral marketing. ACM Transactions on the Web 1(1):article 5

Levy Y, Ellis TJ (2006) A systems approach to conduct an effective literature review in support of information systems research. Informing Science Journal 9:181-212

Libai B, Bolton R, Buegel MS, De Ruyter K, Goetz O, Risselada H, Stephen AT (2010) Customer-to-customer interactions: Broadening the scope of word of mouth research. Journal of Service Research 13(3):267-282

Liu X, Bollen J, Nelson ML, Van De Sompel H (2005) Co-authorship networks in the digital library research community. Information Processing & Management 41(6):1462-1480

Ma H, Yang H, Lyu MR, King I (2008) Mining social networks using heat diffusion processes for marketing candidates selection. In: Proc 17th ACM conference on information and knowledge management, Napa Valley, pp 233-242

Mahajan V, Muller E (1979) Innovation diffusion and new product growth models in marketing. Journal of Marketing 43(4):55-68

Manchanda P, Xie Y, Youn N (2008) The role of targeted communication and contagion in product adoption. Marketing Science 27(6):961-976

Mangold WG, Miller F, Brockway GR (1999) Word-of-mouth communication in the service marketplace. Journal of Services Marketing 13(1):73-89

Mansfield E (1961) Technical change and the rate of imitation. Econometrica 29(4):741-766

Manski CF (1993) Identification of endogenous social effects: The reflection problem. The Review of Economic Studies 60(3):531-542

Manski CF (2000) Economic analysis of social interactions. Journal of Economic Perspectives 14(3):115-136

McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. Annual Review of Sociology 27:415-444

Mintzberg H (1983) Power in and around organizations. Prentice-Hall, Englewood Cliffs

Moffitt R (2001) Policy interventions, low-level equilibria, and social interactions. In: Durlauf SN, Young HP (eds) Social dynamics. MIT Press, Cambridge, pp 45-82

Moreno JL (1934) Who shall survive. Beacon House, New York

Nair HS, Manchanda P, Bhatia T (2010) Asymmetric social interactions in physician prescription behavior: The role of opinion leaders. Journal of Marketing Research 47(5):883-895

Narayan V, Rao VR, Saunders C (2011) How peer influence affects attribute preferences: A Bayesian updating mechanism. Marketing Science 30(2):368-384

Narayanam R, Narahari Y (2011) A Shapley value-based approach to discover influential nodes in social networks. IEEE Transactions on Automation Science and Engineering 8(1):130-147

Newman MEJ (2003) The structure and function of complex networks. SIAM Review 45(2):167-256

Nitzan I, Libai B (2011) Social effects on customer retention. Journal of Marketing 75(6): 24-38

Oinas-Kukkonen H, Lyytinen K, Yoo Y (2010) Social networks and information systems: Ongoing and future research streams. Journal of the Association for Information Systems 11(2):61-68

Pallis G, Zeinalipour-Yazti D, Dikaiakos MD (2011) Online social networks: Status and trends. In: Vakali A, Jain LC (eds) New Directions in Web Data Management 1. Springer, Berlin, pp 213-234

Peres R, Muller E, Mahajan V (2010) Innovation diffusion and new product growth models: A critical review and research directions. International Journal of Research in Marketing 27(2):91-106

Poeppelbuss J, Niehaves B, Simons A, Becker J (2011) Maturity models in information systems research: Literature search and analysis. Communications of the Association for Information Systems 29:505-532

Quercia D, Ellis J, Capra L, Crowcroft J (2011) In the mood for being influential on Twitter. In: 3rd international conference on social computing, Boston, pp 307-314

Rapoport A (1952) “Ignition” phenomena in random nets. Bulletin of Mathematical Biology 14(1):35-44

Rapoport A (1953) Spread of information through a population with socio-structural bias: I. assumption of transitivity. Bulletin of Mathematical Biology 15(4):523-533

Rapoport A, Rebhun LI (1952) On the mathematical theory of rumor spread. Bulletin of Mathematical Biology 14(4):375-383

Rayport J (1996) The virus of marketing. Fast Company. http://www.fastcompany.com/ magazine/06/virus.html, accessed 2012-06-25

Richardson M, Domingos P (2002) Mining knowledge-sharing sites for viral marketing. In: Proc 8th ACM SIGKDD international conference on knowledge discovery and data mining, Edmonton, pp 61-70

Richter D, Riemer K, vom Brocke J (2011) Internet social networking: Research state of the art and implications for Enterprise 2.0. Business & Information Systems Engineering 3(2): 89-101

Rogers EM (1962) Diffusion of innovations. Free Press, New York

Rogers EM, Cartano DG (1962) Methods of measuring opinion leadership. Public Opinion Quarterly 26(3):435-441

Ryan B, Gross N (1943) The diffusion of hybrid seed corn in two Iowa communities. Rural Sociology 8(1):15-24

Saito K, Kimura M, Ohara K, Motoda H (2009) Learning continuous-time information diffusion model for social behavioral data analysis. In: Proc 1st Asian conference on machine learning: Advances in machine learning, Nanjing, pp 322-337

Saito K, Kimura M, Ohara K, Motoda H (2010) Behavioral analyses of information diffusion models by observed data of social network. Lecture Notes in Computer Science 6007: 149-158

Saito K, Kimura M, Ohara K, Motoda H (2012) Efficient discovery of influential nodes for SIS models in social networks. Knowledge and Information Systems 30(3):613-635

Scandura TA, Williams EA (2000) Research methodology in management: current practices, trends and implications for future research. Academy of Management Journal 43(6): 1248-1264

Schmitt P, Skiera B, Van den Bulte C (2011) Referral programs and customer value. Journal of Marketing 75(1):46-59

Scott J (2000) Social network analysis: A handbook. SAGE, London

Sein MK, Henfridsson O, Purao S, Rossi M, Lindgren R (2011) Action design research. MIS Quarterly 35(1):37-56

Smith AN, Fischer E, Yongjian C (2012) How does brand-related user-generated content differ across YouTube, Facebook, and Twitter? Journal of Interactive Marketing 26(2):102‑113

Strang D, Tuma NB (1993) Spatial and temporal heterogeneity in diffusion. American Journal of Sociology 99(3):614-639

Subramani MR, Rajagopalan B (2003) Knowledge-sharing and influence in online social networks via viral marketing. Communications of the ACM 46(12):300-307

Sundararajan A (2006) Network seeding. Workshop on Information Systems Economics, Evanston, pp 1-5

Sundararajan A, Provost F, Oestreicher-Singer G, Aral S (2012) Information in digital, economic and social networks. Information Systems Research (forthcoming)

Susarla A, Oh JH, Tan Y (2012) Social networks and the diffusion of user-generated content: Evidence from YouTube. Information Systems Research 23(1):23-41

Tang X, Yang CC (2010) Identifying influential users in an online healthcare social network. In: Proc IEEE international conference on intelligence and security informatics, Vancouver, pp 43-48

Trusov M, Bodapati A, Bucklin R (2010) Determining influential users in Internet social networks. Journal of Marketing Research 47(4):643-658

Trusov M, Bucklin RE, Pauwels K (2009) Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. Journal of Marketing 73(5):90-102

Valente TW (1995) Network models of the diffusion of innovations. Hampton Press, Cresskill

Valente TW, Rogers EM (1995) The origins and development of the diffusion of innovations paradigm as an example of scientific growth. Science Communication 16(3):242-273

Van den Bulte C, Joshi YV (2007) New product diffusion with influentials and imitators. Marketing Science 26(3):400-421

Van den Bulte C, Lilien GL (2001) Medical innovation revisited: Social contagion versus marketing effort. American Journal of Sociology 106(5):1409-1435

Van den Bulte C, Lilien GL (2003) Two-stage partial observability models of innovation adoption. Working Paper University of Pennsylvania, Philadelphia

Van den Bulte C, Stremersch S (2004) Social contagion and income heterogeneity in new product diffusion: A meta-analytic test. Marketing Science 23(4):530-544

Van den Bulte C, Wuyts (2007) Social networks and marketing. Marketing Science Institute Cambridge, Cambridge

vom Brocke J, Simons A, Niehaves B, Riemer K, Plattfaut R, Cleven A (2009) Reconstructing the giant: On the importance of rigour in documenting the literature search. In: Proc 17th European conference on information systems, Verona, paper 161

Wasserman S, Faust K (1994) Social network analysis: Methods and applications. Cambridge University Press, Cambridge

Watts DJ (2004) Small worlds: The dynamics of networks between order and randomness. Princeton University Press, Princeton

Watts DJ, Dodds P (2007) Influentials, networks, and public opinion formation. Journal of Consumer Research 34(4):441-458

Webster J, Watson RT (2002) Analyzing the past to prepare for the future: Writing a literature review. MIS Quarterly 26(2):xiii-xxiii

Weimann G (1991) The influentials: Back to the concept of opinion leaders? Public Opinion Quarterly 55(2):267-279

Weimann G, Tustin D, van Vuuren D, Joubert JR (2007) Looking for opinion leaders: Traditional vs. modern measures in traditional societies. International Journal of Public Opinion Research 19(2):173-190

Wellman B, Salaff J, Dimitrova D, Garton L, Gulia M, Haythornthwaite C (1996) Computer networks as social networks: Collaborative work, telework, and virtual community. Annual Review of Sociology, 22:213-238

Wu S, Hofman JM, Mason WA, Watts DJ (2011) Who says what to whom on Twitter. In: Proc 20th international conference on World Wide Web, Hyderabad, pp 705-714

Zhang Y, Wang Z, Xia C (2010) Identifying key users for targeted marketing by mining online social network. In: Proc 24th international conference on advanced information networking and applications, Perth, pp 644-649

Zhang Y, Zhou J, Cheng J (2011) Preference-based top-k influential nodes mining in social networks. In: Proc 10th international conference on trust, security and privacy in computing and communications, Changsha, pp 1512-1518