Customer Relationship Management in a Digitally Connected World

Dissertation

der Wirtschaftswissenschaftlichen Fakultät

der Universität Augsburg

zur Erlangung des Grades eines Doktors

der Wirtschaftswissenschaften

(Dr. rer. pol.)

vorgelegt

von

Florian Probst

(Master of Science with honors)

Augsburg, Januar 2013

Erstgutachter:	Prof. Dr. Hans Ulrich Buhl
Zweitgutachter:	Prof. Dr. Heribert Gierl
Vorsitzender der mündlichen Prüfung:	Prof. Dr. Marco C. Meier
Datum der mündlichen Prüfung:	18. Februar 2013

Table of Contents

Please note: Pages are consecutively numbered per chapter and within chapter III per section, each representing one research paper. References are provided at the end of each chapter or each research paper, respectively.

Index of Research Papers

This dissertation contains the following research papers that are either already published or conditionally accepted for publication:

Research Paper 1:

Heidemann J, Klier M, Probst F (2012) Online Social Networks: A Survey of a Global Phenomenon. Computer Networks 56(18):3866-3878
(Impact Factor 2011: 1.200)

Research Paper 2:

Probst F, Grosswiele L, Pfleger R (2013) Who will lead and who will follow: Identifying Influential Users in Online Social Networks - A Critical Review and Future Research Directions. Accepted with minor revisions for Special Issue 3/2013 "BISE and Marketing" of Business & Information Systems Engineering (BISE) by Spann M, Hinz O, and Ramachandran V
(VHB-JOURQUAL 2: 7.29 points, category B; Impact Factor 2011: 0.810)

Research Paper 3:

Heidemann J, Klier M, Probst F (2010) Identifying Key Users in Online Social Networks: A PageRank Based Approach. In: Proceedings of the 31st International Conference on Information Systems, St Louis, MO, USA, paper 79
(VHB-JOURQUAL 2: 8.48 points, category A)

Research Paper 4:

Probst F (2011) Predicting Users' Future Level of Communication Activity in Online Social Networks: A First Step towards More Advertising Effectiveness. In: Proceedings of the 17th Americas Conference on Information Systems, Detroit, MI, USA, paper 39
(VHB-JOURQUAL 2: 5.92 points, category D)

I Introduction

Already in 1982, Peters and Waterman emphasized in their book *In search of excellence* that being "close to the customer" is a core principle of excellent organizations (Peters and Waterman 1982). At that time, their claim was in line with numerous marketing scholars propagating a shift from a rather short-term and product-oriented transactional marketing to a more long-term-oriented relationship marketing (e.g., Arndt 1979; Bagozzi 1974; 1978; Day and Wensley 1983; Dwyer et al. 1987; Levitt 1983). In this research context, relationship marketing has often been defined in accordance to Berry (1983, p. 25) with a strict focus on customers as "[…] attracting, maintaining and […] enhancing customer relationships"[I-1]. Nowadays, customer relationships are thereby perceived as valuable (intangible) assets (Berger et al. 2002, p. 40; Srivastava et al. 1998, p. 2). In some companies, particularly in the service sector (Heidemann 2009, p. I-1), customer relationships even account for a considerable share of the firm value (Hogan et al. 2002, p. 4; Gupta et al. 2004, p. 7 f.; Kumar et al. 2004, p. 63).

Since the beginning of the 1990s, the widely acknowledged paradigm of value-based management (Coenenberg and Salfeld 2007, p. 3), which is based on the shareholder value principle (cf. e.g., Rappaport 1986), postulates "[…] the maximization of the long-term sustainable enterprise value as a guideline for all business activities" (Buhl et al. 2011, p. 164). As "[…] without customer value there can be no shareholder value" (Rappaport 1998, p. 76), customer relationships need to be constantly and actively managed (Berger et al. 2002, p. 39 ff.; Doyle 2002, p. 235; Hogan et al. 2002, p. 4). That is, they "[…] need to be invested and de-invested in just like any other (tangible) asset of the firm" (Wübben 2008, p. 19). To measure the value of customer relationships and to account for the principles of value-based management when making such investment decisions, the customer lifetime value (CLV) has become an intensively researched and widely accepted concept (Pepe 2012, p. 2). In its basic

[I-1] Today, some authors apply a broader definition of relationship marketing (e.g., Grönroos 1991, p. 8) including not only customers but also other parties such as suppliers or even competitors (Hippner et al. 2011, p. 19; Wübben 2008, p. 12).

form[I-2], the CLV is "[…] the present value of the net contribution associated with a customer's purchases/transactions over a 'lifetime'" (Weinberg and Berger 2011, p. 328; cf. e.g., also Berger and Nasr 1998; Blattberg et al. 2001; Dwyer 1997). As prior research found that not all customers contribute equally to a firm's value and some might even have a negative impact (Ang and Taylor 2005, p. 301; Reinartz and Kumar 2000, p. 19 f.), a differentiated customer relationship management (CRM) is required. Here, CRM is defined in accordance to Payne and Frow (2005, p. 168) as:

> "[…] a strategic approach that is concerned with creating improved shareholder value through the development of appropriate relationships with key customers and customer segments. CRM unites the potential of relationship marketing strategies and IT to create profitable, long-term relationships with customers and other key stakeholders. CRM provides enhanced opportunities to use data and information to both understand customers and co-create value with them. This requires a cross-functional integration of processes, people, operations, and marketing capabilities that is enabled through information, technology, and applications".

Traditionally, researchers and practitioners engaged in the field of CRM perceived customer relationships solely as bidirectional connections between firms and customers, whereby firms were able to control the conversations to a large extent (Band et al. 2010, p. 5). However, as already anticipated in Tapscott's vision of *The Digital Economy: Promise and Peril in the Age of Networked Intelligence* (Tapscott 1996), the increasing digital connectedness changed the behavior of customers dramatically (Hennig-Thurau et al. 2010, p. 311). Especially with the rise of Online Social Networks (OSN), such as Facebook, customers nowadays vividly exchange their experiences and opinions with previously unknown intensity, reach, and speed among each other (Rosemann et al. 2012, p. 2 f.). Thus, it is even claimed that "[…] companies are no longer in control of the [customer] relationship. Instead, customers and their highly influential virtual networks are now driving the conversation […]" (Baird and Parasnis 2011, p. 30). In the following, OSN – which impressively represent such virtual networks in the digitally connected world we live in – are defined according to Boyd and Ellison (2007, p. 211) as[I-3]:

[I-2] For reviews of further CLV models cf. e.g., Gupta et al. (2006) and Villanueva and Hanssens (2007).

[I-3] Boyd and Ellison (2007) use the term Social Networking Site. In this dissertation OSN is used synonymously.

> "[…] 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".

By offering numerous technical functionalities that enable users to communicate easily and in real-time with acquaintances, friends, or even strangers (cf. e.g., Boyd and Ellison 2007; Huber et al. 2012; Kim et al. 2010), OSN "[…] constitute powerful communication platforms that allow for presenting oneself and exchanging information in an efficient and timely manner" (Heidemann et al. 2012, p. 3868). For example, solely on Facebook – the most popular OSN in the world – one billion active users are connected in a network of 140 billion friendship connections as of October 2012 (Facebook 2012). These users share 684,478 pieces of content and "like" 34,722 brands or organizations – every single minute (Tepper 2012). Consequently, also customers became to a large extent digitally connected (Weinberg and Berger 2011, p. 328), creating an enormous amount of user-generated content and closely related electronic word-of-mouth (Kaplan and Haenlein 2010, p. 63; Smith et al. 2012, p. 102). The latter can be defined as "[…] any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet" (Hennig-Thurau et al. 2004, p. 39). Customers thereby rate products, report their experiences with products, services, and customer support, or even create advertising on their own (Berthon et al. 2008, p. 6; Fader and Winer 2012, p. 369; Hanna et al. 2011, p. 265). Moreover, prior research indicates that customers consider recommendations and opinions of like-minded others to be more credible than information provided by companies (Chen and Xie 2008, p. 478; Moon et al. 2010, p. 108). Therefore, "[…] the early part of the 21st century has become the era of social commerce" (Fader and Winer 2012, p. 369).

To account for these changes, users of OSN should be perceived as important communication partners who provide valuable feedback on brands, products, and services (Li and Bernoff 2011, p. 194; Segrave et al. 2011, p. 4). Such feedback can be already integrated during the *research and development* phase of new products or services (Richter et al. 2011, p. 98). Furthermore, companies can leverage OSN within the business areas *marketing and sales*. For instance, users can be motivated to advertise products and services on their own by drawing on targeted and viral marketing strategies that leverage network effects (Hill et al. 2006, p. 257; Libai et al. 2010, p. 270; Trusov et al. 2009, p. 90). By creating higher levels of awareness as well as motivating and incentivizing users to increase the number of positive user-generated ratings and comments, companies can thus influence customer behavior and significantly grow their sales (Chevalier and Mayzlin 2006, p. 349 ff.; Lin and Goh 2011, p. 9 ff.; Moon et al. 2010, p. 118; Zhu and Zhang 2010, p. 135 ff.). Companies also try themselves to participate actively in conversations about their brands, products, and services (Hennig-Thurau et al. 2010, p. 313 f.). For instance, many companies set up fan pages in OSN, which can be described as corporate profile (Kim et al. 2010, p. 228). Jahn and Kunz (2012) show that users' intensive fan page usage and engagement lead to higher levels of brand loyalty, which can in turn increase brand commitment, brand-related word-of-mouth, and brand purchase. Besides using OSN within the business areas marketing and sales, companies can also use OSN to provide *customer services* such as customer support (Bonchi et al. 2011, p. 4; Weinberg and Berger 2011, p. 342). Taken together, companies can use OSN to be closer to the customer along the whole value chain (cf. Figure I-1).

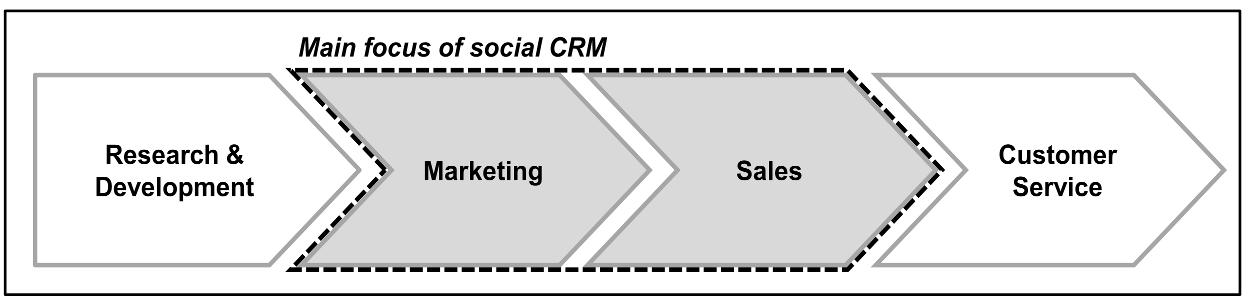

Figure I-1. Potential OSN usage in business areas along the value chain[I-4]
Source: Own illustration based on Bonchi et al. (2011, p. 4) and Mattern et al. (2012, p. 9)

To account for these new possibilities and related challenges, numerous researchers and practitioners propagate the necessity of a further developed "social" CRM (e.g., Alt and Reinhold 2012; Ang 2010; Baird and Parasnis 2011; Band et al. 2010; Faase et al. 2011; Rapp and Panagopoulos 2012; Rosemann et al. 2012; Reinhold and Alt 2012;

[I-4] Since the focus of this work is on customers, management and support processes such as human resources or internal applications for knowledge management and collaboration, which can also be supported by OSN (Bonchi et al. 2011, p. 4; Mattern et al. 2012, p. 8 f.), are not further considered.

Woodcock et al. 2011), which incorporates data from OSN when managing customer relationships (Myron 2010, p. 4). While OSN can be leveraged along the entire value chain (Mattern et al. 2012, p. 8 f.), the main focus of social CRM is on the business areas marketing and sales (Bonchi et al. 2011, p. 4) (cf. Figure I-1). Particularly in these areas, the connected customer can exercise his or her power to a large extent by promoting brands, endorsing the purchase of products and services, or by advising against them. Hence, through their influence on other existing or potential customers and the related customer values, the firm value can be increased or decreased (Hogan et al. 2003, p. 196; Nitzan and Libai 2011, p. 24 f.; Weinberg and Berger 2011, p. 328). Tirunillai and Tellis (2012) further show, that the mere volume of positive (negative) user-generated content can lead to positive (negative) abnormal stock market returns. Taken together, companies are therefore highly interested in (1) understanding the specific characteristics of OSN and their users, (2) learning more about the influence of customer-to-customer interactions in OSN, and (3) being actually able to identify the most influential users in OSN, who can be targeted to initiate and control the diffusion process of user-generated content, such as electronic word-of-mouth (Bonchi et al. 2011, p. 21; Hinz et al. 2011, p. 55; Libai et al. 2010, p. 271). However, even though OSN provide a huge amount of user data that can be leveraged to do so (Bonchi et al. 2011, p. 2; Katona et al. 2011, p. 425 f.; Subramani and Rajagopalan 2003, p. 301), companies "[…] are seeking guidance on effective principles and mechanisms for analysing and leveraging these data in an impactful way" (Weinberg and Berger 2011, p. 342). Hereby, this dissertation shall contribute to research and practice. The following section I.1 pinpoints its objectives and structure. In the subsequent section I.2, the corresponding research papers are embedded in the research context and the fundamental research questions are highlighted.

I.1 Objectives and Structure of the Dissertation

The main objective of this dissertation is to contribute to the field of CRM with a particular focus on OSN and the identification of influential users within these networks. Figure I-2 provides an overview of the dissertation's pursued objectives and its structure.

I Introduction	
Objective I.1:	Pinpointing the objectives and the structure of the dissertation
Objective I.2:	Embedding the corresponding papers into the research context of the dissertation and motivating the fundamental research questions
II Business Value of Online Social Networks along the Value Chain (Research Paper 1)	
Objective II.1:	Defining the concept of Online Social Networks and reviewing their development over time
Objective II.2:	Demonstrating the impact and value as well as major risks and challenges of Online Social Networks from a business perspective
III Influential Users in Online Social Networks (Research Papers 2, 3, and 4)	
Objective III.1:	Outlining fundamental research on social influence, influential people, and their identification in social networks before the rise of Online Social Networks
Objective III.2:	Analyzing and synthesizing the growing number of publications on the identification of influential users in Online Social Networks and deriving a research agenda by identifying research gaps
Objective III.3:	Developing a novel approach for the identification of influential users in Online Social Network bringing together main findings from prior research
Objective III.4:	Evaluating the novel PageRank based approach against existing approaches using objective data
Objective III.5:	Proposing an approach for predicting users' communication activity in Online Social Networks to improve the effectiveness of advertising strategies by addressing the most active users deliberately
IV Summary and Future Research	
Objective IV.1:	Summarizing the key findings
Objective IV.2:	Highlighting starting points for future research

Figure I-2. Objectives and structure of the dissertation

I.2 Research Context and Research Questions

In the following section, the corresponding research papers included in this dissertation are embedded in the research context with respect to the above stated objectives and the respective research questions are motivated.

I.2.1 Business Value of Online Social Networks along the Value Chain

Research Paper 1*: "Online Social Networks: A Survey of a Global Phenomenon"*

During the last decade, research in the fields of Business & Information Systems Engineering (BISE) and Marketing has been substantially influenced by the rise of OSN (Bonchi et al. 2011, p. 2; Fader and Winer 2012, p. 369; Rosemann et al. 2012, p. 1). Today, the most popular OSN Facebook connects one billion active users (Facebook 2012) and there are also numerous "smaller" OSN with millions of users (Kim et al. 2010, p. 118). These users vividly exchange information and express their opinions and emotions (Lin and Goh 2011, p. 2), whereby an enormous volume of user-generated content is created (Kaplan and Haenlein 2010, p. 63; Smith et al. 2012, p. 102). Taken together with information about the underlying network structure, this so far unknown rich amount of customer data can be leveraged to improve companies' CRM, especially with respect to the business areas marketing and sales (Bonchi et al. 2011, p. 2 ff.; Katona et al. 2011, p. 425 f.; Subramani and Rajagopalan 2003, p. 301). However, before integrating data from OSN to achieve a more "social" CRM (Myron 2010, p. 4; Rosemann et al. 2012, p. 5), it is necessary to understand the specific characteristics of OSN and their users. Therefore, the goal of the first research paper is to contribute to a better understanding of OSN by addressing the following research questions (Heidemann et al. 2012, p. 3866):

- What are OSN and why are they used?
- What are the structural characteristics that form the backbone of OSN?
- How did OSN emerge and develop over time?
- How can the large number of OSN be classified?
- What are the impact and value of OSN from a business perspective?
- What are major risks and challenges in the context of OSN?

I.2.2 Influential Users in Online Social Networks

***Research Paper 2**: "Who will lead and who will follow: Identifying Influential Users in Online Social Networks - A Critical Review and Future Research Directions"*

Along with the above mentioned rise of the phenomenon OSN, identifying influential users in OSN is receiving a great deal of attention in both, BISE and Marketing literature (Bonchi et al. 2011, p. 21; Hinz et al. 2013; Katona et al. 2011, p. 426). Besides the previously unknown technical possibilities and the large amount of available data, this is especially due to the decreasing impact of traditional marketing techniques (Clemons 2009, p. 48 f.; Hinz et al. 2011, p. 55; Trusov et al. 2009, p. 90) and the already noted trust of customers in recommendations of other users (Chen and Xie 2008, p. 478; Moon et al. 2010, p. 108). Therefore, more and more companies try to leverage the effect of social influence on product adoption (cf. e.g., Godes and Mayzlin 2009; Goldenberg et al. 2009; Hinz et al. 2013; Iyengar et al. 2011), by targeting the most influential people in social networks (Bonchi et al. 2011, p. 21; Hinz et al. 2011, p. 55; Libai et al. 2010, p. 271). However, as for instance noted by Richter et al. (2011, p. 98), "[…] the development of practical approaches for the identification of influential users in OSN is still in its infancy […] and researchers face numerous challenges" (Probst et al. 2013). Therefore, the second research paper aims at analyzing and synthesizing the current state of the art on the identification of influential users in OSN, in order to stimulate and guide further research at this interface of BISE and Marketing. Derived from a brief overview of fundamental research on social influence, influential people, and their identification in social networks before the rise of OSN, the following research questions are investigated and – based on the results and the identified research gaps – a research agenda is postulated (Probst et al. 2013):

- How are influential users characterized in the context of OSN?
- Which approaches have been developed and applied for the identification of influential users in OSN?
- How have these approaches been evaluated and which implications have been derived?

***Research Paper 3**: "Identifying Key Users in Online Social Networks: A PageRank Based Approach"*

Despite the enormous popularity of OSN, the question of how their providers can yield sustainable revenues is still not fully answered (Clemons 2009; Lu and Hsiao 2010, p. 150). Therefore, not only companies are interested in identifying influential users as mentioned above, but also OSN providers consider so-called key users as a potential source of revenue (Xu et al. 2009, p. 17). Such key users can be exceptionally influential users, who can be addressed for marketing purposes. OSN providers, who are able to identify such users, can for instance develop significantly improved advertising models (cf. e.g., Bao and Chang 2010). However, key users can also be very loyal users, who are unlikely to leave the OSN. Since only remaining users can be leveraged for marketing purposes or to gain membership revenues (Heidemann et al. 2010, p. 12 f.), such users are "[…] crucial to growth and survival of large online social networks" (Nazir et al. 2009, p. 65). Finally, key users can also be users with a relatively high willingness to pay for premium services in OSN (Heidemann et al. 2010, p. 2). Prior literature indicates that especially users' connectivity and communication activity should be considered when identifying these different types of key users (e.g., Algesheimer and von Wangenheim 2006; Bampo et al. 2008; Xu et al. 2009). Existing approaches, however, often identify either well-connected users (e.g., Hinz et al. 2011) or particularly active users (e.g., Goldenberg et al. 2009). Therefore, the third research paper aims at developing a novel approach bringing concepts and findings from research on users' connectivity (by applying an adapted PageRank based centrality measure) and users' communication activity (by deriving a weighted activity graph considering the communication intensity among users) together. With users' retention as the evaluation criterion, the approach is evaluated regarding its applicability and practical utility using a publicly available dataset of Facebook. Thus, two research questions are addressed:

- How can key users in OSN be identified by a novel approach bringing concepts and findings from research on users' connectivity and activity together?
- Is the proposed novel approach better suited for identifying key users in OSN than existing approaches based on either users' connectivity or communication activity?

***Research Paper 4**: "Predicting Users' Future Level of Communication Activity in Online Social Networks: A First Step towards More Advertising Effectiveness"*

As already indicated within this introduction, both users' connectivity and communication activity are essential when trying to increase advertising effectiveness in terms of product endorsement in OSN (cf. e.g., Wen et al. 2009). In OSN, users connect by sending and confirming friendship requests leading to static and undirected social connections (Heidemann et al. 2010, p. 3 f.; Heidemann et al. 2012, p. 3867). However, only a small share of these connections can be considered as actually active, as on average 20% of a user's connected "friends" account for 70% of his or her communication activity (Wilson et al. 2009, p. 210). Hence, Wilson et al. (2009, p. 210) conclude that "[…] not all social links are equally useful". This is in line with Xu et al. (2008, p. 14) who also highlight that "[…] interaction information is invaluable to marketers, more important than the static links". Consequently, knowledge about users' future communication activity in OSN is highly valuable (Trusov et al. 2010, p. 643 f.; Willinger et al. 2009, p. 49). To increase advertising effectiveness, companies need to address especially users who stay active (Cheung and Lee 2010, p. 28; Ganley and Lampe 2009, p. 273), as only active users become potentially aware of advertising and virally spread the advertising message. Prior research suggests relying on users' records of past communication activity to identify these active users (Xiang et al. 2010, p. 981). This rather straightforward approach, however, is contrary to previous findings, which show that high levels of communication activity cannot be taken for granted and past communication behavior must not necessarily indicate future levels of users' communication activity (Cummings et al. 2002, p. 107; Viswanath et al. 2009, p. 39 f.). Therefore, the fourth research paper proposes to transfer a probability-based method, which has been primarily developed by Fader et al. (2005) to forecast purchasing behavior of customers, to the context of users' communication activity in OSN. Its suitability for predicting users' future level of communication activity in OSN is then evaluated using publicly available data from Facebook (Probst 2011, p. 2). Hence, the following research question is investigated:

- Can users' individual future level of communication activity in OSN be predicted by a method that has originally been developed to forecast purchasing behavior?

The research context is summarized in Figure I-3, which highlights the business areas along the value chain that are mainly addressed by the respective chapters and research papers.

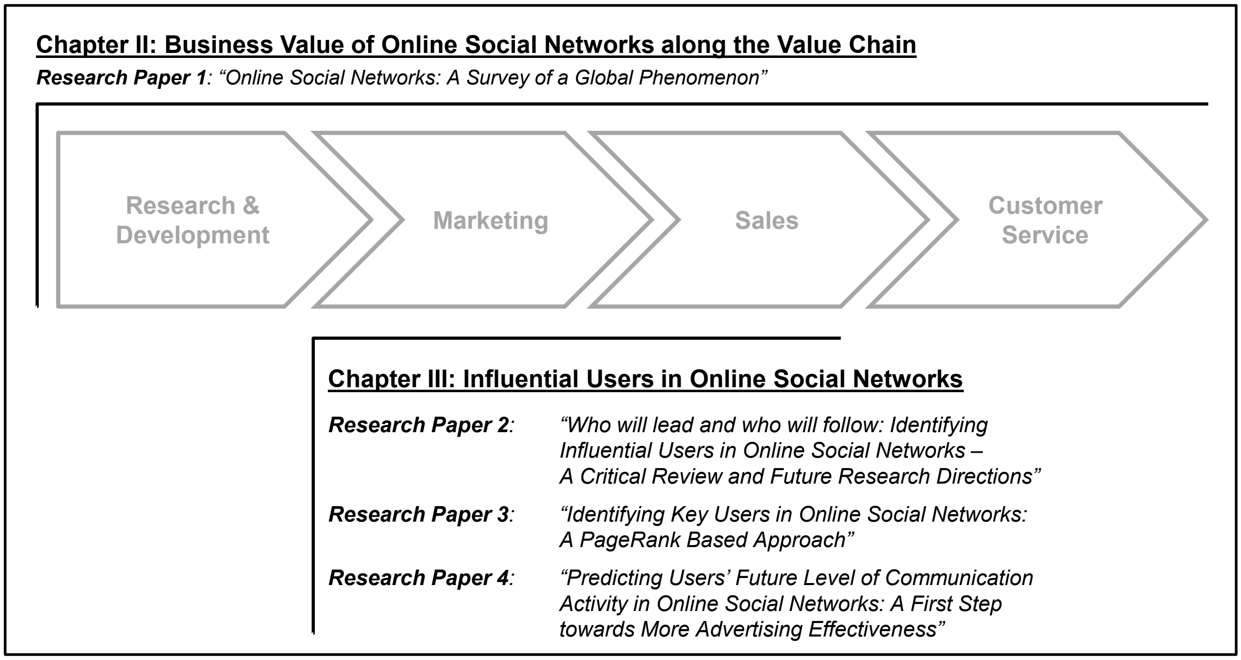

Figure I-3. Chapters and research papers in the research context

After this introduction pinpointing the objectives and the structure of the dissertation as well as motivating the research context and the addressed research questions, the respective research papers are presented in chapters II and III. Subsequently, the key findings are summarized and starting points for future research are highlighted in chapter IV.

References (Chapter I)

Algesheimer R, von Wangenheim F (2006) A network based approach to customer equity management. Journal of Relationship Marketing 5(1):39-57

Alt R, Reinhold O (2012) Social customer relationship management (social CRM) - Application and technology. Business & Information Systems Engineering 4(5):287-291

Ang L (2010) Community relationship management and social media. Journal of Database Marketing & Customer Strategy Management 18(1):31-38

Ang L, Taylor B (2005) Managing customer profitability using portfolio matrices. Database Marketing & Customer Strategy Management 12(4):298-304

Arndt J (1979) Toward a concept of domesticated markets. Journal of Marketing 43(4):69-75

Bagozzi RP (1974) Marketing as an organized behavioral system of exchange. Journal of Marketing 38(4):77-81

Bagozzi RP (1978) Marketing as exchange: A theory of transactions in the market-place. American Behavioral Scientist 21(4):535-555

Baird CH, Parasnis G (2011) From social media to social customer relationship management. Strategy & Leadership 39(5):30-37

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

Band W, Petouhoff NL, Moore C, Magarie A (2010) Topic overview: Social CRM goes mainstream. Forrester Research Inc., Cambridge, MA, USA

Bao H, Chang EY (2010) AdHeat: An influence-based diffusion model for propagating hints to match ads. In: Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, pp. 71-80

Berger PD, Bolton RN, Bowman D, Briggs E, Kumar V, Parasuraman A, Terry C (2002) Marketing actions and the value of customer assets: a framework for customer asset management. Journal of Service Research 5(1):39-54

Berger PD, Nasr NI (1998) Customer lifetime value: Marketing models and applications. Journal of Interactive Marketing 12(1):17-30

Berry LL (1983) Relationship marketing. In: Berry LL, Shostack GL, Upah GD (eds) Emerging perspectives on services marketing. American Marketing Association, Chicago, IL, USA, pp. 25-28

Berthon P, Pitt L, Campbell C (2008) Ad Lib: When customers create the ad. California Management Review 50(4):6-30

Blattberg RC, Getz G, Thomas JS (2001) Customer equity: Building and managing relationships as valuable assets. HBS Press, Boston, MA, USA

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

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

Buhl HU, Röglinger M, Stöckl S, Braunwarth KS (2011) Value orientation in process management. Business & Information Systems Engineering 3(3):163-172

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

Cheung CMK, Lee MKO (2010) A theoretical model of intentional social action in online social networks. Decision Support Systems 49(1):24-30

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

Coenenberg AG, Salfeld R (2007) Wertorientierte Unternehmensführung. Vom Strategieentwurf zur Implementierung. Schäffer-Poeschel, Stuttgart, Germany

Cummings JN, Butler B, Kraut R (2002) The quality of online social relationships. Communications of the ACM 45(7):103-108

Day GS, Wensley R (1983) Marketing theory with a strategic orientation. Journal of Marketing 47(4):79-89

Doyle P (2000) Valuing marketing’s contribution. European Management Journal 18(3):233-245

Dwyer FR (1997) Customer lifetime valuation to support marketing decision making. Journal of Direct Marketing 11(4):6-13

Dwyer FR, Schurr PH, Oh S (1987) Developing buyer-seller relationships. Journal of Marketing 51(2):11-27

Faase R, Helms R, Spruit M (2011) Web 2.0 in the CRM domain: Defining social CRM. International Journal of Electronic Customer Relationship Management 5(1):1-22

Facebook (2012) One billion fact sheet. Facebook. http://newsroom.fb.com/image library/downloadmedia.ashx?MediaDetailsID=4227&SizeId=-1, accessed 2012-12-28

Fader PS, Hardie BGS, Lee KL (2005) “Counting your customers” the easy way: An alternative to the Pareto/NBD model. Marketing Science 24(2):275-284

Fader PS, Winer RS (2012) Introduction to the special issue on the emergence and impact of user-generated content. Marketing Science 31(3):369-371

Ganley D, Lampe C (2009) The ties that bind: Social network principles in online communities. Decision Support Systems 47(3):266-274

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

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

Grönroos C (1991) The marketing strategy continuum: toward a marketing concept for the services marketing. Management Decision 29(1):7-13

Gupta S, Hanssens DM, Hardie BGS, Kahn W, Kumar V, Lin N, Ravishanker N, Sriram S (2006) Modeling customer lifetime value. Journal of Service Research 9(2): 139-155

Gupta S, Lehmann D, Stuart J (2004) Valuing customers. Journal of Marketing Research 41(1):7-18

Hanna R, Rohm A, Crittenden VL (2011) We’re all connected: The power of the social media ecosystem. Business Horizons 54(3):265-273

Heidemann J (2009) Wertorientierte Gestaltung von Kundenbeziehungen. Dissertation University of Augsburg. OPUS, http://opus.bibliothek.uni-augsburg.de/opus4/files/1437/Dissertation_Heidemann.pdf

Heidemann J, Klier M, Probst F (2010) Identifying key users in online social networks: A PageRank based approach. In: Proceedings of the 31st International Conference on Information Systems, St Louis, MO, USA, paper 79

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

Hennig-Thurau T, Edward MC, Friege C, Gensler S, Lobschat L, Rangaswamy A, Skiera B (2010) The impact of new media on customer relationships. Journal of Service Research 13(3):311-330

Hennig-Thurau T, Gwinner KP, Walsh G, Gremler DG (2004) Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? Journal of Interactive Marketing 18(1):38-52

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

Hippner H, Hubrich B, Wilde KD (2011) Grundlagen des CRM: Strategie, Geschäftsprozesse und IT-Unterstützung. Gabler Verlag, Wiesbaden, Germany

Hogan JE, Lemon KN, Libai B (2003) What is the true value of a lost customer? Journal of Service Research 5(3):196-208

Hogan JE, Lemon KN, Rust RT (2002) Customer equity management: Charting new directions for the future of marketing. Journal of Service Research 5(1):4-12

Huber J, Landherr A, Probst F, Reisser C (2012) Stimulating user activity on company fan pages in online social networks. In: Proceedings of the 20th European Conference on Information Systems, Barcelona, Spain, paper 188

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

Jahn B, Kunz W (2012) How to transform consumers into fans of your brand. Journal of Service Management 23(3):344-361

Kaplan AM, Haenlein M (2010) Users of the world, unite! The challenges and opportunities of social media. Business Horizons 53(1):59-68

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

Kim W, Jeong O-R, Lee S-W (2010) On social web sites. Information Systems 35(2):215-236

Kumar V, Ramani G, Bohling T (2004) Customer lifetime value approaches and best practice applications. Journal of Interactive Marketing 18(3):60-72

Levitt T (1983) After the sale is over. Harvard Business Review 61(5):87-93

Li C, Bernoff J (2011) Groundswell: Winning in a world transformed by social technologies. Harvard Business Press, Boston, MA, USA

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

Lin Z, Goh KY (2011) Measuring the business value of online social media content for marketers. In: Proceedings of the 32nd International Conference on Information Systems, Shanghai, China, paper 16

Lu H-P, Hsiao K-L (2010) The influence of extro/introversion on the intention to pay for social networking sites. Information & Management 47(3):150-157

Mattern F, Huhn W, Perrey J, Dörner K, Lorenz J-T, Spillecke D (2012) Turning buzz into gold - How pioneers create value from social media. McKinsey & Company

Moon S, Bergey PK, Iacobucci D (2010) Dynamic effects among movie ratings, movie revenues, and viewer satisfaction. Journal of Marketing 74(1):108-121

Myron D (2010) Social CRM is growing up fast. Customer Relationship Management 14(6):4

Nazir A, Raza S, Gupta D, Chuah C-N, Krishnamurthy B (2009) Network level footprints of Facebook applications. In: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, Chicago, IL, USA, pp. 63-75

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

Payne A, Frow P (2005) A strategic framework for customer relationship management. Journal of Marketing 69(4):167-176

Pepe M (2012) Customer lifetime value: A vital marketing/financial concept for businesses. Journal of Business & Economics Research 10(1):1-10

Peters T, Waterman RH (1982) In search of excellence: Lessons from America's best-run companies. Harper & Row, New York, NY, USA

Probst F (2011) Predicting users' future level of communication activity in online social networks: A first step towards more advertising effectiveness. In: Proceedings of the 17th Americas Conference on Information Systems, Detroit, MI, USA, paper 39

Probst F, Grosswiele L, Pfleger R (2013) Who will lead and who will follow: Identifying influential users in online social networks - A critical review and future research directions. Accepted with minor revisions for Special Issue 3/2013 "BISE and Marketing" of Business & Information Systems Engineering (BISE) by Spann M, Hinz O, and Ramachandran V

Rapp A, Panagopoulos NG (2012) Perspectives on personal selling and social media: introduction to the special issue. Journal of Personal Selling & Sales Management 32(3):301-304

Rappaport A (1986) Creating shareholder value: The new standard for business performance. Free Press, New York, NY, USA

Rappaport A (1998) Creating shareholder value: A guide for managers and investors. Free Press, New York, NY, USA

Reinartz W, Kumar V (2000) On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. Journal of Marketing 64(4):17-35

Reinhold O, Alt R (2012) Social customer relationship management: State of the art and learnings from current projects. In: Proceedings of the 25th Bled eConference, Bled, Slovenia, pp. 155-169

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

Rosemann M, Eggert M, Voigt M, Beverungen D (2012) Leveraging social network data for analytical CRM strategies – the introduction of social BI. In: Proceedings of the 20th European Conference on Information Systems, Barcelona, Spain, paper 95

Segrave J, Carson C, Merhout JW (2011) Online social networks: An online brand community framework. In: Proceedings of the 17th Americas Conference on Information Systems, Detroit, MI, USA, paper 249

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

Srivastava RK, Shervani TA, Fahey L (1998) Market-based assets and shareholder value: A framework for analysis. Journal of Marketing 62(1):2-18

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

Tapscott D (1996) The digital economy: Promise and peril in the age of networked intelligence. McGraw-Hill, New York, NY, USA

Tepper A (2012) How much data is created every minute? mashable.com/2012/06/22/ data-created-every-minute/, accessed 2012-12-28

Tirunillai S, Tellis GJ (2012) Does chatter really matter? Dynamics of user-generated content and stock performance. Marketing Science 31(2):198-215

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

Villanueva J, Hanssens DM (2007) Customer equity: Measurement, management and research opportunities. Foundations & Trends in Marketing 1(1):1-95

Viswanath B, Mislove A, Cha M, Gummadi KP (2009) On the evolution of user interaction in Facebook. In: Proceedings of the 2nd ACM Workshop on Online Social Networks, Barcelona, Spain, pp. 37-42

Weinberg BD, Berger PD (2011) Connected customer lifetime value: The impact of social media. Journal of Direct, Data and Digital Marketing Practice 12(4):328-344

Wen C, Tan BCY, Chang KT-T (2009) Advertising effectiveness on social network sites: An investigation of tie strength, endorser expertise and product type on consumer purchase intention. In: Proceedings of the 30th International Conference on Information Systems, Phoenix, AZ, USA, paper 151

Willinger W, Rejaie R, Torkjazi M, Valafar M, Maggioni M (2010) Research on online social networks: Time to face the real challenges. ACM SIGMETRICS Performance Evaluation Review 37(3):49-54

Wilson C, Boe B, Sala A, Puttaswamy KPN, Zhao BY (2009) User Interactions in social networks and their Implications. In: Proceedings of the 4th ACM European Conference on Computer Systems, Nuremberg, Germany, pp. 205-218

Woodcock N, Green A, Starkey M (2011) Social CRM as a business strategy. Journal of Database Marketing & Customer Strategy Management 18(1):50-64

Wübben M (2008) Analytical CRM: developing and maintaining profitable customer relationships in non-contractual settings. Dissertation Technical University of Munich. Gabler Verlag, Wiesbaden, Germany

Xiang R, Neville J, Rogati M (2010) Modeling relationship strength in online social networks. In: Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, pp. 981-990

Xu Y, Lu X, Goh KY, Jiang Z, Zhu X (2009) The impact of online social network on consumer loyalty: An empirical study of an online dining community. In: Proceedings of the 30th International Conference on Information Systems, Phoenix, AZ, USA, paper 17

Xu Y, Zhang C, Xue L, Yeo LL (2008) Product adoption in online social network. In: Proceedings of the 29th International Conference on Information Systems, Paris, France, paper 200

Zhu F, Zhang X (2010) Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. Journal of Marketing 74(2):133-148

II Business Value of Online Social Networks along the Value Chain (Research Paper 1: "Online Social Networks: A Survey of a Global Phenomenon")

Authors:	Dr. Julia Heidemann[a], Dr. Mathias Klier[a], Florian Probst[b] [a] Department of Management Information Systems, University of Regensburg, Germany julia.heidemann@wiwi.uni-regensburg.de mathias.klier@wiwi.uni-regensburg.de [b] FIM Research Center, Department of Information Systems Engineering & Financial Management (Prof. Dr. Hans Ulrich Buhl), University of Augsburg, Germany florian.probst@wiwi.uni-augsburg.de

| Published in: | Computer Networks 56(18):3866-3878 |

Abstract:

Online social networks became a global phenomenon with enormous social as well as economic impact within a few years. Alone, the most popular online social network, Facebook, counts currently more than 850 million users worldwide. Consequently, online social networks attract a great deal of attention among practitioners as well as researchers. The goal of this article is to provide an overview of online social networks in order to contribute to a better understanding of this worldwide phenomenon. In this context, we address for example the following questions: What are the major functionalities and characteristics of online social networks? What are the users' motives for using them and how did online social networks emerge and develop over time? What is the impact and value of online social networks from a business perspective and what are corresponding challenges and risks?

II.1 Introduction

Since the launch of the first recognizable network SixDegrees in 1997 (Boyd and Ellison 2007), multiple Online Social Networks (OSN) such as Facebook, LinkedIn, or Google+ have become popular Internet platforms, where people around the world congregate and get connected. The use of OSN has reached an enormous scale: The fraction of Internet users visiting OSN at least once a month is expected to grow from 41% in 2008 to over 65% in 2014 (Williamson 2010). The OSN Facebook, for instance, already outperformed Google as the most frequently visited website of the week in the US in March 2010 (Dougherty 2010) and counted 845 million active users in December 2011 (Facebook 2011). Although originally designed for private use (Bughin and Manyika 2007), more and more companies aim at presenting their brands and products within OSN to leverage their popularity (Wen et al. 2009). Worldwide advertisement spending on OSN is therefore expected to grow from US$ 5.2 billion in 2011 to US$ 11.9 billion in 2014 (eMarketer 2012). Taken together with the immense value of information that OSN hold (Beer 2008), numerous OSN have been consequently valued at billions of dollars. Hence, this technical and social phenomenon has evolved into a global mainstream medium with increasing social and economic impact.

In this article we give an overview of the phenomenon OSN. However, we do not present a full survey, but aim at providing the reader with the most relevant information to follow up on any of the subareas covered. Thereby, we address the following questions: (1) What are OSN and why are they used? (2) What are the structural characteristics that form the backbone of OSN? (3) How did OSN emerge and develop over time? (4) How can the large number of OSN be classified? (5) What is the impact and value of OSN from a business perspective? (6) What are major risks and challenges in the context of OSN? The remainder of the paper is organized as follows: In the subsequent section, we focus on the definition of OSN and highlight the main functionalities and major motives for using them. Section II.3 describes the structural characteristics of OSN while section II.4 briefly summarizes their genesis and development over time. Section II.5 is dedicated to the classification of the large number of existing OSN. After that general characterization of the phenomenon OSN, we provide a discussion of the impact and value of OSN from a business perspective in section II.6 and point out major risks and challenges in section II.7. Finally, we conclude with a brief summary.

II.2 Definition, Functionalities, and Usage of Online Social Networks

Especially in social sciences, the collective desire to take part in a community has been a well-explored phenomenon for a long time (Bagozzi and Dholakia 2006). Already about 400 years before Christ, Aristotle described human beings as *zoon politicon* – a character with the fundamental need of searching and creating communities (Buhl 2008). Therefore, the general idea of social networks is not really new. With the emergence of the World Wide Web (WWW) and the development of information technologies, however, social networks reached a new dimension. Thanks to numerous types of social software (cf. e.g., Boyd 2006), including blogs, user-generated content sites, and countless virtual communities across the WWW, people started connecting and communicating online with one another (Bernoff and Li 2008). Along with these changes, formerly passive information users were becoming actors, creating the content of the WWW themselves (Gneiser et al. 2012). Aroused by this development, also known as the emergence of the Web 2.0 (O’Reilly 2005), particularly OSN have evolved into a new, mostly free of cost mass medium where users present themselves to a wide public.

II.2.1 Definition of Online Social Networks

OSN are a particular type of virtual community (Dwyer et al. 2007) and of social software (Richter et al. 2011). However, as it is common for rather new phenomena related to the Web 2.0, there is neither one generally accepted term nor one well-established definition for OSN. There rather exist numerous similar terms such as social networking service, social networking site, or social network site. Table II-1 provides some selected terms and corresponding definitions.

Table II-1. Selected terms and definitions with respect to online social networks

Term	Authors	Definition
Online Social Network	Schneider et al. (2009, p. 35)	“OSNs form online communities among people with common interests, activities, backgrounds, and/or friendships. Most OSNs are Web-based and allow users to upload profiles (text, images, and videos) and interact with others in numerous ways.”
Social Networking Service	Adamic and Adar (2005, p. 188)	“Social networking services gather information on users’ social contacts, construct a large interconnected social network, and reveal to users how they are connected to others in the network.”
Social Network Site	Boyd and Ellison (2007, p. 211)	“We define social network sites 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.”

These different terms for OSN are often used synonymously, even though they do not share a common definition of the object under consideration. Boyd and Ellison (2007), for instance, point out that they deliberately did not choose the term social networking site since “[…] ’[n]etworking’ emphasizes relationship initiation, often between strangers. While networking is possible on these sites, it is not the primary practice on many of them […]” (Boyd and Ellison 2007, p. 211). Examples for such content-oriented sites are YouTube, Twitter, or Flickr. Beer (2008) criticizes the definition of social network site provided by Boyd and Ellison (2007) as being too broad. Therefore, we define OSN according to Boyd and Ellison (2007) but focus on user-oriented sites.

II.2.2 Functionalities of Online Social Networks

While the culture that emerges around different OSN varies, the maintenance of individual contacts and most of the key technological features are fairly consistent (Boyd and Ellison 2007). The core of an OSN consists of personalized user profiles, which usually contain identifying information (e.g., name and photo), interests (e.g., subscribed interest groups), and personal contacts (e.g., list of connected users, so-called “friends”). Users acquire new friends by searching offline as well as online friends or acquaintances within the OSN and by sending requests to be added as a friend. The completeness of provided user data has been intensively researched under the label of (self-)disclosure (cf. e.g., Nosko et al. 2010). In this context, several studies found that users of OSN intensively share private information (cf. e.g., Gross and Acquisti 2005; Lampe et al. 2007). Therefore, OSN provide a basis for "[…] maintaining social relationships, for finding users with similar interests, and for locating content and knowledge that has been contributed or endorsed by other users" (Mislove et al. 2007, p. 29).

To enable communication among users, OSN usually offer common messaging functionalities such as private messages or chats. Besides, most user profiles in OSN incorporate a kind of message board (often called "wall"). When creating a message on his or her own or on another user's message board, one can choose between a broad range of media types (e.g., status, link, photo, or app) in order to spread information the most adequate way (Yu et al. 2011). Moreover, users can comment on such messages. Comments are usually listed directly below the corresponding message in reverse chronological order. Within Facebook, for example, users can also endorse such wallposts by liking them and thereby pushing them in real time into the news feeds of their friends (Debatin et al. 2009). Besides, users can actively and virally spread wallposts among their friends via functionalities to "share" content with only a single click. In the context of Facebook it has been shown that 70% of all likes on wallposts happen within 4 hours and about 95% are received within 22 hours (Miller 2011). These figures underline the fact that OSN constitute powerful communication platforms that allow for presenting oneself and exchanging information in an efficient and timely manner.

II.2.3 Usage of Online Social Networks

Existing literature intensively deals with the users' motives for using OSN (cf. e.g., Dwyer et al. 2007; Hu and Kettinger 2008). While the majority of studies focus on the most popular and well-known OSN such as Facebook, it is important to keep in mind that a generalization of these findings for all possible kinds of OSN (cf. classification of OSN in section II.5) is hardly possible due to their different foci (Hargittai 2007). However, prior research suggests that particularly "identity management", i.e., constructing and maintaining a personal profile to present oneself to other users, is a major motive to use OSN (Kreps 2008; Lampe et al. 2007). Larsen (2007) found that especially “self-construction”, i.e., users providing information on their own profile and “co-construction”, i.e., others adding information about users (e.g., on their message board) play an important role. Thereby, some users try to create an ideal self that describes more how the person wants to be perceived (Zhao et al. 2008). In that context, Krasnova et al. (2008) identified the satisfaction of the needs for belongingness and esteem through self-presentation among others as main motives for using OSN. Prior work also indicates that having a lot of friends in one’s contact list itself can be a motive for using OSN (Donath and Boyd 2004). Vom Brocke et al. (2009) moreover identify “contact management”, i.e., maintaining personal contacts by means of OSN, as a major motive for using OSN. Thereby, they differentiate “social motives”, i.e., maintaining and searching for personal contacts, and “interest motives”, i.e., interest in a certain type of contacts. In this context, literature indicates that particularly the management and maintenance of existing contacts are major motives for using OSN (Raacke and Bonds-Raacke 2008). This is supported by earlier findings of Lampe et al. (2007), who distinguish the users’ motives “social searching”, i.e., learning more about contacts with whom they share an offline connection, and “social browsing”, i.e., finding new contacts online. Based on a survey with 1,440 freshmen students the study’s results indicate that the users of OSN such as Facebook are primarily motivated by social searching.

To sum up, the underlying idea of OSN is that users can first act independently from each other and build an own virtual identity by setting up a user profile. Afterwards, the creation and use of already existing and new relationships to other users becomes the central motive for using OSN. Thus, users can create a personal network consisting of hundreds of direct and indirect connections to friends, acquaintances, colleagues, and other like-minded users.

II.3 Structural Characteristics of Online Social Networks

The users and the structural characteristics of the network, i.e., the connections among the users (cf. Howard 2008; Oinas-Kukkonen et al. 2010), are key aspects of OSN. In general, structural characteristics have been extensively studied for instance to understand and explain human behavior in multiple social networks (cf. e.g., Shapiro and Varian 1999). In contrast to traditional social networks, which usually contain a small number of rather similar members, OSN and their structure are much more heterogenic as well as complex (Krasnova et al. 2010). For example, while in traditional social networks the number of close relationships is about 10 to 20 (Parks 2007), an average user of Facebook counts 130 so-called "friends" (Facebook 2011). Taken together with new possibilities to collect data by technological means, the previously unimagined availability and size of social network data led to the emergence of a new research stream (Bonchi et al. 2011).

In so-called "computational social science" (Lazer et al. 2009; Watts 2004), the structure invoked by the connections among users in OSN is mostly perceived as a set of nodes (users), and a set of directed or undirected edges (ties) connecting pairs of nodes (Adamic and Adar 2003; Bampo et al. 2008). These nodes and edges determining the network structure can be represented by a graph (Wasserman and Faust 1994). In most cases, the graph of OSN is based on the binary and rather static social links among users, i.e., friendship relationships, irrespective of these users' actual interactions. This graph is usually called the social graph (Benevenuto et al. 2009; Wilson et al. 2009). Its visualization especially highlights so-called hubs, i.e., users who have an exceedingly large number of social links to other users. Users who are in such a hub position are characterized by a great potential for communication and interaction within networks. However, not only the users' social links, but also the users' actual communication activity, i.e., the exchange of information for instance via messages or wallposts, is highly relevant. Prior research emphasizes the importance of users' communication activity: "No matter what resources are available within a structure, without communication activity those resources will remain dormant, and no benefits will be provided for individuals" (Butler 2001, p. 350). Recent work in the context of OSN indicates that the value of OSN stems from the communication activity between users (Willinger et al. 2009; Xu et al. 2008). Therefore, current studies also focus on the network that is based on users who actually interact rather than on users connected by mere social links. This network is usually called the activity network (Viswanath et al. 2009). While previous work on activity networks often examined instant messengers or telecommunication networks (e.g., Leskovec and Horvitz 2008; Onnela et al. 2007), there are some initial studies in the context of OSN as well (cf. e.g., Chun et al. 2008; Wilson et al. 2009). The graph resulting from such an activity network is usually referred to as the activity graph (Heidemann et al. 2010; Nazir et al. 2008), whereby nodes represent users and directed or undirected edges (activity links) represent communication activity between pairs of users.

In a basic activity graph of an OSN, all edges between nodes are the same, regardless of whether the corresponding users have a strong connection (i.e., interact frequently) or a weak connection (i.e., interact infrequently). However, literature highlights that there may be stronger and weaker connections between users (Gilbert and Karahalios 2009; Kahanda and Neville 2009; Wen et al. 2009; Xiang et al. 2010). In general, strong connections (also called strong ties) between users are for instance more likely to be activated for information flow and more influential (Brown and Reingen 1987). In contrast, weak connections (weak ties) provide people with access to information and resources beyond those available in their social circle (Granovetter 1973; 1983) and bridge cliques of strong connections. In the context of OSN, Wen et al. (2009, p. 2) conclude that the strength of connections “[…] denotes an irresistible element for [...] advertising”. In order to distinguish between strong and weak connections, a few studies started to examine each connection’s communication activity level (cf. e.g., Chun et al. 2008; Heidemann et al. 2010; Kiss and Bichler 2008). Thus, a weighted activity graph considering the strength of activity links can be built (cf. e.g., Barrat et al. 2004; Heidemann et al. 2010).

Based on a single or multiple snapshots of the social graph, the activity graph, or the weighted activity graph, OSN can be analyzed in detail by applying Social Network Analysis (Bonchi et al. 2011). Thereby, many researchers have verified similarities between traditional social networks and OSN. For instance, usually social networks as well as the social and activity graphs of OSN are scale-free (Chun et al. 2008; Watts and Strogatz 1998), i.e., follow power-law degree distributions. That is, in OSN many users have only few connections and some hubs create short cuts between users which otherwise would be far away from each other. Even though there might be gaps between users within large OSN, i.e., there are no direct links among all users, well-connected users tie together sub-networks. A number of experiments, constructing paths through social networks to distant target individuals (cf. e.g., Dodds et al. 2003; Korte and Milgram 1970) and current studies (cf. e.g., Leskovec and Horvitz 2008) lend credence to the six degrees of separation hypothesis, i.e., that everyone is just a few steps apart in the global social network (Milgram 1967). This so-called "small world" effect is also typical for modern networks such as OSN (Schnettler 2009; Watts and Strogatz 1998; Wilson 2009). Hence, OSN allow users to draw on resources from others in the network and to leverage connections from multiple social and geographically dispersed contexts (Haythornthwaite 2002).

In this regard, prior research emphasizes the importance of the size and the density of the network, as "[…] people are more likely to become active users, if they enter a dense […] network" (Howard 2008, p. 16). Furthermore, the whole network structure, i.e., direct and indirect connections, plays a decisive role. Findings by Kiss and Bichler (2008), for example, underline that a connection to a user with many social links is more valuable than to a user with only one or no further social link to other users. Benevenuto et al. (2009) showed that users do not only interact with directly connected users, but also have significant exposure to users "[…] that are 2 or more hops away […]" (Benevenuto et al. 2009, p. 50). A user's connectivity in the whole network constitutes a significant factor that may impact for instance advertising effectiveness in OSN (Wen et al. 2009). This is underpinned by further studies, which illustrate that well-connected users are particularly important for OSN, as they can be highly relevant for the promotion of brands, products, and viral marketing campaigns (Domingos and Richardson 2001; Staab et al. 2005; de Valck et al. 2009). Moreover, well-connected users tend to be more loyal, as for example every additional direct or indirect social link raises a user's barrier to leave the network (Algesheimer and von Wangenheim 2006; Ridings and Wasko 2010; Xu et al. 2009). Therefore, quantifying the interconnectedness of users in OSN is of great interest in theory and practice.

Approaches for quantifying the interconnectedness of users can be found not only in Social Network Analysis but also in many other fields for instance in scientometrics for the ranking of scientific journals (e.g., Bollen et al. 2006). For the specific context of social networks, several measures have been suggested to identify influential and prestigious nodes (Bonacich 1972; 1987; Wasserman and Faust 1994). The three most common centrality measures to quantify the centrality of a certain node in social networks are presented in Freeman's article "Centrality in Social Networks: Conceptual Clarification" (Freeman 1979): degree centrality, closeness centrality, and betweenness centrality. A fourth popular centrality measure, namely eigenvector centrality, is proposed by Bonacich (1972). The underlying primary eigenvector has been applied extensively to rank nodes in all types of networks. For instance, it has been used for the ranking of web pages (e.g., Brin and Page 1998) or to evaluate the influence of scientific journals (e.g., Bollen et al. 2006). These approaches acknowledge explicitly that not all connections are equal, as connections to nodes that are themselves influential are assumed to lend a node more influence than connections to less influential nodes (Newman 2003). Therefore, approaches such as PageRank have been used to identify particularly influential users in OSN (cf. e.g., Heidemann et al. 2010).

To sum up, the structural characteristics in terms of social as well as activity links among users form the backbone of OSN. The visibility and searchability of the users' social networks and the viral diffusion of information are distinctive features of OSN that allow the creation of "[…] substantial value for the individuals who participate in them, the organizations that sponsor them, and the larger society in multiple ways" (Agarwal et al. 2008, p. 243).

II.4 Genesis and Development of Online Social Networks

In the course of their development and because of their enormous usage and high potential, many OSNs have evolved over the last few years. While some became well established and are known around the world, also the rise and failure of some OSNs could be observed. In the following, we present one perspective on the genesis and development of the phenomenon OSNs and discuss major changes.

The Beginning of Online Social Networks: 1997 – 2002

New York 1997: Andrew Weinreich founded the first remarkable OSN SixDegrees was named after the six degrees of separation concept. Only one year later, SixDegrees already attracted one million registered users (Bedell 1998). However, the OSN was not able to create a sustainable business model (Boyd and Ellison 2007). The main reasons that contributed to the failure of the OSN in 2000 were the poorly developed web technology as well as the fact that the advertising industry was not mature enough (Prall 2010). According to the founder of SixDegrees it “[…] was simply ahead of its time” (as cited in Boyd and Ellison 2007, p. 214). Despite its fall, SixDegrees marked the beginning of a new era. In the following years a couple of further OSN as for example AsianAvenue, Black-Planet, MiGente, or LiveJournal began to support combinations of various technical functionalities, for example creating profiles, lists of friends, or guest books. While these early networks focused primarily on private networking, in 2001, the first business network designed to link business professionals, Ryze, was founded by Adrian Scott in San Francisco. Indeed, Ryze served as a role model for the subsequent business networks (e.g., LinkedIn). However, Ryze never enjoyed great popularity (Boyd and Ellison 2007). In 2002, the well-known OSN Friendster launched as a competitor to the online dating platform Match (Boyd 2004). Friendster was created to set up friends-of-friends, based on the assumption that friends-of-friends are more likely to build romantic relationships than strangers would (Boyd and Heer 2006). Therefore, Friendster restricted the access to other users within four degrees distance (Boyd 2004). Until the beginning of 2004, Friendster had been the largest OSN. However, in the following years it lost many of its early users due to technical problems (e.g., the site was not able to handle the rapid growth) and social problems (e.g., users found themselves contacting their bosses and classmates) (Boyd and Ellison 2007). Although many of these early networks failed, they marked a new era and built the foundation for future OSN.

The Growth of Online Social Networks and the Rise to Popularity: 2003 – 2009

A new wave of social networking began with the rise of MySpace in California in 2003. At the beginning, MySpace primarily focused on attracting frustrated Friendster users. Thus, MySpace was able to grow quickly. According to Jonathan Abrams, the founder of Friendster, "[…] the real reason that Friendster got supplanted by MySpace in the U.S. was that MySpace's website just worked and Friendster's didn't" (as cited in Milan 2009). Although MySpace did not start with a focus on bands in mind, one of the first user groups were musicians who appreciated this new possibility to present themselves to their fans (Boyd and Ellison 2007). The symbiotic band and fan relationship helped MySpace to attract particularly younger users beyond the Friendster network. From 2003 onwards, many new OSN were launched trying to replicate the early success of Friendster. The social software analyst Clay Shirky (2003) described this development with the term YASNS: "Yet Another Social Networking Service". In that context, several new OSN launched focusing on niche demographics or special interests, explicitly seeking narrower audiences. Professional sites such as XING and LinkedIn were founded in order to gain access to a new group of users, i.e., business people. In contrast, elite sites like aSmallWorld, activity-centered sites like Couchsurfing, or religion-focused sites like MyChurch tried to gain a competitive advantage by limiting their target groups (Boyd and Ellison 2007). Also one of the most popular OSN to date started to support niche demographics before expanding to a broader audience: Facebook was launched in early 2004 by Mark Zuckerberg and began as a Harvard-only OSN, while its mission today is "[…] to make the world more open and connected" (Facebook 2012). From 2005 onwards, Facebook was open for students from other schools and shortly after membership was possible for a broader audience as well. Mark Zuckerberg was definitely not the first person who built an OSN. However, he was one of the first with enormous and sustainable success. With the growth of Facebook as well as the success of OSN in countries all over the world – like StudiVZ in Germany, Hyves in the Netherlands, Renren in Asia, and Orkut in Brazil – more and more people paid attention to OSN. At the same time, with the growing number of users, OSN generated an increasing economic interest among investors. In 2005, for example, the media company News Corporation acquired the OSN MySpace for US$ 580 million (BBC 2005). Two years later, Microsoft paid US$ 240 million for a 1.6% minority interest in the OSN Facebook (MSNBC 2007). In 2008, AOL acquired the OSN Bebo for US$ 850 million (BBC 2010). In other countries investors paid considerable amounts for acquiring OSN as well. These facts demonstrate that between 2003 and 2009, OSN evolved to a global phenomenon with an increasing social and economic impact.

Online Social Networks – A Global Phenomenon: 2010 – Present

With the beginning of 2010, the major share of the market was dominated by a handful of OSN (Richter et al. 2011). Facebook was the most popular platform available in 70 languages with over 800 million users worldwide (Facebook 2012). However, besides the success of Facebook, there are several further OSN that launched from 2010 onwards and concentrate either on niches to survive as a complement or rival to Facebook. Examples for such niche services are Audimated or Folksdirect – the latter promised to offer a privacy-focused environment (Richter et al. 2011). Similar to Folksdirect, Unthink started as an “anti-Facebook” social network in 2011, distinguishing itself from Facebook by focusing on the easy control of privacy (Perez 2011). One of the biggest attempts to attack Facebook up to this time was the launch of Google+ in 2011. Google+ was founded to bring friends together, but in comparison to previous OSN, users could organize their contacts around “circles” that enable users to share specific information with particular user groups. The period from 2010 onwards is characterized by the emergence of further OSN. Many existing OSN also face the challenge of how to build a sustainable business model by leveraging the potential of their fast-growing user base in order to remain financial viable. As a consequence, many OSN had to reassess their business models (Clemons 2009; Lu and Hsiao 2010). Friendster, for example, repositioned itself from an OSN to a social entertainment and gaming site with its strongest market in Asia in June 2011. Since then, the number of registered users has reached over 100 million. However, there are also examples of OSN that did not achieve a renaissance and declined. In 2010, for example, AOL sold the OSN Bebo for a sum probably less than US$ 10 million after just two years (BBC 2010). MySpace may serve as another famous example how quickly OSN can rise and fall. In 2011, News Corporation sold MySpace for US$ 35 million six years after acquiring the network for US$ 580 million (Vascellaro et al. 2011). Nevertheless, there are successful examples how to cope with the challenge of building sustainable business models. Facebook, for example, generated US$ 3.7 billion in revenues in 2011 and is therefore the most successful OSN to this time. To sum up, nowadays OSN are no longer a niche phenomenon for young people. It is a global phenomenon with a still increasing economic and social impact that reaches all demographic groups all over the world. A timeline of the market appearances of selected OSN between 1997 and 2011 is shown in in Figure II-1.

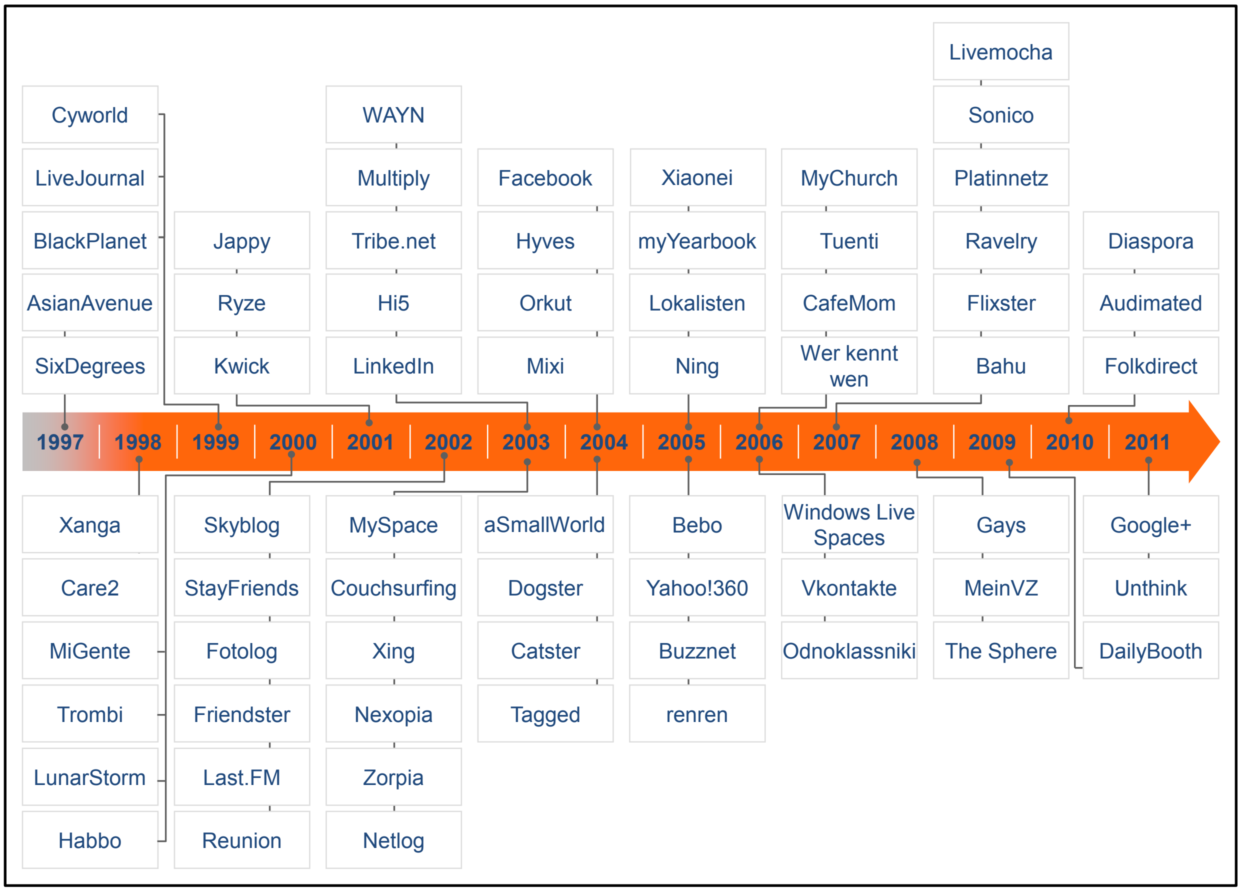

Figure II-1. A timeline of the foundation of selected online social networks from 1997 – 2011

II.5 Classification of Online Social Networks

The genesis and development of OSN illustrate that OSN exist for many target groups and fields of interest. They can be particularly distinguished according to their primary range of usage between “private networks” (e.g., Facebook, MySpace) and “business networks” (e.g., LinkedIn, Xing) (cf. Mesch and Talmud 2006). Since OSN were originally designed for private use (Bughin and Manyika 2007), it is not surprising that private networks such as Facebook are among the most popular and well-known OSN in the world (Thadani and Cheung 2011). In contrast to private networks, business networks “[…] specialise in maintaining professional contacts and searching for new jobs” (Bonneau and Preibusch 2010, p. 125). Besides the usual information provided in OSN, business networks usually also incorporate a curriculum vitae (e.g., current position, job title). Furthermore, many business networks include additional details in user profiles, such as registration date or an index indicating a user’s activity within the business network (cf. Strufe 2010).

Another criterion for categorization is restriction of focus: There are “general networks”, without any particular focus (e.g., Facebook), as well as “special interest networks“, with specific focuses (e.g., Bottletalk). Drawing on Boyd and Ellison (2007) as well as Leimeister et al. (2004), such special interest networks could be defined as technical online platforms that have a particular focus and aim at specific target groups of users who interact socially. In line with OSN in general, these platforms allow users to construct public or semi-public profiles and to articulate lists of other users with whom they share a connection. Special interest networks, however, are aligned to the particular focus and particularly enable and support the users’ interactions that help to build trust and a common feeling among its users. Due to their relatively narrow focus, special interest networks are similar to so-called “communities of interest”, where individuals interact with one another on specific topics (Armstrong and Hagel III 1996). Thereby, communities of interest are solely organized around interests. Special interest networks, however, are also organized around their users (Boyd and Ellison 2007). Thus, in contrast to communities of interest, where users do not share intensely personal information (Armstrong and Hagel III 1996), i.e., not much attention is paid to socializing; special interest networks provide functionalities for finding and maintaining social contacts. Prior work indicates a high potential for special interest networks, particularly regarding the chance to successfully establish supplementary offerings in addition to general networks such as Facebook (cf. e.g., Richter et al. 2011). This is underlined by Chris Anderson, Editor-in-Chief of the magazine Wired, who said that “[w]e don’t need another giant social-network site. The world needs an infinite number of micro social networks about specific issues" (as cited in Costa 2008). Altogether, OSN can be differentiated between "private", "business", "general", and "special interest" as highlighted in Figure II-2.

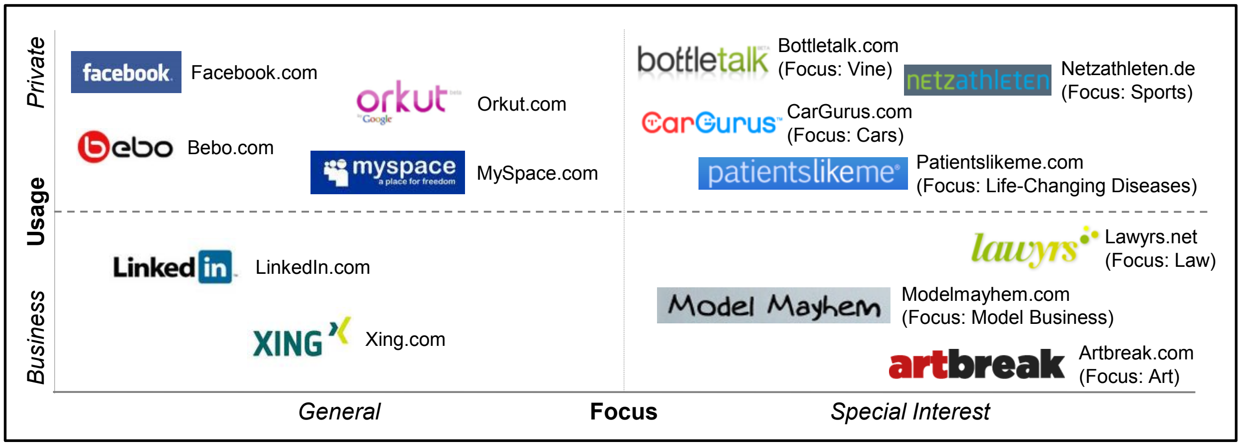

Figure II-2. Classification of online social networks (in accordance to Heidemann et al. 2011, p. 104)

II.6 Potential Business Value of Online Social Networks along the Value Chain

The increasing importance of OSN also has an impact on companies. Industry experts believe that OSN will create a significant change in consumer behavior and have a substantial impact on traditional industries (Pallis et al. 2011). However, companies often remain uncertain about the actual benefits of using these networks in a business context (Kettles and David 2008). Therefore, major questions from today's view are: (1) what are potential business benefits of OSN from a company's perspective? (2) In which business functions may OSN effectively be used? In the following, we present potential business benefits of OSN along the value chain, point out selected activities, and provide some practical examples.

First, OSN can be leveraged in the context of *research and development*. Users of OSN may develop and design products and services, discuss new innovative ideas, and evaluate them. For example in "open source spaces" users develop new software, share information, look for social support, and collaborate on generating innovative ideas (e.g., Kozinets et al. 2008). This open innovation approach follows the statement of Bill Joy (co-founder of Sun Microsystems), that "No matter who you are, most of the smartest people work for someone else" (as cited in Haller et al. 2011, p. 103).

Integrating customers as innovators into the product development process allows companies to receive valuable insights about their customers’ needs and to unleash their creativity and potential for innovation resulting in reduced costs for product development (Kettles and David 2008). There are a lot of successful practical examples on how companies used the “crowd” in OSN for social product development: The automotive company Fiat received over 170.000 design concepts for free by integrating customers into the development process of the model Fiat 500. Lego took advantage of the worldwide creativity of its Lego factory community (which was accessed over a million times) by calling for new Lego design models that were evaluated and commented upon by other users (Heidemann 2010). Due to the volume of available data in news feeds, groups, etc., OSN can be leveraged for market research as well. Casteleyn et al. (2009), for example, highlight that data in OSN can be seen as a “crystal ball” for future consumer intentions and showed how Facebook can be used for market research.

Second, academic studies pay considerable attention to the role of OSN in the context of *marketing and sales* (Cheung and Lee 2010; Trusov et al. 2009). In this area OSN can be leveraged for various business activities such as conducting marketing campaigns and word-of-mouth marketing (cf. e.g., Algesheimer and von Wangenheim 2006; Bernoff and Li 2008; Brown et al. 2007) or targeted advertising (cf. e.g., Enders et al. 2008). Bernoff and Li (2008), for example, illustrate how OSN can be used as a new channel for effective and efficient marketing campaigns using the case of Chevrolet. The major idea behind this is to benefit from the fact that people talk to their friends (word-of-mouth marketing) and in doing so may mobilize thousands of users who become aware of a company’s product or service. In the context of viral marketing campaigns, research recommends taking into account the social structure of an OSN to optimize the campaign performance (Bampo et al. 2008). Prior literature shows that customers tend to trust more in recommendations by other customers than in marketing messages originated by companies (Ermecke et al. 2009). Thus, it is promising to attract the interest of (satisfied) customers and facilitate positive word-of-mouth in order to generate business value (Bernoff and Li 2008). This change demonstrates a fundamental shift in marketing from sales marketing via relationship marketing to social network marketing (Hill et. al. 2006). In that context, Algesheimer and von Wangenheim (2006) state that the more central a customer is positioned within the network, the higher is his or her customer network lifetime value for a company. Against this background, identifying influential users with respect to connectivity and activity in OSN (cf. section II.3) is an important mean to enable successful social network marketing. Moreover, OSN show great potential towards targeted advertising purposes, especially against the background that traditional advertising is increasingly losing its impact (Clemons 2009; Enders et al. 2011). Another emerging area for the application of OSN is social Customer Relationship Management (CRM). The addition “social” to CRM includes for example trend-analysis for future business opportunities as well as reputation monitoring (Bonchi et al. 2011). Faase et al. (2011), for instance, found based on empirical research that social CRM allows generation of substantial business value. Finally, more and more companies are using OSN as a new sales channel – some experts call it “F-commerce”. In December 2010, for instance, the company JC Penney launched a full e-commerce store within its Facebook page.

Furthermore, many companies leverage OSN to yield specific benefits in the field of *Customer Service*. According to Libai et al. (2010) companies are using OSN as a service channel to reduce service support costs as a result of customer-to-customer interactions and to receive valuable real-time customer feedback. The German telecommunication provider Telekom, for example, established Facebook (“Telekom-hilft”) as a new channel which is almost entirely dedicated to providing customer service (not only from Telekom-to-customer but also from customer-to-customer). Besides, Bonchi et al. (2011) point out that with an internal social network in place, customer calls and emails can be routed more effectively to experts.

In the field of *human resources*, OSN are gaining increasing importance as well. Initial studies underline the potential of using OSN to recruit business professionals (c.f. e.g., Bonchi et al. 2011). Users typically reveal a large amount of private information about themselves in OSN (Stutzmann 2006) which can be helpful in recruiting new expertise. In this context, OSN that address a professional audience, such as LinkedIn or XING, can be viewed as marketplaces for the exchange of skills (Richter et al. 2011). Besides the purpose of identifying and contacting potential employees (cf. e.g., Bonneau and Preibusch 2010), companies have also started to use OSN to support the selection process and to make hiring decisions. Kluemper and Rosen (2008), for example, showed that recruiters are able to accurately distinguish high from low performers solely based on viewing their OSN profiles. Beyond recruiting, companies can leverage OSN to develop their employer branding (Girard and Fallery 2009). Companies, such as BMW or Bertelsmann, for example, have their own career fan pages on Facebook. Furthermore, IBM has started a XING-group “The greater IBM connection” in 2006 which currently counts more than 13,000 users and aims at connecting employees and Alumni.

Finally, complementary to the external use of public OSN, companies increasingly engage in setting up OSN for *internal applications* to support networking among their employees. As knowledge workers in organizations collaborate more and more as virtual teams in distributed setups (Breu and Hemingway 2004), internal OSN offer attractive means to create social structures and can serve as channel for information transfer between individuals. Di Micco and Millen (2007), for example, showed that internal OSN can help employees to identify topics of common interest and create a basis for communication between distant workers. Other studies emphasized that internal OSN open up new possibilities for skill-based staffing of knowledge intensive projects (Richter et al. 2011). In this context, “IBM blue pages” may serve as a prominent example. Agarwal et al. (2008, p. 244) even point out that leading companies are using the power of OSN to transform their internal organizations from “[…] command-and-control to connect-and-coordinate”. Thereby, the value of internal OSNs is “[…] determined not by the tools but by how tools are harnessed for value creation” (Majchrzak et al. 2009, p. 103).

To sum up, for companies there are promising fields of application of OSN along the whole value chain. The use of OSN may be beneficial in multiple ways, including generating innovation, providing social support, enhancing knowledge, or boosting sales by marketing campaigns (cf. Table II-2). Given, for instance, the huge amount of data provided in OSN and the fast diffusion of information enabled by OSN, their effective usage can lead to lower costs (e.g., Bonchi et al. 2011) and increased revenues (e.g., Demailly and Silman 2008).

Table II-2. Exemplary areas of application of online social networks

Business area	Selected activities	Exemplary references
1. Research and development	Product development Market research	Bernoff and Li (2008); Casteleyn et al. (2009); Heidemann (2010); Kettles and David (2008); Kozinets et al. (2008)
2. Marketing and Sales	Marketing campaigns Word-of-mouth marketing Targeted advertising Social CRM	Algesheimer and von Wangenheim (2006); Bampo et al. (2008); Bernoff and Li (2008); Bonchi et al. (2011); Brown et al. (2007); Enders et al. (2008); Ermecke et al. (2009); Faase et al. (2011); Heidemann et al. (2010); Trusov et al. (2009); Xu et al. (2009)
3. Customer Service	Customer support After sales support	Bonchi et al. (2011); Libai et al. (2010)
4. Human Resources	Recruiting Employer branding	Girard and Fallery (2009); Kluemper and Rosen (2008); Richter et al. (2011)
5. Internal applications	Expert search Collaboration in virtual teams Knowledge management	DiMicco and Millen (2007); Majchrzak et al. (2009); Richter et al. (2011)

II.7 Challenges and Risks of Online Social Networks

On the one hand, OSN promise numerous business potentials for companies (cf. section II.6). On the other hand, the usage of OSN in the business context may go along with several challenges and risks as well. In the following, we will provide an overview and illustrative examples of selected challenges and risks companies are confronted with.

One major challenge for companies is to understand within which business functions and for which activities OSN can be leveraged reasonably. Looking at the press coverage, it seems that OSN are often seen as an “all-purpose tool”. Clemons (2009), for instance, pointed out that companies, marketing agencies, and media giants believe that OSN solve their problems by simply using them as a new channel. However, as a first step, companies have to deeply analyze which goals they pursue by using OSN. They need to understand in which business functions OSN can be used in order to create business value. This is all the more important as the usage of OSN causes costs; for instance often a handful of people is needed as support staff as in the case of “Telekom-hilft” (cf. section II.6). Thereby, it is also important to note that the successful use of OSN requires the right skills of employees and a good management of the operational execution. Possible consequences of the mismanagement of a company’s OSN activities can be illustrated using the example of the German railway company Deutsche Bahn: In 2010, Deutsche Bahn launched the sales campaign “Chefticket” – a cheap railway ticket that was sold via the company’s Facebook fan page. At the beginning, the campaign attracted thousands of customers. But within a short time, the Facebook fan page developed into a melting pot for customer criticism. Deutsche Bahn was overstrained with the flood of user-generated negative comments. Instead of responding to the users’ feedback the company ignored the users and their complaints. Experts blame Deutsche Bahn for this behavior and the missing communication strategy that put many customers off.

Another major challenge for companies is the loss of control in the context of OSN. The organizational transformation from “[…] command-and-control to connect-and-coordinate” (Agarwal et al. 2008, p. 244) itself opens the door to several risks. This is a result of the fact that many companies are not prepared for such a cultural change. Thereby, the loss of control can lead, for instance, to reputation risks and unexpected results. In 2010, for example, Greenpeace conducted a campaign against the chocolate bar “KitKat” of the Swiss food company Nestlé. Thereby, Greenpeace used the chocolate bar’s Facebook fan page to convince the fan community that Nestlé is responsible for the death of monkeys in the primeval forest. Within a short time, independently of the truth of the story, the campaign succeeded and a large part of the fan community turned against Nestlé. The protest against UEFA sponsor Adidas on its Facebook fan page in 2011 may serve as another example. Adidas fans protested against killing of stray dogs in the Ukraine before the upcoming soccer championship. The protests forced the sponsor of the event Adidas to cooperate with the Ukrainian government to pass a new legislation that averts the killing. This example underlines the importance of quickly detecting and responding to externally enforced negative word-of-mouth in OSN.

Another critical aspect of OSN refers to the privacy risks they involve (Krasnova et al. 2009). Indeed, data privacy and security concerns are a challenge for companies as well. According to a study of Fraunhofer, many of the most popular OSN (e.g., Facebook, LinkedIn) have enormous data privacy problems (Galdy 2008). Gross and Acquisti (2005), for example, showed that based on the personal information users provide online, they expose themselves to various physical and cyber risks. The same holds true for companies if employees participate in OSN. As a consequence, many companies restrict the usage of OSN as they are afraid that confidential information may be disclosed via OSN. Another issue in the context of data privacy concerns the ownership of data provided within OSN. Many OSN providers are convinced that fan pages and the data provided there are solely under the responsibility of the OSN. Therefore, for example a German Data Protection Authority recommended that institutions shut down Facebook fan pages and remove “like” buttons from their websites (ULD 2011). Moreover, companies can become victims of “fake profiles” in OSN. Such fake profiles can for instance lead to enormous negative implications such as spam or negative publicity. Thus, aroused by the enormous digital availability of personal information about users (e.g., profiles, photos, friend list) and the accompanying potential risks, most of the research in this field concentrates on the development of techniques for privacy protection (Pallis et al. 2011).

Finally, OSN provide a huge data base for knowledge discovery, for example with regard to customer habits, churn prediction, or new product trends. However, there is a lack of knowledge about the use of these enormous amounts of data for potential business applications. While there is a lot of research regarding different problems and methods for Social Network Analysis, “[…] there is a gap between the techniques developed by the research community and their deployment in real world applications” (Bonchi et al. 2011, p. 1). Thus, even though there are pioneers, for example in the telecommunication industry using Social Network Analysis to gain customer insights (cf. e.g., Kiss and Bichler 2008), most companies still face the challenge of how to discover knowledge from OSN and how to use Social Network Analysis and mining techniques. Another related challenge is the question of how to measure and quantify the actual business benefit of OSN. Fisher (2009, p. 189), for example, points out that the return on investment (ROI) has become the “[…] Holy Grail of social media”.

To sum up, in the context of OSN there are still numerous challenges and risks companies are confronted with. The aspects mentioned above may serve as illustrative examples. Overall, it is most important to be aware of these challenges and risks and to account for a company’s specific situation, know-how, resources, and culture when deriving a stringent and goal-oriented OSN strategy. These challenges and risks, however, should not discourage companies from exploring and leveraging OSN step by step.

II.8 Conclusion

The goal of this article was to provide an overview of OSN. In literature different terms are used to describe this current phenomenon. We defined OSN according to Boyd and Ellison (2007) (who use the term Social Network Site) but focus on user-oriented sites that are mainly used for networking purposes. Afterwards, we introduced the main functionalities that enable users to set up a personal profile and to extensively communicate with each other. These functionalities particularly support identity and contact management, which have been identified as users’ major motives for using OSN. Subsequently, we discussed the structural characteristics that form the backbone of OSN. In this context, we briefly introduced concepts and findings related to the application of Social Network Analysis to the graphs that can either represent users’ friendship relationships or users’ actual communication activity within the network. Having introduced the basic characteristics, we shed light on the genesis and the emergence of OSN over time. Besides the history of OSN, the large number of existing OSN aiming at different target groups and topics became apparent. Therefore, we also provided a classification of OSN differentiating between a rather private or business-oriented usage and a rather general or specialized focus. Finally, we investigated the impact and value of OSN for companies by outlining possible fields of application along the value chain. Moreover, we critically discussed major challenges and risks companies are confronted with in the context of OSN.

Taken together, we highlighted the – from our point of view – most relevant information and sufficient references to follow up on any of the above mentioned topics. However, OSN constitute a very large, interdisciplinary area of research that is rather young but tremendously fast evolving. Therefore, our analysis is by no means complete. Furthermore, as mentioned above, we solely focused on user-oriented sites. Research on content-oriented sites such as YouTube or Flickr has consequently been omitted. Future work could on the one hand widen the scope by addressing further aspects of OSN and on the other hand focus on going deeper into specific subareas. Nevertheless, we hope that our article can contribute to a better understanding of the current phenomenon of OSN and provide starting points for future research.

References (Chapter II)

Adamic LA, Adar E (2003) Friends and neighbors on the Web. Social Networks 25(3):211-230

Adamic LA, Adar E (2005) How to search a social network. Social Networks 27(3): 187-203

Agarwal R, Gupta AK, Kraut R (2008) The interplay between digital and social networks. Information Systems Research 19(3):243-252

Algesheimer R, von Wangenheim F (2006) A network based approach to customer equity management. Journal of Relationship Marketing 5(1):39-57

Armstrong A, Hagel III J (1996) The real value of online communities. Harvard Business Review 74(3):134-141

Bagozzi RP, Dholakia UM (2006) Open source software user communities: a study of participation in Linux user groups. Management Science 52(7):1099-1115

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

Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. Proceedings of the National Academy of Sciences of the United States of America 101(11):3747-3752

BBC (2005) News Corp in $580m Internet buy. http://news.bbc.co.uk/2/hi/business/4695495.stm, accessed 2012-04-20

BBC (2010) Bebo sold by AOL after just two years. http://www.bbc.co.uk/news/10341413, accessed 2012-04-20

Bedell D (1998) Meeting your new best friends Six Degrees widens your contacts in exchange for sampling Web sites. http://www.dougbedell.com/sixdegrees1.html, accessed 2012-04-20

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

Benevenuto F, Rodrigues T, Cha M, Almeida V (2009) characterizing user behavior in online social networks. In: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, Chicago, IL, USA, pp. 49-62

Bernoff J, Li C (2008) Harnessing the power of the oh-so-social Web. MIT Sloan Management Review 49(3):36-42

Bollen J, Rodriguez MA, van De Sompel H (2006) Journal status. Scientometrics 69(3):669-687

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

Bonacich P (1987) Power and centrality: A family of measures. American Journal of Sociology 92(5):1170-1182

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

Bonneau J, Preibusch S (2010) The privacy jungle: On the market for data protection in social networks. In: Moore T, Pym D, Ioannidis C (eds) Economics of information security and privacy. Springer, New York, NY, USA, pp. 121-167

Boyd DM (2004) Friendsters and publicly articulated social networking. In: Proceedings of the Conference on Human Factors in Computing Systems, Vienna, Austria, pp. 1279-1282

Boyd DM (2006) The significance of social software. In: Burg TN, Schmidt J (eds) Blog talks reloaded: Social software research & cases. Books on Demand, Norderstedt, Germany, pp. 15-30

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

Boyd DM, Heer J (2006) Profiles as conversation: Networked identity performance on Friendster. In: Proceedings of the 39th Hawaii International Conference on System Sciences, Kauai, HI, USA, pp. 1-10

Breu K, Hemingway CJ (2004) Making organizations virtual: The hidden costs of distributed teams. Journal of Information Technology 19(3):191-202

Brin S, Page L (1998) The Anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems 30(1-7):107-117

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 PH (1987) Social ties and word-of-mouth referral behavior. Journal of Consumer Research 14(3):350-362

Bughin J, Manyika J (2007) How businesses are using the web 2.0: A McKinsey global survey. The McKinsey Quarterly March 32-39

Buhl HU (2008) Online communities. WIRTSCHAFTSINFORMATIK 50(2):81-84

Butler BS (2001) Membership size, communication activity, and sustainability: A resource-based model of online social structures. Information Systems Research 12(4):346-362

Casteleyn J, Mottart A, Rutten K (2009) How to use Facebook in your market research. International Journal of Market Research 51(4):439-447

Cheung CMK, Lee MKO (2010) A theoretical model of intentional social action in online social networks. Decision Support Systems 49(1):24-30

Chun H, Kwak H, Eom Y, Ahn Y, Moon S, Jeong H (2008) comparison of online social relations in volume vs. interaction: A case study of Cyworld. In: Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement, Vouliagmeni, Greece, pp. 57-70

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

Costa D (2008) The micro threat to big social networks. http://www.pcmag.com/article2/0,2817,2307688,00.asp, accessed 2012-04-20

de Valck K, van Bruggen GH, Wierenga B (2009) Virtual communities: A marketing perspective. Decision Support Systems 47(3):185-203

Debatin B, Lovejoy JP, Horn A, Hughes BN (2009) Facebook and online privacy: Attitudes, behaviours, and unintended consequences. Journal of Computer-Mediated Communication 15(1):83-108

Demailly C, Silman M (2008) The business impact of social networking (AT&T whitepaper), http://www.business.att.com/content/whitepaper/WP-soc_17172_v3_11-10-08.pdf, accessed 2012-04-20

DiMicco JM, Millen DR (2007) Identity management: multiple presentations of self in Facebook. In: Proceedings of the International ACM Conference on Supporting Group Work, Sanibel Island, FL, USA, pp. 383-386

Dodds PS, Muhamad R, Watts DJ (2003) An experimental study of search in global social networks. Science 301(5634):827-829

Domingos P, Richardson M (2001) Mining the network value of customers. In: Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, pp. 57-66

Donath J, Boyd DM (2004) Public displays of connection. BT Technology Journal 22(4):71-82

Dougherty H (2010) Facebook reaches top ranking in US, http://weblogs.hitwise.com/heather-dougherty/2010/03/facebook_reaches_top_ranking_i.html, accessed 2012-04-20

Dwyer C, Hiltz S, Passerini K (2007) Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. In: Proceedings of the 13th Americas Conference on Information Systems, Keystone, CO, USA, paper 339

eMarketer (2012) Total worldwide social network ad revenues continue strong growth. http://www.emarketer.com/Article.aspx?R=1008862, accessed 2012-04-20

Enders A, Hungenberg H, Denker H-P, Mauch S (2008) The long tail of social networking. Revenue models of social networking sites. European Management Journal 26(3):199-211

Ermecke R, Mayrhofer P, Wagner S (2009) Agents of diffusion – Insights from a survey of Facebook users. In: Proceedings of the 42nd Hawaii International Conference on System Sciences, Big Island, HI, USA, pp. 1-10

Faase R, Helms R, Spruit M (2011) Web 2.0 in the CRM domain: Defining social CRM. International Journal of Electronic Customer Relationship Management 5(1):1-22

Facebook (2011) Facebook statistics. Facebook. https://www.facebook.com/press/info.php?statistics, accessed 2012-04-20

Facebook (2012) Facebook newsroom. Facebook. http://newsroom.fb.com/content/default.aspx?NewsAreaId=22, accessed 2012-04-20

Fisher T (2009) ROI in social media – A look at the arguments. Journal of Database Marketing & Customer Strategy Management 16(3):189-195

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

Galdy A (2008) Soziale Netzwerke: Datenschutz mangelhaft. http://www.cio.de/knowledgecenter/security/859286/, accessed 2012-04-20

Gilbert E, Karahalios K (2009) Predicting tie strength with social media. In: Proceedings of the 27th SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, pp. 211-220

Girard A, Fallery B (2009) E-Recruitment: New practices, new issues an exploratory study. In: Proceedings of the 3rd International Workshop on Human Resource Information Systems, Milan, Italy, pp. 39-48

Gneiser M, Heidemann J, Klier M, Landherr A, Probst F (2012) Valuation of online social networks taking into account users’ interconnectedness. Information Systems and e-Business Management 10(1):61-84

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

Granovetter MS (1983) The strength of weak ties: A network theory revisited. Sociological Theory 1(1):201-233

Gross R, Acquisti A (2005) Information revelation and privacy in online social networks. In: Proceedings of the ACM Workshop on Privacy in the Electronic Society, Alexandria, VA, USA, pp. 71-80

Haller J, Bullinger AC, Möslein (2011) KM innovation contests – An IT-based tool for innovation management. Business & Information Systems Engineering 53(2):103-106

Hargittai E (2007) Whose space? Differences among users and non-users of social network sites. Journal of Computer-Mediated Communication 13(1):276-297

Haythornthwaite C (2002) Building social networks via computer networks: Creating and sustaining distributed learning communities. In: Renninger KA, Shumar W (eds) Building virtual communities: Learning and change in cyberspace. Cambridge University Press, Cambridge, United Kingdom, pp. 159-190

Heidemann J (2010) Online Social Networks – Ein sozialer und technischer Überblick. Informatikspektrum 33(3):262-271

Heidemann J, Klier M, Landherr A, Probst F, Calmbach F (2011) Special Interest Networks: Eine Fallstudie am Beispiel von Netzathleten.de. HMD - Praxis der Wirtschaftsinformatik 48(6):103-112

Heidemann J, Klier M, Probst F (2010) Identifying key users in online social networks: A PageRank based approach. In: Proceedings of the 31st International Conference on Information Systems, St. Louis, MO, USA, paper 79

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

Howard B (2008) Analyzing online social networks. Communications of the ACM 51(11):14-16

Hu T, Kettinger WJ (2008) Why people continue to use social networking services: developing a comprehensive model. In: Proceedings of the 29th International Conference on Information Systems, Paris, France, paper 89

Kahanda I, Neville J (2009) Using transactional information to predict link strength in online social networks. In: Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media, San Jose, CA, USA, pp. 74-81

Kettles D, David S (2008) The business value of social network technologies: A framework for identifying opportunities for business value and an emerging research. In: Proceedings of the 14th Americas Conference on Information Systems, Toronto, ON, Canada, paper 379

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

Kluemper DH, Rosen PA (2009) Future employment selection methods: Evaluating social networking web sites. Journal of Managerial Psychology 24(6):567-580

Korte C, Milgram S (1970) Acquaintance networks between racial groups: Application of the small world method. Journal of Personality and Social Psychology 15(2): 101-108

Kozinets RV, Hemetsberger A, Schau HJ (2008) The wisdom of consumer crowds. Journal of Macromarketing 28(4):339-354

Krasnova H, Hildebrand T, Günther O (2009) Investigating the value of privacy in online social networks: Conjoint analysis. In: Proceedings of the 30th International Conference on Information Systems, Phoenix, AZ, USA, paper 173

Krasnova H, Hildebrand T, Günther O, Kovrigin A, Nowobilska A (2008) Why participate in an online social network: an empirical analysis. In: Proceedings of the 16th European Conference on Information Systems, Galway, Ireland, paper 33

Krasnova H, Koroleva K, Veltri NF (2010) Investigation of the network construction behavior on social networking sites. In: Proceedings of the 31st International Conference on Information Systems, St. Louis, MO, paper 182

Kreps D (2008) My Facebook profile: Copy, resemblance or simulacrum. In: Proceedings of the 16th European Conference on Information Systems, Galway, Ireland, paper 141

Lampe C, Ellison N, Steinfield CA (2007) A familiar Face(book): Profile elements as signals in an online social network. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, pp. 435-444

Larsen MC (2007) Understanding social networking: On young people's construction and co-construction of identity online. In: Proceedings of the Internet Research 8.0: Let’s Play Conference of the Association of Internet Researchers, Vancouver, BC, Canada, pp. 1-18

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

Leimeister JM, Sidirask P, Krcmar H (2004) Success factors of virtual communities from the perspective of members and operators: An empirical study. In: Proceedings of the 37th Hawaii International Conference on System Sciences, Big Island, HI, USA, pp. 1-10

Leskovec J, Horvitz E (2008) Planetary-scale views on an instant-messaging network. In: Proceedings of the 17th International Conference on World Wide Web, Beijing, China, pp. 915-924

Libai B, Bolton R, Bügel MS, Götz 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

Lu H-P, Hsiao K-L (2010) The influence of extro/introversion on the intention to pay for social networking sites. Information & Management 47(3):150-157

Majchrzak A, Cherbakov L, Ives B (2009) harnessing the power of the crowds with corporate social networking tools: How IBM does it. MIS Quarterly Executive 8(2):103-108

Mesch G, Talmud I (2006) The quality of online and offline relationships: The role of multiplexity and duration of social relationships. The Information Society 22(3): 137-148

Milan M (2009) Friendster founder on the rise and fall of America’s first big social network. http://latimesblogs.latimes.com/technology/2009/07/friendster.html, accessed 2012-04-20

Milgram S (1967) The small world problem. Psychology Today 1(1):61-67

Miller A (2011) Media makeover: Improving the news one click at a time. TED Books, New York, NY, USA

Mislove A, Marcon M, Gummadi KP, Druschel P, Bhattacharjee B (2007) Measurement and analysis of online social networks. In: Proceedings of the 17th ACM SIGCOMM Conference on Internet Measurement, San Diego, CA, USA, pp. 29-42

MSNBC (2007) Microsoft invests $240 Million in Facebook. http://www.msnbc.msn.com/id/21458486, accessed 2012-04-20

Nazir A, Raza S, Chuah C-N (2008) Unveiling Facebook: A measurement study of social network based applications. In: Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement, Vouliagmeni, Greece, pp. 43-56

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

Nosko A, Wood E, Molema S (2010) All about me: Disclosure in online social networking profiles: The case of FACEBOOK. Computers in Human Behavior 26(3):406-418

O’Reilly T (2005) What is Web 2.0? http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html, accessed 2012-04-20

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

Onnela J-P, Saramäki J, Hyvönen J, Szabó G, de Menezes MA, Kaski K, Barabási A-L, Kertész J (2007) Analysis of a large-scale weighted network of one-to-one human communication. New Journal of Physics 9(6):article 179

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. Springer, Berlin, Germany, pp. 213-234

Parks MR (2007) Personal relationships and personal networks. Lawrence Erlbaum, Mahwah, NJ, USA

Perez S (2011) Anti-Facebook social network “unthink” launches to public. http://techcrunch.com/2011/10/25/anti-facebook-social-network-unthink-launches-to-public/, accessed 2012-04-20

Prall L (2010) Sixdegrees.com – Social networking in its infancy. http://blog.afridesign.com/2010/09/sixdegrees-com-social-networking-in-its-infancy, accessed 2012-04-20

Raacke J, Bonds-Raacke J (2008) MySpace and Facebook: Applying the uses and gratifications theory to exploring friend-networking sites. Cyberpsychology & Behavior 11(2):169-174

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

Ridings C, Wasko MM (2010) Online discussion group sustainability: Investigating the interplay between structural dynamics and social dynamics over time. Journal of the Association for Information Systems 11(2):95-121

Schneider F, Feldmann A, Krishnamurthy B, Willinger W (2009) Understanding online social network usage from a network perspective. In: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, Chicago, IL, USA, pp. 35-48

Schnettler S (2009) A structured overview of 50 years of small-world research. Social Networks 31(3):165-178

Shapiro C, Varian HR (1999) Information rules: A strategic guide to the network economy. Harvard Business School Press, Boston, MA, USA

Shirky C (2003) People on page: YASNS, http://many.corante.com/archives/2003/05/12/people_on_page_yasns.php, accessed 2012-04-20

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

Strufe T (2010) Profile popularity in a business-oriented online social network. In: Proceedings of the Workshop on Social Network Systems, Paris, France, article 2

Stutzmann F (2006) An evaluation of identity sharing behavior in social network communities. Journal of the International Digital Media and Arts Association 3(1):10-18

Thadani DR, Cheung CMK (2011) Exploring the role of online social network dependency in habit formation. In: Proceedings of the 32nd International Conference on Information Systems, Shanghai, China, paper 34

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

ULD (2011) Dialog with Facebook does not hinder data protection enforcement. https://www.datenschutzzentrum.de/presse/20110930-facebook-enforce-privacy.html, accessed 2012-04-20

Vascellaro J, Steel E, Adams R (2011) News Corp. sells MySpace for a song. http://online.wsj.com/article/SB10001424052702304584004576415932273770852.html, accessed 2012-04-20

Viswanath B, Mislove A, Cha M, Gummadi KP (2009) On the evolution of user interaction in Facebook. In: Proceedings of the 2nd ACM Workshop on Online Social Networks, Barcelona, Spain, pp. 37-42

vom Brocke J, Richter D, Riemer K (2009) Motives for using social network sites (SNSs) – An analysis of SNS adoption among students. In: Proceedings of the 22nd Bled eConference, Bled, Slovenia, paper 40

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

Watts DJ (2004) The “new” science of networks. Annual Review of Sociology 30: 243-270

Watts DJ, Strogatz SH (1998) Collective dynamics of “small-world” networks. Nature 393:440-442

Wen C, Tan BCY, Chang KT-T (2009) Advertising effectiveness on social network sites: An investigation of tie strength, endorser expertise and product type on consumer purchase intention. In: Proceedings of the 30th International Conference on Information Systems, Phoenix, AZ, USA, paper 151

Williamson DA (2010) Social network demographics and usage. http://www.emarketer.com/ Reports/ All/Emarketer_2000644.aspx, accessed 2012-04-20

Willinger W, Rejaie R, Torkjazi M, Valafar M, Maggioni M (2010) Research on online social networks: Time to face the real challenges. ACM SIGMETRICS Performance Evaluation Review 37(3):49-54

Wilson C, Boe B, Sala A, Puttaswamy KPN, Zhao BY (2009) User interactions in social networks and their implications. In: Proceedings of the 4th ACM European Conference on Computer Systems, Nuremberg, Germany, pp. 205-218

Xiang R, Neville J, Rogati M (2010) Modeling relationship strength in online social networks. In: Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, pp. 981-990

Xu Y, Lu X, Goh KY, Jiang Z, Zhu X (2009) The impact of online social network on consumer loyalty: An empirical study of an online dining community. In: Proceedings of the 30th International Conference on Information Systems, Phoenix, AZ, USA, paper 17

Xu Y, Zhang C, Xue L, Yeo LL (2008) Product adoption in online social network. In: Proceedings of the 29th International Conference on Information Systems, Paris, France, paper 200

Yu B, Chen M, Kwok L (2011) Toward predicting popularity of social marketing messages. Lecture Notes in Computer Science 6589:317-324

Zhao S, Grasmuck S, Martin J (2008) Identity construction on Facebook: Digital empowerment in anchored relationships. Computers in Human Behavior 24(5):1816-1836

III Influential Users in Online Social Networks

In this chapter, influential users and approaches for their identification in OSN are researched in detail. As outlined in chapter I and chapter II, the connected customer can exercise considerable influence on other existing or potential customers, particularly in the business areas marketing and sales. By promoting brands, endorsing the purchase of products and services, or by advising against them, even the value of firms can be increased or decreased (Hogan et al. 2003, p. 196; Nitzan and Libai 2011, p. 24 f.; Weinberg and Berger 2011, p. 328). Therefore, the three research papers presented in the following intend to contribute to a better understanding of the influence of customer-to-customer interactions in OSN and to develop approaches that actually allow for identifying the most influential users in OSN, who can be targeted to initiate and control the diffusion process of user-generated content, such as electronic word-of-mouth (Bonchi et al. 2011, p. 21; Hinz et al. 2011, p. 55; Libai et al. 2010, p. 271).

The first research paper within this chapter entitled “Who will lead and who will follow: Identifying Influential Users in Online Social Networks - A Critical Review and Future Research Directions” (section III.1) presents an overview of fundamental research on social influence, influential people, and their identification in social networks before the rise of OSN. On that basis, the current state of the art on the identification of influential users in OSN is analyzed and synthesized. Finally, a research agenda is postulated.

In the second research paper presented in this chapter named “Identifying Key Users in Online Social Networks: A PageRank Based Approach” (section III.2) a novel approach bringing together concepts and findings from research on users’ connectivity and communication activity is developed. With users’ retention as evaluation criterion, the approach is evaluated regarding its applicability and practical utility by using a publicly available dataset of Facebook.

Finally, the third research paper included in this chapter called “Predicting Users’ Future Level of Communication Activity in Online Social Networks: A First Step towards More Advertising Effectiveness” (section III.3) focuses on forecasting users’ communication activity.

For the references cited in this introduction to chapter III please see the list of references provided within chapter I.

III.1 Research Paper 2: "Who will lead and who will follow: Identifying Influential Users in Online Social Networks - A Critical Review and Future Research Directions"

Authors:	Florian Probst, Laura Grosswiele, Regina Pfleger FIM Research Center, Department of Information Systems Engineering & Financial Management (Prof. Dr. Hans Ulrich Buhl), University of Augsburg, Germany florian.probst@wiwi.uni-augsburg.de laura.grosswiele@wiwi.uni-augsburg.de regina.pfleger@wiwi.uni-augsburg.de

Accepted with minor revisions for:	Business & Information Systems Engineering (BISE)

Abstract:

Along with the explosive growth of the phenomenon Online Social Networks (OSN), identifying influential users in OSN 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. Moreover, we outline two major research streams. It becomes apparent that most marketing-oriented articles draw on real-world datasets of OSN, while rather technical-oriented papers have a more theoretical approach and mostly evaluate their artifacts by 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.

III.1.1 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 in 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 very large part of the network should 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 (Hinz et al. 2011, p. 55 f.; Iyengar et al. 2011b, p. 195).

Traditionally, self-designation, that is, people report 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 datasets 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 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 that was not available before and can now be leveraged for marketing purposes (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 data and the consequently required scalability of existing approaches for the identification of influential people are not trivial (cf. e.g., Watts 2004). Second, several sources of bias might confound the identification of influential users (cf. section III.1.2.1). Third, findings from research on viral marketing and the identification of influential users 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; 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), “[…] where, to a certain extent, networking is the main preoccupation” (Beer 2008, p. 518). Thus, we do not incorporate research on content-oriented sites such as YouTube, Twitter, or Flickr (Pallis et al. 2011) that exhibit some features of OSN but rather are microblogging sites or content communities with different characteristics than OSN (Richter et al. 2011, p. 90; Smith et al. 2012, p. 103). Based on this definition, we aim at informing two particular perspectives (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 the approaches.

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 the identification of influential people in social networks and delineate three research questions: (1) How are influential users characterized in the context of OSN? (2) Which approaches have been developed and applied for the identification of influential users in OSN? (3) How have these approaches been evaluated and which implications have been derived? In section III.1.3, we outline the procedure of our structured literature search. In the subsequent section III.1.4, 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 III.1.5. Thereby, also an audience from practice, who adopt approaches for the identification of influential users, can benefit. Finally, in section III.1.6 we draw an overall conclusion and explicate limitations.

III.1.2 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 allows for reaching 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 needs to be at work, (2) there actually need 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. Thereby, we also derive our research questions that are addressed in the subsequent structured literature review.

III.1.2.1 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 for 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 ground-breaking 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, it generally needs to be distinguished 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 like 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 reducing search efforts and uncertainty (cf. e.g., Iyengar et al. 2011a). Third, normative pressures might lead to discomfort when not adopting a new product or innovation, 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 hypotheses 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 rising 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; Manski 2000; Moffitt 2001). Therefore, recent studies have been controlling 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). Taken together, even though also heterogeneity and several other factors play an important role in the diffusion process, the presence of social influence could be confirmed and is generally acknowledged today (Iyengar et al. 2011a).

III.1.2.2 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 upon the fact 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 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 (Eccleston and Griseri 2008, p. 595; Gladwell 2000, p. 70). 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 (Eccleston and Griseri 2008; Gladwell 2000). 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 contextual 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 III.1.2.1) 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, moderate 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 of how influential people can be characterized 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?

III.1.2.3 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 based on its structural characteristics (cf. e.g., Bampo et al. 2008; Borgatti 2006, p. 21; Kiss and Bichler 2008) (cf. third factor that characterizes influential people, section III.1.2.2). Structural characteristics are thereby 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 (Wasserman and Faust 1994; Watts 2004).

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, several 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., Hinz and Spann 2008; Rayport 1996). 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 lend a node more 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). Taken together, 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).

The second research stream on 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. 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 in 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.

Taken together, the first major research stream on 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 outlined within the introduction, these findings may not be transferred 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?

III.1.3 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 through (Webster and Watson 2002). Second, the precise *search strategy* needs to be defined, that is, relevant search terms, search fields, and an appropriate time period (Cooper 1998; Levy and Ellis 2006). Finally, we outline the (number of) included and excluded articles and the selection procedure to allow for comprehensibility (vom Brocke et al. 2009).

III.1.3.1 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, IS, or Computer Science, we conducted an extensive query in quality scholarly literature databases (cf. Table III.1-1) (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).

III.1.3.2 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 terminology 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.com launched in 1997 (Boyd and Ellison 2007), we chose a six-teen year time period for our search spanning from 1997 to 2012.

Table III.1-1. Summary of the underlying sources and the search systematic

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

III.1.3.3 Search Results

In order to determine the relevant articles with respect to our research questions (cf. section III.1.2), all search results have been screened by at least two authors. Only such articles have been selected, that in essence provide a clear proposition on how influential users can be identified. Thereby, also 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 query 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 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. Thereby, we excluded articles that were obviously not subject to some kind of formalized peer-review or quality verification (Levy and Ellis 2006, p. 185). Besides journals, also conferences[III.1-1] were considered (Webster and

[III.1-1] If workshop or conference papers were identified that have been published also in a journal, only the journal article has been considered when in essence the key findings remained the same.

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), and professional magazines, newspapers, or patents were excluded (Levy and Ellis 2006, p. 185). As the field of research on OSN is quite young (Richter et al. 2011, p. 89), we also excluded books, as methods and theories need some time to be established and verified before being generally accepted. By this means, we obtained 12 mere 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 content analysis.

III.1.4 Findings and Critical Discussion

In the following, we analyze the relevant articles with respect to the delineated 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 for OSN (Libai et al. 2010). As outlined in section III.1.2.2, 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 communication than others (Weimann et al. 2007, p. 175). Second, users in OSN like Facebook have up to several hundred of 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 for 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 III.1-2 illustrates the findings.

Table III.1-2. 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)	○	○	●	○
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*** ◑

Overall, the majority of the relevant articles relies on rather broad definitions of influential users or stays imprecise about which characteristics are taken into account. Surprisingly, two factors (“who one is” and “what one knows”) are hardly considered, although 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 people in an OSN. In summary, we observe that current approaches barely consider user specific attributes as well as users’ knowledge on certain topics.

After the synthesis of how influential users are characterized within our set of articles, we examine the articles with respect to the proposed methods along with their evaluation and implications 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 have been derived?

With respect to the two outlined major research streams (cf. section III.1.2.3), six of the relevant articles apply approaches that are generally based on the strategic location of nodes in a graph (cf. Table III.1-3). 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 were links between users do not represent friendship connections but the activity of 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 centralities generally adopt earlier due to their large number of connections to other users. Furthermore, a user’s 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 (who compromise the rest) 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 undirected and weighted graph, influential users are identified by means of high rankings among all users’ PageRank scores. The authors apply their approach to a Facebook dataset 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 dataset, 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 dataset (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 eigenvalue centrality the number of identified influential neighborhoods and users can be increased by applying PCC. The authors further find that the tendency of eigenvalue 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 score and the highest activity index calculated as the 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 higher number of new adopter than when addressing mediocrities (Kim and Han 2009). Table III.1-3 summarizes the approaches and findings.

Table III.1-3. Articles focusing on the strategic location of users in Online Social Networks

References	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** dataset 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 customer networks. 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 second 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** dataset 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 III.1-3), another six of all relevant articles focus on solving the influence maximization problem (top-*k* nodes problem) by different approximation algorithms (cf. Table III.1-4). In contrast to the former ones, 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 III.1.2.3) 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 III.1.2.2) than 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 performs often 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 on the basis of an appropriately defined cooperative game. The authors show that their algorithm can not only solve the top-*k* nodes problem investigated in all articles displayed in Table III.1-4 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) differs as well from the previously discussed approaches. Instead of using a SIR model, the authors model diffusion by a heat diffusion process. Thus, the approach can not only capture users that diffuse positive information but also 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. Besides Ma et al. (2008), also Saito et al. (2012) take into account the time factor. Therefore, the authors apply a susceptible/infected/susceptible (SIS) model and define a final-time and an integral-time maximization problem. While the first problem cares only about 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 nodes 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 III.1-4 summarizes the approaches and findings.

Table III.1-4. Articles focusing on the solution of the influence maximization problem

References	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 under the greedy hill climbing algorithm on the basis of **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** dataset 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 the **Shaply value** based influential nodes (SPIN) algorithm on the basis of 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 nodes** can be discovered than by approaches based on **centrality measures** and that the **identified influential users differ remarkably** depending on the chosen 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** dataset and show that the graph built by “trust” and “review-rate” includes more information on the social network. Thus, a **degree-centrality-based algorithm performs** often 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 a CRLF 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) propose 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. Taken together, 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) as well 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 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 dataset, 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 login time). By applying their approach to a synthetic and MySpace dataset, the authors find that influential users that might have been missed 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 have actually 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 III.1-5 summarizes the approaches and findings.

Table III.1-5. Articles focusing on further approaches

References	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 **susceptible decreases with age**, susceptibility **increases** with increasing **relationship commitment** until marriage, **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 nodes. Differentiate between "**sources**" and "**targets**". Apply their approach to a **YouTube** and **Flickr** dataset to show that it is robust and effective. Find that their approach allows to **identify more targets** and **sources** than when applying in-degree centrality.

Eirinaki et al. (2012)	Propose to identify influential nodes by **selecting and combining** a set of **profile-based characteristics** representing popularity and activity. Apply their approach to a **synthetic** and **MySpace** dataset. Find that their approach allows for identifying influential users that might have been missed 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 nodes 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 have actually 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.

III.1.5 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 III.1.2.1). 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. 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 steam either from the scientific Business & Information Systems Engineering (BISE) or Marketing community. Taken together with our findings presented in section III.1.4, it becomes apparent that rather marketing-oriented articles extensively draw on rich real-world datasets of OSN and even collaborate with OSN providers (cf. e.g., Trusov et al. 2010). In contrast, rather technical-oriented papers from the field of Computer Science and Engineering have a more theoretical approach and evaluate their artifacts in most cases by formal proofs, for instance regarding efficiency, run-time, or in a few cases apply synthetic 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.

A human being is not just a node in a graph.

The majority of the articles do neither incorporate personal information on users that allow for assessing “who one is” or “what one knows” (cf. Table III.1-2). For instance, Trusov et al. (2010, p. 645) and Hinz et al. (2011, p. 68) find that having many friends (i.e., social links) does not make users influential per se. 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 influence 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 influence can hardly be found. A first study by Aral and Walker (2012) finds that 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, influence is often overestimated, as homophily actually accounts for a large share of social contagion (cf. section III.1.2.3). 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.

Not just positive information might be propagated.

Besides the article by Ma et al. (2008) (cf. Table III.1-4), 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 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 or doubtless spread positive information.

The one who leads might not follow.

Most of the discussed approaches (cf. section III.1.4) 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 susceptibles hypothesis is for instance well represented in theoretical threshold-based models (cf. section III.1.2.3), which are also used by some of the approaches discussed in section III.1.4 (cf. Table III.1-4). However, besides Aral and Walker (2012) and partly Canali and Lancellotti (2012), none of the discussed articles analyzes the role of susceptibles in depth. Particularly behind the backdrop of the findings of Aral and Walker (2012) outlined in section III.1.4, it still seems to be promising for practitioners to address influential users in OSN, but further research is needed to enrich our understanding of the role of susceptibles 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. 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 III.1-4). 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 III.1-3). This finding is also verified by Zhang et al. (2010), who show that degree centrality-based algorithms perform often 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). Also Tang and Yang (2010) find in a similar context that a simple degree centrality based algorithm performs almost as good 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 decreases with a rising number of contacts (Katona et al. 2011; Narayan et al. 2011). Moreover, some articles indicate that users with high degree centrality scores do not have higher conversion rates due to a higher persuasiveness but are rather more active (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) for the diffusion of viral marketing campaigns, which 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 it is hardly the case that there is only a single influential neighborhood 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. However, since not all studies confirm the assumptions 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 is crucial.

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 III.1-4), 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 simpler centrality measures are often favorable, even though more sophisticated algorithms might be more accurate (cf. e.g., Zhang et al. 2011). Taken together, practitioners and researchers need to be aware of the trade-off between high accuracy and sufficient efficiency for large-scale datasets of OSN. Further research could address questions of optimal levels of accuracy and efficiency from an economical perspective when identifying influential users for marketing purposes in OSN.

III.1.6 Conclusion

Who will lead and who will follow? The question of identifying those people that mobilize and propagate influence in networks and society 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, identifying influential users in OSN received a great deal of attention 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 subject to limitations: First, despite 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, by our focus on user-oriented sites 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. Additionally, the focus on influential users in OSN could be broadened in the future 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 get a first overview and better understanding of the body of knowledge regarding the identification of influential users in OSN. Additionally, we hope to provide directions for future research in this field.

References (Section III.1)

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, PA, USA, 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: Proceedings of the 19th European Conference on Information Systems, Helsinki, Finland, paper 221

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

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, IA, USA

Beal GM, Bohlen JM (1957) The diffusion process. Cooperative Extension Service Report 18. Ames, IA, USA

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: Proceedings of the 3rd International Workshop on Internet and Network Economics, San Diego, CA, USA, 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, MA, USA

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: Proceedings of the 9th International Conference on Electronic Commerce, Minneapolis, MN, USA, pp. 351-360

Cha M, Haddadi H, Benevenuto F, Gummadi KP (2010) Measuring user influence in Twitter: The million follower fallacy. In: Proceedings of the 4th Annual Conference on Weblogs and Social Media, Barcelona, Spain, pp. 10-17

Chen W, Wang Y, Yang S (2009) Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 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, NY, USA

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

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: Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 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. Facebook. http://newsroom.fb.com/image library/downloadmedia.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, MI, USA, 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, United Kingdom

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, CA, USA. 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: Proceedings of the 31st International Conference of Information Systems, St. Louis, MO, USA, paper 79

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

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, CA, USA

Ilyas MU, Radha H (2011) Identifying influential nodes in online social networks using principal component centrality. In: Proceedings of the International Conference on Communications, Budapest, Hungary, 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, PA, USA

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, NJ, USA

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: Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 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: International Conference on Advances in Social Network Analysis and Mining, Athens, Greece, pp. 41-46

Kimura M, Saito K, Nakano R (2007) Extracting influential nodes for information diffusion on a social network. In: Proceedings of the 22nd National Conference on Artificial intelligence, Vancouver, BC, Canada, 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, China, pp. 1358-1363

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

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

Lerman K, Ghosh R (2010) Information contagion: An empirical study of the spread of news on Digg and Twitter social networks. In: Proceedings of the 4th International AAAI Conference on Weblogs and Social Media, Washington D.C., USA, 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: Proceeding of the 17th ACM Conference on Information and Knowledge Management, Napa Valley, CA, USA, 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, NJ, USA

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

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

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. New Directions in Web Data Management 331:213-234

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, MA, USA, 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: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, AB, Canada, 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, NY, USA

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: Proceedings of the 1st Asian Conference on Machine Learning: Advances in Machine Learning, Nanjing, China, 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, United Kingdom

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, IL, USA, 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: Proceedings of the IEEE International Conference on Intelligence and Security Informatics, Vancouver, BC, Canada, 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, NJ, USA

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, PA, USA

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, MA, USA 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: Proceedings of the 17th European Conference on Information Systems, Verona, Italy, paper 161

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

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

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

Zhang Y, Wang Z, Xia C (2010) Identifying key users for targeted marketing by mining online social network. In: Proceedings of the 24th International Conference on Advanced Information Networking and Applications, Perth, WA, Australia, pp. 644-649

Zhang Y, Zhou J, Cheng J (2011) Preference-based top-k influential nodes mining in social networks. In: Proceedings of the 10th International Conference on Trust, Security and Privacy in Computing and Communications, Changsha, China, pp. 1512-1518

III.2 Research Paper 3: “Identifying Key Users in Online Social Networks: A PageRank Based Approach”

Authors*:	Dr. Julia Heidemann[a], Dr. Mathias Klier[b], Florian Probst[a] [a] FIM Research Center, Department of Information Systems Engineering & Financial Management (Prof. Dr. Hans Ulrich Buhl), University of Augsburg, Germany julia.heidemann@wiwi.uni-augsburg.de florian.probst@wiwi.uni-augsburg.de [b] Department of Information Systems, Production and Logistics Management, University of Innsbruck, Austria mathias.klier@uibk.ac.at
Published in:	Proceedings of the 31st International Conference on Information Systems, St Louis, MO, USA, paper 79

Abstract:

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

* At the time of writing this paper, the authors have been affiliated with these institutions.

III.2.1 Introduction

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

Despite the rising number of users, the purchase prices for OSN are also being considered critically. For instance, Martin Sorrell, CEO of the WWP Group, seriously questioned the valuation of Facebook.com at US$ 15 billion (Andrews 2009). In fact, OSN face the question how to leverage on their fast growing popularity to achieve sustainable revenues. For example, many OSN are not sure how to generate adequate revenues through advertising and membership fees (Clemons 2009; Lu and Hsiao 2010). This is critical, since nowadays the majority of OSN relies on the advertisement based and/or the two-tiered business model, the latter meaning that basic services are offered for free and premium services are provided for a fee (Riggins 2003). Particularly these business models pose major challenges to OSN providers: On the one hand, more effective advertising strategies are needed in order to remain financially viable (Wen et al. 2009). Even though worldwide advertisement spending on OSN are expected to grow from US$ 2.0 billion in 2008 to US$ 3.5 billion in 2013

(Williamson 2009a), OSN often do not know how to unleash this potential. Consequently, there are already indicators for unexpected low advertising sales (Delany et al. 2008). MySpace.com for instance, recently "[…] has fallen 'significantly' short of expectations and is jeopardising a critical US$ 900 million [...] agreement with Google" (Edgecliffe-Johnson and Li 2009). On the other hand, OSN need to foster users' retention, i.e., they need to ensure that users don't leave the OSN or become inactive, since "[…] retention of users and virality are crucial to growth and survival of large online social networks" (Nazir et al. 2009, p. 65). Especially for OSN operating under the two-tiered business model, acquiring and retaining users that are willing to pay fees for premiums services is essential.

To overcome these challenges and to tap the enormous potential originated by the dramatic increase in the popularity of OSN, key users play a decisive role (Bampo et al. 2008; Xu et al. 2008; Xu et al. 2009). In our context, a key user is characterized by one or more of the following aspects: (1) He or she can affect a large number of his or her friends, acquaintances, or other users in an OSN. Such a user can for instance be addressed in marketing campaigns to achieve a high awareness of a product or service (Zahng et al. 2010). This strategy is very promising, since Ray et al. (2010) found "[…] that people in the US generate more than 500 billion online impressions *on each other* regarding products and services" and that only "[…] 16% of online consumers generate 80% of these impressions". (2) He or she is very unlikely to leave the OSN or to become inactive. Such a loyal user can also be helpful to increase stickiness, i.e., the ability to attract and hold users' interest (Bhat et al. 2002), which is for instance an important success factor for web-based advertisement (Wang and Fesenmaier 2006). (3) He or she is more likely to be willing to pay for premium services in an OSN, which are provided for a fee. Such a user is particularly interesting for OSN operating under the two-tiered business model. To enable more effective and user centric advertising strategies as well as sophisticated customer loyalty programs by addressing users deliberately, approaches for the identification of such key users in OSN are needed. For the identification of key users, users' connectivity and communication activity are particularly important regarding advertisement in OSN (Cheung and Lee 2010; Ganley and Lampe 2009; Staab et al. 2005; Wen et al. 2009;

Xu et al. 2008), users' loyalty (Algesheimer and von Wangenheim 2006; Xu et al. 2009), and users' willingness to pay for services in OSN (Oestreicher-Singer and Zalmanson 2009). However, even though studies emphasize the importance of both a user's connectivity *and* communication activity (Ganley and Lampe 2009; Staab et al. 2005; de Valck et al. 2009; Willinger et al. 2009), quantitative approaches for the identification of key users in OSN merging both aspects are missing. Therefore, we propose a novel PageRank based approach for identifying key users in OSN bringing together concepts and findings from both research streams. In addition, we demonstrate the practical applicability by using a publicly available dataset of Facebook.com and evaluate our novel PageRank based approach in comparison to existing approaches, which could also be used to identify key users in OSN.

The paper is based on the design science research paradigm and in particular on the guidelines for conducting design science research by Hevner et al. (2004). Since Hevner et al. (2004) do not propose an approach for structuring and organizing design science research contributions, we follow Peffers et al. (2008) and their nominal process model for the conduct of design science research, which is based on the guidelines by Hevner et al. (2004) and contains six activities. Hence, after the discussion of the general relevance of the problem and its motivation within this introduction (activity 1: *"problem identification and motivation"*), we specify the problem context for which the novel approach is relevant and review prior research on users' connectivity and communication activity in OSN. Afterwards, we identify the research gap (activity 2: *"define the objectives for a solution"*). In section III.2.3, we develop our artifact as a novel PageRank based approach for the identification of key users in OSN (activity 3: *"design and development"*). The penultimate section illustrates the applicability of the artifact (activity 4: *"demonstration"*) by using a publicly available dataset of Facebook.com. Furthermore, the artifact's utility (activity 5: *"evaluation"*) is extensively assessed in comparison to *"competing artifacts"*. Finally, the last section summarizes our results and provides an outlook on future steps (activity 6: *"communication"*).

III.2.2 Problem Context and Related Work

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

III.2.2.1 Problem Context

Boyd and Ellison (2007, p. 211) define OSN as "[…] web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system". Aroused by the web 2.0 boom, OSN have evolved into a mass medium, where users present themselves to a broad public and establish or maintain connections to other users. Hence, OSN provide a basis for "[…] maintaining social relationships, for finding users with similar interests, and for locating content and knowledge that has been contributed or endorsed by other users" (Mislove et al. 2007, p. 29). Particularly the aspect of networking, i.e., establishing and maintaining connections between users, plays a decisive role. Thereby, the visibility and searchability of the users' social network of friends, or at least acquaintances, is a distinctive feature of OSN. Thus, OSN can "[…] create substantial value for the individuals who participate in them, the organizations that sponsor them, and the larger society in multiple ways" (Agarwal et al. 2008, p. 243). However, the majority of OSN relying on advertisement based and/or two-tiered business models face the challenge to tap the enormous potential originated by the dramatic increase in the popularity of OSN in order to generate sustainable revenues (Clemons 2009; Lu and Hsiao 2010). Therefore, approaches for the identification of key users in OSN are needed to enable for instance more effective advertising strategies (e.g., viral marketing campaigns, targeted marketing) and sophisticated customer loyalty programs by addressing users deliberately (Bampo et al. 2008; Xu et al. 2008; Xu et al. 2009). In this context, literature indicates that particularly users' connectivity based on social structures in the network and users' communication activity are essential (Ganley and Lampe 2009; Kiss and Bichler 2008; Staab et al. 2005; de Valck et al. 2009; Willinger et al. 2009).

III.2.2.2 Users' Connectivity in Online Social Networks

Users' connectivity in OSN is primarily based on the structural characteristics of the network, i.e., patterns of connections among users (cf. Oinas-Kukkonen et al. 2010). Prior research suggests that a user's connectivity plays a decisive role for the identification of key users in OSN. Wen et al. (2009) for instance point out that a user's connectivity in the whole network could be a significant factor that may impact advertising effectiveness in OSN. This is underpinned by further studies, which illustrate that well-connected users, i.e., users with many direct and indirect connections to other users, are particularly important for OSN, as they can be highly relevant for the promotion of brands, products, and viral marketing campaigns (Domingos and Richardson 2001; Kiss and Bichler 2008; Staab et al. 2005; de Valck et al. 2009). Moreover, well-connected users tend to be more loyal, as for example every additional direct or indirect connection raises a user's barrier to leave the network (Algesheimer and von Wangenheim 2006; Xu et al. 2009). Thus, a user's connectivity based on the structural characteristics of the network needs to be considered when identifying key users in OSN.

In general, structural characteristics have been extensively studied for instance to understand and explain human behavior in multiple social networks (Monge and Contractor 2003; Nohria and Eccles 1992; Shapiro and Varian 1999). Thereby, particularly interesting elements in the context of OSN include social capital (Burt 1992; Granovetter 1974) and embeddedness (Saxenian 1994; Uzzi 1997). The structure invoked by the binary connections among users in OSN is mostly perceived as a set of nodes (users), and a set of undirected edges (ties or in the following social links) connecting pairs of nodes (Adamic and Adar 2003; Bampo et al. 2008). These nodes and undirected edges determining the network structure can be represented by a graph (Wasserman and Faust 1994), as shown in Figure III.2-1.

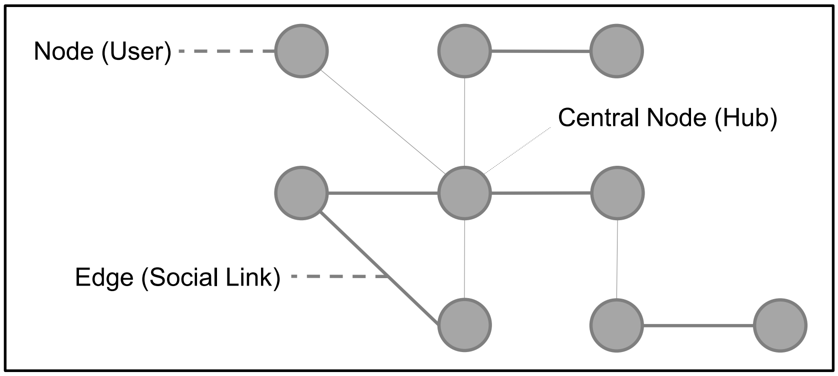

Figure III.2-1. Visualization of the network structure (social graph)

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

Approaches for the identification of important nodes that consider direct and/or indirect connections in networks can be found not only in social network analysis, but also in many other fields for instance in biology for the identification of genes (e.g., Özgür et al. 2008) or in scientometrics for the ranking of scientific journals (e.g., Bollen et al. 2006). These approaches' interpretations highly depend on the particular context (Borgatti 2005; Borgatti and Everett 2006; Freeman et al. 1980). For the specific context of social networks, several measures have been suggested to identify influential and prestigious nodes (Bonacich 1972; 1987; Scott 2000; Wasserman and Faust 1994). Additional measures indicate the social influence of nodes on other nodes in a network (Friedkin 1991) or assess a node's integration into a network (Valente and

Foreman 1998). The three most common centrality measures to quantify the centrality of a certain node in social networks are presented in Freeman's article "Centrality in Social Networks: Conceptual Clarification" (Freeman 1979): Degree centrality, closeness centrality, and betweenness centrality. The first centrality measure called degree centrality represents the simplest instantiation of centrality, assuming that a node with many direct connections to other nodes is central to the network. The second measure named closeness centrality expands the definition of degree centrality by focusing on how close a node is to all other nodes in the network. The idea behind the third measure referred to as betweenness centrality is that if a node is more often on the shortest paths between other nodes, it is more central to the network. A fourth popular centrality measure, namely eigenvector centrality, is proposed by Bonacich (1972). Eigenvector centrality extends the logic of degree and closeness centrality, since a node's connectivity in the whole network is incorporated (Bolland 1988). Thus, eigenvector centrality tries to quantify the centrality of a node in terms of the global or overall structure of the network, and pays less attention to local patterns (Hanneman and Riddle 2005). To calculate the centralities of the nodes in the network, eigenvector centrality uses the primary eigenvector of a graph's adjacency matrix (Rodriguez 2008). Thereby, the adjacency matrix represents, which nodes of the graph are adjacent, i.e., connected by an edge (the formal representation of a graph's adjacency matrix can be found in section III.2.3). For a detailed description of how to calculate eigenvector centrality and the primary eigenvector see for instance Kiss and Bichler (2008) or Newman (2003b). The primary eigenvector has been applied extensively to rank nodes in all types of networks. It has been used for instance for the ranking of web pages (Brin and Page 1998; Kleinberg 1998; Xing and Ghorbani 2004) and to evaluate the influence of scientific journals (Bollen et al. 2006; Pinski and Narin 1976), articles, and authors (Ding et al. 2009; Liu et al. 2008). These approaches acknowledge explicitly that not all connections are equal, as connections to nodes that are themselves influential are assumed to lend a node more influence than connections to less influential nodes (Newman 2003b). Therefore, the concept underlying eigenvector centrality qualifies particularly for the quantification of a user's connectivity in OSN. Thus, approaches based on the primary eigenvector can be conducive to the identification of key users in OSN.

III.2.2.3 Users' Communication Activity in Online Social Networks

Latest studies show that not only the structural characteristics underlying a user's connectivity, but also the user's communication activity, i.e., the exchange of information for instance via messages or wall posts, is highly relevant for advertising effectiveness, a user's loyalty, and a user's willingness to pay for services in OSN (Cheung and Lee 2010; Ganley and Lampe 2009; Oestreicher-Singer and Zalmanson 2009). Hence, users' communication activity among each other plays an important role for the identification of key users in OSN. Prior research emphasizes the importance of users' communication activity: "No matter what resources are available within a structure, without communication activity those resources will remain dormant, and no benefits will be provided for individuals" (Butler 2001, p. 350). Ridings and Wasko (2010) further illustrate, how users' retention in online discussion groups increases as communication activity rises. Moreover, recent work in the context of OSN indicates that the value of OSN lies in the communication activity between users (Krasnova et al. 2009; Willinger et al. 2009). Xu et al. (2008, p. 14) for instance emphasize "[…] that interaction information is invaluable to marketers, more important than the static links". Thus, a user's communication activity should be considered when identifying key users in OSN.

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

In the activity graph of an OSN all edges between nodes are the same, regardless whether the corresponding users have a strong connection (i.e., interact frequently) or a weak connection (i.e., interact infrequently). However, literature states that there may be stronger and weaker connections between users in social networks (Newman 2004) and in OSN particularly (Gilbert and Karahalios 2009; Kahanda and Neville 2009; Wen et al. 2009; Xiang et al. 2010). In general, strong connections between users are for instance more likely to be activated for information flow and more influential (Brown and Reingen 1987). In contrast, weak connections provide people with access to information and resources beyond those available in their social circle (Granovetter 1973; 1983) and bridge cliques of strong connections (Constant et al. 1996). Further studies emphasize that the strength of connections facilitates awareness in the context of electronic referrals (de Bruyn and Lilien 2008) and that the influence of a reference group and word of mouth recommendations strongly depends on the strength of connections (de Valck et al. 2009). In the context of OSN, for instance Wen et al. (2009, p. 2) conclude that the strength of connections "[…] denotes an irresistible element for [...] advertising". Nevertheless, previous work on activity graphs in OSN does often not distinguish between strong and weak connections and leaves exploration of this facet to future work (Nazir et al. 2008; Viswanath et al. 2009; Wilson et al. 2009). Only a few authors consider the strength of connections based on users' activity when identifying important nodes in customer networks (Kiss and Bichler 2008) or when comparing structural characteristics of social graphs and weighted activity graphs in OSN (Chun et al. 2008). In order to distinguish between strong and weak connections, these studies started to examine each connection's communication activity level. In this context, communication activity can be any sort of interaction among users facilitated by methods provided by OSN, for example messages or wall posts (cf. Schneider et al. 2009). Since almost every OSN provides such infrastructure for communication and transfer of information, the record of communication activities between users can be used to identify which activity link can be considered as strong and weak, respectively (Xiang et al. 2010). Thus, the strength of a user's activity link can be a measure of intensity, duration, intimacy, or exchange of information between users (Barrat et al. 2004; Granovetter 1973). Furthermore, in accordance to the above mentioned findings from research on users' connectivity, Benevenuto et al. (2009, p. 50) discovered that users do not only interact with directly connected users, but also have significant exposure to users "[…] that are 2 or more hops away […]". Therefore, not only a user's communication activity represented by the activity graph but also the strength of a user's direct and indirect activity links based on each activity link's communication activity level should be incorporated when identifying key users in OSN.

III.2.2.4 Research Gap

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

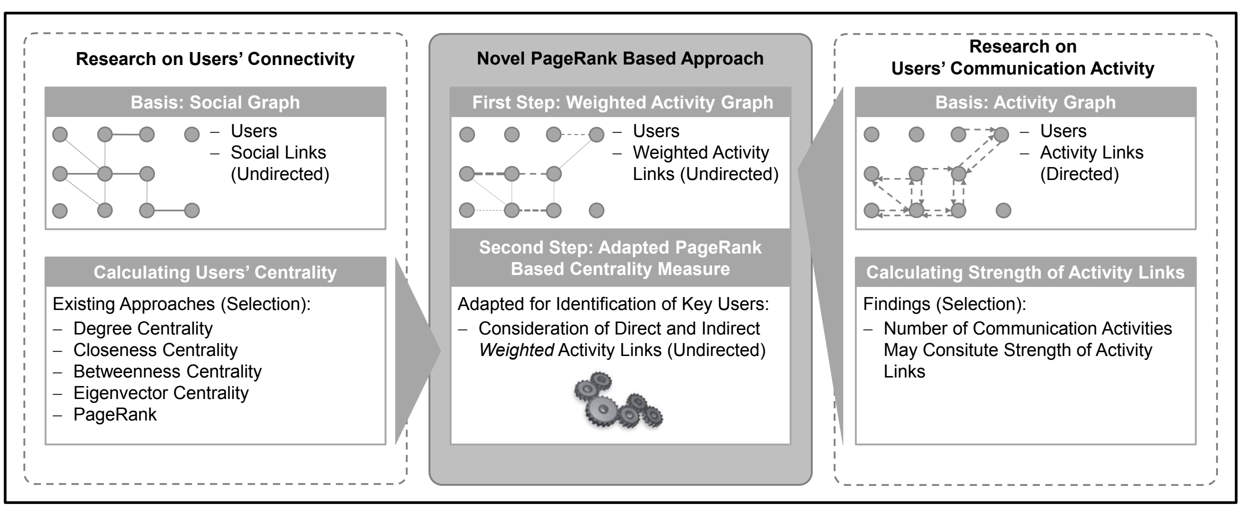

Figure III.2-2. Novel PageRank based approach

III.2.3 Novel PageRank Based Approach

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

III.2.3.1 First Step: Deriving the Weighted Activity Graph

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

First of all, we define the activity graph as a graph that is based on users who actually communicate with each other instead on users who are connected by a static social link (cf. Chun et al. 2008; Nazir et al. 2008; Wilson et al. 2009). In the activity graph, a node represents a user and an edge (activity link) represents communication activity (e.g., a wall post, a message) between a pair of users. Thus, the activity graph differs from the social graph, as inactive social links are not considered in the activity graph. However, users who are not connected by a social link in the social graph can be connected by an activity link in the activity graph, if there has been communication activity between this pair of users. An example in Figure III.2-3 highlights the possible differences between a social graph and an activity graph.

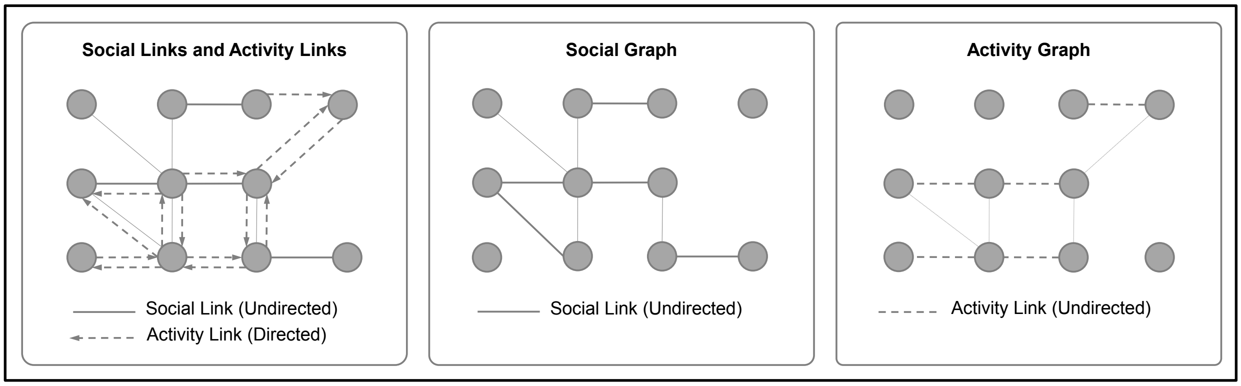

Figure III.2-3. Example: social graph vs. activity graph

For the identification of key users in OSN, we need to adapt the basic concept of activity graphs. As illustrated in the left picture of Figure III.2-3, activity links are usually assumed to be directed, since communication activity needs an initiator and a receiver. However, the direction of influence (e.g., word of mouth or peer pressure) through communication activity, which can lead to higher advertising effectiveness, users' loyalty, and users' willingness to pay for services in OSN, can be bidirectional. Theoretically, this influence can be classified according to social influence literature as informational social influence and normative social influence (Deutsch and Gerard 1955). While informational social influence means that users rely on information provided by others, normative social influence describes the pressure or assumed need to align the own attitude with that of some other valued users (Bass 1969; Kraut et al. 1998; Wen et al. 2009). In the special case of OSN however, it is hard to tell if the initiator or the receiver of communication activity is more likely to be affected by each type of social influence. For instance, a user who writes a message on another user's wall can either point attention to a brand, product, or service himself or he or she can be influenced by an advertisement placed on the other user's profile (e.g., the user is member of a brand community, i.e., he or she declares himself as a fan of a certain brand). Or a user who receives a lot of messages can be more loyal and likely to stay in a network in almost the same manner than a user who sends a lot of messages. Thus, we model communication activity as undirected activity links to cover bidirectional social influence. Moreover, since pairs of users usually perform reciprocal communication activity, modeling undirected activity links represents to a great extent users' communication behavior in OSN (Chun et al. 2008; Wilson et al. 2009). For instance, for 65% of the users in the largest Facebook.com regional networks all interactions via wall posts were reciprocated (Wilson et al. 2009). Therefore, the loss of information by modeling undirected activity links is limited and the advantages of a bidirectional interpretability of social influence prevail. Hence, we model the activity graph for the identification of key users in OSN by using undirected activity links.

Formally we define the activity graph according to graph theory as $G = (V, E)$, where V denotes a set of nodes (users) and E a set of undirected edges (activity links) (cf. Albert and Barabási 2002; Wassermann and Faust 1994). Thereby, $|V| = n$ represents the number of users in the OSN and $|E| = m$ the number of undirected activity links between them. Two nodes i and j are called adjacent, if and only if they are connected by an activity link $\{i, j\} \in E$. Thus, the activity graph can be represented by its symmetric adjacency matrix $A = (a_{ij}) \in \{0; 1\}^{n \times n}$, whose elements take the value 1 if an undirected activity link connects the nodes i and j, and 0 otherwise. Furthermore, we let t (with $t = 1, 2, ...$) determine a window of time, during which at least once communication activity between two nodes i and j must have occurred in order to create an activity link between them. Thereby, t denotes the number of periods (e.g., days) counted backwards from the point in time when the activity graph is constructed. To account for stronger and weaker activity links (cf. section III.2.2), we further extend our activity graph to include weights of the undirected activity links. Thereby, c_{ij} (with $c_{ij} = 0, 1, ...$) denotes the number of communication activities initiated by node i and received by node j during the time interval stipulated by t (Chun et al. 2008; Onnela et al. 2007). Respectively, c_{ji} (with $c_{ji} = 0, 1, ...$) constitutes the number of communication activities initiated by node j and received by node i. Thus, we define the weight w_{ij} of an undirected activity link between two users i and j as the number of communication activities between that pair of users:

$$w_{ij} = c_{ij} + c_{ji}. \qquad \text{(III.2-1)}$$

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

Thus, in contrast to the activity graph, our weighted activity graph does not only contain binary information about whether communication activity occurred at least once between two users i and j during the time interval stipulated by t (existence of an activity link), but also indicates the strength of activity links between users (cf. w_{ij} in formula (III.2-1)). Based on this definition, the weighted activity graph derived in the first step provides the basis for the second step of our novel PageRank based approach towards the identification of key users in OSN.

III.2.3.2 Second Step: Determining Users’ Centrality Scores

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

For the determination of users’ centrality scores based on users’ connectivity in the weighted activity graph, we consider particularly approaches based on the primary eigenvector of a graph’s adjacency matrix. These approaches acknowledge explicitly that not all connections are equal (cf. section III.2.2). Connections to nodes that are themselves influential are rather assumed to lend a node more influence than connections to less influential nodes (Newman 2003b). Since the nodes’ connectivity in the whole network is incorporated (Bolland 1988), approaches based on the primary eigenvector try to find well-connected nodes in terms of the global or overall structure of the network, and pay less attention to local patterns (Hanneman and Riddle 2005). Thus, these approaches qualify particularly to rank nodes in a network. Consequently, approaches based on the primary eigenvector of a graph’s adjacency matrix have been applied extensively to calculate centrality scores in all types of networks. In single-relational networks, i.e., networks with a data structure that can only represent a single type of relationship, such as social links or undirected activity links, the primary eigenvector can be computed using the power method. Thereby, the power method simulates the behavior of random walkers traversing the network. Hence, the nodes that have a higher probability of being traversed are the most central or important nodes in the network and gain consequently a higher centrality score (Brandes and Erlebach 2005). Single-relational networks can result in different types of graphs. First, there can be strongly connected, aperiodic graphs, i.e., graphs that contain paths from all nodes to all other nodes, whose lengths are sufficiently long (Kemeny and Snell 1960). In this type of graphs, for instance eigenvector centrality can be used to rank nodes (Bonacich 1987). However, graphs as our weighted activity graph *G*' do not certainly fulfill these properties, as they are not always strongly connected or are even periodic, i.e., there exist isolated nodes (cf. Figure III.2-3). For this second type of graphs, the network's topology can be altered, such that a "teleportation network" is overlaid with the graph *G*' to construct an irreducible and aperiodic network (Rodriguez 2008). This "teleportation network" introduces an artificial activity link with equal weights between all possible pairs of nodes, even if they are not connected according to our weighted activity graph *G*'. Thus, when there exists a non-zero probability of "teleportation" to every node in *V*', the network becomes strongly connected (cf. Rodriguez 2008). This idea was introduced by Brin and Page (1998) who developed the well-known random web surfer model of the PageRank algorithm to rank web pages in the World Wide Web (WWW) (Brin and Page 1998; Page et al. 1999). PageRank interprets the web pages as nodes and directed edges represent the links between them. Thus, PageRank uses the link structure of the WWW as an indicator of an individual web page's importance relative to other web pages by interpreting a link from web page *A* to web page *B* as a vote by web page *A* for web page *B*. Following Langville and Meyer (2004), the PageRank *PR*(*i*) for a web page *i* can be defined as:

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

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

Due to its characteristics, the general concept of PageRank seems to be appropriate regarding the identification of key users in OSN. However, for our context we need to adapt the PageRank formula by two modifications. First, a general difference between the WWW and our weighted activity graph in OSN exists. While links in the WWW are directed, the activity links in our weighted activity graph are considered to be undirected. To account for this distinction when identifying key users in OSN, we have to adapt the original PageRank formula accordingly by substituting the set B_i (set of web pages connected to i by ingoing, i.e., directed links) by a set F_i, which represents a set of users connected to i by undirected activity links. The second modification concerns the activity links' weights. A reduction of the activity links' weights to binary values as in the original PageRank formula would entail a severe loss of information (Newman 2004). We therefore have to define a modification of PageRank, which considers the undirected activity links' weights. Our second modification is based on an adaption of the original PageRank's assumption, that a node transfers its centrality evenly to all the web pages it connects to (cf. Xing and Ghorbani 2004). However, in our weighed activity graph the distribution should be determined by the level of communication activity between user i and the users it connects to (cf. section III.2.2). Therefore, we need to consider the weights w_{ij} of each undirected activity link as defined in formula (III.2-1). Thus, we remove the dominator O_j and the undirected activity link's weight w_{ij} is added to account for strong and weak connections among users. Finally, we define the formula of our adapted PageRank based centrality measure $S(i)$ for a user i as:

$$S(i) = \frac{(1-d)}{N} + d \cdot \sum_{j \in F_i} S(j) \cdot w_{ij} \ , \qquad \text{(III.2-3)}$$

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

III.2.4 Demonstration and Evaluation

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

[III.2-1] Erratum: The formula has been stated wrongly in the original publication. The correct formula is: $S(i) = \frac{(1-d)}{N} + d \cdot \sum_{j \in F_i} \frac{S(j) \cdot w_{ij}}{\sum_{k \in F_j} w_{jk}}$.

III.2.4.1 Facebook.com New Orleans Network Dataset

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

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

III.2.4.2 Characteristics of the Facebook.com New Orleans Network Dataset

To validate that the New Orleans Network dataset exhibits the OSN specific characteristics, we examine the social graph as well as the activity graph and compare them to graphs used in prior research on OSN. For that purpose, we draw on the social graph derived from the first and the activity graph derived from the second part of the dataset. As described in the previous section, the social graph consists of 63,731 users connected by 817,090 undirected social links. To analyze whether the social graph is characteristic of an OSN, we determine the average path length, the average clustering coefficient, and the assortativity coefficient. Table III.2-1 provides an overview of the social graph's statistics compared to social graphs from prior research on OSN.

Table III.2-1. Comparison of social graphs' statistics

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

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

To derive the corresponding activity graph, we use the wall posts contained in the second part of the dataset, which represent the most popular form of communication activity between users in OSN (Benevenuto et al. 2009). As described before, the social graph was crawled during December 29, 2008 and January 3, 2009. For our activity graph, we use a fraction of 832,277 wall posts spanning from September 14, 2004 to January 3, 2009. Thus, the end of the considered period of communication activity equals the date when the crawl of the underlying network structure ended. The remainder of 44,410 wall posts spanning from January 4, 2009 to January 22, 2009 were written and received after the social structure was crawled. In the section after next, we evaluate our novel PageRank based approach in comparison to alternative approaches for the identification of key users in OSN, which are based on the social graph. Since we do not want to discriminate these approaches, we do not consider the remainder of wall posts for our activity graph. The activity graph $G = (V, E)$ contains the same set of users V as the social graph, with $|V| = 63{,}731$ (cf. Figure III.2-3 for an example). These users are connected by a set of undirected activity links E, with $|E| = 171{,}711$. Thereby, an undirected activity link between a user A and a user B exists if and only if the users A and B interacted during September 14, 2004 and January 3, 2009 at least one time directly with each other, in a way that a wall post was initiated by user A and received by user B, or vice versa. 6,392 (3.7%) of these undirected activity links in our activity graph do not have a corresponding social link in the social graph. This equals to 191,980 (23.1%) wall posts exchanged via these activity links. This finding is in line with prior research on users communication activity in OSN. Benevenuto et al. (2009) for instance discovered that 22.0% of users' wall posts in their Orkut.com dataset were exchanged between users, which were not connected by a social link in the social graph. To further examine the activity graph, we determine again the average path length (5.39), the average clustering coefficient

(0.109), and the assortativity coefficient (0.220). The activity graph's statistics are in line with the little prior research on activity graphs in OSN. Wilson et al. (2009) for instance display average path lengths in the range of 5.00 to 7.00, average clustering coefficients between 0.030 and 0.080, and assortativity coefficients around 0.200. Chun et al. (2008) show similar properties and comparable correlations between their social graph's and activity graph's measurements. To sum it up, the activity graph derived from our New Orleans Network dataset is in line with prior studies on activity graphs in OSN. Since both the social and the activity graph exhibit the OSN specific characteristics, the New Orleans Network dataset provides a sound basis for the demonstration and evaluation of our novel PageRank based approach for the identification of key users in OSN.

III.2.4.3 Demonstration of the Novel PageRank Based Approach

We demonstrate the applicability of our novel PageRank based approach developed in section III.2.3 by using the New Orleans Network dataset. Thereby, we conduct the two major steps of the approach. In the first step, we derive the weighted activity graph as a basis for the identification of key users in the network. In the second step, we determine each user's centrality score in terms of his or her connectivity in the weighted activity graph. Hence, we apply the PageRank based centrality measure developed in the previous section. Sorting users by their centrality scores in descending order allows us to define a ranking of key users based on the New Orleans Network dataset.

First, we build the weighted activity graph on the basis of the collected dataset of the Facebook.com New Orleans Network. Therefore, we use the activity graph derived in the previous section and extend it to include weights for the undirected activity links. Consequently, the weighted activity graph contains 63,731 users, which are connected by 171,711 undirected activity links. Since the activity graph is based on wall posts spanning from September 14, 2004 to January 3, 2009, we set the parameter t of our weighted activity graph to t = 1,573 days. Thus, all wall posts initiated and received by the 63,731 users in the activity graph during that period of time are covered. To calculate each undirected activity link's weight w_{ij}, we apply formula (III.2-1). Thereby, c_{ij} (respectively c_{ji}) denotes the number of wall posts initiated by user i and received by user *j* (respectively initiated by user *j* and received by user *i*) between September 14, 2004 and January 3, 2009. We represent our weighted activity graph as a symmetric adjacency matrix, where $A' = (a'_{ij})^{63,731 \times 63,731}$, with $a'_{ij} = \begin{cases} w_{ij} & \text{if } \{i, j\} \in E \\ 0 & \text{otherwise} \end{cases}$. Based on the weighted activity graph, we calculate the centrality score $S(i)$ of each user *i* in the second step.

Therefore, we apply the PageRank based centrality measure defined in formula (III.2-3). For that purpose, we need to choose the dampening factor *d* first. When *d* takes a value close to 1, the measure places greater emphasis on the structure of the weighted activity graph and less on the teleportation network modeled in the first part of formula (III.2-3). However, higher values of *d* slow down the convergence of the power method (Langville and Meyer 2004). Moreover, Boldi et al. (2005) provided a mathematical analysis of different values for *d*, finding that values close to 1 do not give a more meaningful ranking than other high damping factors. Pretto (2002) further found that when *d* changes, the top section of the ranking changes only slightly. As we are especially interested in users with high centrality scores, i.e., the top section, the impact of the dampening factor's choice is limited. Thus, we set the dampening factor to $d = 0.85$. This value is favorable in terms of computational performance and is also often considered as the default value for PageRank calculations in literature (cf. Langville and Meyer 2004). Finally, we calculate the centrality scores $S(i)$ applying the PageRank based centrality measure. For that purpose, we use the software package "NetworkX" for the exploration and analysis of networks and network algorithms (cf. Hagberg et al. 2008). In conclusion, we derive a centrality score $S(i)$ for every user *i* included in our weighted activity graph. By sorting these centrality scores in descending order, we receive a ranking of users. Based on this ranking of all users included in the New Orleans Network dataset, the key users in the network can be identified by choosing a designated top segment of the ranking.

III.2.4.4 Evaluation of the Novel PageRank Based Approach

Building on the ranking of identified key users in the network, we evaluate our novel PageRank based approach. As we highlighted in the introduction, in our context the term key user stands for users who can affect a large number of other users in terms of marketing, users who are unlikely to leave an OSN or to become inactive, and/or users who are more likely to be willing to pay for premium services in an OSN. Here, we use users' retention as evaluation criterion, since in particular "[…] retention of users […] [is] crucial to growth and survival of large online social networks" (Nazir et al. 2009, p. 65). Thereby, we define that a user is retained, if he or she stays active in the network. A user's retention strongly affects the retention of other users in the network, since every additional connection raises users' barrier to leave the network (Algesheimer and von Wangenheim, 2006). In addition, retained users are particularly valuable, as they support a sense of familiarity and community (Figallo 1998; Hagel III and Armstrong 1997; Wellman and Gulia 1999). Gan et al. (2009, p. 14) further illustrate, that as individuals become more involved in online communities, their "[…] habit effect strengthens". Thus, users who are continuously retained, have a higher probability to remain involved "[…] as participation becomes more automatic" (Gan et al. 2009, p. 14). Finally, retained users are particularly important for OSN providers, since they can only leverage users, for instance for targeted marketing or premium services, if they stay active in the network.

Based on users' retention, we compare our novel PageRank based approach to existing approaches, which could also be used to identify key users in OSN. This comparison to alternative approaches, so-called *"competing artifacts"*, is integral to design science research (Hevner et al. 2004, p. 100). For our context, we consider the common centrality measures degree centrality, closeness centrality, and betweenness centrality. We do not employ eigenvector centrality, since graphs as social graphs and activity graphs are usually not connected and aperiodic graphs, as required for the calculation of eigenvector centrality (cf. section III.2.3). Even though the same holds true for closeness centrality, we computed closeness centrality for each connected part of the graphs separately for comparison reasons. However, the results indicate the bias when identifying key users based on closeness centrality in not connected and aperiodic graphs. Applying the common centrality measures to the social graph derived from the New Orleans Network dataset allows us to identify key users based on their connectivity in the network as it is common practice in social network analysis. Hence, we first evaluate our novel PageRank based approach in comparison to the application of common centrality measures to the social graph (evaluation step 1). However, the application of common centrality measures to the social graph derived from the New Orleans Network dataset focuses solely on users' connectivity, but does not incorporate users' communication activity. Second, we evaluate our novel PageRank based approach in comparison to an approach, which is solely based on users' prior communication activity in the network, but does not incorporate users' connectivity (evaluation step 2). So far, we consider existing approaches taking either users' connectivity or users' communication activity into account. However, in contrast to the common centrality measures applied to the social graph and users' prior communication activity, our novel PageRank based approach merges concepts from research on users' connectivity and users' communication activity. Even though existing approaches do not incorporate both aspects, we finally compare our novel PageRank based approach to the common centrality measures applied to the activity graph of the New Orleans Network dataset (evaluation step 3). Thus, we extend our evaluation by approaches, which are also based on both users' connectivity and users' communication activity.

As basis for our three evaluation steps, we use the fraction of wall posts in the New Orleans Network dataset spanning from January 4, 2009 to January 22, 2009 to determine users' retention. Thereby, following Java et al. (2007) and Kolari et al. (2007), we consider a user retained, if he or she wrote at least one wall post during this period. For evaluation step 1, we calculate the common centrality measures degree centrality, closeness centrality, and betweenness centrality for every user in the social graph of the New Orleans Network dataset by using the software package "NetworkX". For each common centrality measure, we are therefore able to rank users based on each user's corresponding centrality score, which represents his or her connectivity in the social graph. For evaluation step 2, we determine users' prior communication activity. Therefore, we draw on users' wall posts between September 14, 2004 and January 3, 2009. Thus, we rank users solely based on the number of wall posts, i.e., their prior communication activity. Gan et al. (2009, p. 7) refer to this cumulative number as "rank" in order to determine status in the context of online communities. To compare our novel PageRank based approach to the alternative approaches for the identification of key users in OSN (evaluation steps 1-3), we use a method, which has been similarly applied in biology to evaluate competing approaches for the identification of genes (Özgür et al. 2008). Thereby, we create top segments of *u* identified key users in every ranking, which has been either derived by applying our novel PageRank based approach, by the common centrality measures applied to the social graph (evaluation step 1), by users' prior communication activity (evaluation step 2), and by the common centrality measures applied to the activity graph (evaluation step 3). Afterwards, we compare the percentages of retained users in these segments to evaluate how many identified key users were actually retained. In Table III.2-2, we display segments of top *u* identified key users and the corresponding percentages of actually retained users for the common centrality measures applied to the social graph (evaluation step 1) and for the ranking by users' prior communication activity (evaluation step 2). Table III.2-2 highlights that by applying our novel PageRank based approach, 92% of the top 100 identified key users were actually retained. However, the application of the common centrality measures applied to the social graph leads to 48% retained top 100 identified key users for degree centrality, 43% for closeness centrality, and 54% for betweenness centrality. Ranking users solely based on users' prior communication activity resulted in 90% retained top 100 identified key users. Thus, Table III.2-2 illustrates that our PageRank based approach leads to better results for all top segments of identified key users than the common centrality measures applied to the social graph (evaluation step 1). Furthermore, also the percentages of retained identified key users compared to the ranking solely based on users' prior communication activity are higher for every top segment (evaluation step 2).

Table III.2-2. Percentages of the actually retained top *u* identified key users (common centrality measures applied to social graph)

Top *u* Identified Key Users	PageRank Based Approach	Degree Centrality	Closeness Centrality	Betweenness Centrality	Prior Communication Activity
100	92%	48%	43%	54%	90%
500	87%	61%	55%	60%	84%
637 (1%)	82%	62%	55%	58%	80%
1000	86%	62%	55%	58%	83%
6373 (10%)	65%	51%	48%	50%	61%

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

Table III.2-3. Percentages of the actually retained top *u* identified key users (common centrality measures applied to activity graph)

Top *u* Identified Key Users	PageRank Based Approach	Degree Centrality	Closeness Centrality	Betweenness Centrality	Prior Communication Activity
100	92%	89%	80%	83%	90%
500	87%	80%	72%	83%	84%
637 (1%)	82%	76%	68%	78%	80%
1000	86%	78%	71%	81%	83%
6373 (10%)	65%	59%	52%	62%	61%

Table III.2-3 illustrates that the common centrality measures identify more key users that are retained when they are applied to the activity graph. Nevertheless, our novel PageRank based approach still leads to better results than the common centrality measures applied to the activity graph (evaluation step 3). In order to test whether our results are significant, we ran a paired t-test. Thus, we came to the result that the novel PageRank based approach is significantly better than each of the other approaches in comparison (e.g., for the top 10% identified key users and $\alpha = 0.05$). In addition, we evaluated our novel PageRank based approach in comparison to the common centrality measures applied to the weighted activity graph derived from the New Orleans Network dataset. Therefore, we applied adapted common centrality measures, which have been extended to account for the activity links’ weights (cf. Barrat et al. 2004). However, the application of weighted degree centrality, closeness centrality, and betweenness centrality to the weighted activity graph did not lead to better results than the ones of our novel PageRank based approach.

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

III.2.4.5 Discussion and Limitations of the Novel PageRank Based Approach

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

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

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

Finally, besides the discussion on how OSN can create value, there is an ongoing debate about the privacy risks they involve (cf. e.g., Gross and Acquisti 2005;

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

III.2.5 Conclusion

OSN face the challenge to tap the enormous potential originated by the dramatic increase in the popularity of OSN in order to generate sustainable revenues. In that context, particularly more effective advertising strategies and sophisticated customer loyalty programs to foster users’ retention are necessary. Therefore, quantitative approaches for the identification of key users in OSN are needed to address users deliberately and to enable for instance more effective and user centric marketing campaigns. In this paper, we propose a novel PageRank based approach bringing together concepts and findings from research on users’ connectivity and users’ communication activity in OSN. Related to the seven guidelines for conducting design science research articulated by Hevner et al. (2004), we can summarize as follows: We propose an *“artifact”* (cf. guideline 1) that is a method in terms of a PageRank based approach, which is composed of two steps. In the first step, a weighted activity graph is derived as basis for the identification of key users in OSN. In the second step, users’ centrality scores are determined by using a novel PageRank based centrality measure. For the design of our artifact, we specified our *“problem context”* and focused on relevant literature regarding the identification of key users in OSN. Thereby, statements in literature support that the identification of key users in OSN is an *"important and relevant business problem"* (cf. guideline 2). Moreover, we reviewed prior research on users' connectivity and communication activity in OSN. Drawing on these two research streams, we identified the research gap: Quantitative approaches for the identification of key users in OSN bringing together concepts and findings from research on users' connectivity and users' communication activity were missing. Thus, we developed a novel PageRank based approach to "[…] address an important organizational problem" (Hevner et al. 2004, p. 82). We believe that our artifact contributes as a first, but essential step to overcome the challenges faced by the majority of OSN. We *"evaluated"* our novel PageRank based approach (cf. guideline 3) regarding its applicability and its practical utility by using a publicly available dataset of Facebook.com. For the evaluation, we chose users' retention as evaluation criterion and compared our novel PageRank based approach with *"competing artifacts"*, which could also be used to identify key users in OSN. Thus, we illustrated the advantages of our *"research contribution"* (cf. guideline 4), i.e., of our novel PageRank based approach. According to literature, we highlighted the importance of both users' connectivity and users' communication activity when identifying key users in OSN. We incorporated users' communication activity and the strength of users' connections in the first step of our approach by deriving a weighted activity graph. For the second step of our approach, we designed a PageRank based centrality measure to determine users' centrality scores in terms of their connectivity in the weighted activity graph. Hence, we developed a first quantitative approach for the identification of key users in OSN bringing together concepts and findings from research on users' connectivity *and* users' communication activity in OSN and addressed the research gap stated above. The evaluation based on the Facebook.com New Orleans Network dataset illustrates that the novel PageRank based approach leads to (significantly) better results regarding the retained users than all other approaches in comparison. Therefore, the proposed approach, which allows to identify key users in OSN, seems to be quite promising and may contribute to overcome current challenges of OSN (e.g., regarding their monetization by enabling more effective advertising strategies etc.). Nevertheless, future work is needed and intended to further evaluate the approach.

To support a *"rigorous"* definition and presentation of our artifact (cf. guideline 5), we denoted it formally. Thereby, we drew on Hevner et al. (2004, p. 88), who state: "[…] to be mathematically rigorous, important parts of the problem may be abstracted […]". This implicates assumptions and limitations of the approach, which were critically discussed. Future work should address these issues either by confirming our assumptions or by relaxing the assumptions when developing further approaches for the identification of key users in OSN. Furthermore, upcoming challenges, for instance due to changing privacy practices, need to be carefully observed and considered. Thus, the *"search process"* (cf. guideline 6) can be distinguished in present and future steps. In this paper we presented the initial design of a PageRank based approach for the identification of key users in OSN, which may represent a starting point for OSN to overcome the described challenges. Thereby, the design process was guided by existing literature and the identified main factors of influence, namely users' connectivity and communication activity in OSN. Certainly, we abstracted quite strongly when initially designing our novel PageRank based approach. Future iterations need to relax assumptions and particularize and enhance the artifact accordingly. We are currently collaborating with a German OSN provider to additionally analyze our approach "in depth in business" (Hevner et al. 2004, p. 86) and to extend our basic approach for the identification of key users in OSN. Regarding the *"communication"* of our results (cf. guideline 7), we chose a formal, mathematical presentation in order to be able to demonstrate and evaluate our artifact in a rigorous and unambiguous way. However, we also tried to attract a managerial audience by means of the extensive explanations of the used concepts and formulas as well as detailed description of the application and the practical utility of our novel PageRank based approach.

References (Section III.2)

Adamic LA, Adar E (2003) Friends and neighbors on the Web. Social Networks 25(3):211-230

Agarwal R, Gupta AK, Kraut R (2008) The interplay between digital and social networks. Information Systems Research 19(3):243-252

Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. Reviews of Modern Physics 74(1):47-97

Algesheimer R, von Wangenheim F (2006) A network based approach to customer equity management. Journal of Relationship Marketing 5(1):39-57

Andrews A (2009) Facebook and other social networking sites face a critical year. Telegraph. http://www.telegraph.co.uk/technology/facebook/4436417/Facebook-and-other-social-networking-sites-face-a-critical-year.html, accessed 2010-12-04

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

Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. Proceedings of the National Academy of Sciences of the United States of America 101(11):3747-3752

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

BBC (2005) News Corp in $580m Internet buy. http://news.bbc.co.uk/2/hi/business/4695495.stm, accessed 2010-12-04

Benevenuto F, Rodrigues T, Cha M, Almeida V (2009) Characterizing user behavior in online social networks. In: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, Chicago, IL, USA, pp. 49-62

Bhat S, Bevans M, Sengup S (2002) Measuring users' web activity to evaluate and enhance advertising effectiveness. Journal of Advertising 31(3):97-106

Boldi P, Santini M, Vigna S (2005) PageRank as a function of the damping factor. In: Proceedings of 14th International Conference on World Wide Web, Chiba, Japan, pp. 557-566

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

Bollen J, Rodriguez MA, van De Sompel H (2006) Journal status. Scientometrics 69(3):669-687

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

Bonacich P (1987) Power and centrality: A family of measures. American Journal of Sociology 92(5):1170-1182

Borgatti SP (2005) Centrality and network flow. Social Networks 27(1):55-71

Borgatti SP, Everett MG (2006) A graph-theoretic perspective on centrality. Social Networks 28(4):466-484

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

Brandes U, Erlebach T (2005) Network analysis: Methodological foundations. Springer, Berlin, Germany

Brin S, Page L (1998) The anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems 30(1-7):107-117

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

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

Butler BS (2001) Membership size, communication activity, and sustainability: A resource-based model of online social structures. Information Systems Research 12(4):346-362

Cheung CMK, Lee MKO (2010) A theoretical model of intentional social action in online social networks. Decision Support Systems 49(1):24-30

Chun H, Kwak H, Eom Y, Ahn Y, Moon S, Jeong H (2008) Comparison of online social relations in volume vs. interaction: A case study of Cyworld. In: Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement, Vouliagmeni, Greece, pp. 57-70

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

Constant D, Sproull L, Kiesler S (1996) The kindness of strangers: The usefulness of electronic weak ties for technical advice. Organization Science 7(2):119-135

Cummings JN, Butler B, Kraut R (2002) The quality of online social relationships. Communications of the ACM 45(7):103-108

de Bruyn A, Lilien GL (2008) A multi-stage model of word-of-mouth influence through viral marketing. International Journal of Research in Marketing 25(3):151-163

de Valck K, van Bruggen GH, Wierenga B (2009) Virtual communities: A marketing perspective. Decision Support Systems 47(3):185-203

Delaney KJ, Steel E, Vara V, Vascellaro JE (2008) Social sites don't deliver big ad gains. Wall Street Journal-Eastern Edition, February 5

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

Ding Y, Yan E, Frazho A, Caverlee J (2009) PageRank for ranking authors in co-citation networks. Journal of the American Society for Information Science and Technology 60(11):2229-2243

Domingos P, Richardson M (2001) Mining the network value of customers. In: Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, pp. 57-66

Dougherty H (2010) Facebook reaches top ranking in US, http://weblogs.hitwise.com/heather-dougherty/2010/03/facebook_reaches_top_ranking_i.html, accessed 2010-04-12

Edgecliffe-Johnson A, Li K (2009) News Corp says sluggish MySpace jeopardises $900m Google deal. Financial Times, November 5. http://www.ft.com/cms/s/0/4cf7d554-c9ab-11de-a071-00144feabdc0.html#axzz2HBvbcJ9R, accessed 2010-04-12

Facebook (2010) Six years of making connections. Facebook. http://blog.facebook.com/blog.php?post=287542162130, accessed 2010-04-12

Figallo C (1998) Hosting Web communities: Building relationships, increasing customer loyalty, and maintaining a competitive edge. John Wiley & Sons, Inc., New York, NY, USA

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

Freeman LC, Roeder D, Mulholland RR (1980) Centrality in social networks: Experimental results. Social Networks 2(2):119-141

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

Gan L, Gu B, Jarvenpaa S, Yang G (2009) Habit formation in online communities. In: Proceedings of the 30th International Conference on Information Systems, Phoenix, AZ, USA, paper 189

Ganley D, Lampe C (2009) The ties that bind: Social network principles in online communities. Decision Support Systems 47(3):266-274

Gilbert E, Karahalios K (2009) Predicting tie strength with social media. In: Proceedings of the 27th SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, pp. 211-220

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

Granovetter MS (1974) Getting a job: A study of contacts and careers. University Of Chicago Press, Chicago, IL, USA

Granovetter MS (1983) The strength of weak ties: A network theory revisited. Sociological Theory 1(1):201-233

Gross R, Acquisti A (2005) Information revelation and privacy in online social networks. In: Proceedings of the ACM Workshop on Privacy in the Electronic Society, Alexandria, VA, USA, pp. 71-80

Hagberg AA, Schult DA, Swart PJ (2008) Exploring network structure, dynamics, and function using NetworkX. In: Proceedings of the 7th Python in Science Conference, Pasadena, CA, USA, pp. 11-15

Hagel III J, Armstrong AG (1997) Net gain: Expanding markets through virtual communities. Harvard Business School Press, Boston, MA, USA

Hanneman RA, Riddle M (2005) Introduction to social network methods. University of California, Riverside, CA, USA. http://www.faculty.ucr.edu/~hanneman/nettext/, accessed 2010-08-10

Haythornthwaite C (2002) Building social networks via computer networks: Creating and sustaining distributed learning communities. In: Renninger KA, Shumar W (eds) Building virtual communities: learning and change in cyberspace. Cambridge University Press, Cambridge, United Kingdom, pp. 159-190

Hevner AR, March ST, Park J, Ram S (2004) Design science in information systems research. MIS Quarterly 28(1):75-105

Java A, Song S, Finin T, Tseng B (2007) Why we twitter: Understanding microblogging usage and communities. In: Proceedings of the 9th WebKDD and First SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis, San Jose, CA, USA, pp. 56-65

Kahanda I, Neville J (2009) Using transactional information to predict link strength in online social network. In: Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media, San Jose, CA, USA, paper 106

Kemeny JG, Snell JL (1960) Finite Markov chains. Van Nostrand Co., Inc., Princeton, NJ, USA

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

Kleinberg JM (1998) Authoritative sources in a hyperlinked environment. In: Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms, San Francisco, CA, USA, pp. 668-677

Kolari P, Finin T, Yesha Y, Yesha Y, Lyons K, Perelgut S, Hawkins J (2007) On the structure, properties and utility of internal corporate blogs. In: Proceedings of the International Conference on Weblogs and Social Media, Boulder, CO, USA, pp. 1-8

Krasnova H, Hildebrand T, Günther O (2009) Investigating the value of privacy in online social networks: Conjoint analysis. In: Proceedings of the 30th International Conference on Information Systems, Phoenix, AZ, USA, paper 173

Kraut RE, Rice R, Cool C, Fish R (1998) Varieties of social influence: The role of utility and norms in the success of a communication medium. Organizational Science 9(4):437-453

Lampe C, Ellison N, Steinfield CA (2006) A Face(book) in the crowd: Social searching vs. social browsing. In: Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work, Banff, AB, Canada, pp. 167-170

Langville AN, Meyer CD (2004) Deeper inside PageRank. Internet Mathematics 1(3):335-380

Leskovec J, Horvitz E (2008) Planetary-scale views on an instant-messaging network. In: Proceedings of the Seventeenth International Conference on World Wide Web, Beijing, China, pp. 915-924

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

Lu H-P, Hsiao K-L (2010) The influence of extro/introversion on the intention to pay for social networking sites. Information & Management 47(3):150-157

Milgram S (1967) The small world problem. Psychology Today 1(1):61-67

Mislove A, Marcon M, Gummadi KP, Druschel P, Bhattacharjee B (2007) Measurement and analysis of online social networks. In: Proceedings of the 17th ACM SIGCOMM Conference on Internet Measurement, San Diego, CA, USA, pp. 29-42

Monge PR, Contractor NS (2003) Theories of communication networks. Oxford University Press, New York, NY, USA

MSNBC (2007) Microsoft invests $240 Million in Facebook. http://www.msnbc.msn.com/id/21458486, accessed 2010-04-12

Nazir A, Raza S, Chuah C-N (2008) Unveiling Facebook: A measurement study of social network based applications. In: Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement, Vouliagmeni, Greece, pp. 43-56

Nazir A, Raza S, Gupta D, Chuah C-N, Krishnamurthy B (2009) Network level footprints of Facebook applications. In: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, Chicago, IL, USA, pp. 63-75

Newman MEJ (2002) Assortative mixing in networks. Physical Review Letters 89(20): article 208701

Newman MEJ (2003a) Mixing patterns in networks. Physical Review E 67(2): article 026126

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

Newman MEJ (2004) Analysis of weighted networks. Physical Review E 70(5): article 056131

Nohria N, Eccles RG (1992) Networks and organizations. Harvard Business School Press, Boston, MA, USA

Oestreicher-Singer G, Zalmanson L (2009) "Paying for content or paying for community?" The effect of social involvement on subscribing to media Web sites. In: Proceedings of the 30th International Conference on Information Systems, Phoenix, AZ, USA, paper 9

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

Onnela J-P, Saramäki J, Hyvönen J, Szabó G, de Menezes MA, Kaski K, Barabási A-L, Kertész J (2007) Analysis of a large-scale weighted network of one-to-one human communication. New Journal of Physics 9(6):article 179

Özgür A, Vu T, Erkan G, Radev DR (2008) Identifying gene-disease associations using centrality on a literature mined gene-interaction network. Bioinformatics 24(13): 277-285

Page L, Brin S, Motwani R, Winograd T (1999) The PageRank citation ranking: Bringing order to the Web. Technical Report. Stanford InfoLab, Stanford, CA, USA. http://ilpubs.stanford.edu:8090/422/, accessed 2010-04-12

Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S (2008) A design science research methodology for information systems research. Journal of Management Information Systems 24(3):45-77

Pinski G, Narin F (1976) Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. Information Processing and Management 12(5):297-312

Pretto L (2002) A theoretical analysis of Google's PageRank. In: Laender AH, Oliveira AL (eds) Proceedings of the 9th International Symposium on String Processing and Information Retrieval. Lecture Notes In Computer Science, 2476:131-144

Ray A, Bernoff J, Wise J (2010) Peer influence analysis - Mass influencers are the key to achieving scale in social media marketing. Forrester. http://www.forrester.com/rb/Research/peer_influence_analysis/q/id/56766/t/2, accessed 2010-04-12

Ridings C, Wasko MM (2010) Online discussion group sustainability: Investigating the interplay between structural dynamics and social dynamics over time. Journal of the Association for Information Systems 11(2):95-121

Riggins FJ (2003) Market segmentation and information development costs in a two-tiered fee-based and sponsorship-based Web site. Journal of Management Information Systems 19(3):69-81

Rodriguez MA (2008) Grammar-based random walkers in semantic networks. Knowledge-Based Systems 21(7):727-739

Saxenian AL (1994) Regional advantage: Culture and competition in Silicon Valley and Route 128. Harvard University Press, Cambridge, MA, USA

Schneider F, Feldmann A, Krishnamurthy B, Willinger W (2009) Understanding online social network usage from a network perspective. In: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, Chicago, IL, USA, pp. 35-48

Schnettler S (2009) A structured overview of 50 years of small-world research. Social Networks 31(3):165-178

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

Shapiro C, Varian HR (1999) Information rules: A Strategic guide to the network economy. Harvard Business School Press, Boston, MA, USA

Staab S, Domingos P, Mika P, Golbeck J, Ding L, Finin T, Joshi A, Nowak A, Vallacher RR (2005) Social networks applied. IEEE Intelligent Systems 20(1):80-93

Uzzi B (1997) Social structure and competition in interfirm networks: The paradox of embeddedness. Administrative Science Quarterly 42(1):35-67

Valente TW, Foreman RK (1998) Integration and radiality: Measuring the extent of an individual's connectedness and reachability in a network. Social Networks 20(1): 89-105

Viswanath B, Mislove A, Cha M, Gummadi KP (2009) On the evolution of user interaction in Facebook. In: Proceedings of the 2nd ACM Workshop on Online Social Networks, Barcelona, Spain, pp. 37-42

Wang Y, Fesenmaier DR (2006) Identifying the success factors of Web-based marketing strategy: An investigation of convention and visitors bureaus in the United States. Journal of Travel Research 44(3):239-249

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

Watts DJ, Strogatz SH (1998) Collective dynamics of "small-world" networks. Nature 393:440-442

Wellman B, Gulia M (1999) Virtual communities as communities. In: Smith MA, Kollocks P (eds) Communities in cyberspace. Routledge, New York, NY, USA, pp. 167-194

Wen C, Tan BCY, Chang KT-T (2009) Advertising effectiveness on social network sites: An investigation of tie strength, endorser expertise and product type on consumer purchase intention. In: Proceedings of the 30th International Conference on Information Systems, Phoenix, AZ, USA, paper 151

Williamson DA (2009a) Social networking worldwide: Ad spending and usage. eMarketer. http://www.emarketer.com/Reports/All/Emarketer_2000567.aspx, accessed 2010-04-12

Williamson DA (2009b) Social networks: Five consumer trends for 2009. eMarketer. http://www.emarketer.com/Report.aspx?code=emarketer_2000566, accessed 2010-04-12

Willinger W, Rejaie R, Torkjazi M, Valafar M, Maggioni M (2010) Research on online social networks: Time to face the real challenges. ACM SIGMETRICS Performance Evaluation Review 37(3):49-54

Wilson C, Boe B, Sala A, Puttaswamy KPN, Zhao BY (2009) User interactions in social networks and their implications. In: Proceedings of the 4th ACM European Conference on Computer Systems, Nuremberg, Germany, pp. 205-218

Xiang R, Neville J, Rogati M (2010) Modeling relationship strength in online social networks. In: Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, pp. 981-990

Xing W, Ghorbani A (2004) Weighted PageRank algorithm. In: Proceedings of the 2nd Annual Conference on Communication Networks and Services Research, Washington D.C., USA, pp. 305-314

Xu Y, Lu X, Goh KY, Jiang Z, Zhu X (2009) The impact of online social network on consumer loyalty: An empirical study of an online dining community. In: Proceedings of the 30th International Conference on Information Systems, Phoenix, AZ, USA, paper 17

Xu Y, Zhang C, Xue L, Yeo LL (2008) Product adoption in online social network. In: Proceedings of the 29th International Conference on Information Systems, Paris, France, paper 200

Zhang Y, Wang Z, Xia C (2010) Identifying key users for targeted marketing by mining online social network. In: Proceedings of the 24th International Conference on Advanced Information Networking and Applications, Perth, WA, Australia, pp. 644-649

III.3 Research Paper 4: "Predicting Users' Future Level of Communication Activity in Online Social Networks: A First Step towards More Advertising Effectiveness"

Author:	Florian Probst FIM Research Center, Department of Information Systems Engineering & Financial Management (Prof. Dr. Hans Ulrich Buhl), University of Augsburg, Germany florian.probst@wiwi.uni-augsburg.de

Published in:	Proceedings of the 17th Americas Conference on Information Systems, Detroit, MI, USA, paper 39

Abstract:

Many Online Social Networks do not generate sustainable revenues through advertising, even though active usage has reached enormous scales. To enable more effective advertising strategies in Online Social Networks, it is essential to identify users who can affect a large number of friends, acquaintances, or other users in the network. In this context, especially users' future level of communication activity in the Online Social Network plays an important role. A highly active past, however, does not guarantee high levels of future communication activity. Thus, approaches for the prediction of users' future level of communication activity are needed. Therefore, we transfer a probability-based method that has been primarily developed to forecast purchasing behavior of customers to the context of users' communication activity in Online Social Networks. In addition, we demonstrate the method's applicability and suitability by using a publicly available dataset of Facebook.com.

III.3.1 Introduction

In the last couple of years, Online Social Networks (OSN) have become popular Internet platforms that connect people around the globe. Thereby, the active usage of OSN has reached enormous scales: In March 2010, the OSN Facebook.com surpassed Google.com to become the most visited website of the week in the US (Dougherty 2010), while a few months later, the number of active users of Facebook.com exceeded 500 million (Facebook 2010). According to a recent study, two-thirds of the US Internet users already visit OSN each month, with 43% of them using OSN more than once a day (Alison 2010). Thus, the phenomenon OSN has evolved into a global mainstream medium that generates an increasing social and economic impact.

However, many OSN face the question of how to leverage on their fast growing popularity to achieve sustainable revenues (Heidemann et al. 2010). Nowadays, the majority of OSN relies on an advertisement-based business model (Gnyawali et al. 2010). Nevertheless, many OSN do not generate sustainable revenues through advertising: Even though the worldwide advertisement spending on OSN is expected to grow from US$ 2.0 billion in 2008 to US$ 3.5 billion in 2013 (Williamson 2009), OSN often do not know how to unleash this potential (Clemons 2009; Lu and Hsiao 2010). Therefore, more effective advertising strategies in OSN are needed in order to remain financially viable (Wen et al. 2009; Xu et al. 2008).

To tap the enormous potential originated by the dramatic increase in the popularity of OSN, particularly the identification of users who can affect a large number of friends, acquaintances, or other users in an OSN is essential (Heidemann et al. 2010; Hill et al. 2006). Such users can for example be addressed in marketing campaigns to achieve a high awareness of a product or service. This strategy is often referred to as network-based marketing, word-of-mouth marketing, or viral marketing (cf. e.g., Brown et al. 2007; Hill et al. 2006; Lee et al. 2009). The underlying key assumption is that users propagate “positive” information about a product or service after they have either been made aware by traditional marketing techniques or experienced it by themselves (Hill et al. 2006). In this context, Ray et al. (2010) found “[…] that people in the US generate more than 500 billion online impressions on each other regarding products and services”. However, as only “[…] 16% of online consumers generate 80% of these impressions” (Ray et al. 2010), only a subset of users is particularly valuable for marketers (Trusov et al. 2010). For the identification of these users, especially users’ future level of communication activity, i.e., each user’s number of future communication activities, plays an important role (Cheung and Lee 2010; Hoffman and Fodor 2010; Willinger et al. 2010; Xu et al. 2008). But even though some users might have been highly active in the past, high levels of future communication activity cannot be taken for granted (Cummings et al. 2002; Viswanath et al. 2009). Hence, approaches for the prediction of users’ future level of communication activity are needed, which might serve as a first step towards more effective advertising strategies in OSN. However, to the best of our knowledge, approaches for predicting users’ future level of communication activity in OSN are missing. Therefore, we transfer a probability-based method, which has been primarily developed by Fader et al. (2005), to forecast purchasing behavior of customers, to the context of users’ communication activity in OSN. In addition, we demonstrate the practical applicability of the method and evaluate its suitability for predicting users’ future level of communication activity in OSN by using a dataset of Facebook.com.

After the discussion of the general relevance of the problem and its motivation within this introduction, we specify the problem context and review prior research on users’ communication activity in OSN. Thereby, we identify our research gap. Afterwards, we propose our artifact as a probability-based method. In the penultimate section, we illustrate the method’s practical applicability and suitability to predict users’ future level of communication activity in OSN. Finally, we conclude with a summary of the results and an outlook on future steps.

III.3.2 Problem Context and Related work

OSN are a particular type of virtual communities[III.3-1] (Dwyer et al. 2007). According to Boyd and Ellison (2007, p. 211), we define OSN as “[…] web-based services that allow

[III.3-1] A definition of virtual communities can be found in Leimeister et al. (2004). While Dwyer et al. (2007) and Boyd and Ellison (2007) use the term Social Networking Site, we are using the term OSN throughout the paper synonymously.

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”. Current OSN are primarily used for maintaining existing relationships from an offline context, although they also allow for creating pure online relationships (Ellison et al. 2007; Lampe et al. 2006). Especially the visibility and searchability of the users’ relationships is a distinctive feature of OSN. Thus, OSN can “[…] create substantial value for the individuals who participate in them, the organizations that sponsor them, and the larger society in multiple ways” (Agarwal et al. 2008, p. 243). The majority of OSN that rely on the advertisement-based business model, however, face the challenge to tap the enormous potential originated by the dramatic increase in the popularity of OSN in order to generate sustainable revenues (Clemons 2009; Lu and Hsiao 2010). Therefore, more effective advertising strategies are needed (Wen et al. 2009; Xu et al. 2008). Word-of-mouth marketing, for example, “[…] is known to be the most effective form of advertising, but, until recently, was the most expensive […]” (Jacks and Salam 2009, p. 2). In contrast, word-of-mouth and targeted marketing can be much more efficient and cost-effective in OSN (Trusov et al. 2009; Zhang et al. 2010). However, only a small subset of users has actually significant influence on other users (Trusov et al. 2010). Thus, the identification of the most influential users is necessary to enable more effective advertising strategies in OSN (Heidemann et al. 2010; Trusov et al. 2010; Zhang et al. 2010). In this context, literature indicates that particularly the identification of users with high levels of future communication activity is essential (Cheung and Lee 2010; Hoffman and Fodor 2010; Willinger et al. 2009; Xu et al. 2008).

Communication activity in OSN can be any sort of interaction among users facilitated by methods provided by OSN, such as messages or wallposts (cf. Wilson et al. 2009). Prior work emphasizes the importance of users’ communication activity: “No matter what resources are available within a structure, without communication activity those resources will remain dormant, and no benefits will be provided for individuals” (Butler 2001, p. 350). Recent work supports that the value of OSN lies in the communication activity between users (Krasnova et al. 2009; Willinger et al. 2009). Latest studies further show that user’s communication activity is highly relevant for advertising effectiveness in OSN (e.g., Cheung and Lee 2010; Ganley and Lampe 2009). The record of communication activities between users in OSN can be used to identify users’ with high levels of communication activity in the past (Xiang et al. 2010). However, prior research found that users who have been highly active in the past are not necessarily highly active in the future (Cummings et al. 2002; Viswanath et al. 2009). Hence, approaches for the prediction of each user’s future level of communication activity need to “[…] abandon the traditional treatment of OSNs as static networks […]” (Willinger et al. 2010, p. 49) and incorporate the dynamic of communication activity in OSN.

Plenty of research addressing the dynamic nature of OSN can be found with respect to network structures in OSN. Thereby, previous studies focus on the evolution of network structures in general (for an overview cf. Dorogovtsev and Mendes 2003) and the establishment of static social links, i.e., friendship relationships, between users in particular (e.g., Liben-Nowell and Kleinberg 2007). However, Xu et al. (2008, p. 14) emphasize “[…] that interaction information is invaluable to marketers, more important than the static links”. Consequently, some studies take into account the dynamic nature of OSN and users’ communication activity: De Choudhury et al. (2007), for example, determine the intent to communicate and the communication delay between users based on several contextual factors in OSN, such as the relevance of a topic. Therefore, it is first assumed that a user receives a message. Second, the likelihood that the receiver will communicate with the sender on a particular topic and the delay in communication are predicted. However, to the best of our knowledge existing approaches do not allow for determining conditional expectations about users’ future level of communication activity in OSN on an individual-level, i.e., making predictions about each user’s future level of communication activity given information about his or her past communication activity. Thus, approaches for predicting users’ future level of communication activity in OSN are missing.

III.3.3 Method

For predicting users' future level of communication activity, we draw on a probability-based method that has been primarily developed by Fader et al. (2005) to address a very similar problem, i.e., to forecast purchasing behavior of customers. This beta-geometric/negative binomial distribution (BG/NBD)-based method goes back to the highly regarded "counting your customers" framework introduced by Schmittlein et al. (1987). In the following, we discuss the BG/NBD-based method as possibility to predict users' future level of communication activity in OSN.

First, it is assumed that a user $i \in \{1,2,...,N\}$ is active at $t_{0_i}=0$. This will generally be satisfied if we take t_{0_i} as the point of time at which the user's initial communication activity occurred. Furthermore, the BG/NBD-based method requires three pieces of information about each user, represented by $(X_i=x_i, t_{x_i}, T_i)$. Thereby, $x_i \in IN$ denotes the "frequency", i.e., the number of communication activities after the initial communication activity within the observation period $(0,T_i]$, $t_{x_i} \in IR^+$ (with $0 \leq t_{x_i} \leq T_i$) is the "recency", i.e., the point of time of the last communication activity (if $x_i=0$ then $t_{x_i}=0$), and $T_i \in IR^+$ represents the length of the observation period, i.e., the time between the initial communication activity and the end of the observation period. Notice, that the relation between the period $(0,T_i]$ and calendar time will vary from user to user depending on when the user's initial communication activity occurred. Based on this information, we aim to predict each user's future level of communication activity during the forecasting period of length t_i (cf. Figure III.3-1).

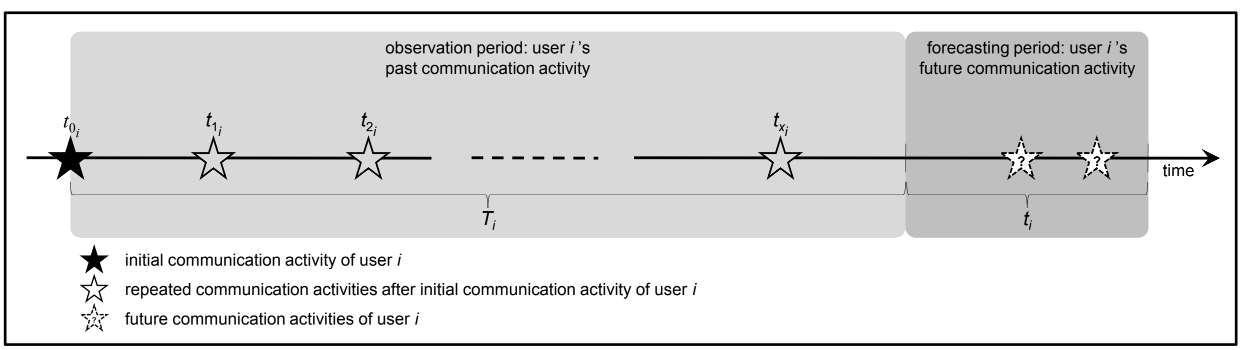

Figure III.3-1. A user's communication activity over time

To predict the future level of a user’s communication activity, assumptions about the communication activity process and the time that users stay active are needed (cf. Fader et al. 2005; Schmittlein et al. 1987):

Assumption 1 (Users’ communication activity):

While being active, each user i’s communication activity during a forecasting period of length t_i follows a homogeneous Poisson process with rate $\lambda_i \in IR^+$. This is equivalent to assuming that the time between communication activities is distributed exponentially with rate λ_i and mean $\lambda_i t_i$ (Fader et al. 2005).

Even though the Poisson assumption has a long validated history for frequently purchased consumer goods (Ehrenberg 1972) and is often used to describe aspects of human activity (Malmgren et al. 2008), the transferability to the context of online communication behavior is also critically discussed (Malmgren et al. 2009). Particularly circadian cycles, i.e., for example some people’s habit to answer e-mails only in the morning and the evening, are reported to lead to heavy-tailed power-law distributions of inter-communication times (cf. Malmgren et al. 2008). However, many studies confirm the applicability of a homogeneous Poisson process for modeling repeated communication behavior within these cycles, i.e., referring to our example stated above, while being in one session of answering e-mails (e.g., in the morning) (Malmgren et al. 2009). In a first approximation, we assume that users in OSN are in only one cycle while being active and thus a homogeneous Poisson process qualifies for modeling users’ communication activity in OSN. After each communication activity, there is a certain probability that a user becomes inactive and does not continue to communicate within the OSN:

Assumption 2 (Users’ probability of becoming inactive):

After any communication activity, user i becomes inactive with probability $p_i \in [0;1]$. The point at which the user “drops out” is distributed across communication activities according to a (shifted) geometric distribution with probability mass function:

P(inactive immediately after jth communication activity)=$p_i(1-p_i)^{j-1}$, with j=1,2,3,....

This assumption is supported by prior work on customer retention in general (cf. Fader and Hardie 2007) and the dropout process of users in OSN in particular (cf. Ahmed et al. 2010). As there are users with high and users with low levels of communication activity as well as users with high probability to drop out and vice versa, some heterogeneity assumptions are mandated. The gamma distribution is a flexible distribution and can capture the spirit of most of the reasonable distributions on λ_i and p_i (cf. Fader et al. 2005; Schmittlein et al. 1987):

Assumption 3 (Heterogeneity in users' communication activity):

Heterogeneity in λ_i follows a gamma distribution with shape parameter $r \in IR^+$, scale parameter $\alpha \in IR^+$, and probability density function:

$$f(\lambda_i|r,\alpha)=(\alpha^r\lambda_i^{r-1}e^{-\lambda_i\alpha})/\Gamma(r), \text{ with } \lambda_i>0. \qquad \text{(III.3-1)}$$

Assumption 4 (Heterogeneity in users' probability of becoming inactive):

Heterogeneity in p_i follows a beta distribution with probability density function:

$$f(p_i|a,b)=p_i^{a-1}(1-p_i)^{b-1}/B(a,b), \text{ with } 0\leq p_i\leq 1, \qquad \text{(III.3-2)}$$

where $B(a,b)$ with $a \in IR^+$ and $b \in IR^+$ is the beta function, which can be expressed in terms of gamma functions, i.e.:

$$B(a,b)=\Gamma(a)\Gamma(b)/\Gamma(a+b).$$

Finally, there is no a priori reason to favor a positive correlation between users' future communication activity and their probability to drop out over a negative correlation. On the one hand, users with high levels of communication activity may have more frequent opportunities to be disenchanted by the OSN (e.g., due to privacy concerns, system malfunctions) and drop out. On the other hand, these users are probably more strongly attached to the OSN and hence less easily disenchanted. Hence, neglecting interdependencies seems to be a reasonable first approximation:

Assumption 5 (Independence of user's communication activity and probability of becoming inactive):

The transaction rate λ_i and the dropout probability p_i vary independently across users.

With the random variable $Y(t_i)$ denoting the number of a user *i*'s future communication activities initiated in the forecasting period of length t_i, we finally aim to derive the expected number of a user’s future communication activities by computing the conditional expectation $E(Y(t_i)|X_i=x_i,t_{xi},T_i)$ for a user i with observed behavior ($X_i=x_i,t_{xi},T_i$). However, λ_i and p_i (cf. assumptions 1 and 2) are unobserved. While there is usually not enough observed user-specific behavior to reliably estimate these parameters for each user, there is generally enough information to estimate the distribution of (λ_i,p_i) over all users. Hence, we can derive the desired probabilities for a randomly chosen user by taking the expectation of the individual-level results over the mixing distributions for λ_i and p_i as given in formulas (III.3-1) and (III.3-2) (cf. Fader et al. 2005). Thus, the four BG/NBD parameters (r, α, a, b) (cf. assumptions 3 and 4) can be estimated via the method of maximum likelihood. For N users, the sample log-likelihood function is given by (cf. Fader et al. 2005):

$$LL(r,\alpha,a,b) = \sum_{i=1}^{N} \ln\left[L\left(r,\alpha,a,b \middle| X_i = x_i, t_{x_i}, T_i\right)\right], \qquad \text{(III.3-3)}$$

which can be maximized by using standard numerical optimization routines. Thereby, $L(r,\alpha,a,b|X_i=x_i,t_{xi},T_i)$ represents the likelihood function for a single user i, which can be derived according to formula (6) in Fader et al. (2005). Afterwards, we can calculate the expected number of each user’s communication activities in the forecasting period of length t_i by (for a detailed derivation of the formula cf. Fader et al. 2005):

$$E\left(Y(t_i)\middle|X_i = x_i, t_{x_i}, T_i, r, \alpha, a, b\right) = \frac{\dfrac{a+b+x_i-1}{a-1}\left[1-\left(\dfrac{\alpha+T_i}{\alpha+T_i+t}\right)^{r+x_i} {}_2F_1\left(r+x_i, b+x_i; a+b+x_i-1; \dfrac{t_i}{\alpha+T_i+t_i}\right)\right]}{1+\delta_{x_i>0}\dfrac{a}{b+x_i-1}\left(\dfrac{\alpha+T_i}{\alpha+t_{x_i}}\right)^{r+x_i}}, \qquad \text{(III.3-4)}$$

where ${}_2F_1(\cdot)$ is the Gaussian hypergeometric function, which can be closely approximated with a polynomial series, and $\delta_{xi>0}=1$ if $x_i>0$, 0 otherwise (cf. Fader et al. 2005). By calculating the expected number of each user i’s future communication activities, users’ future level of communication activity in OSN can be predicted. In the following, we demonstrate the method’s practical applicability and evaluate its suitability for predicting users’ future level of communication activity.

III.3.4 Demonstration and Evaluation

As many other OSN, Facebook.com allows users to set up personal profiles and to establish undirected social links by entering virtual friendship relationships. One of the most popular mechanisms for communication activity within OSN in general and Facebook.com in particular is a message board called “wall” that is included in each profile (Wilson et al. 2009).

For the demonstration and evaluation of the BG/NBD-based method for predicting users’ future level of communication activity, we use a publicly available dataset provided by Viswanath et al. (2009) that has also been used and described in detail in Heidemann et al. (2010). It contains 63,731 users of the Facebook.com New Orleans Network connected by 817,090 undirected social links and exhibits the OSN specific characteristics (cf. Heidemann et al. 2010). The dataset also contains information on users’ communication activity in terms of 876,687 wallposts initiated and received by the users covered by the dataset. Each wallpost includes information about the initiator, the receiver, and the time at which the wallpost was made. Overall, the wallposts span 227 weeks from September 14, 2004 to January 22, 2009. In the following, we use these wallposts to represent users’ communication activity. To account for the potential bias induced by the strong growth of the number of active users after week 175, we chose two scenarios, i.e., a low volatility scenario spanning nine weeks from week 150 to 158 (scenario 1) and a high volatility scenario spanning from week 216 to 224 (scenario 2). Figure III.3-2 displays the development of the number of wallposts and the number of distinct active users covered by the dataset over time.

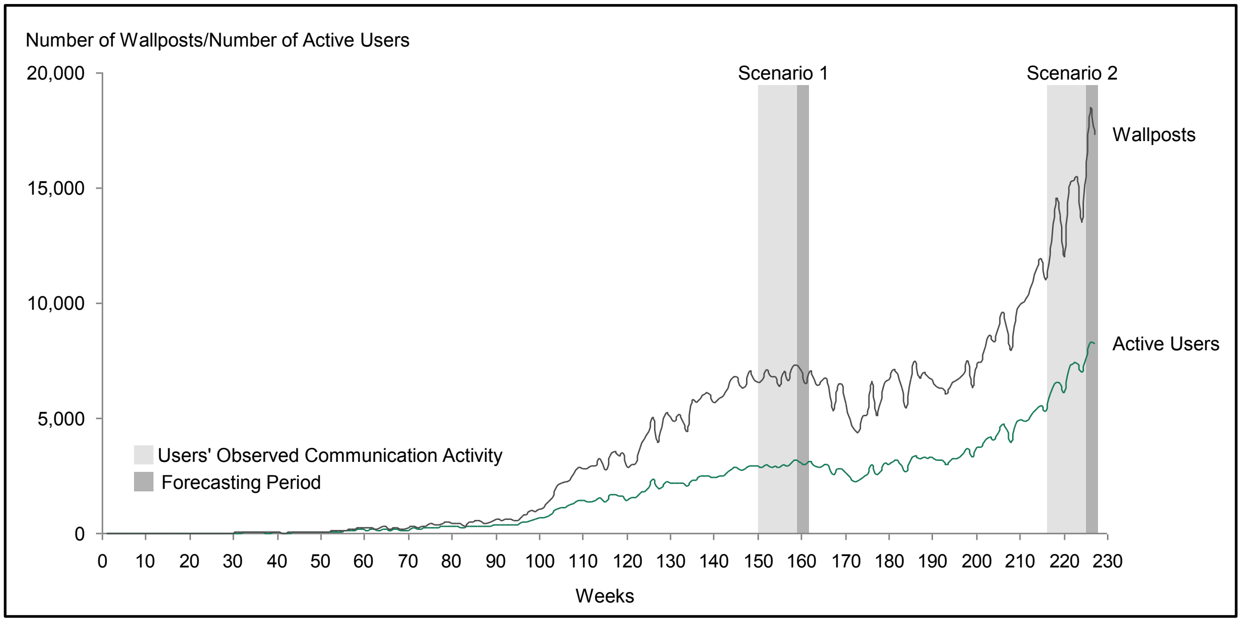

Figure III.3-2. Number of wallposts and number of active users over time

Table III.3-1 summarizes both scenarios' characteristics.

Table III.3-1. Overview of scenarios' characteristics

Characteristic	Scenario 1	Scenario 2
Total number of active users	9,815	25,182
Total number of wallposts	60,138	130,824
Average number of active users per week	2,994	6,965
Standard deviation of active users per week	100	894
Average number of wallposts per week	6,682	14,536
Standard deviation of wallposts per week	296	2,182

For each scenario, we aim to predict the users' number of wallposts during a forecasting period of three weeks, i.e., from week 159 to 161 (scenario 1) and week 225 to 227 (scenario 2), respectively. Therefore, we first derive each user *i*'s observed behavior (X_i=x_i,t_{xi},T_i) and estimate the parameters (r, α, a, b) by applying formula (III.3-3). Table III.3-2 summarizes the results.

Table III.3-2. Estimation results

Parameter	Scenario 1	Scenario 2
r	0.485	0.420
α	0.332	0.293
a	0.433	0.742
b	4.346	5.555

Second, we calculate for both scenarios the expected number of each user’s wallposts in the forecasting period according to formula (III.3-4). To evaluate the suitability of the BG/NBG-based method for predicting users’ future level of communication activity in OSN, we apply the evaluation approach suggested by Fader et al. (2005). In Figure III.3-3, we report the average of the predicted along with the average of the actual number of wallposts that took place in the forecasting period broken down by the number of wallposts in the users’ observation periods for scenario 1.

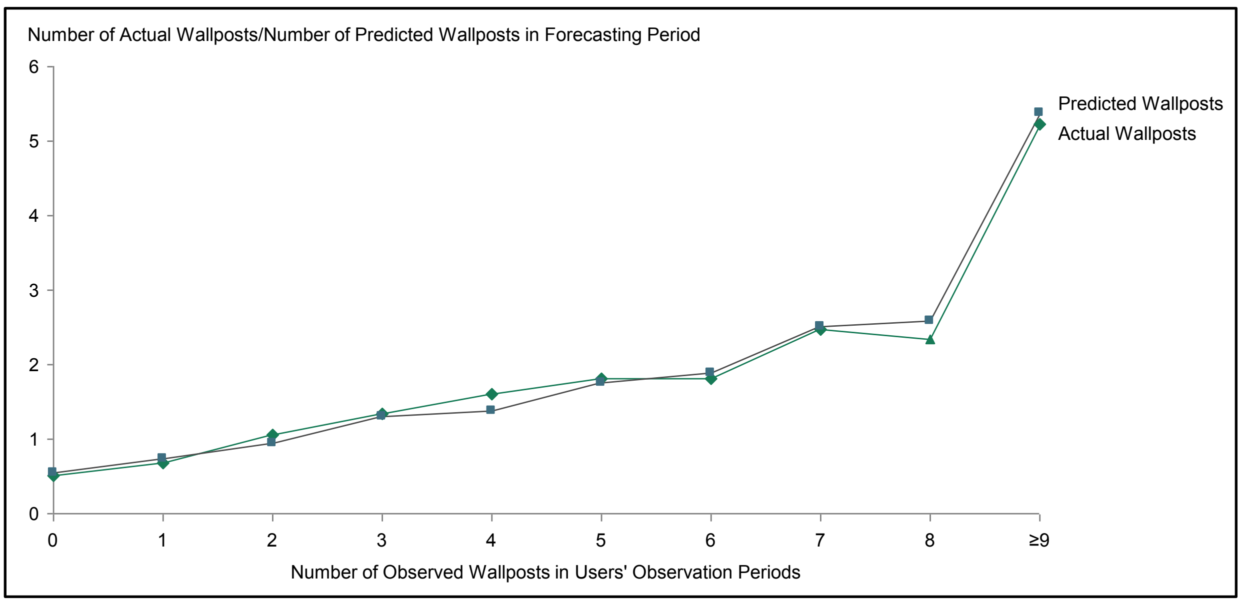

Figure III.3-3. Evaluation scenario 1

The virtually absent deviation between the users’ predicted and actual wallposts highlights that in scenario 1 the BG/NBG-based method provides excellent predictions of the users’ number of wallposts in the forecasting period. Figure III.3-4 displays the evaluation’s results for scenario 2.

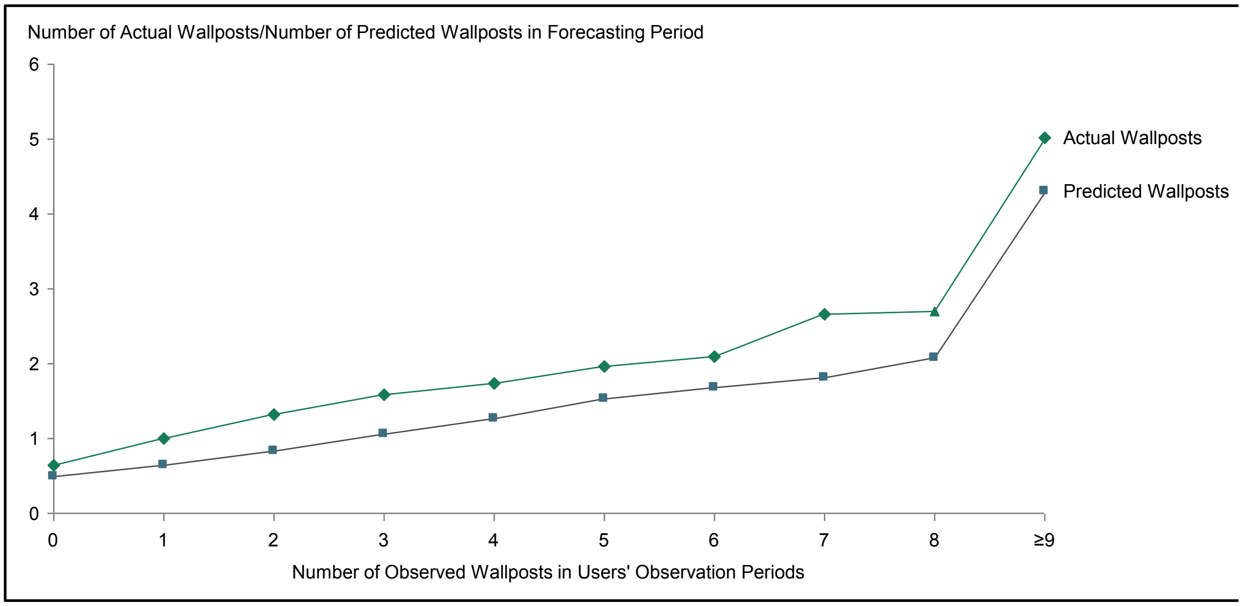

Figure III.3-4. Evaluation scenario 2

Here, the evaluation reveals that the users' actual level of communication activity is slightly underestimated. However, the deviation of 0.4 on average is quite small and constant. Thus, we think future work will be able to account for high volatility in the data by slightly adapting the method. Taken together, the BG/NBG-based method seems to be suitable for the prediction of users' future level of communication activity in OSN.

III.3.5 Conclusion

Even though active usage has reached enormous scales, the majority of OSN relying on the advertisement-based business model face the challenge of generating sustainable revenues. Particularly the identification of users with high levels of future communication activity plays an important role when developing more effective advertising strategies by addressing users deliberately. For the identification of these users, we transferred a probability-based method, which has been primarily developed to forecast purchasing behavior of customers, to the context of users' communication activity in OSN. The application and evaluation illustrated that the BG/NBG-based method seems to be suitable for predicting users' future level of communication activity in OSN. Even though the method seems to be rather complex at first sight, the necessary parameters can be derived easily and the calculation can be automated and even be done in Excel, as pointed out by Fader et al. (2005). Nevertheless, future work is needed to further evaluate the approach (e.g., regarding further sample groups), for instance by using other datasets and by considering also economic aspects. The method’s formal denotation implicates assumptions and limitations, for instance regarding the underlying probability distributions. Future work needs to address these issues either by confirming or by relaxing these assumptions. Moreover, upcoming challenges, for instance due to changing privacy practices, need to be carefully observed and considered when developing approaches for the identification of users with high levels of future communication activity. Finally, in line with prior research on the identification of influential users in OSN (e.g., Trusov et al. 2010) we did not address the question how responsive highly active users are to certain marketing strategies (e.g., word-of-mouth marketing). Even though this question is subject to future research, the BG/NBG-based method serves as a first step towards more effective advertising strategies in OSN.

References (Section III.3)

Agarwal R, Gupta AK, Kraut R (2008) The interplay between digital and social networks. Information Systems Research 19(3):243-252

Ahmed NK, Berchmans F, Neville J, Kompella R (2010) Time-based sampling of social network activity graphs. In: Proceedings of the 8th Workshop on Mining and Learning with Graphs, Washington D.C., USA, pp. 1-9

Alison D (2010) Social media up 230% since 2007. http://www.informationweek.com/news/software/web_services/showArticle.jhtml?articleID=225701600&subSection=News, accessed 2010-10-20

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

Butler BS (2001) Membership size, communication activity, and sustainability: A resource-based model of online social structures. Information Systems Research 12(4):346-362

Cheung CMK, Lee MKO (2010) A theoretical model of intentional social action in online social networks. Decision Support Systems 49(1):24-30

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

Cummings JN, Butler B, Kraut R (2002) The quality of online social relationships. Communications of the ACM 45(7):103-108

de Choudhury M, Sundaram H, John A, Seligmann DD (2007) Contextual prediction of communication flow in social networks. In: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, Silicon Valley, CA, USA, pp. 57-65

Dorogovtsev SN, Mendes JFF (2003) Evolution of networks: From biological networks to the Internet and WWW. Oxford University Press, Oxford, United Kingdom

Dougherty H (2010) Facebook reaches top ranking in US, http://weblogs.hitwise.com/heather-dougherty/2010/03/facebook_reaches_top_ranking_i.html, accessed 2010-10-20

Dwyer C, Hiltz S, Passerini K (2007) Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. In: Proceedings of the 13th Americas Conference on Information Systems, Keystone, CO, USA, paper 339

Ehrenberg ASC (1972) Repeat buying. North-Holland, Amsterdam, Netherlands

Ellison N, Steinfield CA, Lampe C (2007) The benefits of Facebook friends: Social capital and college students use of online social network sites. Journal of Computer-Mediated Communication 12(4):1143-1168

Facebook (2010) 500 million stories. Facebook. http://blog.facebook.com/blog.php?post= 409753352130, accessed 2010-10-20

Fader PS, Hardie BGS (2007) How to project customer retention. Journal of Interactive Marketing 21(1):76-90

Fader PS, Hardie BGS, Lee KL (2005) “Counting your customers” the easy way: An alternative to the Pareto/NBD model. Marketing Science 24(2):275-284

Ganley D, Lampe C (2009) The ties that bind: Social network principles in online communities. Decision Support Systems 47(3):266-274

Gnyawali DR, Fan W, Penner J (2010) Competitive actions and dynamics in the digital age: An empirical investigation of social networking firms. Information Systems Research 21(3):594-613

Heidemann J, Klier M, Probst F (2010) Identifying key users in online social networks: A PageRank based approach. In: Proceedings of the 31st International Conference on Information Systems, St Louis, MO, USA, paper 79

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

Hoffman DL, Fodor M (2010) Can you measure the ROI of your social media marketing? MIT Sloan Management Review 52(1):41-49

Jacks T, Salam AF (2009) Computer-mediated friendship networks. In: Proceedings of the 30th International Conference on Information Systems, Phoenix, AZ, USA, paper 115

Krasnova H, Hildebrand T, Günther O (2009) Investigating the value of privacy in online social networks: Conjoint analysis. In: Proceedings of the 30th International Conference on Information Systems, Phoenix, AZ, USA, paper 173

Lampe C, Ellison N, Steinfield CA (2006) A Face(book) in the crowd: Social searching vs. social browsing. In: Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work, Banff, AB, Canada, pp. 167-170

Lee J, Lee J-H, Lee D (2009) Impacts of tie characteristics on online viral diffusion. Communications of the Association for Information Systems 24(1):article 32

Leimeister JM, Sidirask P, Krcmar H (2004) Success factors of virtual communities from the perspective of members and operators: An empirical study. In: Proceedings of the 37th Hawaii International Conference on System Sciences, Big Island, HI, USA, pp. 1-10

Liben-Nowell D, Kleinberg J (2007) The link-prediction problem for social networks. Journal of the American Society for Information Science and Technology 58(7):1019-1031

Lu H-P, Hsiao K-L (2010) The influence of extro/introversion on the intention to pay for social networking sites. Information & Management 47(3):150-157

Malmgren RD, Hofman JM, Amaral LAN, Watts DJ (2009) characterizing individual communication patterns. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, pp. 607-616

Malmgren RD, Stouffer DB, Motter AE, Amaral LAN (2008) A Poissonian explanation for heavy tails in e-mail communication. Proceedings of the National Academy of Sciences of the United States of America 105(47):18153-18158

Ray A, Bernoff J, Wise J (2010) Peer influence analysis - Mass influencers are the key to achieving scale in social media marketing. Forrester. http://www.forrester.com/rb/Research/peer_influence_analysis/q/id/56766/t/2, accessed 2010-10-20

Schmittlein DC, Morrison DG, Colombo R (1987) Counting your customers: Who are they and what will they do next? Management Science 33(1):1-24

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

Viswanath B, Mislove A, Cha M, Gummadi KP (2009) On the evolution of user interaction in Facebook. In: Proceedings of the 2nd ACM Workshop on Online Social Networks, Barcelona, Spain, pp. 37-42

Wen C, Tan BCY, Chang KT-T (2009) Advertising effectiveness on social network sites: An investigation of tie strength, endorser expertise and product type on consumer purchase intention. In: Proceedings of the 30th International Conference on Information Systems, Phoenix, AZ, USA, paper 151

Williamson DA (2009) Social networking worldwide: Ad spending and usage. eMarketer. http://www.emarketer.com/Reports/All/Emarketer_2000567.aspx, accessed 2010-10-20

Willinger W, Rejaie R, Torkjazi M, Valafar M, Maggioni M (2010) research on online social networks: Time to face the real challenges. ACM SIGMETRICS Performance Evaluation Review 37(3):49-54

Wilson C, Boe B, Sala A, Puttaswamy KPN, Zhao BY (2009) User interactions in social networks and their implications. In: Proceedings of the 4th ACM European Conference on Computer Systems, Nuremberg, Germany, pp. 205-218

Xiang R, Neville J, Rogati M (2010) Modeling relationship strength in online social networks. In: Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, pp. 981-990

Xu Y, Zhang C, Xue L, Yeo LL (2008) Product adoption in online social network. In: Proceedings of the 29th International Conference on Information Systems, Paris, France, paper 200

Zhang Y, Wang Z, Xia C (2010) Identifying key users for targeted marketing by mining online social network. In: Proceedings of the 24th International Conference on Advanced Information Networking and Applications, Perth, WA, Australia, pp. 644-649

IV Summary and Future Research

In this chapter, the key findings are summarized (section IV.1) and potential starting points for future research are presented (section IV.2).

IV.1 Summary

The main objective of this dissertation was to contribute to the field of CRM with a particular focus on OSN – which impressively represent the digitally connected world we live in – and the identification of influential users within these networks. After the introduction of the foundations on OSN and their impact and value from a business perspective, the dissertation focused on the business areas marketing and sales. In this context, particularly the identification of influential users in OSN has been addressed, which is of great interest in both, research and practice. In the following, the key findings of the corresponding research papers included in this dissertation are summarized.

- The objectives of chapter II were twofold: First, the concept of OSN should be defined and their development over time should be reviewed. Second, the impact and value as well as major risks and challenges of OSN should be demonstrated from a business perspective. Therefore, the first research paper started with a discussion of different existing terms and definitions that are used in the relevant literature on OSN. After setting the focus on user-oriented sites, the underlying concept including these sites' functionalities, users' motives to use them, and the structural characteristics forming the backbone of OSN has been introduced. Thereby, particularly findings related to the application of Social Network Analysis (SNA) to the graphs that can either represent users' rather static connections or users' actual communication activity in OSN have been discussed. The following review of the development of OSN over time and their classification regarding usage and focus revealed that the visibility and searchability of the users' social networks and the viral diffusion of information are the main success factors for all types of OSN. Against this background, the phenomenon OSN has been critically reflected from a business perspective. It became apparent that the specific characteristics of OSN and the enormous amount of data (e.g., information about the graph structure, user-generated content) can be leveraged by companies to generate business value along the value chain. As the discussion of prior research and selected examples from business practice revealed, especially the business areas marketing and sales can benefit from utilizing OSN for marketing campaigns or further areas of applications in the context of social CRM (e.g., reputation monitoring). However, besides these promising opportunities to create business value also risks and challenges, such as privacy issues or the potential spread of negative word-of-mouth, have been critically reflected to increase companies' awareness when engaging in OSN.

In chapter III, influential users and approaches for their identification have been researched in three sections.

- In section III.1, the first objective was to outline fundamental research on social influence, influential people, and their identification in social networks before the rise of OSN. The second objective was to analyze and synthesize the growing number of publications on the identification of influential users in OSN. To achieve these objectives, three research questions have been derived in the second research paper based on related fundamental work: (1) How are influential users characterized in the context of OSN? (2) Which approaches have been developed and applied for the identification of influential users in OSN? (3) How have these approaches been evaluated and which implications have been derived? By means of a structured literature search, it has been found that the majority of existing studies characterizes influential users as particularly well-connected and active users. The analysis further revealed that one leading stream of research on the identification of influential users focuses on users' strategic location, for instance by applying well-known centrality measures originating from SNA. A second major stream of research aims at solving the influence maximization problem by applying diffusion models and (greedy) algorithms to identify influential users in OSN. Regarding the evaluation of approaches it became apparent that most marketing-oriented articles (mostly from the first research stream) draw on real-world datasets of OSN, while rather technical-oriented papers (mostly from the second research stream) have a more theoretical approach and usually evaluate their artifacts by formal proofs. Finally, based on the findings, a research agenda has been proposed to motivate and guide future research. Amongst others, it has thereby been suggested that a stronger collaboration between the scientific BISE and Marketing community than observed today could be mutually beneficial.

- In section III.2, the first objective was to develop a novel approach for the identification of influential users in OSN bringing main findings from prior research together. Subsequently, the second objective was to evaluate the novel approach against existing approaches that could alternatively be used by means of objective data. Based on relevant literature, the third research paper therefore highlighted concepts and findings from existing studies and pointed out the crucial role of users' connectivity and communication activity in OSN. Following the design science research paradigm, a new artifact has been proposed taking into account these previously derived findings, that is, a method in terms of the novel PageRank based approach, which is composed of two steps. In the first step, a weighted activity graph has been derived as the basis for the identification of key users in OSN. In the second step, a novel PageRank based centrality measure has been suggested to rank users in accordance to their centrality scores. By using a publicly available dataset of Facebook, the novel PageRank based approach has been evaluated against competing artifacts, that is, common approaches that could have been alternatively used to identify key users in OSN (e.g., solely connectivity or activity based approaches). With users' retention as the evaluation criterion, it has been illustrated that bringing concepts and findings from research on users' connectivity and users' communication activity together allows for achieving (significantly) better results than using solely connectivity or activity based approaches.
- Finally, in section III.3, the focus has been on users' communication activity in OSN. The objective was to propose an approach for predicting users' communication activity in OSN to improve the effectiveness of advertising strategies by being able to address the most active users deliberately. Therefore, a probability-based method, which has been primarily developed to forecast

purchasing behavior of customers, has been transferred to the context of users' communication activity in OSN. The application and evaluation using a publicly available dataset of Facebook illustrated that the method seems also suitable for predicting users' future level of communication activity in OSN.

Taken together, it can be concluded that the corresponding research papers included in this dissertation contributed to the existing literature in the field of CRM with a particular focus on OSN and the identification of influential users within these networks. Despite the presented findings, however, further challenges remain and offer starting points for future research.

IV.2 Future Research

In the following, potential starting points for future research are highlighted for each research paper included in the chapters or sections, respectively.

- The overview of OSN and their potential business value along the value chain presented in the first research paper (cf. chapter II) did not intend to present a complete survey based on a structured literature review, but aimed at providing the most relevant information to follow up on each covered subarea. As OSN constitute a young but at the same time very large and interdisciplinary area of research, which evolves rapidly (Richter et al. 2011, p. 89), there is still room for further research:

  1. Future studies could select specific topics and conduct detailed and structured state of the art analyses, as for instance already done for the subareas Enterprise 2.0 (cf. Richter et al. 2011) or the identification of influential users in OSN (cf. section III.1).

  2. As the focus was on user-oriented sites and prior research on content-oriented sites such as YouTube or Flickr has consequently been omitted, future work could apply a broader definition of OSN to investigate content-oriented types of OSN as well.

  3. Finally, management and support processes such as human resources, information technology, or financial resources, which can also be supported by

OSN (Bonchi et al. 2011, p. 4), have only been briefly discussed or have not been considered so far. Thus, future research could on the one hand put a stronger focus on business areas beyond marketing and sales. On the other hand, specific subareas and related opportunities and challenges such as the application of social CRM and social Business Intelligence (BI) could be investigated in more detail (cf. e.g., Rosemann et al. 2012).

- When conducting the structured literature search on the identification of influential users in OSN in the second research paper (cf. section III.1), the focus has again been on user-oriented OSN. Besides the possibility that not all relevant studies have been identified even though a rigorous research approach has been applied, certain findings that have only been derived in content-oriented OSN and sites for microblogging such as Twitter or the influence of offline interactions have not been considered so far.

  1. Future research could therefore focus on all types of OSN and beyond, considering all possible sources of user-generated content and platforms that support customer-to-customer interactions (cf. e.g., Libai et al. 2010; Smith et al. 2012).

  2. Likewise, the focus on influential users in OSN could be broadened in order to discuss commonalities and differences of social influence in online and offline settings. Further studies might particularly investigate questions at the interface of online and offline worlds, for instance, how data available in OSN can be used to learn more about social influence that disseminates from online to offline settings.

- Considering the influence of offline interactions could be beneficial when developing approaches for the identification of key users in OSN, too. For instance, the novel PageRank based approach presented in the third research paper (cf. section III.2) did not account for user interactions through other media or face to face. However, also users without or with only low levels of communication activity in OSN could influence advertisement effectiveness, users' loyalty, and users' willingness to pay for services in OSN through social influence exerted through other channels.

1. At this point, future research could investigate user behavior by conducting behavioral studies such as surveys or field experiments (cf. e.g., Fischbach et al. 2009) to get a more holistic view on users' communication behavior and social influence.
2. While the focus has been on a centrality based approach so far, diffusion models should be analyzed and evaluated in more detail as well (cf. e.g., Garg et al. 2011). Thereby, the suggested novel PageRank based approach could for instance be also evaluated against approaches belonging to the second research stream, which aims at solving the influence maximization problem by applying diffusion models and (greedy) algorithms (cf. section III.1).
3. Furthermore, not only users' connectedness and communication activity in terms of communication frequency could be incorporated. Future approaches for the identification of key users in OSN should also consider the actual content of user interactions in order to analyze if positive or negative and actually brand or product-related information is exchanged (cf. e.g., Lin and Goh 2011).
4. When analyzing this user-generated content and electronic word-of-mouth in more detail, the quality of this data should be critically examined. Prior research indicates that data quality in OSN varies considerably (Chai et al. 2009, p. 791). From a private user perspective, for instance, users might face the problem of information overload if they are swamped with outdated information (Chai et al. 2009, p. 791). From a professional user perspective, especially marketers depend on accurate up-to-date data when applying targeted advertising campaigns in OSN (Evans 2011) and recruiters rely on data presented within business networks such as XING or LinkedIn (Lin and Stasinskaya 2002). Hence, data quality in OSN should be considered in future research.
5. Finally, new privacy practices and novel privacy protection directives might come up and reduce the available amount of data to conduct analyses in OSN. Against this background, the data requirements of approaches for the identification of key users in OSN have to be critically discussed. Moreover, users' acceptance of marketing campaigns based on data from OSN should be considered carefully, as there might be also negative effects if users feel that

their right for privacy is neglected (cf. e.g., Krasnova et al. 2009; Krasnova et al. 2010).

- With respect to the fourth research paper (cf. section III.3), the proposed approach for predicting users' communication activity in OSN has not been evaluated against competing approaches (e.g., regression models) and regarding the actual effect of users' communication activity on marketing effectiveness.

    1. Prior research found that in contexts such as predicting customer retention probability based models outperform "curve-fitting" regression models (Fader and Hardie 2007, p. 76). Hence, the proposed probability based method for predicting users' communication activity in OSN should also be evaluated against approaches that could be alternatively used to compare their accuracy and usefulness in the context of users' communication activity in OSN.

    2. In line with prior research on the identification of influential users in OSN based on users' communication activity (e.g., Trusov et al. 2010), the question of how responsive highly active users actually are to certain marketing strategies (e.g., word-of-mouth or targeted marketing) has not yet been satisfactorily answered. Therefore, future research revealing the actual (monetary) benefit of being able to identify users with high future levels of communication activity should be encouraged.

    3. To improve existing approaches such as the PageRank based approach presented in section III.2, findings from research on predicting user behavior (e.g., the proposed method allowing to forecast users' communication activity) could be used to identify key users in OSN not only based on users' past but anticipated future communication behavior.

Taken together, selected topics from the field of CRM with a particular focus on OSN and the identification of influential users within these networks have been researched within the research papers presented in this dissertation. Even though some related questions could thus be answered, being close to the customer will remain a hot topic in research and practice over the next years: As for instance the IBM global chief executive officer (CEO) study 2012 just recently revealed, "CEOs are investing in customer insights more than any other functional area [...]” to achieve a competitive advantage by developing a better understanding of individual customer needs (IBM 2012, p. 7). With respect to a value-based (social) CRM, particularly the (economic) impact of our digitally connected world and OSN as maybe the most prominent representation should consequently be investigated in more detail. It is hoped, that this dissertation can contribute to this endeavor by offering new insights and starting points for future research.

References (Section IV)

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

Chai K, Potdar V, Dillon T (2009) Content quality assessment related frameworks for social media. Lecture Notes in Computer Science 5593:791-805

Evans S (2011) CBR data governance survey 2010. Computer Business Review. http://bi.cbronline.com/features/cbr-data-governance-survey-dataflux-140111, accessed 2012-12-27

Fader PS, Hardie BGS (2007) How to project customer retention. Journal of Interactive Marketing 21(1):76-90

Fischbach K, Schoder D, Gloor P (2009) Analysis of informal communication networks – A case study. Business & Information Systems Engineering 1(2):140-149

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

IBM (2012) Leading through connections. Insights from the Global Chief Executive Officer Study 2012. IBM Corporation, Somers, NY, USA

Krasnova H, Hildebrand T, Günther O (2009) Investigating the value of privacy in online social networks: Conjoint analysis. In: Proceedings of the 30th International Conference on Information Systems, Phoenix, AZ, USA, paper 173

Krasnova H, Spiekermann S, Koroleva K, Hildebrand T (2010) Online social networks: Why we disclose. Journal of Information Technology 25(2):109-125

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

Lin B, Stasinskaya VS (2002) Data warehousing management issues in online recruiting. Human Systems Management 21(1):1-8

Lin Z, Goh KY (2011) Measuring the business value of online social media content for marketers. In: Proceedings of the 32nd International Conference on Information Systems, Shanghai, China, paper 16

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

Rosemann M, Eggert M, Voigt M, Beverungen D (2012) Leveraging social network data for analytical CRM strategies – The introduction of social BI. In: Proceedings of the 20th European Conference on Information Systems, Barcelona, Spain, paper 95

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

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