RESEARCH PAPER

Automated identification of different lead users regarding the innovation process

Isabel Schmid[1] · Janik Wörner[1] · Susanne Leist[1]

Abstract

Lead users are often established in an organizational innovation process to attenuate the difficulties a company faces, such as high costs or the obscurity of customers' needs. But to benefit from these lead users a major challenge is to characterize and identify them especially in the fast-moving world of social media. Therefore, we aim to design a tool to identify lead users automatically for the two innovation phases ("Idea generation" and "Development") by combining different approaches such as social network analysis, topic modeling and sentiment analysis. Thus, we consulted the design science approach and applied our artifact to 11,481 contributions of an online digital platform. The technical realization of the six different characteristics and their respective weighting according to the different phases of the innovation process resulted in different lead users and showed the necessity of distinguishing between them. Our results were evaluated and confirmed by the identified lead users and an expert. Hence, our investigation contributes to both practice and theory (kernel theories and design theory) alike.

Keywords Lead user · Innovation process · Digital platforms

Introduction

In today's dynamic environment, innovation is central for the competitiveness of companies. To create competitive advantages a profound understanding of the sources of innovation is necessary (Von Hippel, 2007). New and innovative ideas can be released by both internal and external sources. According to Innovation Theory generating a new idea by consulting the corporate research and development (R&D) department can be seen as the internal way of innovation creation to advance a company's technology (Freeman & Soete, 1997; Marxt & Hacklin, 2005). The external way of innovation creation, however, consults innovation ideas initialized from parties outside the company such as customers, suppliers, universities or individuals (West & Bogers, 2014). Integrating an individual in the innovation process can take place in different ways e.g. in terms of co-creation (Ramaswamy, 2010), crowdsourcing (Poetz & Schreier, 2012) or open innovation (Martínez-Torres, 2014).

In current research literature, there are a lot of documented examples of successful collaboration with external parties in the innovation process. General Electrics, for example, has banded together with a number of venture capital companies to arrange the "Ecomagination Challenge," a $200 million fund for identifying and investing in innovative ideas and business models regarding renewable energy, grid efficiency and energy consumption. They created a platform where different external stakeholders submitted their ideas and in total they attracted more than 5000 ideas (King & Lakhani, 2013). Moreover, Lilien et al. (2002) have shown that in the company 3 M the inclusion of external individuals in the innovation process results in ideas that have greater commercial potential than ideas without the inclusion of external persons. Whereas TopCoder arranged a two-sided innovation platform to bring software programmers and companies together in order to fix IT-related problems (Lakhani et al., 2010).

✉ Isabel Schmid
  Isabel.Schmid@ur.de

  Janik Wörner
  Janik.Woerner@ur.de

  Susanne Leist
  Susanne.Leist@ur.de

[1] Business Informatics, University of Regensburg, Universitätsstraße 31, 93053 Regensburg, Germany

🍎 Springer

Therefore, new ways of communication such as social media or online communities, a form of digital platforms, provide companies the possibility to access a huge number of users for a new way to innovate (Brem & Bilgram, 2015; Gawer & Cusumano, 2014). Hence, a company can use social media like a magnet to capture customer feedback, improve market research and facilitate innovation (Gallaugher & Ransbotham, 2010). Thus, on the one hand the company benefits from the collaboration with a user as it may result in ideas for extending product varieties, in entirely new products and/or in modifications to existing ones (Al-Zu'bi and Tsinopoulos, 2012). On the other hand, the user also benefits strongly from the innovative products as these are tailored to their own needs (Tuarob & Tucker, 2014; Von Hippel, 1986). In order to realize these benefits, some approaches consult the opinions and suggestions of a crowd of people (cf. open innovation). Although a company thereby receives a lot of input, the ideas are often futile, as they are either not innovative, not feasible or are formulated too superficially. Furthermore, the processing and evaluation of the ideas is very time-consuming as the example of Fiat Mio shows (Saldanha & Pozzebon, 2015). The Fiat Mio team aimed to create a concept car by composing a collaborative website where they received 21,000 ideas and 45,000 comments. The whole process took 15 months and a lot of resources – both human and capital – to screen all the posted ideas and suggestions. To avoid such an intricate and expensive process it is more constructive to concentrate on single persons who are able, due to their individual characteristics, to support a company's innovation process – so called lead users. Hence, a lead user is a user who identifies needs and trends in the market months or years before other people do and who benefit significantly by obtaining a solution to those needs according to the Lead User Theory (Hienerth & Lettl, 2017; Schaarschmidt et al., 2019; Von Hippel, 1986).

These lead users can attenuate the difficulties a company faces during the innovation process, such as high costs or the unsteadiness of customers' acceptance of a company's innovation (Ye & Kankanhalli, 2018). Therefore, a lead user is often established at the beginning and at the end of an innovation process. In the early phases of this process lead users formulate their needs which can result in new ideas. At the end of the process, a lead user can be incorporated to test the product's functionality and durability (Al-Zu'bi and Tsinopoulos, 2012). But in order to benefit from lead users, one major challenge in both research and practice is to characterize and identify them (Ernst et al., 2013). Amongst other factors, the tremendous amount of online community data is responsible for the fact that the identification of lead users is the most difficult and time-consuming aspect within the lead user method (Brem & Bilgram, 2015). In current research literature there are a lot of different lead user identification approaches, but these investigations only covered a limited point of view as they either focus on only one lead user characteristic such as the high level of activity (Martínez-Torres, 2014) or include a very small amount of data (Hau & Kang, 2016). Moreover, various investigations base their approach on observations or online questionnaires (Brandtzaeg et al., 2016; Hung et al., 2011; Tuunanen et al., 2011), resulting in rather low sample efficiency and high costs. Additionally, lead user characteristics are derived based on self-assessments, which may bias the results due to subjective assessments (Hienerth & Lettl, 2017). Finally, most of the aforementioned identification methods are time-consuming, which contradicts to the trend specific short-term construct of lead users (Hienerth & Lettl, 2017).

With this work at hand, we address these aforementioned problems by following the Design Science (DS) approach suggesting an automated and effective approach for the lead user identification. Therefore, this research seeks to answer the following research questions informed by the Lead User- and Innovation Theory and is thus built on a descriptive knowledge base:

RQ1: What different characteristics does a lead user in an online community exhibit?
RQ2: How can the identification of a lead user be supported by a software artifact?

With this investigation we seek to cover the major characteristics of a lead user found in the literature. We aim to identify this user type automatically by means of combining different analysis approaches such as social network analysis (SNA), topic modeling and sentiment analysis. By screening the current research literature, it became obvious that different lead users can be identified in different phases of the innovation process. Thus, we further aim to show the differences between the lead users in these two innovation phases and therefore the legitimacy of the differentiation with respect to the phase-specific characteristics. To cope with this, our goal is to develop a software tool for the automated lead user identification, enabling e.g. the identification of different lead users regarding the two phases of the innovation process, the mapping of all prior identified characteristics in an automated manner and the processing of large amounts of online community data containing relevant information regarding the characteristics of a lead user. To show the applicability of the designed identification approach we apply our artifact to real-world data of an online community for kitesurfing. Kitesurfing is a water sport in which the athlete surfs through the water by pulling a large, controllable kite while standing on a special board. It is a popular example for lead user innovation as this sport was initiated by surfers who – driven by the desire to jump higher and further – experimented with the combination of a surfboard and sails from hang gliding. Moreover, this

area of application is further suitable as these individuals in this area are quite active as innovators and kitesurfing is comprised of a young community, essentially all serious participants are active members in some kind of online community (Franke et al., 2006; Von Hippel, 2005; Wagner & Piller, 2011).

In addition to the instantiated artifact and the results obtained from the demonstration we want to highlight contributions to both practice and theory. Thereby, we want to acknowledge both perspectives of contribution in a design science research project – the artifact school of thought (cf. Hevner et al., 2004) and the design school of thought (cf. Gregor & Jones, 2007). Furthermore, different knowledge contributions will be taken into account to contribute to theory—the descriptive knowledge base in the course of kernel theories and the prescriptive knowledge base in the course of design theory (by deriving the design principles and evaluating them in the course of applying the artifact). To achieve this goal, our investigation follows the third research question:

RQ3: What different contributions for theory and practice can be derived from our Design Science project?

The remainder of this paper is as follows: the following section "Conceptual basics" provides a theoretical background by introducing important definitions and related work regarding lead users and their characteristics. Next, the procedure of the research following the DS approach (Hevner et al., 2004; Peffers et al., 2007) is described in the subsequent section. The section "Design and development" particularly deals with the technical realization and derivation of the design principles to enable the automated identification of lead users regarding the different phases of the innovation process. The following section "Demonstration, evaluation and discussion" shows the application of the demonstrated approach on approximately 12,000 online community data and presents as well as discuss the resulting outcomes, which are additionally evaluated by an interview with our cooperating partner and interviews with the identified lead users. The paper concludes with the contribution for practice and theory and a conclusion.

Conceptual basics

Online communities

Social media are defined as internet-based applications that offer opportunities for interactive and dynamic communication, collaboration and participation (Kaplan & Haenlein, 2010; Obar & Wildman, 2015). Thus, different types of social media can be identified: whereas social network sites

(SNS) especially enable users to connect with other people by creating personal profiles, online communities, as a further hyponym of social media, focus on sharing content between users (Kaplan & Haenlein, 2010). Therefore, online communities can be defined as internet-based platforms for communicating and exchanging content among users who are interested in a given product or technology (Autio et al., 2013; Breitsohl et al., 2018; Preece & Maloney-Krichmar, 2003). Online communities as digital, multisided platforms benefit mainly from so called "network effects": the more users the platform access, the more valuable the platform becomes for both users and companies (de Reuver et al., 2018; Gawer & Cusumano, 2014).

For online communities, which have become increasingly popular due to the rise of social media, various characteristics were defined early and are still relevant today: such as (1) users follow a shared goal, interest or need (Breitsohl et al., 2018; Preece & Maloney-Krichmar, 2003; Tuunanen et al., 2011) (2) users participate actively, interact with each other and build up ties (Dahlander & Frederiksen, 2012; Fisher, 2019; Füller et al., 2007; Preece & Maloney-Krichmar, 2003) and (3) users have access to shared resources (like knowledge or information) (Breitsohl et al., 2015, 2018; Preece & Maloney-Krichmar, 2003). The communication in online communities is established around discussion threads. This means that users initialize new threats in order to start a new discussion, issue or call for advice (Autio et al., 2013). Thus, as online communities often cover one main topic (e.g. mountain biking or kitesurfing), this subtype of social media focus more on connecting people with the same interests than SNS.

Moreover, companies can also benefit from the broad dissemination of digital, multisided platforms in terms of online communities because of social media's reach (via social media a lot of people can be reached) and richness (social media platforms provide various types of information) (Shang et al., 2017). This kind of communication medium gives a company the opportunity to communicate and engage with (potential) customer communities (Fisher, 2019). Thus, as users discuss their experiences, news, improvements or ideas, companies become aware of the customers' needs (Autio et al., 2013; Kaplan & Haenlein, 2010; Tuunanen et al., 2011). Especially in brand communities, excited users group together and share brand-related content (Breitsohl et al., 2015, 2018). However, a company can not only benefit from online communities in terms of nurturing brand commitment and the awareness of customers' needs but also from the fact that these digital platforms can serve as a source of innovation (Dahlander & Frederiksen, 2012; Fisher, 2019). In terms of the discussions taking place in online communities, users also provide new ideas, offer solutions for problems, work out details and test new product ideas (Füller et al., 2007). Thus, these platform-based new

product developments can be consulted to increase product variety, meet diverse customer requirements and business needs (Gawer & Cusumano, 2014).

All in all, online communities allow communication and interaction between users and companies in different ways. Gallaugher and Ransbotham (2010) take this on in their 3-M framework including the three different customer dialog approaches: Megaphone (firm-initiated dialog), Monitor (customer-to-customer dialog) and Magnet (customer-initiated dialog). From this follows that a company can use digital platforms in terms of online communities not only as a megaphone (in terms of spreading marketing messages) but also especially as a monitor to get to know customers' needs. Thus, by monitoring customer-to-customer dialogs companies can gain insights in customers' opinions or market intelligence (Gallaugher & Ransbotham, 2010). Furthermore, a company can also use online communities as a magnet, the customer-initiated dialog, to capture customer feedback, improve market research and facilitate innovation (Dahlander & Frederiksen, 2012; Fisher, 2019; Gallaugher & Ransbotham, 2010).

Lead user innovation

Innovation is a central construct for organizational competitiveness and effectiveness (Wolfe, 1994). It can be seen as an essential process for driving economic growth (Chen et al., 2018). In general, innovation can be defined as a process that includes the generation, adoption and implementation of new ideas, practices, or artifacts in organizations (Axtell et al., 2000; Ye & Kankanhalli, 2018). So, innovation is a complex issue that comprises many theories, each with a different focus (Wolfe, 1994). In addition, there are also many innovation process models that describe how innovations can be implemented step by step. Cooper (1996), for example, established the so-called stage-gate model and divided the innovation process into the following five different phases (stages): 1. preliminary investigation, 2. detailed investigation, 3. development, 4. testing and validation, 5. full production and market launch. The stage-gate model describes a conceptual and operational model for moving new product projects from idea to launch. Other widely spread innovation process models (cf. Crawford, 1994; Herstatt, 1999; Ulrich & Eppinger, 1995) are similar to the process of Cooper's (1996) approach.

By scanning further research literature about innovation and keeping the process models in mind, exemplarily the stage-gate model, it became apparent that most innovation approaches identify two comprising key phases: (A) the idea generation which means the "awareness" of an innovation and incorporates therefore the preliminary and detailed investigation phases of the stage-gate model, and (B) the development of an innovation incorporating the development

as well as the testing and validation phases of the stage model (Amabile, 1988; Axtell et al., 2000; Unsworth et al., 2000). We follow this approach and concentrate on the two phases "Idea generation" and "Development". Consequently, we excluded in our investigation e.g. the step "market launch" as here another user type – the influencer – can be applied to support this phase optimally (Schmid, 2020). According to the definition of an influencer, this user type is applied by companies for disseminating information, for spreading marketing messages, and for changing the opinions and even the purchase decisions of people in its direct and indirect environments (Schmid, 2020). If an influencer would be involved in upstream value creation stages respectively innovation related phases (such as "Idea Generation" or "Development"), the user would feature characteristics of lead users (e.g., ahead of trends, etc.) and would therefore be – in addition of being an influencer – also a lead user. So, as our overall goal of this paper is to identify users who can support a company in their innovation process, we focus the characteristics of a lead user who can also appear in other phases as other type of user.

In the last decades, it has become conventional that consumers or users themselves support one or even both phases of the illustrated innovation process. Hence, this user innovation can be conveyed from the shift of traditional firm-centered innovation to user-centered innovation (Von Hippel, 2005). Prior research highlights that users, rather than firms, are frequently the ones who initiate new product ideas and product developments (Dong & Wu, 2015). Thus, user innovation can be defined as innovative activities undertaken by users who are the source of innovative ideas and who engage actively in developing and modifying products also to meet their own needs (Zheng & Zhou, 2017). These users can invent, prototype, and test new innovations (Roy, 2018). The advantages of user innovations can be mainly attributed above all to the nature of digital innovation platforms.

From a company's point of view, the most important driver for user innovation is to overcome information stickiness. Innovation requires both information about the problem and problem-solving information or, in other words, need-related and solution-related knowledge (Von Hippel, 1994, 2005). Often the information about user's needs and the information used in problem solving is costly to acquire and therefore "sticky" (Piller, 2006; Von Hippel, 1994). Hence, the acquisition as well as transferring costs of the information that is decisive for initiating innovation have tremendous influence on where innovation is created (Idota, 2019). Therefore, as users with highly sticky information can create innovation, organizations should include them in their innovation process to get to know the user's needs, to solve (product) problems and to reduce R&D costs. Thus, User Innovation Theory postulates i.a. that "*innovation among users tends to be concentrated on lead users (people with*

high lead userness) of those products or services" (Ye & Kankanhalli, 2018). This means that those users who carry out user innovation are so-called lead users (Von Hippel, 1986).

In current research literature there is no consistent definition of a lead user, but the Lead User Theory of Von Hippel (1986) is often used as a starting point for defining them: "*Lead users face needs that will be general in a marketplace – but face them months or years before the bulk of that marketplace encounters them, and – Lead users are positioned to benefit significantly by obtaining a solution to those needs*." (Von Hippel, 1986, p.796). Thus Lead User Theory states that lead users can be used as a source of innovative and commercially attractive ideas about products and services and are characterized by two distinct characteristics: ahead of trend and high benefits from innovation (Hau & Kang, 2016; Von Hippel, 1986, 2005). Hence, lead users are able to invent, prototype and field test innovations (Roy, 2018). Therefore, they can either be applied for the entire innovation process (cf. Ye & Kankanhalli, 2018) or can be applied for only one part of the innovation – either need or solution related tasks (cf. Von Hippel & Katz, 2002). As lead users can constrict the challenges a company faces during the innovation process and as they are simultaneously able to disclose new ideas, lead users can be seen as a valuable resource for companies in terms of different phases of the innovation process (Al-Zu'bi and Tsinopoulos, 2012; Ye & Kankanhalli, 2018). Several studies have shown that their involvement in this process, especially in the early and late phases, can enhance product success (Brem et al., 2018; Schreier et al., 2007). Hence, an intensive collaboration with lead users can increase the product variety as well as the rapidness of a new product development process (Al-Zu'bi and Tsinopoulos, 2012).

Furthermore, as the lead user is the only user type who can be applied in terms of user innovation and therefore support a company in their innovation process, we focus on this type of user. To benefit from a lead user, one major challenge in both research and practice is to characterize and identify him/her (Ernst et al., 2013) – the second step in Von Hippel (1986) 4-step process of utilizing lead users (1. identification of the trend, 2. identification of a lead user, 3. analyze lead user need data, 4. project lead user data onto the general market). Here, the identification of adequate lead users is mostly accompanied by horrendous monetary, time and human resources (e.g. Brandtzaeg et al., 2016; Hung et al., 2011; Tuunanen et al., 2011). Therefore, to reduce the devoted resources regarding the characterization and identification of lead users, we aim to design a software tool, enabling the automated identification of lead users based on their descriptive characteristics. This tool is intended to automate the identification process which is described in current research literature as the most difficult

and time-consuming aspect within the Lead User Theory (Brem & Bilgram, 2015). However, to automate the identification process we need to characterize the lead user in detail first.

Characterization of lead users in online communities

In order to characterize lead users in online communities, we conducted an extensive literature search. This resulted in 18 investigations (see Table 1) that focus on the characterization of lead users in terms of online communities. A minority of the 18 investigations (3 out of 18) examines lead users within SNS rather than explicitly in online communities. Nevertheless, since these investigations specify SNS with the same characteristics as online communities and since the authors of these three investigations also base their research primarily on identifying lead users in online communities, these investigations are also included here. Numerous research papers that are not related to the online area were excluded as well as those that do not focus on the identification or characterization process.

However, the description of lead users by Von Hippel (1986) in the course of the description of the Lead User Theory (see section: "Lead user innovation"), were used as a starting point for the characterization of lead user as almost every investigation mentioned in the following relate to these two major lead user characteristics: (1) trend leadership/being ahead of trend and (2) the high expected benefit from innovative solutions, meaning that lead users benefit strongly from adopting new products tailored to their needs (Brem et al., 2018). Prior research about the identification and characterization of lead users in the social media sphere have shown that these two characteristics of the basic model of the Lead User Theory remain valid (Pajo et al., 2017; Schaarschmidt et al., 2019; Tuarob & Tucker, 2014; Ye & Kankanhalli, 2018).

(1) Trend leadership incorporates the degree to which a user can be seen as a leading edge with respect to a certain trend (Franke & von Hippel, 2003). That means lead users have prevailing information and expertise about major trends of products and services as well as future demands for them in the market (Hau & Kang, 2016; Tuarob & Tucker, 2014). Hence, a lead user is a consumer of a product that identifies problems and unmet needs that will later be experienced by the public. This means that the innovations lead users strive for often do not exist on the market (Franke & von Hippel, 2003). Therefore, as lead users recognize what the mass desires months or years before others do, they are ahead of trends (Brandtzaeg et al., 2016; Pajo et al.,

Table 1 Prior research on lead users

		Trend leadership	Sentiment		High Activity Level	High product related knowledge	Opinion leadership
			Dissatis-faction	Enjoy-ment			
Idea generation	(Belz and Baumbach, 2010)	X	X		X	X	X
	(Chen et al., 2019)	X		X	X	X	
	(Miao & Zhang, 2017)				X		
	(Marchi et al., 2011)	X			X	X	
	(Martínez-Torres, 2014)				X		
	(Hau & Kang, 2016)	X	X			X	
	(Hung et al., 2011)	X	X		X		
	(Pajo et al., 2017)	X	X		X	X	X
	(Tuarob & Tucker, 2014)	X				X	
Development	(Brandtzaeg et al., 2016)	X					
	(Chen et al., 2019)	X		X	X	X	
	(Franke and Hippel, 2003)	X				X	
	(Franke et al., 2006)	X				X	
	(Li & Tang, 2016)	X				X	
	(Miao and Zhang, 2017)				X	X	
	(Mahr & Lievens, 2012)	X				X	
	(Pajo et al., 2014)	X	X		X	X	X
	(Pajo et al., 2017)	X	X		X	X	X
	(Schaarschmidt et al., 2019)	X	X			X	
	(Schreier et al., 2007)	X				X	X
	(Tuunanen et al., 2011)	X			X		

2017; Pajo et al., 2014; Tuarob & Tucker, 2014; Ye & Kankanhalli, 2018).

(2) In addition, the characteristic high expected benefit is broken down into further sub-characteristics in the current research literature to make this characteristic more tangible and (especially against the background of the large amount of social media data) more measurable (Ye & Kankanhalli, 2018). We agree with this approach and focus on these sub-characteristics (e.g. dissatisfaction) when defining and characterizing a lead user in the following section. Thus, a lead user does not only come up with attractive innovations to help others but they also benefit strongly from the adoption of new or improved products (cf. high expected benefit) (Schreier et al., 2007; Von Hippel, 1986). Often it is not the financial benefit that motivates a lead user to innovate, but e.g. the chance to execute their sports more effectively. By undertaking their sports, users become aware of the mismatch of expected and experienced performance of the products which can lead to dissatisfaction (Lüthje, 2004). Therefore, the discrepancy between the users' needs and the solutions available on the market leads to dissatisfaction. Given the nature of the kitesurfing or mountain bike community and their genesis, it was through the dissatisfaction of the athletes that the sport emerged, which leads to this proxy measure for users' expected benefit (Belz & Baumbach, 2010; Pajo et al., 2014, 2017; Schaarschmidt et al., 2019). The unmet needs and the relating dissatisfaction of a user lead to the expectation to benefit significantly from an innovative solution (Pajo et al., 2017; Ye & Kankanhalli, 2018). Although this characteristic is prevalent in the current research literature, a discrepancy can be determined. Chen et al. (2019) e.g. introduce a new model (ITF model) for determining a user's index of innovativeness including the three dimensions of involvement, thinking and feeling. The last dimension "feeling" is related to the extent of a user's enjoyment, exploration and creativity, which in turn enables the users to make full use of their potential innovativeness. Therefore, the authors refer to the emotional attachement and the preference for the product by users and therefore state that a lead user exhibits positive sentiment rather than negative sentiment such as dissatisfaction (Chen et al., 2019).

Additionally, with regard to the topic of lead users in online communities, multiple other characteristics, beside the abovementioned, can be identified e.g. the high level of activity with regards to the involvement. According to

various investigations (Martínez-Torres, 2014; Miao & Zhang, 2017; Pajo et al., 2017) lead users are more active in a community than the rest of their members. Moreover, the authors Hung et al. (2011) emphazise the lead user's creative and active participation in order to facilitate effective innovations and to encourage innovation communication. The more a lead user's participation level is, the more they get involved in the community. High involvement of users usually implies that there will be high effort made by the users in interacting with the product (Chen et al., 2019). This active involvement is necessary to disclose the sticky information that resides in a lead user. This information can only be valuable in terms of innovations when a user exhibits a high product related knowledge (Franke et al., 2006; Li & Tang, 2016). According to Schaarschmidt et al. (2019) a lead user differs most "*from `typical` consumers as they also have considerable levels of solution knowledge*" (Schaarschmidt et al., 2019, S. 4). This kind of product-related knowledge contains expertise about the product architecture, the used materials and the technologies as this is the basis for creating new ideas (Franke et al., 2006; Schreier et al., 2007). Only by having high product related knowledge, a lead user is able to formulate the needs into concrete innovation ideas and/or concrete specifications of new products (Chen et al., 2019; Marchi et al., 2011; Pajo et al., 2017; Pajo et al., 2014; Tuarob & Tucker, 2014).

As lead users not only have ideas for realizing innovation but also diffusing them, a lead user can also be described by the characteristic "opinion leadership". Opinion leadership is the ability to enable the flow of information and especially to diffuse it. Strong social relationships and a high level of engagement are premises for a functioning exchange of ideas and innovation (Pajo et al., 2014, 2017).

However, lead users can be defined not only in terms of these different characteristics but also – as already mentioned in the section "Lead user innovation" – in terms of the different phases of the innovation process where a lead user can be applied. Therefore, to support the identification of lead users regarding these different innovation phases, we further allocate the aforementioned characteristics to the respective innovation phase.

- Lead users can be applied in the phase "Idea generation" of the innovation process and are therefore more problem-oriented (Belz & Baumbach, 2010; Miao & Zhang, 2017). Lead users in this phase of the innovation process describe problems and unmet needs with the already existing products (cf. dissatisfaction) (Belz & Baumbach, 2010; Hau & Kang, 2016). Furthermore they bring forward new ideas which might help to fix the problem described before (cf. trend leadership).

These ideas tend to be unique and can possibly be useful for the development of the next generation (Tuarob & Tucker, 2014). In online communities lead users can share their innovative ideas and other community members can comment and evaluate these ideas. The users offer suggestions on the one hand about modifications and adaptions regarding product attributes, positioning, etc.. On the other hand, lead users formulate innovative ideas about completely new products which can be realized afterwards by a company's R&D team (Marchi et al., 2011; Martínez-Torres, 2014). Therefore, lead users are incorporated in a very early phase in the innovation process (Hung et al., 2011). This phase "Idea generation" can be seen as a venue for brainstorming to make the free exchange of ideas possible (Muller et al., 2012; Paulus et al., 2002). When brainstorming, people are encouraged to generate as many ideas as possible and therefore a high participation as well as a high activity level is necessary here (Chen et al., 2019; Hung et al., 2011; Miao & Zhang, 2017))

- Lead users are not only able to provide new ideas but can also be integrated into the "Development" phase of the innovation process. Because of their high product related knowledge and their vast experience lead users are able to suggest concrete solutions instead of describing problems or stating customer needs (Mahr & Lievens, 2012). Hence, scientific articles which characterize and identify lead users in terms of the "Development" phase focus on users e.g. who have already done security-related modifications to a web server software (Franke & von Hippel, 2003) or who have already developed applications for different plattforms (Schaarschmidt et al., 2019). Hence, Mahr and Lievens (2012) summarize it and state that lead users are best suited for improvements pertaining to functionality. Thus, lead users in this second phase of the innovation process are able to support companies in order to develop new products and solutions with the aim of meeting rapidly changing consumer needs and to stay competitive (Pajo et al., 2014). This can diminish failure rates of new product introduction. So utilizing this high-product related knowledge combined with the high level of trend leadership, a lead user can be conducive in strengthening a company's innovation competitive advantage (Li & Tang, 2016).

The assignment of the characteristics to the different phases in the innovation process and thus the difference made by the lead users in the two innovation phases can be detected in Table 1.

Related work

The review of the research literature has shown that, in addition to the two characteristics from the basic model of the Lead User Theory, there are many different characteristics to describe and characterize lead users in the online environment (see Table 1), whereby different approaches such as screening (Brandtzaeg et al., 2016; Hung et al., 2011; Tuunanen et al., 2011), pyramiding (Von Hippel et al., 2009), SNA (Martínez-Torres, 2014), or netnography (Belz & Baumbach, 2010; Mahr & Lievens, 2012) have been used. However, these studies on lead user identification covered only a limited point of view as they either focus only on one characteristic of a lead user, like the high level of activity (Martínez-Torres, 2014), or they include a very small amount of data (Hau & Kang, 2016). Furthermore, investigations are based on observations or (online) questionnaires (Brandtzaeg et al., 2016; Hung et al., 2011; Tuunanen et al., 2011) which results in a low sample efficiency and high costs. In addition, lead user characteristics are thereby based on the self-assessment of respondents, which means that the results can be affected by subjective assessments (Hienerth & Lettl, 2017). Another problem within the current research literature is the aspect of time (Brem et al., 2018; Hienerth & Lettl, 2017). Most of the aforementioned identification methods are time-consuming. This contrasts with the fact that the concept of "lead userness" is not a long-term construct but it is trend specific and can change over time. A lead user today may or may not be a lead user in distant future (Hienerth & Lettl, 2017).

Furthermore, as already mentioned, lead users can be identified in both innovation phases: the "Idea generation" and "Development" (Füller et al., 2007). However, in current research literature it is common practice to identify a lead user for one of the two phases (Marchi et al., 2011; Martínez-Torres, 2014; Schaarschmidt et al., 2019). Moreover, only a minority of the investigations incorporates an innovation process but identifies a lead user for all of its phases (Miao & Zhang, 2017; Pajo et al., 2017). Consequently, the current research literature is incomplete here as there is no approach that identifies different lead users for every phase of the innovation process and so an overall approach is missing. Thus, we aim to show the differences between the lead users in the two innovation phases and the legitimacy of the different identification processes, although we have seen in our summary Table 1 that there are no large differences between the two phases.

All in all, in order to avoid these negative aspects of the prior research literature, we have come to the conclusion that a tool for automated lead user identification is needed. This tool should meet the following design requirements and should therefore be able to:

- map all prior identified characteristics,
- process a large amount of online community data,
- apply objective identification methods,
- repeat the identification process for lead users at any time as lead users are trend specific, and
- identify different lead users regarding the two phases of the innovation process.

Procedure of the research

In order to make the development of a systematic approach for the automated identification of lead user comprehensible, we applied Design Science (DS) research. Research projects that follow the DS paradigm are concerned with the design, development, implementation, use, and evaluation of socio-technical systems in organizational contexts. Design scientists produce and apply knowledge of tasks or situations to create effective artifacts (March & Smith, 1995). These artifacts are delineated in different structured forms such as software, formal logic, and rigorous mathematics to informal natural language descriptions (Hevner et al., 2004).

An important step in DS research is to prove the utility, quality, and efficacy of the artifact via well-executed evaluation methods. Since the artifact's performance is related to the environment in which it is used, an incomplete understanding of the environment can induce inappropriately designed artifacts (March & Smith, 1995). Therefore, Hevner's "design cycle" (Hevner, 2007) substantiates the importance of constructing and evaluating the artifact, and suggests balancing the efforts spent on both activities, which must additionally be convincingly based in relevance and rigor (Hevner, 2007). Consequently, DS research is based on and contributes to scientific knowledge by performing the research process rigorously (e.g., by reflecting the construction or/and evaluation of the artifact) which is represented by Hevner's "rigor cycle". DS research additionally uses practical knowledge and leads to several practical contributions which constitutes Hevner's "relevance cycle" and which can be seen as self-evident objectives of a DS research project (Hevner, 2007).

We followed the DS research paradigm (Gregor & Hevner, 2013; Hevner et al., 2004) and aligned our research activities with the procedure as proposed by Peffers et al. (2007) (see Fig. 1). This procedure provides a commonly accepted framework for conducting research based on DS principles. In addition, Peffers et al. (2007) designed the procedure as a result of a consensus-building approach, which comprises well-agreed process elements (Peffers et al., 2007). As a first step, (1) corresponding problems and drawbacks of already existing approaches to identify lead users in online communities were identified (see sections "Introduction" and "Conceptual basics"). Hence, in
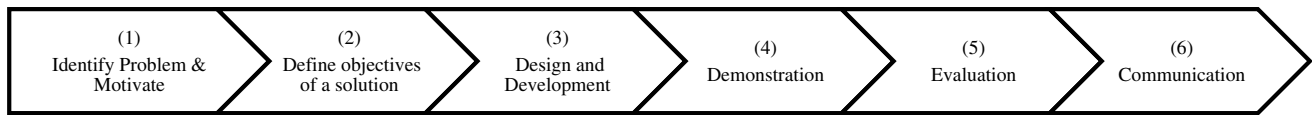
Fig. 1 Design science process

current research literature there are a lot of different lead user identification approaches, but these investigations only cover a limited point of view as they either focus only on one lead user characteristic, they include a very small amount of data or their approach is reliant on the self-assessment of users. Consequently, our (2) objective is to provide and combine a set of methods, based on the characteristics of a lead user, in order to identify this type of user automatically in an online community (see sections "Conceptual basics" and "Design and development"). The third step of the DS process model contains the (3) design and development (see section "Design and development") of a solution or an artifact, respectively. Such artifacts can be constructs, models, methods or instantiations. In order to fill the gaps identified within phase (1), we focus on the design of the technical realization of the tool by means of the combination of different methods such as SNA, topic modeling and sentiment analysis. Thus, our approach was established to support and simplify the lead user identification process and to eliminate the existing disadvantages. In the next step, the (4) demonstration, we show the application of the demonstrated approach on approximately a data set of about 12,000 contributions from an online community about kitesurfing. Kitesurfing is a suitable area of application as individuals in this area are quite active as innovators. Furthermore, kitesurfing comprises a young community, essentially all serious participants are active members in some kind of online communities (Franke et al., 2006; Von Hippel, 2005; Wagner & Piller, 2011). The overall results of the application of the analysis are shown in this chapter, consolidated in a summary table and discussed in detail. These results are additionally evaluated (5) by conducting both interviews with lead users and an in-depth interview with an expert (head of marketing of our cooperating partner) in the field of kitesurfing. In terms of these interviews, we evaluated our artifact and showed that our approach provides an added value. We have further discussed the results of the evaluation as well. Finally, the results are then (6) communicated.

The orientation towards the procedure by Peffers et al. (2007) also makes it possible to align our research with the guidelines of Hevner et al. (2004) or Hevner (2007), respectively. According to the design cycle, we present our artifact as the result that has gone through the process of demonstration (application of our approach to an online community about kitesurfing and evaluation with several interviews (see section "Demonstration, evaluation and discussion")). In view of the relevance cycle, we identified several design requirements (from literature including several case studies (see section "Conceptual basics")) that guided the design of the artifact, and so the practical application of our artifact brought up several contributions for practice (e.g. identifying relevant users for innovation/trends (see section "Discussion of the results of demonstration")). In view of the rigor cycle, we used several methods and techniques to rigorously construct and evaluate our artifact (e.g. topic modeling, SNA, frequency analysis) and derived initial findings as contributions to theory, both kernel theory (Lead User- and Innovation Theory) and Design Theory (see section "Contribution for practice and research"). Thus to contribute to a rather general and abstract knowledge base – "nascent design theory" (Gregor & Hevner, 2013) – and in order to design a purposeful artifact in a comprehensible way, we first established both, a set of meta-requirements and design principles (Gregor & Jones, 2007; Heinrich & Schwabe, 2014). Thus, the design of the lead user identification tool is grounded on design requirements retrieved from seminal works on Lead User- and Innovation Theory. In a next step, we then describe our prototypical implementation that demonstrates the feasibility of the design principles and meta-requirements in the tool.

Design and development

Design principles for a lead user identification tool

First, the composition of meta-requirements (MRs) that describe "*what the system is for*" (Gregor & Jones, 2007, p.325) is based on the purpose and scope of the identification tool that has been discussed in the introduction. Thus, we define the solution objectives based on the class of problems our paper addresses and present them in Fig. 2. These MRs established to be suitable for a class of artifacts and are based on the current research literature (Gregor & Jones, 2007; Heinrich & Schwabe, 2014; Walls et al., 1992). Besides the MRs, the design principles are synthesized in a next step. Design principles are defined as prescriptive statements that show how to do something to achieve a goal (Gregor et al., 2020). The design principles that we dispose fall into the category of action and materiality-oriented design principles that describe what an artifact should enable users to do and how the artifact should be built in order to

Fig. 2 Overview of the design phase
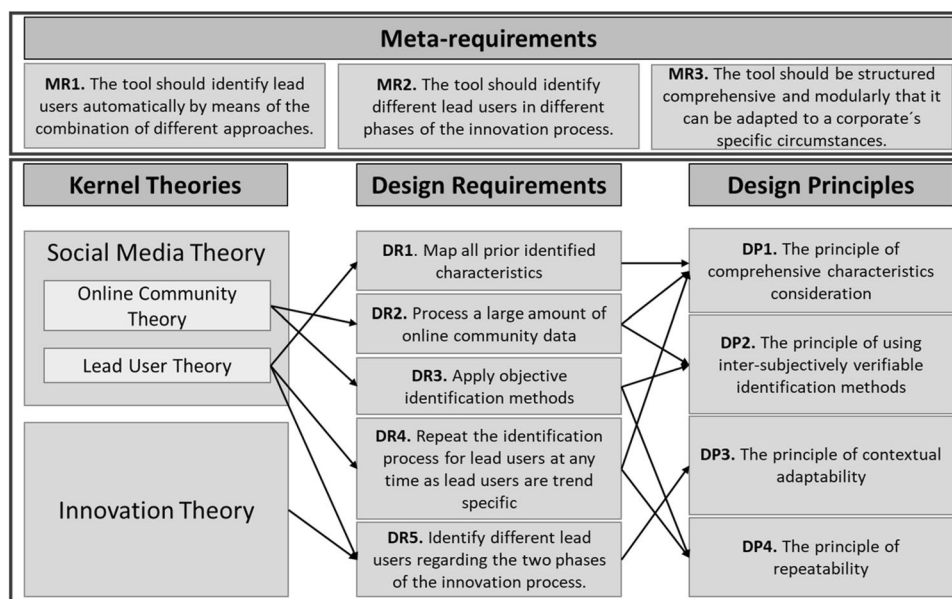
do so (Chandra et al., 2015). Regarding companies (=users) who are interested in identifying lead users in online communities (=boundary conditions) and keeping our design requirements for our artifact in mind, we derive four design principles for (lead) user identification tools:

1. The principle of comprehensive characteristics consideration. In order to identify specific user types in online communities, e.g. a lead user, it is necessary to precisely define and describe their characteristics. Thus, the automated identification of a lead user requires a technical implementation of its characteristics that we have derived from the current research literature. Therefore, the tool should be able to incorporate and technically realize all relevant lead user characteristics (trend leadership, sentiment, high activity level, high product related knowledge and opinion leadership) to obtain precisely targeted results.
2. The principle of using inter-subjectively verifiable identification methods. In order to counteract the subjective self-assessment of respondents of (online) questionnaires different inter-subjective methods should be consulted and combined to identify a lead user. Therefore, the tool should use comprehensible and inter-subjectively verifiable identification methods to make the identification process traceable.
3. The principle of contextual adaptability. Since lead users are applied in terms of innovations in a company, the identification of a lead user must also take into account the different phases of the innovation process the user supports. Therefore, the tool should be able to adapt the weights of the characteristics according to the different

circumstances of the companies and their aim to apply lead user in different stages of the innovation process.
4. The principle of repeatability. As lead userness is a short-term construct, which means that lead users can change over time, the identification process should be executable often and in a resource-saving way. Therefore, the tool should allow repetition of the identification process at any time to react quickly to changing circumstances such as trends.

These design principles are deduced from the design requirements that are further based on kernel theories and prior research literature. Gregor and Jones (2007) state that these kernel theories disclose "*an explanation of why an artifact is constructed as it is and why it works*" (p.328). So, these kernel theories include the body of knowledge that is necessary to meet the design requirements (Böckle et al., 2021). Hence, based on the discussion of the kernel theories and thus the related work, we derive design requirements our tool should meet. These design requirements offer guidance by designing the artifact and advise the design principles (Böckle, et al., 2021; Gregor & Jones, 2007). These principles refer to at least one requirement and serve as an abstract "blueprint" of our artifact (Böckle, et al., 2021; Gregor & Jones, 2007; Heinrich & Schwabe, 2014). By establishing these design principles, we made sure that they follow the value grounding (reference to the requirement) and the explanatory grounding (design principles are based on the current literature and thus on kernel theories) (Heinrich & Schwabe, 2014).

For each of the design principles, its instantiation in the artifact is described in the following sections.

Weighting of the according lead user characterizations

To address the shortcomings of the prevailing research and therefore consider the derived design principle *1. The principle of comprehensive characteristics consideration*, we aim to compose an automatic identification approach including all characteristics identified in literature. Furthermore, to account for the different circumstances of the two innovation phases (see section "Lead user innovation"), we also distinguish between lead users associated to the phase "Idea generation" and lead users associated to the phase "Development" of the innovation process. Therefore, to consider the different relevancies related to the identified characteristics with respect to each innovation phases, the characteristics are weighted accordingly (see Table 2) based on their occurrences within the current research literature (see Table 1). In addition, with respect to the derived design principle *3. The principle of contextual adaptability*, companies are enabled to adapt the respective weights to their circumstances and thus to apply lead users in different stages of the innovation process.

Table 2 summarizes the characteristics' accompanied relevancies in the context of the respective innovation phases. Here, the weights illustrate that the respective focus of the innovation process within the activity of the users (Idea generation) as well as their product-specific knowledge (Development) differs considerably. Additionally, users also differ

in the mood they exhibit. Here, the characteristic dissatisfaction is given greater meaning in the "Idea generation" phase since users express their unmet needs of a product or service within negatively afflicted communication. These differentiations enable to adequately consider the circumstances of the two innovation phases, resulting in the determination of precisely fitting and goal-oriented lead users.

Technical realization

To enable the automated identification of lead users based on the above determined relevancies, the previously identified characteristics (see Table 1) must be mapped in an automatic manner. Therefore, we have implemented each identified characteristic in the programming language *Python*. As the underlying data (e.g. online community posts, network interactions, etc.) are mainly represented in a textual way, we focused on finding computer-based procedures from the research field of text mining to map the identified characteristics. Text mining enables an automatic identification of hidden structures or patterns within a corpus of textual data (Feldman & Sanger, 2007; Heyer et al., 2006). In addition, we conducted the SNA, which best suits the identification of users within a potentially high influence to be solved, as SNA enables us to show the relations in a structured network via nodes and ties to state quantitative characteristics of users. Furthermore, due to the different nature of each characteristic (see Table 3), the values must be normalized to make them comparable. Therefore, we have conducted the Min–Max normalization (Han et al., 2006) to rescale each characteristic into a value range between [0;1]. In the course of normalization, the specific values of all users were related to each other. Thus, the higher the respective value, the more the respective user exhibits the specific characteristic. By this, all values are located at the same scale and therefore can be weighted by their allocated relevance (see Table 2). To give an overview of the characteristics and their technical realization, they are further summarized in Table 3. Here, to consider the derived design principle *2. The principle of using inter-subjectively verifiable identification methods* and therefore ascertain an adequate analysis process,

Table 2 Weighting the relevance of each characteristic for both innovation phases

Characteristics	Weight by occurrence in literature	
	Idea generation	Development
Trend leadership	5	5
Dissatisfaction	3	2
Enjoyment	1	1
High activity level	5	3
High product related knowledge	4	4
Opinion leadership	2	2

Table 3 Technical conception of the identified characteristics

Characteristics	Technical realization	Implementation	Outcome
Trend leadership	Latent Dirichlet Allocation (LDA)	Gensim; Mallet (McCallum, 2002)	Probability
Dissatisfaction	Sentiment Analysis	VADER (Hutto & Gilbert, 2014)	Classification
Enjoyment	Sentiment Analysis	VADER (Hutto & Gilbert, 2014)	Classification
High activity level	Frequency Analysis	Self-created	Frequency
High product related knowledge	Frequency Analysis	Self-created	Frequency
Opinion leadership	Closeness, Betweeness and Degree centrality	NetworkX (Hagberg et al., 2020)	Centrality measure

all technical realizations are based on broadly known and prevalent quantitative and qualitative content analysis methods. To further meet the particular needs associated to the respective characteristic, the methods used have further been adapted as described in the following.

With respect to the characteristic of trend leadership, the aim is the identification of users who talk about trends before they became general, community-wide discussed topics. To meet these requirements, we mainly had to split the automation into two separate sub-sequences: (1) identify trends (e.g. frequently discussed product issues or service properties), based on UGC; (2) identify users who talked about one or multiple of these previously identified trends, before they became discussed community-wide. With regard to these two identified sub-sequences, we focused on the use of text mining methods, enabling the automatic processing of unstructured, unlabeled data such as online community posts. More specifically, as trends represent frequently emerging topics as well as the advantage of topic modeling compared to other text mining techniques to operate directly on the textual data instead of solely comparing their underlying structure (Aggarwal & Zhai, 2012), we have chosen topic modeling for the automatic identification of trends. Topic modeling can project the textual corpus of contributions into a topical space by reducing the dimensionality and attaching different weights, which results in semantically coherent groups of words (topics), which represents our trends (Crain et al., 2012; Xie & Xing, 2013). Specifically, because of LDA's simple applicability but also its satisfactory analysis results within the topic modeling (Eickhoff & Neuss, 2017), the choice was made for LDA. For the implementation of LDA within the automated identification approach, the established python library Gensim was used in combination with Mallet (see Table 3). In order to achieve the highest quality of results possible, we further automatically prepared the data for the analysis by applying tokenization, stop word removal and case folding (cf. Boyd-Graber et al., 2014). Furthermore, in order to take the characteristics of trends into account (1), we adapted LDA to only consider contributions of the last eight weeks to extract the trending topics. By this adaption, the identification of those user who were already talking about these trending topics within their contributions at an earlier point in time than eight weeks ago (2) is feasible. The identification takes place through statistical inference and reflects the cumulated probability with which a user talks about one of the identified trending topics.

Considering the characteristic of dissatisfaction or enjoyment aims to identify users with either negative or positive mood. Therefore, the global mood of each user within their contributions has to be identified. The automatic identification of moods within textual data is summarized under the term "sentiment analysis". Through this, for instance, it is possible to identify users who have unfulfilled expectations and thus show a significant potential for improvement of a product or service (Pajo et al., 2017). To determine the mood of each post by a user, we implemented the "Valene Aware Dictionary for sEntiment Reasoning" (VADER) (Hutto and Gilbert, 2014) technique. VADER is a lexicon and rule-based sentiment analysis technique that is specifically attuned to sentiments expressed in social media and has achieved remarkable results compared to other prominent sentiment analysis techniques (Hutto and Gilbert, 2014). To determine the sentiment value, VADER uses a labelled dictionary adapted to the contextual characteristics of social media data. Hereby, VADER is able to combine the positive and negative inflections and generates a single sentiment score within the range of -1 to + 1. In order to determine the global sentiment value of each user, we further adjusted the technique to build a consolidated sentiment score for each user reflecting its global mood by setting the individual scores of each contribution into relation to the total amount of contributions of a user. This results in the mean value of all mood-bearing contributions of a single user, which reflects their average mood.

To measure the activity level of a user, we further determined the amount of user interactions within the community. For this purpose, the number of posts and transacted comments per user within the analyzed period was identified to attain information about the activity level of a user (Miao & Zhang, 2017).

In the case of high product related knowledge, the aim is to identify users who have an immense knowledge of product specific information. To accomplish this, we considered splitting the determination of the characteristic into two parts. In the first step, a dictionary of product-specific terms was extracted from product and service descriptions e.g. product brochures. Secondly, following the generation of the product-specific dictionary, the occurrence of the extracted product-specific words in the contributions were determined. Therefore, matching word candidates from the contributions are identified and reflected against the product-related dictionary. If an entity matches with a product-specific word, the total sum of the user's usage of product-related words will be increased. After all contributions of the related user have been analyzed, the number of product-specific words is divided by the total number of all words used by the specific user. The resulting value reflects the average use of product-related words by a user and allows conclusions to be drawn about the product knowledge of a user.

With regards to the determination of the user's ability to enable the flow of information and especially diffuse it, which are prerequisites for opinion leadership, we have considered several centrality measurements which best suit the identification of strong social relationships within a social network (Pajo et al., 2017). These measures are those of *degree, closeness* and *betweenness centrality*, and are

fundamentally related to the concept of social influence in terms of the structural effects of different connections within a network of users (Aggarwal, 2011). Degree centrality σ_D is used to determine the number of direct interactions of a participant within the network, which represents an indicator of quality for the member's interconnectedness. Through the use of an adjacency matrix $A = (a_{ij})$, the degree centrality can be formulized as follows:

$$\sigma_D(x) = \sum a_{ix.} \tag{1}$$

As a consequence, the higher the centrality score $\sigma_D(x)$ is, the more contacts a node x has. Thus, by implementing the degree centrality, we are able to identify those users who have the most interactions with other network participants (Aggarwal, 2011). The closeness centrality σ_C is based on the idea that nodes with a short distance to other nodes can disseminate information very productively in the network. To calculate $\sigma_C(x)$ of a node x, the distances between node x and all other nodes in the network are summed up. By using the reciprocal value, the closeness increases when the distance to another node decreases, i.e., when the integration into the network is improved. The closeness centrality can be formulized as follows:

$$\sigma_C(x) = \frac{1}{\sum_{i=1}^{n} d_G(x, i)} \tag{2}$$

In this respect, through the implementation of the closeness centrality, we are able to identify those users who distribute information among other network participants as efficiently as possible (Latora & Marchiori, 2007). In case of the third centrality measure—the betweenness centrality σ_B—a network member is well connected if it is located on as many of the shortest paths as possible between pairs of other nodes. The underlying assumption of this centrality measure is that the interaction between two non-directly connected nodes x and y depends on the nodes between x and y. The betweenness centrality for a node x can therefore be formulized as

$$\sigma_B(x) = \sum_{i=1, i \neq x}^{n} \sum_{j=1, j<i, j \neq x}^{n} \frac{g_{ij}(x)}{g_{ij}} \tag{3}$$

with g_{ij} representing the number of shortest paths from node i to node j, and $g_{ij}(x)$ denoting the number of these paths which pass through the node x. Through this, we are able to identify those situated on the shortest path distance between various actors, showing that a user has fast access to and control over network flows (AlFalahi et al., 2014; Freeman & Soete, 1997). By these centrality measures, we are able to subdivide the users on the basis of their network characteristics. Regarding the calculation of the respective centrality measurements, the well-known and widely used python library *NetworkX* found

application (see Table 3). Besides the plain calculation of the centrality measures of each user, we further adapted the technique to normalize the calculated values into the range of [0;1]. Based on this normalization, it is possible to consolidate the different centrality measures into a single value by calculating their mean. By this, the respective user's position in the network and therefore their ability to enable the flow of information is being represented.

Following the calculation of the individual metrics, the automatic identification of the lead users per phases in the innovation process takes place. Therefore, the result per metric is multiplied by the corresponding weight of the respective phases (see Table 2) and summed up for each specific user. Finally, the calculated sum is divided by the maximum number of points to be achieved (see (4)). Thus, two resulting scores for each user, one each for the two phases in the innovation process, will be generated. These two resulting scores represent the cumulative relevance of a user with respect to the phases in the innovation process. The higher the resulting score for a respective user is, the more highly the user is defined by his characteristics as a lead user for one of the respective phases: "Idea generation" or "Development".

$$\text{score}_i = \frac{\sum (x_i * w_i)}{\sum w} | w = weight; \, x = metric \tag{4}$$

An identification of lead users according to the described procedure enables the determination of users who show particular strength in terms of relevant characteristics such as their influence on other participants within the community, their product related knowledge or their level of activity. In combination with an individual weighting of these characteristics, the two identified phases of "Idea generation" and "Development" are also considered. In addition, with respect to the derived design principle *4. The principle of repeatability*, the artifact is designed in a modular and generic way. Thus, the underlying data and the respective characteristics' weighting can be easily adapted, allowing the identification process to be carried out at any time without further restrictions to e.g., react quickly to changing circumstances such as trends. Finally, as our design principles follow our purpose and scope and found consideration within the designed artifact as described above, the derived meta requirements (see Fig. 2) can be seen as successfully met since they are representing our underlying purpose and scope.

Demonstration, evaluation and discussion

To demonstrate the applicability of the developed artifact – including the identification approach and the corresponding software tool – we have conducted several steps. In order to verify the consideration of the design principles,

the underlying design requirements are examined for their met using our specific use case (see Table 4). Subsequently, the artifact was applied on a real-world kitesurfing dataset to ensure its usability for practice. Further, we conducted interviews with our identified lead users and with an expert from our cooperating partner, a market leader in kite- and watersports to evaluate both the usability and the generated added value for practice.

Review of the identified requirements

In order to verify the derived design principles, we further review whether and how the elicited design requirements of our artifact (see Fig. 2) were met. Therefore, we specify them in more detail in Table 4.

Demonstration of the artifact

In order to facilitate the accessibility concerning the use of the developed artifact, including the designed identification process, all its customizabilities, as well as the monitoring of the analysis, a graphical user interface (GUI) was developed. To ensure the development of a highly responsive, performant and platform-independent interface, the GUI was developed using the standardized and well-known Python library PyQt5. Figures 3 and 4 show the two main interfaces – namely the configuration and the result table view – of the developed GUI.

The configuration (see Fig. 3) represents the initial view when starting the tool and can be used to customize the underlying analysis approach to one's own needs. The layout was designed based on three sections (i)—(iii), following an adaptation of the design principles of Garrett (2010).
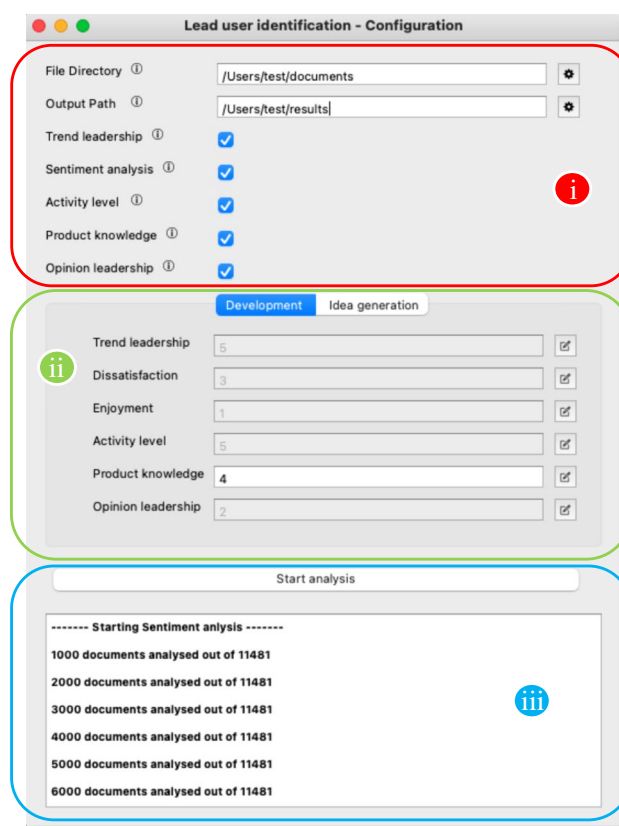

Fig. 3 GUI—Configuration view

In section (i), the user can flexibly specify the data to be analysed as well as the output path for storing the analysis results by selecting the appropriate directories within the native filesystem. Further, in case of not all elicited

Table 4 Met of the previously identified requirements

Design Requirements	Requirements met
Map all prior identified characteristics	As all identified characteristics of lead users are considered within the artifact (see Table 3), we see the requirement as met.
Process a large amount of social media data	The artifact was applied successfully to a real-world dataset of 11,481 contributions (see Table 5). Furthermore, by choosing adequate analysis techniques settled in the field of text mining or SNA respectively (see Table 3), the analysis of larger data sets is easily possible.
Apply objective identification methods	The implemented characteristics were either implemented by well-known text mining or SNA techniques or by self-developed techniques that build on patterns of well-known analysis methods from text mining (see section "Technical realization").
Repeat the identification process for lead users at any time as lead users are trend specific	As the developed artifact is implemented in a generic way, the underlying data and the respective weighting of the characteristics can be easily adapted, allowing the identification to be carried out at any time without further restrictions (see section "Technical realization") to e.g., react quickly to changing circumstances such as trends.
	Additionally, the identification of lead users within the real-world application with 11,481 contributions took 1317 s, which is why a quick identification based on a new dataset is also easily possible.
Identify different lead users regarding the two phases of the innovation process	As the proposed artifact differs between lead users regarding the phase "Idea generation" and lead users regarding the phase "Development", we see the requirement as met.

Fig. 4 GUI—Resulting table view

characteristics being deemed necessary, a subset of them can be individually defined, comprising all specific characteristics relevant to the current circumstances. This ensures that only favoured characteristics are considered in the analysis, resulting in a resource-efficient identification of lead users who show particular strength in terms of significant characteristics such as their influence on other participants within the community (opinion leadership) or their level of activity. In addition, to incorporate the two identified phases "Idea Generation" and "Development", they were implemented modularly using dynamic tabs to enable a distinctive configuration (see Fig. 3, ii). Here, the weightings for each respective phase elicited from literature (see section "Weighting of the according lead user characterizations") are defined as default within the phase's configuration. Nevertheless, to provide maximum flexibility and to be able to react quickly and almost effortlessly to changing circumstances, the pre-defined weighting can also be individualized per phase through the corresponding text fields highlighted in section (ii). The start of the analysis process as well as its monitoring takes place in section (iii). As soon as the process is initiated, all relevant information concerning the process such as the current state or occurring errors will be monitored and logged within the designated text area (see Fig. 3, iii).

Once the process has finished, the results will be consolidated and displayed by a responsive, sortable table (see Fig. 4). Here, the results are subdivided into each characteristic as well as the two calculated phase-specific scores, featuring the dispensation of the users to each characteristic or innovation phase, respectively. To facilitate the selection of relevant lead users, it is further possible to filter the identified users based on each calculated value (see Fig. 4, "Product related knowledge"). This allows companies to select users in an intuitive way based on either a specific

characteristic such as product-related knowledge or the overall scores.

To preserve the obtained results for later usage, two functions were implemented to handle the extraction by use of either a Microsoft Excel or a Comma-separated value (CSV) file. These file formats enable a platform-independent presentation of the results for e.g., marketing campaigns (Excel) as well as the automated processing by a proprietary third-party system, such as importing the generated information into the company's active directory (CSV).

To demonstrate the practical applicability of our developed tool, a representative real-world dataset was needed. Thus, we extracted data from one of the most popular online communities for kiteboards (https://www.seabreeze.com.au/), which comprises a total of 11,481 contributions of 945 users. The data were extracted using the ParseHub extraction tool and span the period from January 1st, 2018 to April 10th, 2020.

Based on these data, the analysis was undertaken to identify the respective lead users. Table 5 presents the top five identified users per phases of the innovation process. The values represent the previously identified characteristics by a normalization within a scale of [0;1]. Thus, a high value implies the respective characteristic is strongly distinctive. The identified users are differentiated regarding the two innovation phases. Accordingly, the weights of the characteristics were adapted to the respective needs of the phase (see section "Weighting of the according lead user characterizations"). The "Overall score" represents the affiliation of the respective user in each phase and is determined through the weighting of the characteristics.

A cursory glance at Table 5 reveals that lead users can be identified in both phases of the innovation process. Thus, the two identified lead users: user #1 and user #2 are identified as lead users exhibiting the highest values compared

Table 5 Top five identified lead users for the specific innovation phases

			Enjoyment	Dissatisfaction	High product related knowledge	Trend Leadership	High Activity Level	Opinion leadership	Overall Score
	User	Weight	1	3	4	5	5	2	
Idea Generation	user #1		0.347	0.653	0.367	0.489	1.0	1.0	*0.662*
	user #2		0.395	0.605	0.438	0.374	0.84	0.925	*0.602*
	user #3		0.346	0.654	0.506	0.304	0.353	0.81	*0.472*
	user #4		0.295	0.705	0.291	0.291	0.497	0.743	*0.46*
	user #5		0.364	0.636	0.36	0.3	0.381	0.79	*0.442*
	User	Weight	1	2	4	5	3	2	
Development	user #1		0.347	0.653	0.367	0.489	1.0	1.0	*0.6*
	user #2		0.395	0.605	0.438	0.374	0.84	0.925	*0.557*
	user #6		0.119	0.881	1.0	0.204	0.005	0.252	*0.51*
	user #7		0.611	0.389	0.92	0.24	0.002	0.257	*0.489*
	user #8		0.59	0.41	0.8	0.203	0.083	0.632	*0.478*

to all users of the innovation process. We assume that the identification of users present in both innovation phases is a rarely occurring exception resulting from extremely conspicuous users. Here, the two identified users have a significantly higher activity level than the lead users identified for a specific phase, which supports this conclusion. User #1 even has the highest activity level (1.0) among all 945 users. In addition to the identification of users who are present in both phases, lead users, who differ significantly in their descriptive characteristics, were further identified for each innovation phase. User #6 e.g., shows an activity level way below average (0.002), but exhibits remarkable results regarding the presence of product related knowledge (0.92). Therefore, the user is considered as lead user regarding the second innovation phase "Development".

To be able to identify lead users scaled to the different circumstances of enterprises in a resource-optimized way, a high degree of generalizability was considered in the implementation of the artifact. Therefore, to be able to adequately react to specific circumstances, the weighting of the respective characteristics can be individualized at the beginning of the analysis process. Thus, the identification approach can be specifically geared to users who dominate a single criterion or a combination of criteria and can therefore be easily adapted to different conditions.

Finally worth to mention, the related lead user characteristics (and therefore the lead users themselves) are validated in an intrinsic way by incorporating different evaluation metrics (e.g., topic coherence) during the identification process. In this way, a high information quality is ensured, supporting the practical applicability of both, the identification process and the retrieved lead users. In this regard, by applying our tool, we revealed promising lead users for the specific innovation phases based on their remarkable characteristics

exhibited (see Table 5, e.g. user #1). However, as the intrinsic evaluation of probabilistic models such as topic modeling (trend leadership) poses various challenges and drawbacks (Chang et al., 2009), it is not sufficient to verify the elicited results. Thus, we evaluate the identified lead users and their characteristics in an extrinsic way by verifying the identified lead users through an interview with an expert of our cooperating partner (a market leader in kite- and watersports) and respective lead users (see section "Evaluation of the artifact"). Therefore, we will evaluate the derived lead users as well as the identification process by applying them to our specific use case, revealing their meaningfulness and potential regarding their practical applicability in a first step.

Discussion of the results of demonstration

Our results have shown that our identification approach and the corresponding software tool are working immaculately. The implementation of the design principles was thus feasible, resulting in the identification of lead users for both innovation phases (see Table 5). Prior research literature is inchoate here as only a minority of the investigations incorporates an innovation process but identifies a lead user for all of the phases in an innovation process (Miao & Zhang, 2017; Pajo et al., 2017). Thus, we provide a new approach that identifies different lead users for every phase of the innovation process. Most of the identified lead users are better suited for one of the two phases but there are also lead users who exhibit very high values in both innovation phases. We have shown with our results that a clear differentiation of the two phases as well as the separated identification and consideration of lead users is necessary as they have different competencies, characteristics and application areas.

The lead users #3 or #4 are according to our results an adequate choice when searching for a lead user in terms of the innovation phase "Idea generation". User #3 e.g. features a high value in the dimensions "trend leadership" (0.304) and "high activity level" (0.353). This means that this lead user can be seen as an active member in the kitesurfing-lifestyle scene. His/her creative and active participation in the online discussions shows that this lead user is highly involved in the kite community. Their active participation and involvement additionally lead to the awareness of unmet needs about existing solutions in the kiteboarding scene (see dissatisfaction: 0.654). Because of his/her high value in the dimensions "trend leadership" (0.304), we can assume that this lead user is able to "translate" his/her dissatisfaction into concrete ideas. Against the background of the fact that a company requires many initial ideas from lead users (as only a few of them can be realized anyway) especially the requirement "repeat the identification process for lead users at any time" is important in this first innovation phase. Approaches that are established and discussed in prior research literature such as screening (Brandtzaeg et al., 2016; Hung et al., 2011; Tuunanen et al., 2011), pyramiding (Von Hippel et al., 2009) and other lead user identification procedures are often based on interviews or (online) questionnaires (Brandtzaeg et al., 2016; Hung et al., 2011; Tuunanen et al., 2011) which makes it almost impossible to repeat the identification process for lead users at any time. However, most business-to-consumer industries are fast-moving and therefore identifying innovative lead users and their ideas repeatedly with little expenditure of time must be focused on. Thus, with our artifact a company is able, on the one hand, to identify lead users who are currently ahead of trends and, on the other hand, to react to changing circumstances such as trends. Furthermore, with this procedure we also counteract the low sample efficiency and the high costs that results from conducting interviews and online questionnaires.

By examining the detailed results of the phase "Development" it is especially interesting that user #6 and user #7 are determined as lead users although they exhibit very low values in the dimension "high activity level" (user #6: 0.005; user #7: 0.002). This shows that due to the medium weight of the dimension "high activity level" in the phase "Development" a lead user does not necessarily exhibit a high active usage behavior which contradicts the results of Martínez-Torres (2014). This means that in our case the lead users #6 and #7 posted only few contents and therefore they do not participate a lot in these online community discussions. But if they did submit a post, it contained very valuable and detailed content including high product-related knowledge (user #6:1.0 user #7: 0.92). These users suggested concrete solutions based on their broad expertise about the products, the components and how they mesh with each other.

The combination of this "high product related knowledge" with the relatively high value for "trend leadership" (0.204) and the simultaneously low value for the "high activity level" (0.005) led to the assignment of user#6 to the phase "Development". The results have shown that this combination of characteristics is more decisive than, for example, the dimension "dissatisfaction" as user #6 is the lead user (compared to all other identified users) with the highest value in the dimension "dissatisfaction" (0.881). This high value of "dissatisfaction" would actually speak for being assigned to the "Idea generation" phase as it is weighted higher (Idea generation: 3, Development: 2) here. However, the high level of product related knowledge and the associated ability to suggest concrete solutions for new products or their improvements is the main factor for a promising cooperation in terms of the "Development" phase. Therefore, observing both the individual results of the characteristics and the overall score has shown that not only is the weighting and selection of the characteristics plausible, but that our software tool is also capable of finding them.

Moreover, we have identified lead users for both phases as they exhibit the highest overall scores: user #1 and user #2. Both users exhibited extraordinary results, especially for the dimensions "high activity level" (user #1: 1.0; user #2: 0.84) and "opinion leadership" (user #1: 1.0; user#2: 0.925). So, their high level of involvement in the kitesurfing scene is characterized by their active participation as well as by their central position in the network. Displaying both strong social relationships and high levels of engagement enable the users to spread information in the online community. Consequently, these users know the overall sentiment and can also identify unmet needs that will later be experienced by the public (see trend leadership and dissatisfaction). Based on this, they formulate and disseminate ideas for new products as well as suggest detailed solutions for the prior identified needs. Thus, they facilitate effective innovation and encourage innovation communication. Only in some dimensions such as "high product related knowledge" other users (e.g. users: #6, #7 and #8) exhibit better results. Nevertheless, the users' overall scores in both phases show their outstanding position as lead users, which we assume, however, that this can be seen rather as an exception. Furthermore, we are convinced, that it makes sense not only to focus on lead users who are suitable for both innovation phases but rather consider for what purpose a lead user should be engaged and to adjust the weighting accordingly. If a lead user is only active in one phase, then s/he can focus on either the objective generating many good and innovative ideas (see Idea generation) or on the objective developing explicit solutions for unmet needs (see Development) and applies his/her strengths accordingly. In other words, if a lead user would

act in both phases and thus focuses on both objectives simultaneously, e.g. knowledge about the realization and development could have a negative impact on the creativity in the idea generation phase. Concentrating on lead users for both phases simultaneously also means that lead users who have extraordinary new ideas but only less product-related knowledge would be excluded. This would lead to a loss of new and potentially successful ideas.

In summary, observing both the individual results of the characteristics and the overall score has shown that not only is the weighting and selection of the characteristics plausible, but that our software tool is also capable of finding them. Our results follow from combining, weighting and considering all relevant lead user characteristics. Previous research literature (e.g., Miao & Zhang, 2017; Martínez-Torres, 2014; Tuarob & Tucker, 2014) has often focused on only few characteristics for identifying lead users and thus resulted in many and ambiguous lead users.

Evaluation of the artifact

To evaluate the generated artifact and our results, we conducted both an in-depth interview with the head of marketing of our cooperating partner and interviews with some of our identified lead users.

The in-depth interview was undertaken by two researchers, recorded, transcribed, and finally condensed to the most important statements. It lasted approximately two hours and we aimed to investigate the artifact's applicability and its generated added value. In the course of the interview, we presented the expert both an excerpt of our results and randomly selected posts from the identified lead users. Thus, we wanted to find out if he can benefit from the lead users identified by the tool and if the expert agrees with the differentiation of the user types. Accordingly, by analyzing the selected posts as well as the excerpt of our results, the interviewee stated first that by means of the software tool he is now able to analyze a huge amount of social media data. Previously, his team only analyzed social media data by hand, which cost a lot of resources (e.g. time, human resources, etc.) and often led to incorrect results.

Second, the expert highlighted the distinction to be beneficial as it allows him to address users for different stages in the innovation process. Ultimately, he was aware of the two lead users that were identified by the tool. As he knew them in advance, he has already incorporated them successfully into the company's innovation process. However, there are also lead users, and therefore also innovative ideas and content, that were unknown to him so far. To reveal what the users are talking about we discussed randomly selected posts with him. Hence, the aim was to detect whether both the analysis of posts in the context of the innovation process and the differentiation of the lead users made sense from

the practitioner's point of view. Consequently, by analyzing and discussing the provided posts, our interviewee already detected some ideas and suggestions for new products or for variations of existing ones. He contemplated the involvement of the users in the company's internal brainstorming/idea-finding process to talk about ideas for new products or about drawbacks of pre-existing ones. Therefore, he is committed to include the dissatisfaction in the identification approach here. To get a better understanding of how the expert came to this decision, we included a short excerpt of a selected post as an example:

> *"[…] The male velcro is facing the wrong way, which means its going to chaff like 60 grit sandpaper if you don't wear a thick rashie or wetsuit. […]"*

According to the interviewee, from this short excerpt, it can already be recognized that the user is pointing to a certain problem (*"male velcro is facing the wrong way"*) and therefore identify an unmet need. Accordingly, the expert identifies here a starting point for improving a specific product. The expert added that the selected posts also show that these identified lead users seem to be "*passionate individuals who are true ambassadors of the kitesurfing scene*" and can therefore be auspicious, prospective partners for the company's internal idea-finding process with the aim of uncovering existing problems, unmet needs and new product ideas.

Furthermore, after showing the interviewee the posts that our tool assigned to the phase "Development", he was enthusiastic about the high level of product-related knowledge. According to our interviewee this and the positiveness of user #7 could, for example, support the engagement of a promising cooperation regarding the development of new products. Including a lead user means the incorporation of the user's vast experience and knowledge. Thus, this cooperation could potentially lead to decreasing failure rates in product innovations. For demonstration purposes we also included here a selected post of a lead user. To exclude potential influence, mentioned competitors were removed here.

> *"Yes it is[,] if you like to ride powered, the early edition was a fave of mine[.] The next imho was completely different, lost all the flex that you want in choppy conditions and a lot of feel by going heavier on build. The newbronq crb 4 is back to the original, been riding the ts from [...], great board but [...] struggle [to] make a non spray board especially in chop conditions, in flat water it isn't an issue, the monk will cover most riders as mentioned. Demo I'd say"*

Phrases such as "*The newbronq crb 4 is back to the original*" led the expert to the assumption that this user exhibits a high product-related knowledge and thus is able to formulate

precise solutions, which are both indicators for assigning this post/user to the "Development" phase. From the analysis of the posts the expert drew the conclusion that these lead users identified here, represent "*progressive riders who are continuously pushing their riding and the sport to new levels*". Therefore, the expert referred to the fact that these users propose to apply new materials and technologies to create constantly better performing products. Moreover, the interviewee stated that not only the identification of promising lead users will be supported with the tool but also analyzing the posts and contents of the respective users, which represents another added value for him.

He further noted that because of these different application areas, it is very constructive to identify lead users regarding the two innovation phases and therefore differentiate between them. The interviewee had already involved some kitesurfers in the company's (innovation) processes and therefore drew the conclusion that he would definitely involve the two lead users who exhibit the highest overall scores in both innovation phases (user#1 and user#2) but he also highlighted that it would be reluctant to focus only on them. He reasoned as follows: First, it is advisable for a company to include more than just two lead users so that multiple perspectives can be included in the company. Second, according to his experience, when a user is part of each innovation phase, the generation of ideas is inhibited if the user always keeps the development and its boundaries in mind. This would limit the venue for brainstorming that should be ensured in the "Idea generation" phase. Moreover, a user can have interesting new ideas, but s/he has too little product related knowledge to implement and develop them. This would also lead to a loss of new and potentially successful ideas. Furthermore, our interviewee noted that the lead users #1 and #2 exhibited good overall scores (and thus he would definitely include them), but he tended to prefer to incorporate the user with the highest level of product-related knowledge (user #6) in the "Development" phase. Overall, the expert confirmed that a clear differentiation between the two phases as well as a separated identification and consideration of lead users is necessary.

In addition to the evaluation with the expert, we discussed our results with three of the identified lead users. To strengthen the results of our tool and to make sure that the identified users are the appropriate lead users for the particular innovation phase we also examine the lead user perspective by conducting short interviews. This should also confirm and, where appropriate, extend the characteristics that we have found in literature. In doing so, the interviewees all confirmed being lead users in this online community and they postulated that they are all aware of making major contributions to idea generation and/or product development in line with our results. In addition, in surprising harmony they all mentioned similar characteristics (enthusiasm for

the sport, high activity level in the online community and experience in the field) as essential for lead users. Only on the time needed to be an experienced kitesurfer there was no agreement. Two of the respondent lead users stated a minimum of 3 and 5 years to be an experienced kitesurfer. The remaining one quantified the respective time by the kitesurfing sessions performed and different kitesurfing locations visited.

With these statements our lead users supported the results of our research and in consequence validated the applicability of our approach. Furthermore, the essential characteristics fit to those we have found in literature and thus confirmed the characteristics we have included in our tool. The characteristic "enthusiasm for the sport " is implemented with "opinion leadership" in our approach. According to the current research literature "opinion leadership" is the ability to enable the flow of information and especially to diffuse it. Strong social relationships and a high level of engagement are premises for a functioning exchange of ideas and innovation (Pajo et al., 2014, 2017). Accordingly, a user who is motivated to build relationships in the community and thus exhibits high centrality scores is highly enthusiastic about the sport. The high activity level in the online community, calculated in our tool by the number of posts and transacted comments per user within the analyzed period, represents the second characteristic the lead users have mentioned. The "Experience in the field " can be partly covered by our characteristic "high product-related knowledge". However, the number of training hours, e.g., could also be included here.

Discussion of the results of evaluation

The evaluation has revealed not only the applicability but also the added value of our artifact in a practical environment. Thus, the in-depth interview with the head of marketing of our cooperating partner has highlighted that he is enthusiastic about the results as he benefits from them in many ways. First, the interviewee was able to assign the innovation potential to many posts by recognizing trends that were talked about in the posts, months before their realization. Moreover, the content of the comments has already made him aware of ideas on how to improve certain products in the company. Second, our expert has also confirmed that the high level of product related knowledge, vast experience and the associated ability to suggest concrete solutions for new products or their improvements is the main factor for a promising cooperation with a lead user in terms of the "Development" phase. The lead users identified with our tool have suggested concrete solutions based on their broad expertise about the products, the components and how they mesh with each other. As our expert has confirmed, this high product related knowledge can lead to decreasing failure rates in new product introductions or improvements

because these users are aware of very specific facts such as every tiny change to a kite's profile can have enormous impact on its flight characteristics. Third, our expert also stated that he would be reluctant to focus only on the lead users who exhibit the highest overall scores in both innovation phases. He further highlighted that he would prefer in the "Development" phase the user with the highest level of product-related knowledge (user #6) and not user #1 or #2 who have higher overall scores. All in all, he confirmed that a clear differentiation of the two phases as well as a separated identification and consideration of lead users are necessary.

Finally, the interviews with the lead user confirm our approach and implicated further interesting perspectives and provided indications on how our approach can be further refined. In future research, these and other possible aspects and characteristics mentioned by the users have to be evaluated additionally.

Contribution for practice and research

Our investigation contributes to research and practice alike. As a contribution to practice, first companies can benefit from our comprehensive and modular approach. By applying our approach companies can resource-efficiently identify lead users which is an important process as the acquisition as well as the transferring costs of the information that are decisive for initiating innovation have tremendous influence on where innovation is created (Idota, 2019). Therefore, as lead users feature highly sticky information and are able to create innovations, organizations benefit from including them in their innovation process in order to overcome their information stickiness and so get to know the user's needs to solve (product) problems and reduce R&D costs. Thus, we stand out against other approaches that follow more resource intensive approaches (e.g. Brandtzaeg et al., 2016; Hung et al., 2011; Tuunanen et al., 2011).

Second, we created an artifact, respectively a tool, that is able to process a large number of social media data which can be repeated at any time as lead users are trend specific. This counteracts i.a. weaknesses of previous approaches that include only a small amount of data in the identification (Hau & Kang, 2016). By means of our tool, companies are able to start and monitor the current state of the identification process, display the analysis results by an intuitive, sortable table to easily enable either the selection of the overall lead users by the respective overall-scores or specific lead users by their identified results of an explicit characteristic and extract and persist the elicited results to various file formats (Excel, CSV) for later usage.

Third, a high degree of generalizability was taken into account to identify a lead user by considering several characteristics depending on the different circumstances of different companies. Thus, they are able to customize the identification process to their own needs by uploading their own dataset and applying all or a selected set of characteristics either following our pre-defined weights for each of the two innovation phases or individualize them as well. Hence, the weight of the respective characteristic is determined in the beginning of the analysis process. The identification process therefore can be specifically geared to users who dominate a single criterion or a combination of criteria. When a company, for instance, wants to focus more on lead users who express a sentiment of enjoyment in the innovation process, then the company can set the weights for dissatisfaction very low (or even to zero) and for enjoyment very high. Thus, we created an extensive, flexible, and resource-saving approach which can be easily applied by companies and which is based on objective traceable characteristics (different to other approaches that include self-assessment of respondents (Hienerth & Lettl, 2017)).

Fourth, the evaluation of our results has shown their contribution for our cooperating partner and therefore for practice. As the expert has highlighted, the tools enable him to turn away from analyzing social media data by hand, which costed a lot of resources (e.g. time, human resources, etc.) and often led to incorrect results. Further he identified lead users and therefore also innovative ideas and content, that had been unknown to him so far. So, our tool also allows to analyze posts and contents of the respective users and is thus able to detect new ideas and suggestions for new products or for variations of existing ones. For practice this can mean decreasing failure rates in product innovations.

In summary, companies aiming to identify different lead users for the particular phases in the innovation process can benefit from our comprehensive and modular artifact, since they are enabled to autonomously analyze large amounts of data and therefore automatically identify respective lead users adapted to the corporate's specific circumstances. Thus, we automated the lead user identification process, the most difficult and time-consuming aspect within the lead user method (Brem & Bilgram, 2015).

Furthermore, as outcomes of our DS research project we achieved theoretical contributions to research that go beyond the technical contribution (i.e., the artifact) and which are explained in more detail in the following. In terms of the DSR knowledge contribution framework of Gregor and Hevner (2013) the enhancements of our artifact over existing ones in the literature can be classified in the group of improvement (development of new solutions for known problems). DSR improvement projects make contributions to both prescriptive theory i.e. Design Theory (Gregor, 2006) and descriptive theory i.e. kernel theories (Gregor & Hevner, 2013). Thus, first, in terms of prescriptive theory our artifact contributes to a rather general and abstract knowledge

base – "nascent design theory" (Gregor & Hevner, 2013). Therefore, design principles based on kernel theories and resulting design requirements were formulated and proposed in the section "Design principles for a lead user identification tool". By applying them in the course of the design and development of the artifact followed by the demonstration and evaluation, an implicit empirical grounding of the design principles was achieved here (Heinrich & Schwabe, 2014). Our design principles capture design-related knowledge and can therefore support the development of further IS (design) theories and new artifacts. For designing further (identification) tools in related areas our design principles can be applied as we have formulated them generally by describing what the artifact should enable users to do and how the artifact should be built. By considering e.g. the design principle 3. Contextual adaptability, the importance of the context is highlighted in which the respective tool should be created. Since the context has a direct impact on the definition and implementation of the requirements, the alignment with the context also will lead to a more targeted identification tool. So, with the compilation of the design principles, we made a first step towards contributing to Design Theory in terms of theory for design and action (Gregor, 2006) as we comply with conditions as March and Smith (1995) and Hevner et al. (2004) pointed out under which a contribution to knowledge in DS has occurred: utility to a community of users, the novelty of the artifact and the persuasiveness of claims that it is effective. To take a next step towards mature Design Theory, according to Gregor and Jones (2007), a total of eight components are necessary. We have shown the "Purpose and scope" by means of the meta-requirements and the "Principles of form and function" by means of the design principles (see both Fig. 2). Furthermore, the latter is based on kernel theories (Lead User Theory and Innovation Theory) which entails the inclusion of a further component – "Justificatory knowledge". Also, the "Constructs" that are described as the most basic levels of the theory are involved with the alignment to the characteristics of a lead user and the two phases of the innovation process. These components resulted in the "Expository instantiation", i.e. in the application of the designed tool in a real world setting. However, the inclusion of the "artifact mutability", the "testable propositions" and the "principles of implementation" explicitly and aligning the investigation on these eight components in general, as for example Böckle et al. (2021) have done, would need to be undertaken as a next step towards a mature contribution to Design Theory.

Beside that, our results also contribute to the kernel theory knowledge base regarding the social media theory as well as the innovation related theory. Moreover, our results contribute to different kernel theories by providing the following useful implications which previous investigations have barely considered until now. First, this study sheds a new light on the lead user construct itself – the core of Lead User Theory – as our investigation has shown that it is meaningful to differentiate lead users according to the different innovation phases as they have different competencies, characteristics and application areas. Until now, no distinction has been made in defining and characterizing lead users in terms of the innovation process. The basic model of Lead User Theory (Von Hippel, 1986) indicates indeed a distinction between lead users, but only against the background of whether the product innovation supported by the lead user is a novelty or one that requires commercially viable modifications and enhancements (Von Hippel, 1986). Our results highlighted that a separated consideration implicates a more targeted identification. If a lead user is active in one phase, then s/he can focus on either the objective generating many good and innovative ideas (see Idea generation) or on the objective developing explicit solutions for unmet needs (see Development) and applies his/her strengths accordingly. When lead users are examined and identified separately in the two phases, the generation of ideas is not inhibited by keeping the development and its boundaries in mind. Additionally, our approach takes also into account who have extraordinary new ideas but only less product-related knowledge and would therefore be excluded from prior identification approaches. Thus, our approach contracts a loss of new and potentially successful ideas. So, our study has revealed a new point by defining a lead user against the background of the purpose of its use (based on the innovation process) whereby we introduce a new dimension to the Lead User Theory. This can constitute an important new implication which includes that the definition of a lead user should not only focus on Von Hippels' characteristics but also on the purpose of its use (Von Hippel, 1986).

Second our investigation contributes to the process of utilizing lead users included in the Lead User Theory. Von Hippel (1986) introduced a 4-step process – which has often been taken up in other studies (cf. Hung et al., 2011) –, including (1) the identification of an important market or technical trend, (2) the identification of a lead user leading that trend, (3) analyzing the lead user need data and (4) project lead user data onto the general market (Von Hippel, 1986). Our approach and results have shown that the (1) identification of a trend before (2) identifying a lead user for that respective trend is no longer deemed necessary as the identification of trend(s) can be included within the identification of corresponding lead users. Thus, the initial step (1) identification of trends is no longer considered as a necessary sequential premise for the (2) lead user identification, since the emerging trend is identified and considered simultaneously, resulting in a more flexible and easier-to-use process. Moreover, our approach provides the opportunity to consider multiple trends simultaneously, rather than being limited to one previously identified trend (Von Hippel,

1986). Therefore, multiple trends reflected in the underlying data can be dynamically considered when identifying lead users, enabling the identification of target-oriented lead users associated with each trend. Thus, the 4-step process can be enhanced in terms of its applicability and ease of use by enabling the automated identification of underlying trends when identifying accompanying lead users, as well as in terms of its functional scope by including multiple trends instead of solely considering the previously, manually identified trend.

Third, this study sheds another light to Lead User Theory and contributes to the automated identification of lead users in online communities in more specific (and thus a further contribution to the identification process). With this work at hand, we provide initially a comprehensive and structured overview of lead user characteristics based on the current research literature. Beyond that, we further technically realized these characteristics by means of an adaption of several machine learning methods (see section "Technical realization") and enriched the related Lead User Theory by establishing synergies of these research areas. Thus, future research in Lead User Theory will benefit from the advantages of automated analysis techniques and will therefore be supported by our concrete techniques for the identification of lead user characteristics. In addition, we distinguish ourselves from investigations that define and identify lead users by including only one or two characteristics (cf. Miao & Zhang, 2017; Tuarob & Tucker, 2014; Tuunanen et al., 2011), as our identification process enables an identification of lead users considering all identified characteristic. This enables the consideration of each relevant characteristic, allowing lead users to be identified in a more target-oriented and fine-grained manner. Moreover, to take a step further in the identification of respective lead users and in order to account the differentiation of them in the innovation process, we have adapted the identification process to incorporate priorities (weights) regarding the characteristics with respect to the different innovation phases. Consequently, contrary to the current research literature which treats all characteristics equally, we assign different weights to different lead user characteristics in the course of the identification process to make this process even more targeted.

Finally, for innovation theories our research identified relevant characteristics of users who can contribute to the different stages of the innovation process. Our results have shown that it is important to consider for what purpose a lead user should be engaged and to adjust the weighting of the identified characteristics accordingly. This has implications for the theories dealing with the process of innovation, e.g. the stage-gate model. By applying specific lead users within the stages preliminary and detailed investigations as well as in development, testing and validation, the rigid sequence of stages and gates can be broken up. By integrating the user's

and therefore the external point of view the assessments at the go/kill checkpoints (i.e. gates) become less elaborate as the alignment with the external requirements is maintained constantly. This results in a more agile and target group-oriented approach. Based on this, the innovation process must be specified more concretely in terms of interactive value creation, especially the open innovation. Thus, including different lead users adds new tasks for companies in the innovation process. These different lead user types can be taken into account by introducing process variants.

Conclusion

In the existing literature there are a lot of different lead user identification approaches, but these investigations only covered a limited point of view as they either focus on only a few lead user characteristics (Martínez-Torres, 2014), include a very small amount of data (Hau & Kang, 2016) or base their approach on the self-assessment of users (Hienerth & Lettl, 2017). This problem is further compounded by the tremendous amount of online community data which makes it even more difficult, costly and time-consuming to identify lead users. We approached this research gap by introducing an automated and – according to our interviewed expert – effective method for identifying lead users. After consulting the research literature, we focused on two main phases of the user innovation process (A) the "Idea generation" of an innovation and (B) the "Development" of an innovation. In both cases (A and B), a lead user is a valuable resource for companies. Furthermore, we have demonstrated that six different characteristics (trend leadership, dissatisfaction, enjoyment, high level of activity, product related knowledge, opinion leadership) are prevalent in existing research literature regarding lead user identification in online communities (see RQ1). Based on this, we further designed and implemented a tool that, on the one hand, combines all of the aforementioned characteristics and, on the other hand, considers the fact that lead users can be applied in different phases of the innovation process (see RQ2). To demonstrate the applicability of our artifact we applied it to 11,481 contributions of 945 users from a popular online forum for kiteboarding. After identifying the lead users, we evaluated our results by interviewing the respective lead users as well as an expert. In conclusion, following the DS research, we derived numerous contributions for both, theory (kernel theories: Innovation- and Lead User Theory; Design Theory: Design Principles) and practice (e.g., the artifact) (see RQ3).

In the previous section (see "Contribution for practice and research") we have shown that companies can benefit from our comprehensive and modular artifact, with which large amounts of data can be analyzed adapted to the corporate's specific circumstances with the aim of identifying

different lead users for the particular phases in the innovation process. Thus, we automated the lead user identification process, the most difficult and time-consuming aspect within the lead user method. Furthermore, we have highlighted how our investigation made a first step towards contributing to Design Theory (theory for design and action (Gregor, 2006)) by formulating four design principles. These design principles (comprehensive characteristics consideration, using inter-subjectively verifiable identification methods, contextual adaptability and repeatability) can support the design of further user identification tools. Beside that we also highlighted our contribution to the kernel theories: Our study has revealed a new point by defining a lead user against the background of the purpose of its use (based on the innovation process) whereby we have introduced a new dimension to the Lead User Theory. Moreover, we enhanced Von Hippel's 4-step lead user utilization process in terms of its applicability and ease of use by enabling the automated identification of underlying trends when identifying accompanying lead users, as well as in terms of its functional scope by including multiple trends instead of solely considering the previously, manually identified trend. Finally with respect to the current Lead User Theory which treats all characteristics equally, we have assigned different weights to different lead user characteristics in the course of the identification process to make it even more targeted. Regarding the Innovation Theory the rigid sequence of stages and gates can be broken up and further parallelized by applying specific lead users within the different stages. However, our research is not without limitations. We have identified the characteristics that are decisive for a lead user in the current research literature. It is possible that there are further characteristics distinctive for a lead user that we have not considered. During our research we came upon areas of further research. In terms of a further evaluation of our results, we are intent on carrying out a study to assess the completeness and usefulness of our approach with other cooperating partners. Further, as the users noted in the interviews that experience in the respective field of application is important and we only partially cover this with the characteristic "product-related knowledge", the question "At what point can a lead user be seen as experienced?" may be subject of future work.

References

Aggarwal, C. C. (2011). An introduction to social network data analytics. In C. Aggarwal (Ed.), *Social network data analytics* (pp. 1–15). Boston. https://doi.org/10.1007/978-1-4419-8462-3_1

Aggarwal, C. C., & Zhai, C. (2012). A survey of text classification algorithms. In *Mining text data* (pp. 163–222). Boston. https://doi.org/10.1007/978-1-4614-3223-4_6

Al-Zu'bi, Z. B. M., & Tsinopoulos, C. (2012). Suppliers versus lead users: Examining their relative impact on product variety. *Journal of Product Innovation Management, 29*(4), 667–680. https://doi.org/10.1111/j.1540-5885.2012.00932

AlFalahi, K., Atif, Y., & Abraham, A. (2014). Models of Influence in Online Social Networks. *International Journal of Intelligent Systems, 29*(2), 161–183. https://doi.org/10.1002/int.21631

Amabile, T. M. (1988). A model of creativity and innovation in organizations. *Research in Organizational Behavior, 10*(1), 123–167.

Autio, E., Dahlander, L., & Frederiksen, L. (2013). Information exposure, opportunity evaluation, and entrepreneurial action: An investigation of an online user community. *Academy of Management Journal, 56*(5), 1348–1371. https://doi.org/10.5465/amj.2010.0328

Axtell, C. M., Holman, D. J., Unsworth, K. L., Wall, T. D., Waterson, P. E., & Harrington, E. (2000). Shopfloor innovation: Facilitating the suggestion and implementation of ideas. *Journal of Occupational and Organizational Psychology, 73*(3), 265–285. https://doi.org/10.1348/096317900167029

Belz, F. M., & Baumbach, W. (2010). Netnography as a method of lead user identification. *Creativity and Innovation Management, 19*(3), 304–313. https://doi.org/10.1111/j.1467-8691.2010.00571.x

Böckle, M., Bick, M., & Novak, J. (2021). Toward a design theory of user-centered score mechanics for gamified competency development. *Information Systems Management,* 1–27. https://doi.org/10.1080/10580530.2021.1975852

Boyd-Graber, J., Mimno, D., & Newman, D. (2014). Care and feeding of topic models: Problems, diagnostics, and improvements. In E. M. Airoldi, D. Blei, E. A. Erosheva, & S. E. Fienberg (Eds.), *Handbook of Mixed Membership Models and Their Applications* (pp. 3–34). CRC Press.

Brandtzaeg, P. B., Haugstveit, I. M., Lüders, M., & Følstad, A. (2016). How should organizations adapt to youth civic engagement in social media? A lead user approach. *Interacting with Computers, 28*(5), 664–679. https://doi.org/10.1093/iwc/iwv041

Breitsohl, J., Kunz, W. H., & Dowell, D. (2015). Does the host match the content? A taxonomical update on online consumption communities. *Journal of Marketing Management, 31*(9–10), 1040–1064. https://doi.org/10.1080/0267257X.2015.1036102

Breitsohl, J., Roschk, H., & Feyertag, C. (2018). *Consumer brand bullying behaviour in online communities of service firms. In: Service business development*. Springer, 289–312. https://doi.org/10.1007/978-3-658-22424-0_13

Brem, A., & Bilgram, V. (2015). The search for innovative partners in co-creation: Identifying lead users in social media through netnography and crowdsourcing. *Journal of Engineering and Technology Management, 37*, 40–51. https://doi.org/10.1016/j.jengtecman.2015.08.004

Brem, A., Bilgram, V., & Gutstein, A. (2018). Involving lead users in innovation: A structured summary of research on the lead user method. *International Journal of Innovation and Technology Management, 15*(03). https://doi.org/10.1142/S0219877018500220

Chandra, L., Seidel, S., & Gregor, S. (2015). Prescriptive knowledge in IS research: Conceptualizing design principles in terms of materiality, action, and boundary conditions. In *48th Hawaii International Conference on System Sciences* (4039–4048). https://doi.org/10.1109/HICSS.2015.485

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems, 22*, 288–296.

Chen, J., Yin, X., & Mei, L. (2018). Holistic innovation: An emerging innovation paradigm. *International Journal of Innovation Studies, 2*(1), 1–13. https://doi.org/10.1016/j.ijis.2018.02.001

Chen, X., Hu, X., Wang, Y., & Tao, D. (2019). Extending Lead Users to Average User Innovation: A Novel Segmentation Framework Based on Users' Innovativeness. *Proceedings of The IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*. https://doi.org/10.1109/ICPHYS.2019.8780274

Cooper, R. G. (1996). Overhauling the new product process. *Industrial Marketing Management, 6*(25), 465–482. https://doi.org/10.1016/S0019-8501(96)00062-4

Crain, S. P., Zhou, K., Yang, S.-H., & Zha, H. (2012). *Dimensionality reduction and topic modeling: From latent semantic indexing to latent dirichlet allocation and beyond. In: Mining text data*. Springer. 129–161. https://doi.org/10.1007/978-1-4614-3223-4_5

Crawford, C. M. (1994). *New products management*. Irwin.

Dahlander, L., & Frederiksen, L. (2012). The core and cosmopolitans: A relational view of innovation in user communities. *Organization Science, 23*(4), 988–1007. https://doi.org/10.1287/orsc.1110.0673

de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: A research agenda. *Journal of Information Technology, 33*(2), 124–135. https://doi.org/10.1057/s41265-016-0033-3

Dong, J. Q., & Wu, W. (2015). Business value of social media technologies: Evidence from online user innovation communities. *The Journal of Strategic Information Systems, 24*(2), 113–127. https://doi.org/10.1016/j.jsis.2015.04.003

Eickhoff, M., & Neuss, N. (2017). Topic modelling methodology: its use in information systems and other managerial disciplines. *Proceedings of the European Conference of Information Systems (ECIS)*. 1327–1347.

Ernst, M., Brem, A., & Voigt, K.-I. (2013). Innovation Management, Lead Users and Social Media: Introduction of a Conceptual Framework for Integrating Social Media Tools in Lead User Management. *Social media in strategic management, Advanced Series in Management, 11,* 169–195. https://doi.org/10.1108/S1877-6361(2013)0000011013.

Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge University Press.

Fisher, G. (2019). Online communities and firm advantages. *Academy of Management Review, 44*(2), 279–298. https://doi.org/10.5465/amr.2015.0290

Franke, N., Von Hippel, E., & Schreier, M. (2006). Finding commercially attractive user innovations: A test of lead-user theory. *Journal of Product Innovation Management, 23*(4), 301–315. https://doi.org/10.1111/j.1540-5885.2006.00203.x.

Franke, N., & von Hippel, E. A. (2003). *Finding Commercially Attractive User Innovations: A Performance Evaluation of the "lead User Construct"*. https://doi.org/10.2139/ssrn.367140

Freeman, C., & Soete, L. (1997). *The economics of industrial innovation*. Psychology Press.

Füller, J., Jawecki, G., & Mühlbacher, H. (2007). Innovation creation by online basketball communities. *Journal of Business Research, 60*(1), 60–71. https://doi.org/10.1016/j.jbusres.2006.09.019

Gallaugher, J., & Ransbotham, S. (2010). Social media and customer dialog management at Starbucks. *MIS Quarterly Executive, 9*(4), 197–212.

Garrett, J. J. (2010). *The elements of user experience: user-centered design for the web and beyond*. Pearson Education.

Gawer, A., & Cusumano, M. A. (2014). Industry Platforms and Ecosystem Innovation. *Journal of Product Innovation Management, 31*(3), 417–433. https://doi.org/10.1111/jpim.12105

Gregor, S. (2006). The nature of theory in information systems. *MIS quarterly*, *30*(3), 611–642. https://doi.org/10.2307/25148742

Gregor, S., & Jones, D. (2007). The anatomy of a design theory. *Journal of the Association for Information Systems, 8*(5), 313–335.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly, 37*(2), 337–355. https://doi.org/10.25300/MISQ/2013/37.2.01

Gregor, S., Chandra Kruse, L., & Seidel, S. (2020). Research perspectives: The anatomy of a design principle. *Journal of the Association for Information Systems*, 21(6). https://doi.org/10.17705/1jais.00649

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In G. Varoquaux, T. Vaught, & J. Millman (Eds.), *Proceedings of the 7th Python in Science Conference* (SciPy2008) (pp. 11–15).

Han, J., Pei, J., & Kamber, M. (2006). *Data Mining*. Elsevier Science.

Hau, Y. S., & Kang, M. (2016). Extending lead user theory to users' innovation-related knowledge sharing in the online user community: The mediating roles of social capital and perceived behavioral control. *International Journal of Information Management, 36*(4), 520–530. https://doi.org/10.1016/j.ijinfomgt.2016.02.008

Herstatt, C. (1999). Theorie und Praxis der frühen Phasen des Innovationsprozesses. *Management Zeitschrift Industrielle Organisation, 68*(10), 80–91.

Heinrich, P., Schwabe, G. (2014). Communicating nascent design theories on innovative information systems through multi-grounded design principles. In M. C. Tremblay, D. VanderMeer, M. Rothenberger, A. Gupta, V. Yoon (eds.), *Advancing the impact of design science: Moving from theory to practice* (vol. 8463, pp. 148–163). DESRIST 2014. Lecture Notes in Computer Science. Cham: Springer. https://doi.org/10.1007/978-3-319-06701-8_10

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian Journal of Information Systems, 19*(2), 87–92.

Hevner, A. R., Salvatore, M. T., Jinsoo, P., & Sudha, R. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75–105. https://doi.org/10.2307/25148625

Heyer, G., Quasthof, U., & Wittig, T. (2006). *Text mining: Wissensrohstoff tex*. W3l, Herdecke, 18

Hienerth, C., & Lettl, C. (2017). Perspective: Understanding the nature and measurement of the lead user construct. *Journal of Product Innovation Management, 34*(1), 3–12. https://doi.org/10.1111/jpim.12318

Hung, C.-L., Chou, J.C.-L., & Dong, T.-P. (2011). Innovations and communication through innovative users: An exploratory mechanism of social networking website. *International Journal of Information Management, 31*(4), 317–326. https://doi.org/10.1016/j.ijinfomgt.2010.12.003

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Paper presented at the Proceedings of the International AAAI Conference on Web and Social Media*.

Idota, H. (2019). Empirical Study on Consumer Innovation by Using Social Media in Japan. In *Paper presented at the UK Academy for Information Systems Conference Proceedings 2019*.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons, 53*(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

King, A., & Lakhani, K. R. (2013). Using open innovation to identify the best ideas. *MIT Sloan Management Review, 55*(1), 41–48.

Lakhani, K. R., Garvin, D. A., & Lonstein, E. (2010). Topcoder (a): *Developing software through crowdsourcing* (January 15, 2010). Harvard Business School General Management Unit Case No. 610-032, Available at SSRN:https://ssrn.com/abstract=2002884

Latora, V., & Marchiori, M. (2007). A measure of centrality based on network efficiency. *New Journal of Physics, 9*(6), 188–199.

Li, Z., & Tang, H. (2016). Identifying Lead User in Mass Collaborative Innovation Community: Based on Knowledge Supernetwork. In *Paper presented at the International Symposium on Knowledge and Systems Sciences*.

Lilien, G. L., Morrison, P. D., Searls, K., Sonnack, M., & Hippel, E. V. (2002). Performance assessment of the lead user idea-generation process for new product development. *Management Science, 48*(8), 1042–1059.

Lüthje, C. (2004). Characteristics of innovating users in a consumer goods field: An empirical study of sport-related product consumers. *Technovation, 24,* 683–695. https://doi.org/10.1016/S0166-4972(02)00150-5.

Mahr, D., & Lievens, A. (2012). Virtual lead user communities: Drivers of knowledge creation for innovation. *Research Policy, 41*(1), 167–177. https://doi.org/10.1016/j.respol.2011.08.006

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems, 15*(4), 251–266. https://doi.org/10.1016/0167-9236(94)00041-2

Marchi, G., Giachetti, C., & De Gennaro, P. (2011). Extending lead-user theory to online brand communities: The case of the community Ducati. *Technovation, 31*(8), 350–361. https://doi.org/10.1016/j.technovation.2011.04.005.

Martínez-Torres, M. R. (2014). Analysis of open innovation communities from the perspective of social network analysis. *Technology Analysis & Strategic Management, 26*(4), 435–451. https://doi.org/10.1080/09537325.2013.851378

Marxt, C., & Hacklin, F. (2005). Design, product development, innovation: all the same in the end? A short discussion on terminology. *Journal of Engineering Design, 16*(4), 413–421.

McCallum, A. K. (2002). A machine learning for language toolkit. *MALLET, 15*(2), 131–136.

Miao, Y., & Zhang, H. (2017). A biclustering-based lead user identification methodology applied to xiaomi. *Proceedings of the Fourth International Forum on Decision Sciences*. https://doi.org/10.1007/978-981-10-2920-2_80

Muller, M., Ehrlich, K., Matthews, T., Perer, A., Ronen, I., & Guy, I. (2012). Diversity among enterprise online communities: Collaborating, teaming, and innovating through social media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2815–2824,*. https://doi.org/10.1145/2207676.2208685

Obar, J. A., & Wildman, S. S. (2015). Social media definition and the governance challenge-an introduction to the special issue. *Telecommunications Policy, 39*(9), 745–750. https://doi.org/10.2139/ssrn.2663153

Pajo, S., Vandevenne, D., & Duflou, J. R. (2017). Automated feature extraction from social media for systematic lead user identification. *Technology Analysis & Strategic Management, 29*(6), 642–654. https://doi.org/10.1080/09537325.2016.1220517.

Pajo, S., Verhaegen, P.-A., Vandevenne, D., & Duflou, J. R. (2014). Lead User Identification through Twitter: Case Study for Camera Lens Products. In *Proceedings of Nord Design 2014*.

Paulus, P. B., Putman, V. L., Dugosh, K. L., Dzindolet, M. T., & Coskun, H. (2002). Social and cognitive influences in group brainstorming: Predicting production gains and losses. *European Review of Social Psychology, 12*(1), 299–325. https://doi.org/10.1080/14792772143000094.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems, 24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Piller, F. T. (2006). User Innovation: Der Kunde kann's besser. In O. Drossou, S. Krempl & A. Poltermann (Eds.), *Die wunderbare Wissensvermehrung: Wie Open Innovation unsere Welt revolutioniert* (pp. 85–97). Heise.

Poetz, M. K., & Schreier, M. (2012). The value of crowdsourcing: can users really compete with professionals in generating new product ideas? *Journal of Product Innovation Management, 29*(2), 245–256. https://doi.org/10.1111/j.1540-5885.2011.00893.x

Preece, J., & Maloney-Krichmar, D. (2003). Online communities. In J. Jacko, & A. Sears (Eds.), *Handbook of human-computer interaction* (pp. 596–620). Lawrence Erlbaum Associates Inc.

Ramaswamy, V. (2010). Competing through co-creation: Innovation at two companies. *Strategy & Leadership, 37*(2), 32–37.

Roy, R. (2018). Role of relevant lead users of mainstream product in the emergence of disruptive innovation. *Technological Forecasting and Social Change, 129*, 314–322. https://doi.org/10.1016/j.techfore.2017.09.036

Saldanha, F. P., & Pozzebon, M. (2015). *Fiat Mio: The project that embraced open innovation, crowdsourcing and creative commons in the automotive industry* (Vol. 13, no. 1). Harvard Business School Press.

Schaarschmidt, M., Stol, K.-J., Walsh, G., & Bertram, M. (2019). Lead Users' Innovative Work Behavior in Digital Platform Ecosystems: A Large Scale Study of App Developers. *Proceedings of International Conference on Information Systems*.

Schmid, I. M. (2020). INFLUENTIAL USERS IN SOCIAL MEDIA NETWORKS: A LITERATURE REVIEW, *Proceedings of the European Conference on Information Systems (ECIS)*.

Schofield, A., & Mimno, D. (2016). Comparing apples to apple: The effects of stemmers on topic models. *Transactions of the Association for Computational Linguistics, 4*, 287–300. https://doi.org/10.1162/tacl_a_00099

Shang, S. S., Wu, Y. L., & Li, E. Y. (2017). Field effects of social media platforms on information-sharing continuance: Do reach and richness matter? *Information & Management, 54*(2), 241–255. https://doi.org/10.1016/j.im.2016.06.008

Schreier, M., Oberhauser, S., & Prügl, R. (2007). Lead users and the adoption and diffusion of new products: Insights from two extreme sports communities. *Marketing Letters, 18*(1–2), 15–30. https://doi.org/10.1007/s11002-006-9009-3

Tuarob, S., & Tucker, C. S. (2014). Discovering next generation product innovations by identifying lead user preferences expressed through large scale social media data. *Paper Presented at the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. https://doi.org/10.1115/DETC2014-34767

Tuunanen, T., Bragge, J., Haivala, J., Hui, W., & Virtanen, V. (2011). A method for recruitment of lead users from virtual communities to innovate it enabled services for consumers in global markets. *Pacific Asia Journal of the Association for Information Systems, 3*(2), 31–56. https://doi.org/10.17705/1pais.03202

Ulrich, K. T., & Eppinger, S. D. (1995). *Product Design and Development*. McCraw-Hill. Inc.

Unsworth, K. L., Brown, H., & McGuire, L. (2000). Employee innovation: The roles of idea generation and idea implementation. *Proceedings in SIOP Conference*.

Von Hippel, E. (1986). Lead users: A source of novel product concepts. *Management Science, 32*(7), 791–805. https://doi.org/10.1287/mnsc.32.7.791

Von Hippel, E. (1994). "Sticky information" and the locus of problem solving: Implications for innovation. *Management Science, 40*(4), 429–439. https://doi.org/10.1287/mnsc.40.4.429

Von Hippel, E. (2005). *The democratization of innovation*. Mass.

von Hippel, E. (2007). The Sources of Innovation. In C. Boersch, & R. Elschen (Eds), *Das Summa Summarum des Management* (pp. 111–120). Gabler. https://doi.org/10.1007/978-3-8349-9320-5_10

Von Hippel, E., Franke, N., & Prügl, R. (2009). Pyramiding: Efficient search for rare subjects. *Research Policy, 38*(9), 1397–1406. https://doi.org/10.1016/j.respol.2009.07.005

Von Hippel, E., & Katz, R. (2002). Shifting innovation to users via toolkits. *Management Science, 48*(7), 821–833.

Wagner, P., & Piller, F. T. (2011). *Mit der Lead-User-Methode zum Innovationserfolg: ein Leitfaden zur praktischen Umsetzung* (1–28). CLIC.

Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (1992). Building an information system design theory for vigilant EIS. *Information Systems Research, 3*(1), 36–59. https://doi.org/10.1287/isre.3.1.36

West, J., & Bogers, M. (2014). Leveraging external sources of innovation: A review of research on open innovation. *Journal of Product Innovation Management, 31*(4), 814–831. https://doi.org/10.1111/jpim.12125

Wolfe, R. A. (1994). Organizational innovation: Review, critique and suggested research directions. *Journal of management studies, 31*(3), 405–431. https://doi.org/10.1111/j.1467-6486.1994.tb00624.x

Xie, P., & Xing, E. P. (2013). Integrating document clustering and topic modeling. *arXiv preprint* arXiv:1309.6874.

Ye, H., & Kankanhalli, A. (2018). User Service Innovation on Mobile Phone Platforms: Investigating Impacts of Lead Userness, Toolkit Support, and Design Autonomy. *MIS quarterly, 42*(1). 165–187. https://doi.org/10.25300/MISQ/2018/12361

Zheng, T.-T., & Zhou, Q.-J. (2017). Analysis of the Motive Mechanism of User Innovation Based on System Dynamics. *Proceedings of the Annual International Conference on Management, Economics and Social Development, 21*, 17–23.