UNIVERSITY OF TECHNOLOGY SYDNEY

UTS:

School of Management

Working Paper Series

http://www.business.uts.edu.au/management/index.htm

I'm Not Chatting, I'm Innovating!
Locating Lead Users in
Open Source Software Communities

Geoff Breach

2008

Working Paper No: 2008/7

I'm Not Chatting, I'm Innovating! Locating Lead Users in Open Source Software Communities

Geoff Breach

School of Management, University of Technology, Sydney, Australia.

Abstract
The Lead User Method recognizes that certain end users of products are a valuable source – in some industries the only source – of new product innovations. Organizations that seek out these lead users and invite them to assist with the new product development process will produce products that perform far better in the marketplace than those developed by traditional methods based on research and development and market research. The individuals who author Open Source Software (OSS) – complex and high-quality computer software that is typically distributed free of charge and without restrictions on use – appear to share many common characteristics with lead users. These skilled computer programmers develop their software collaboratively, and often form large communities that congregate in Internet 'chat rooms' to discuss and manage their software development projects. In response to a call for further research to improve the performance of the lead user method, this paper integrates concepts from the fields of innovation management, data visualization and open source software to a proposal to improve the process of identifying lead users in OSS communities

Key Words:
innovation, lead user method, open source software, Internet, chat rooms

Correspondence To:

Geoff Breach. School of Management. University of Technology, Sydney, NSW, Australia.
Tel +61 2 9514 3049 Fax +61 2 9514 3602
Email Address: Geoff.Breach@uts.edu.au

1.0 Introduction

Recent research in a range of areas has demonstrated that product manufacturers1 are no longer the only developers of new innovations (von Hippel 1988). Product users also innovate, and in some fields almost all innovation is done by users (Shah 2000). Certain user innovators, called lead users, are of particular interest to product manufacturers for the insight that they can provide into new trends and techniques. These innovative individuals are "the unique few who experience the future in advance of the mass market" (Nortel Networks 2006)#:3) They expect to make such significant gains from solutions to their unique problems that they are likely to invent solutions where none are available in the marketplace (von Hippel 1986), and they are often happy to share their solutions (von Hippel 2005).

The lead user method is a process that identifies lead users and involves them in the development of new product concepts (Herstatt & von Hippel 1992). The lead user method has been criticised for its high resource cost – substantial commitments of time and effort are required (Luthje & Herstatt 2004)(Olson & Bakke 2001). The process of identifying lead users is one of the areas that has attracted this criticism. Two methods have been used for identifying lead users – the screening approach and the networking approach – but little empirical data exists regarding the relative merits of the two methods (Luthje et al. 2004). This study responds to Luthje & Herstatt's (2004) call for further research comparing the efficiency and effectiveness of the two approaches.

A new screening method is proposed that relies on certain special characteristics of the individuals that participate in Open Source Software projects, and of the projects themselves, to derive the identities of lead users by observing the patterns of their Internet-based interactions. A form of social network graph is constructed from an analysis of the Internet Relay Chat conversations of group members (Mutton 2004b), and a rank ordered list of individuals who exhibit leading characteristics is produced. Because the new screening method is to be entirely automated, applications of this new approach will be inherently less resource intensive than the interview-based networking approach. The lead user method's networking approach (Herstatt, Lüthje & Lettl 2001) – referred to hereafter as the reference method – is used to produce a second list of leading users from the same population.

The new screening method – hereafter referred to as the IRC method – is to be evaluated against the reference method.

2.0 Literature Review

2.1 Perspectives on Innovation

Joseph Schumpeter took a broad and all-encompassing perspective, positing that innovation was borne of those with an entrepreneurial spirit, and that it was a key driver of economies (Robertson 1967). Where Schumpeter's view was broad, Everett Rogers and Eric von Hippel have taken far narrower and more specific perspectives. Rogers' defined innovation as an "idea, practice or object that is perceived as new" (Rogers 2003)#:12) and focused on the processes by which innovations are disseminated, or diffused (Rogers 1962). Recent work by von Hippel has directed focus to the identity of the innovator and specifically, the sources of innovation (von Hippel 1988), and it is these sources that are the focus of this paper.
The concept of a functional source of innovation is a means to categorise companies and individuals by "the functional relationship through which they derive benefit from a given product, process, or service innovation" (von Hippel 1988) #:3). Manufacturer innovators derive

1 The term 'manufacturer' in this area of the innovation literature refers to the producer of innovations in a broader sense than the literal definition that the operations management literature applies.

benefit from producing the innovation for sale, user innovators benefit from direct use of the innovation, and supplier innovators2 benefit from supplying the materials required to use or to produce the innovation. One or more of these classifications may apply to any one innovator. von Hippel (1988) gives the example of the Boeing company: it is a manufacturer innovator if it develops an new type of braking equipment to fit to its aircraft for sale, and it is a user innovator if it develops a new type of machine tool to speed production of aircraft.

The term users is used herein to describe individuals or companies that hope to gain from using a product or service.

2.2 Users as Innovators

von Hippel (1988) observes that while it has previously been assumed that the vast majority of new product innovations were developed by manufacturers, there are now many fields where this is not the case. User innovators are now known to be an important source of innovations (Lettl, Herstatt & Gemuenden 2006) (Lettl & Gemünden 2005). Commercialised product innovations were developed by users in 82% of cases of scientific instruments studied (Mantel & Meredith 1986)# cited in (Luthje et al. 2004) and 67% in the field of semiconductor assembly equipment (von Hippel 1976). Different numbers of users develop solutions for their own use in different fields (Luthje et al. 2004), and in some fields, most or even all new product innovations derive from user innovation (Shah 2000). Users, be they organisations or individuals, are motivated to innovate by their expectation that substantial benefit will flow to them as a result (von Hippel 1988).

Innovating users will, contrary to what might be expected, often freely reveal the technical details of their innovations (von Hippel 2005) and they stand to benefit significantly – for example by reputational gains – in doing so (Harhoff, Henkel & von Hippel 2003). The lead user method, discussed in section 0 below, takes advantage of this preparedness to reveal details of innovations by inviting lead users to participate in new product development.

2.3 Lead Users

Lead Users are users who exhibit two key characteristics:

1. They are trend leaders who experience strong present needs that are pre-indicators of needs that will be experienced by the general market at a later point in time, and
2. they stand to profit substantially from innovations that resolve those needs, so much so that they will often invent their own solutions (von Hippel 1986).

These lead users are of special interest to new product developers because they can offer insights that cannot be gained via traditional market research techniques that use input from existing product users as guidance for developing new products. The process of canvassing existing users is flawed because those users have experientially based pre-existing conceptions about what a particular product should do (Luthje et al. 2004) and are unable to avoid using their pre-existing knowledge even when instructed to do so (Marsh, Ward & Landau 1999). Their experience interferes with their ability to conceive novel solutions (von Hippel 1986). Lead users on the other hand do not need to imagine that they are in a new situation because they, being ahead of the field as they are, are already in the new situation, and will apply their advanced knowledge to the development of solutions for the problems they actually face (von Hippel 1986).

Having recognised the importance of the lead user, a process is required to identify and engage these innovative users. The Lead User Method is such a process.

---

2 von Hippel identifies the supplier innovator, but gives little further treatment to this class of innovator, apparently preferring to focus on 'manufacturer' and 'user' innovation.

2.4 The Lead User Method

As perhaps the only formal technique for identifying innovators (Intrachooto 2004), the Lead User Method is a technique designed to find lead users amongst communities of users, and to apply their skills to the development of new product ideas (von Hippel & Sonnack 1999a).

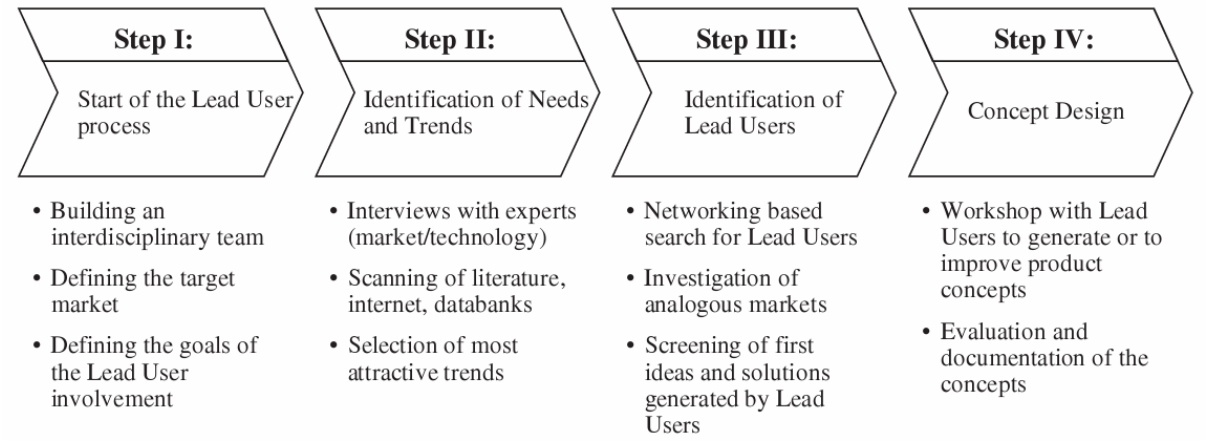


Figure 1: The Process of the Lead User Method Luthje et al. 2004) #:561)

The lead user method has four main steps: First, work is done to define intended targed markets and the types of innovations desired by stakeholders. Second, research is conducted to identify current trends in the field under study. Third, the research team must identify lead users, and Fourth, the lead users and company experts meet to develop ideas for innovative new products (von Hippel, Thomke & Sonnack 1999b).

The third stage may be approached in one of two ways, screening or networking.

2.5 The Screening and Networking Approaches

Two approaches to the third – lead user identification – stage have been used: A *screening* approach, applied in the HILTI study (Herstatt *et al.* 1992), tested a large number of existing product users for lead user qualities, and a *networking* approach, applied in the Johnson & Johnson study (Herstatt *et al.* 2001) begins by interviewing a small number of users in the target market and networks towards experts by asking the interviewees to refer the interviewer to other innovative individuals. Both approaches have been shown to be useful, but there is almost no empirical evidence of the relative merits of the two processes (Luthje *et al.* 2004). The following research question has been proposed:

> *If both approaches (screening vs. networking) are explored in comparative studies: What approach has higher performance in terms of efficiency (search time, search cost) and effectiveness (identification of leading-edge users)?* (Luthje & Herstatt 2004:564)

This study proposes a new approach to screening for lead users that is designed to address some of the criticisms of the lead user method, and validates it using the established networking method as the *reference method.*

2.6 Criticisms of the Lead User Method

Despite significant successes in many fields (von Hippel *et al.* 1999b)(Nortel Networks 2006), the lead user method is not without its problems. The process places onerous demands of time and commitment on participants, to the extent that it distracts from

routine duties (Luthje *et al.* 2004).

> "*In particular, the effort required to find, qualify, and recruit experts for trend analysis and lead users for concept generation were seen as the most burdensome task*" (Olson & Bakke 2001:391)

The process is so difficult that organisations may revert to their old methods of new product development, despite initial successes (Olson et al. 2001). The extensive interview processes have also been criticised for their intrusiveness: a trade-off must be made between interviewing a wide range of individuals and stressing the good will of the community, or focusing on a smaller group and risking a failure to locate the most appropriate experts (Kautz, Selman & Shah 1997).

2.7 A New Approach to Screening for Lead Users

This study proposes a new approach to the established screening technique. Social network analysis and graph theory are applied to a programmatic analysis of the Internet based conversations of communities where Open Source Software is developed, supported and discussed.

In the following sections, the fields of Social Network Analysis, Open Source Software, and Internet Relay Chat are discussed as discrete fields before being assembled in Section 0 to construct a new screening method, called the IRC method.
The IRC method is designed to improve the efficiency of the lead user identification process, and to reduce the degree of intrusion into communities that would be caused if interviews were used as in the case of the reference method.

2.8 Social Network Analysis

By studying the structure of social networks – in context of this study, networks of individuals – a researcher can expect to develop a clearer view of important linkages and key players in the networks (Angot & Josserand 2001). Network analysis techniques have been applied with success to a broad range of disciplines that include human resource management (Josserand & Teo 2004), infectious disease control (Rothenberg et al. 1998) (Rothenberg & Narramore 1996), control of unsolicited bulk email3 (Golbeck & Hendler 2004), and identification of experts in Internet-based groups (Kautz et al. 1997).

Networks of referral (Brin & Page 1998) and trust (Golbeck 2005) (Dasgupta 1986) are strong indicators of reputation (Sabater & Sierra 2002). Newton's shoulders of giants, the academic tradition of citation (Garfield 1979) (Goffman 1971) and the PageRank (Page, Brin, Motwani & Winograd 1998) algorithm at the core of the popular search engine Google all exemplify the same principle: in the absence of the possibility of an individual evaluating all sources of information himself, the recommendations of his peers are an excellent (and possibly the only practical) substitute. It should also be noted that not all networks are reliable – evidence exists of apparent corporate abuse of Internet-based trust networks (David & Pinch 2005).

2.9 Open Source Software

Open Source Software (OSS) is computer software that is distributed under a license endorsed by the Open Source Initiative (Open Source Initiative 2006). OSS is distinguished by the terms of the licenses under which it is released. OSS licenses grant a series of rights, called freedoms, to the users of the software. Those rights typically include the freedom to observe the inner workings of the software, the freedom to make changes to the software, and the freedom to distribute copies of the software (Stallman 2001). In many instances the freedoms are balanced by

3 Colloquially known as 'SPAM'

restrictions that prevent the software from becoming 'closed' – details of any changes made to the software are required to be distributed along with the software itself (Stallman 2001). OSS is not Freeware or Shareware, and it is not in the public domain – it is copyrighted, and strictly licensed (Perens 2001).

Another group, the Free Software Foundation (FSF), operates on a similar ethical footing but a different political basis (Stallman 2001). The FSF is also associated with the production of OSS, called Free Software. Much of the computer software that is 'Free Software' is also 'Open Source Software', released under licensing terms authored by the FSF and approved by the OSI. The distinction between the two is considered immaterial to this study, and the term 'OSS' is used to encompass both Free and Open Source Software.

2.9 OSS Communities

OSS projects are embedded within strong communities (Lakhani & Wolf 2005). Large numbers of programmers work together to produce complex pieces of software that are often of very high quality (Raymond 2001).

Lakhani and Wolf's (2005) study revealed that a substantial proportion (58%) of the developers of OSS are experienced professional programmers and system administrators. Students (19.5%) and academic researchers (7%) also make up a significant proportion. At least half of all contributors have formal university-level training in information technology or computer science, and many more are vocationally or commercially trained computer programmers (9%). These findings are consistent with Raymond's (2001) claim that OSS culture only accepts the most talented programmers.

Raymond (2001) also observes that individual programmers begin by authoring personal solutions to a particular technical problem: "every good work of software starts by scratching a developer's personal itch" (p23). Empirical evidence supports this claim, 58% of OSS developers write their software to meet their needs – some at work and some outside of their jobs (Lakhani et al. 2005). The largest single gain for individuals who write OSS is intellectual stimulation (44.9%), and human capital (personal skill) development follows closely (41.8%) (Lakhani et al. 2005). Raymond acknowledges that programmers who submit their work to OSS projects benefit from a process of peer review of that work, and claims that their programming skills improve as a result (2001). Their technical solutions spread to the community at large "because the problem turns out to be typical for a large class of users" (Raymond 2001:49).

Raymond (2001) goes on to argue that the open source movement is a "very strict meritocracy" (p89) and observes that the programmers who make the greatest contributions to the community are the ones that are most respected by the community and who rise to take on leadership roles. The IRC method recognises the meritocratic nature of the community in highlighting lead users.

2.11 Internet Relay Chat

Of further interest is the means by which many of the more geographically diverse OSS communities communicate. Many communities hold formal project meetings, social gatherings and provide technical support to end users via an Internet based mechanism called Internet Relay Chat, or simply IRC.

IRC was created by Jarkko Oikarinen in 1988 to add a multi-participant capability to the then ubiquitous 'talk' command that was installed on most Unix computer systems of the day. While talk was eventually modified to provide basic capability for more than two people to communicate at once, it lacked the functionality to convene large scale communication.

Where talk made use of direct point-to-point connections to facilitate exchange, IRC is highly scalable. IRC servers have two key roles: they offer a point for individual participants to connect,

and they relay conversations to other servers that are in turn hosting their own participants. It is usual for a network of five or ten or more IRC servers to host hundreds of thousands of individual participants, each participating in conversations that cross multiple servers.

IRC is a text-based means of exchange. Individual participants use a piece of software, called an IRC client, to connect to an IRC server. The names and Internet addresses of public IRC servers are widely publicised, and many IRC clients come pre-configured with large lists of servers. In most cases, access to IRC is free of charge, unrestricted, and largely anonymous.

Participants join an IRC channel – a term analogous to the concept of a citizens band radio channel - and immediately begin to receive a rolling textual conversation on their screen. Individuals make their contributions to the conversation by way of a typed message that appears at the bottom of all participant's screens. Earlier contributions are displaced upwards, eventually scrolling off the top of the screen.

IRC was initially recognised by academics as a useful tool for collaboration amongst groups of geographically disconnected individuals, but a common use of the tool has been entertainment (Rheingold 1995). Private IRC servers are created by some companies for the purpose of providing technical support or to encourage communities to form around their products. Some organisations use IRC internally to allow groups of software developers to communicate in real time with their peers, be it in the next office or on the other side of the world (Urbanec 2005).

2.12 Nature of Analysis of IRC Communications

The usual mode of operation for IRC client software is interactive – an individual uses the tool to join channels, and participate in conversations. Another type of IRC client exists that operates without the direct supervision of a human operator. An automatic, non-interactive IRC client is called a bot – a contraction of 'robot'. This study uses a bot, a derivative of the PieSky Social Network Bot (Mutton 2006), to monitor and conduct initial analysis of IRC conversations. The processes of monitoring and inference of social relationships are automated and data is presented in a form suitable for combinatoric analysis.

IRC conversations are particularly suitable for computerised analysis (Tuulos & Tirri 2004), and many examples exist of analysis that has been made for the purpose of entertaining and informing the participants themselves (Donath 2002). Mutton's (2004) process identifies relationships between participants by detecting when they address each other by name, and infers relationships when two or more inviduals speak with close temporal proximity. Despite its apparent simplicity, the process has been shown to produce social network graphs that actual participants agreed were reliable (Mutton 2004b).

All participants in IRC communications have a self-chosen name, typically a pseudonym, called a 'nick' (a contraction of 'nickname'). Nicknames are typically persistent over extended periods of time and while the identity of an individual can not usually be inferred from the name they choose, it is relatively safe to assume that the person using a given nickname today is the same person who used it on previous days (Rheingold 1995). This characteristic is relied upon to ensure that the same individuals are observed from day to day, and that those identified by the reference method and by the new screening method are able to be compared.

2.13 Common Characteristics

We have seen that lead users are individuals who experience needs that are later experienced by a community of users. They are compelled to develop their own solutions because no suitable solutions are already available and because they stand to gain substantially by doing so. The Lead User Method's networking based approach searches within communities with a view to identifying these most innovative individuals.

OSS is produced by communities of skilled individuals who recognise and are led by those who contribute solutions to needs experienced by the community at large. These individuals are innovators who set out to "scratch a personal itch" and develop their solutions into computer software that can resolve problems for themselves and for other community members. They are rewarded for their innovative contributions with intellectual stimulation, skill development and the respect of their peers.

It is clear that the case of the OSS developer shares many common characteristics with that of the lead user. These common characteristics are central to the IRC method and to the claim that the individuals identified by the new method are in fact lead users.

2.14 The IRC Method

The IRC method relies upon the characteristics of OSS developers that are common to those of lead users to identify individuals within OSS communities who are likely to be lead users. A form of social network graph is constructed by observing the IRC interactions of individuals in the OSS communities, and lead users are identified by applying established methods of social network analysis. The implementation of the IRC method is discussed in detail in the Methodology section, below.

3.0 Aims and Hypothesis

This study constructs and evaluates a new screening method – the IRC method – that is designed to be less resource intensive than the existing approaches to lead user identification that are used in stage 3 of the lead user method.

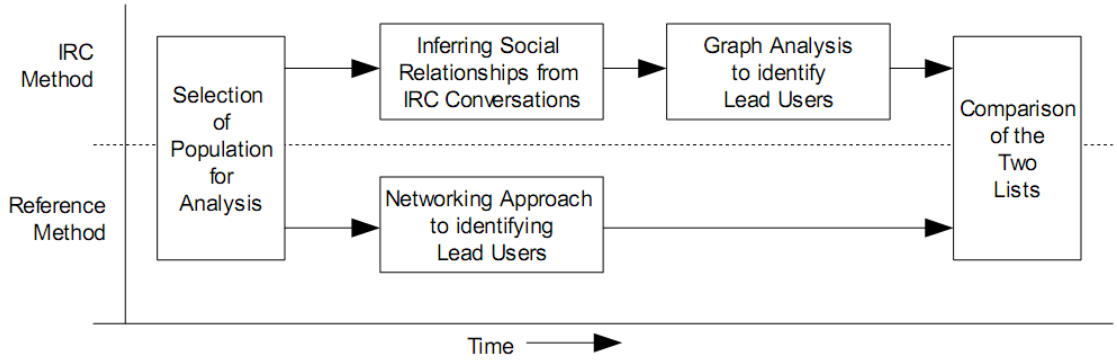
Individuals within OSS communities share common characteristics with the individuals sought by the lead user method, and it is in the nature of those communities to recognise and privilege the individuals who exhibit those characteristics. The IRC method seeks to identify these individuals. Therefore, it is hypothesised that the IRC method will validly identify lead users in OSS communities.

4.0 Method

This study responds to Luthje & Herstatt's (2004) call for empirical comparison of the different approaches to identifying lead users by applying both the networking approach and the IRC method to the same user community and comparing the outcomes.

Subsequent to selection of a population for analysis, the methodology for this study has three key components: First, the IRC method was used to programmatically generate a social network graph from the IRC interactions of OSS project members and to identify central users. Second, the interview-based networking approach to lead user identification (the reference method) was

Figure 2: Steps in data collection and analysis



conducted amongst the same community. Third, the two lists of identified individuals were compared to assess the degree of correlation between the two.

Four key steps occur in the collection and analysis of data. The time-based ordering of those steps is shown in Figure 2. All data collection steps, including the interview process, were completed before data analysis began to ensure that the researcher's knowledge of the outcome of the IRC method did not contaminate the interview process. Details of each step are discussed in the following sections.

4.1 Selection of Population for Analysis

An appropriate population for this study are the members of a large OSS project that specifies IRC as a primary means for community interaction and that has a visibly active IRC channel. The Gentoo Linux project and its corresponding #gentoo channel on the Freenode IRC network was chosen for examination in this study because it fits these criteria.

4.1.1 About the Gentoo Linux Project

Gentoo is a popular variant of the well known Linux computer operating system that serves as a powerful and free alternative to commercial products such as Microsoft Windows and Apple OS X. IRC is a primary means of communication for the Gentoo project. Gentoo is OSS made by an OSS project that is consistent with those described in the literature review sections 0 and 0 above. A total of eighty distinct IRC channels typical of those discussed in section 0 above exist for official Gentoo activities, many are devoted to specific technologies and in some cases specific spoken languages.

The main #gentoo channel, used to hold technical discussions and to provide assistance to Gentoo users, is the focus of this study.

4.2 The IRC Method

The IRC method constructed a social network graph by inferring relationships between participants on the #gentoo channel and applied basic combinatoric technique to identify the most significant participants. The following sections detail the process of deriving a list of the most significant participants from the IRC data.

4.2.1 Inferring Social Relationships from IRC Conversations

Computer software directly derived from the PieSky (Mutton 2004a) was used to automatically generate social network graphs from IRC conversations. Details of the data collection software and data analysis software are discussed in sections 0 and 0 below.

Mutton's (2006) PieSky software preceded to a destination of producing visualisations of the social network graphs (see Figure 3). Whilst the generated images are useful visual indicators they are not believed to be sufficient on their own to sustain rigorous analysis. The modified version of the software to be used in this study generates graph data in a form suitable for formal combinatoric analysis.

Mutton's (2006) software uses three techniques - direct addressing of exchanges, temporal proximity of exchanges, and temporal density of exchanges – to infer relationships between individuals.

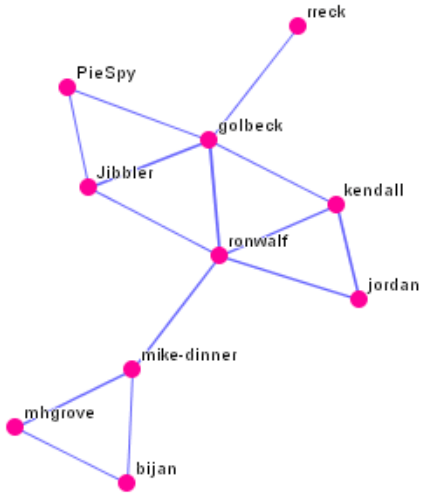


Figure 3: A simple social network Visualisation (Mutton 2004a)

4.2.2 Names of Participants

It is a feature of the IRC protocol that a complete list of the nicknames of participating individuals is provided to all participants (Oikarinen & Reed 1993). This list is used by the bot software as a starting point for the construction of the social network graph. As we have seen, individuals typically use the same nickname at all times (Rheingold 1995) and so the individual's chosen nickname is used as the persistent identifier during data analysis.

4.2.3 Direct Addressing

Direct addressing occurs when an individual precedes his message with the name of the intended

```
<geoff> Can someone tell me the time?
<douglas> geoff: It is 1:00pm
```

Figure 3: An example of direct addressing

recipient (Mutton 2004b).

A simple pattern matching approach is used to compare the text of each message to the names of participants.

4.2.4 Temporal Proximity and Temporal Density of Exchanges

Not all exchanges in IRC are directly addressed. The relational inferences that might be drawn from physical proximity in a 'real life' conversation are also not available in the IRC context – participants are often widely distributed around the globe (Schwartz & Wood 1993). It is reasonable, however, to infer relationships from exchanges that are temporally close, particularly when periods of silence occur pre and post an exchange (Mutton 2004b)(Tuulos et al. 2004).

Initial data analysis demonstrated that members of the #gentoo channel used direct addressing consistently, with 51% of all exchanges on the channel being explicitly addressed to another participant. The remainder of exchanges were observed to be general chatter and undirected questions and so not considered to be contributory to an analysis designed to identify key individuals. Mutton's techniques for inferring relationships by temporal proximity and temporal density were therefore not employed in the analysis made in this study.

4.2.5 Temporal Decay

Because individuals can, and do, join and part an IRC channel at any time, the social graph of the channel varies constantly with time. PieSky (Mutton 2004a) included a function to cause the strengths of relationships to decay as time passed so that the generated visualisations would represent the current state of the channel at the time they were produced.

Whilst Mutton demonstrated that his technique generated accurate visualisations in real time, the decay functionality introduced severe distortions to analysis designed to construct graphs over extended periods of time and so it was excluded from the analysis made in this study.

4.2.6 Data Collection and Preparation

Computer software was written specifically for this study that connected to the Freenode IRC network and recorded the IRC conversations that took place on the #gentoo channel. Over a period of sixty-four days, 618,191 lines of data including 357,027 lines of conversation were

recorded. The data was stored in a PostgreSQL relational database and subsequently processed off-line.

In preparation for processing, individual's nicknames were obfuscated by substituting the names for a four digit hexadecimal code that is persistent for each individual throughout the data. This unique identifying code is used in this paper to protect individual identities.

4.2.7 Data Analysis Software

In some cases the analysis of conversation data resulted in generated graphs with as many as 9000 vertices and 120,000 edges. Both Mutton's (2004) PieSpy software and Petr Baudis' perl language interpretation of PieSpy, piespy.pl (Baudis 2007), exhibited severely degraded performance when working with very large graphs. A new program that derived functionality from Mutton's program and textual analysis and graph drawing features from Baudis' program was written specifically for this study.

The new program was designed to draw data from a database store whereas the predecessor implementations drew data from live conversations on the Internet in real time. A particular focus on efficient handling of graphs was developed, and capabilities for more complex combinatoric analysis were excluded to improve performance.

4.2.8 Social Network Graphs

An IRC channel C consists of a series of U utterances made by participants P at specific times T and so the sequence of conversation may be shown as:

$$C = (u_1, p_1, t_1), \ldots, (u_n, p_n, t_n)$$

Relationships R are inferred between individuals on the basis of the time and content of their utterances, and so exchanges between participants such as those shown in Figures 4 and may be represented as:

$$c_1 = (u_1, p_1, t_1), (u_2, p_2, t_2)$$

and the inferred relationship:

$$r_1 = (p_1, p_2)$$

The list of all individuals present in a channel P_{all} is provided by the IRC server, and updated periodically as channel membership changes. Some members of the channel, lurkers, do not participate however. Non-participants are not of interest to a study that uses inferred relationships, and so:

$$P \subseteq P_{all}$$

A graph G is constructed of a pair (P, R) where P is the set of nodes representing active participants to the IRC conversation, and R, a set of two element subsets of P, are edges representing inferred relationships between participants.

$$G = (P, R)$$

The nature of the inferred relationships is such that they may be represented as directed edges $(p_1, p_2) \neq (p_2, p_1)$ if analysis requires. Because it is the cumulative weight of a number of exchanges is of most interest in the proposed social network analysis, directionality is not considered, and it is considered that $(p_1, p_2) = (p_2, p_1)$ when judging the importance of particular actors in this study.

4.2.9 Graph Analysis to Identify Key Actors

Simple degree centrality – the number of edges connecting to and from a node in an undirected graph – was used as the measure of importance of individual nodes in the IRC method portion of this study, consistent with Mutton's (2004) observation that the undirected approach produced

accurate graphs.

4.3 The Reference Method

Initial candidates were selected for interview from the #gentoo IRC channel by selection at random from those who were active in the channel at the time the interview was to commence. Once a number of initial interviews had begun to identify the same apparent experts, the focus of the interviews was turned to those identified individuals in the IRC channel. Interviews were conducted according to the following process.

4.3.1 Conduct of Interviews

In an approach derived from that discussed by von Hippel, Thomke & Sonnack (1999), and also used at Johnson & Johnson (Herstatt et al. 2001), individuals were asked questions derived from the interview schedule (Appendix **Error! Reference source not found.**) to assess their level of innovativeness. Individuals were also asked to identify other individuals who they considered to be experts. The individuals who were identified by their peers were in turn interviewed according to the same process. By then asking the same or similar questions of the peer-identified experts, the interviewer works his way up a "pyramid of expertise" (von Hippel et al. 1999b).

According to the Lead User Method, the interview process ceases when repeated identifications of the same experts indicates that the most expert individuals have all been identified. A list of lead-users is produced by identifying those that were most often referred by their peers and who exhibited the characteristics of lead users when interviewed.

In accordance with the lead user method, the interview schedule was taken as a guide, and not followed strictly. The interview stages of the lead user method are flexible, and adjust to suit whatever directions conversation may make, so as to allow new information that researchers were not aware of to arise and to allow the first direction-finding stage of the process to continue on through the interview process (Churchill 1999)(von Hippel et al. 1999a)(von Hippel 1986)

4.3.2 Departures from the Reference Method

Due to the time consuming nature of the interview process, questions designed to establish trends in the field and to measure trend leadership status of the interviewee were excluded from this study.

To ensure that results from the IRC method and the interview process are comparable, individual interviewees in this study were asked to identify experts within the #gentoo IRC channel. This is contrary to the Lead User Method where experts in other relevant fields are expected to be identified during the interview process. This containment within a single field is an acknowledged limitation of the IRC method.

As is acknowledged by Luthje and Herstatt (2004) and Olson and Bakke (2001), the interview process is lengthy and time consuming. The number of interviews that would be required to conclusively identify the lead users amongst a group the size of the Gentoo project is beyond the scope of this study, and so the interview process was concluded when interview data sufficient for analysis had been collected.

5.0 Results

Data collected from the IRC channel was stored until after the process of interviews was completed. Results from the interview process – the reference method – are presented here, followed by the outcomes of the IRC method and a comparison of the results of both processes.

5.1 The Reference Method - Interview Process

A total of 15 individuals, selected as described in section 4.3 above, were approached either by private message (colloquially 'pm') on IRC or by a public request within the #gentoo channel for permission to contact by 'pm'. Those who agreed to be interviewed were first asked to speak to their own interest and involvement with the Gentoo Project and to categorise themselves according to their involvement. They were then asked questions designed to indicate their level of innovativeness and finally they were asked to identify others who they considered to be experts. Two individuals who were identified as experts by their peers did not participate in interviews. Three individuals who did not respond to requests for interviews are not shown in Table 2.

5.1.1 Interview Responses

Whilst the interviews included questions designed to ascertain – consistent with the lead user method – the level of innovativeness of the individual participants, those questions are beyond the scope of this study and so the responses to that part of the interviews are not discussed herein. Of relevance to this study are the responses that participants gave when they were asked to nominate #gentoo channel members who they believed to be experts. All interviewees identified at least one expert, and some identified as many as four or five. Where possible, nominated individuals were also interviewed and in turn asked to nominate others who they considered to be leaders. Table 2 lists the obfuscated nicknames of the individuals who participated in interviews, and those of the individuals they nominated as experts. Individuals marked with superscript '1' were identified more than once. Identified individuals marked with superscript '2' were subsequently interviewed.

5.1.2 Reference List of Identified Experts

Table 1 lists individuals who were nominated as experts by interviewees and shows the number of times each was nominated. This list of Identified Experts constitutes the *reference list* for comparison to that produced by the IRC method.

Nominee	Count
26B6	2
26C3	2
26C9	2
26CC	2
26CE	2
0886	1
0D5F	1
182B	1
1C75	1
23C9	1
2658	1
2689	1
2695	1
26AA	1
26B0	1
26B9	1
26BE	1
26CA	1
26CB	1

Table 1: The Reference List

5.1 The IRC Method

IRC chat data was analysed using the software authored specifically for this study. Social network graphs were constructed for the #gentoo channel on the basis of direct addressed conversation between channel members. The primary outputs from the software are lists of individuals ordered by their respective levels of simple degree centrality – hereafter '*centrality*' – within the social network graph.

5.2.1 Data Collected by the IRC Method

Five consecutive weeks of IRC data were analysed. An observation of periodicity was made to determine the appropriate time period – epoch – for subsequent tests. Reliability is tested to ascertain the time at which the accumulated data achieves a sufficient level of internal reliability to support further analysis. Stability is tested to observe the degree to which sample of data take at different times will produce similar results. Finally, the reference list is compared to that produced by the IRC method – the IRC list – in section 0 below.

5.22 Periodicity

The tests conducted in the IRC method portion of this study rely on the accumulation of centrality scores of each participant over time. Initial analysis revealed a prominent periodicity in the data, consistent with individuals joining and leaving the IRC channel at different times of their day. Because the IRC channel is populated by individuals from many different time zones, there is no single 'busy' period at any given time of the day or night, and no persistent periodicity was evident in data collected in 24 hour blocks. All of the tests conducted hereafter are conducted on day-long blocks of data therefore.

Chart 1 gives an example of non-accumulated centrality data for a typical participant over a period of ten days. Grid lines on the X axis are at 24 hour intervals, and centrality is represented

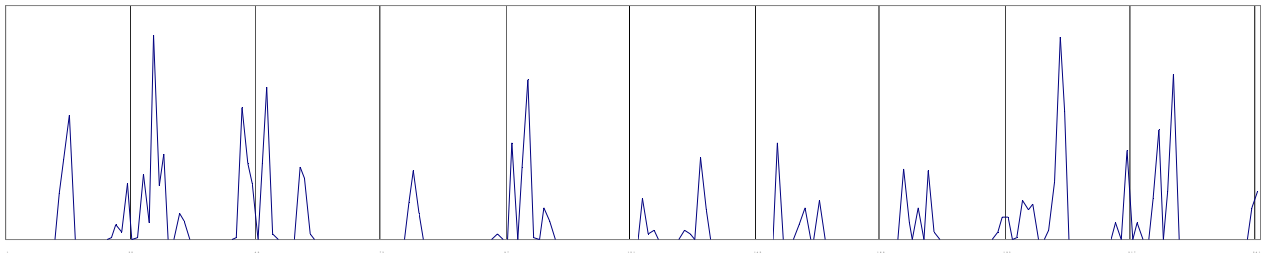


Chart 1: Exemplar non-accumulated centrality data for one participant over ten consecutive days

on the Y axis. A 24 hour periodicity is clearly visible in the data, and to a lesser extent a 7 day periodicity is also evident.

5.2.3 Reliability

A split-half approach was taken to gauge the parallel form reliability (Miner 1992) of the collected data. The centrality score of each individual participant was accumulated for two interleaved but separate runs over the data. Set A consists of odd numbered days and Set B consists of even numbered days. Corresponding days from each set were compared to identify the correlation between the two sets.

This test is designed to establish the number of time periods – epochs – required to produce a reliable aggregated measure of degree centrality of participants. The correlation between the two sets is high (Rowntree 1981) ($r = 0.8$, $p < 0.01$) after seven sets (days). Subsequent tests in this study use the entire set of data that was collected, but a practical application of the IRC method can be expected to produce adequate results in about seven days.

Chart 2 shows correlation between corresponding pairs of data accumulated over 17 consecutive days. Pair 1 indicates correlation between day 1 of Set A and Set B, pair two indicates correlation between day 2 of Set A and Set B, and so on.

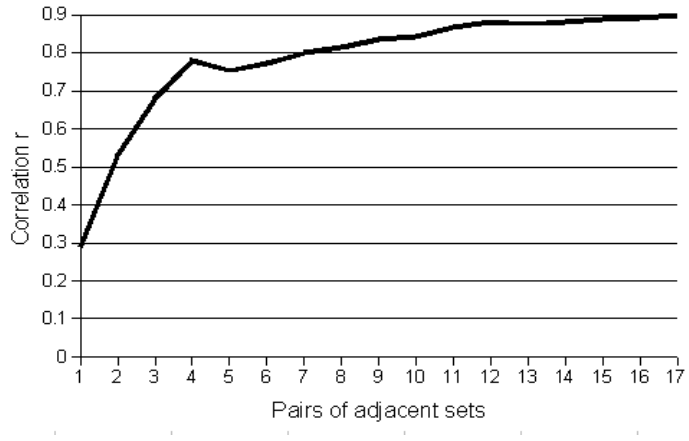


Chart 2: Correlation r between adjacent pairs

#	Participant	Identified Experts
1	2549	26C9[1]
2	19C8	182B
3	26B4	26AA[2] 26CA[2] 2695
4	265B	26BE[2] 26B9[2] 26CE[1] 26CC[1]
5	2004	1C75
6	26C2	26C9[1] 26B0 26B6[1 2] 26CB[2] 0886 26C3[1]
7	26B9[2]	26C3[1]
8	26AA[2]	26CC[1]
9	26CA[2]	*Declined invitation to participate*
10	26CB[2]	26B6[1 2]
11	26BE[2]	26CE[1] 23C9 2689 2658 0D5F
12	26B6[1 2]	*Did not respond to invitations to participate*

Table 2: Tabulation of Interview Results

5.2.4 Stability

To test the stability of the collected data over time, raw degree centrality scores for each individual participant were accumulated for five consecutive seven day periods numbered Week 1 to Week 5. Each weekly accumulation was tested for correlation with all other weeks. Table 3 shows the correlation coefficient, r, for each test.

	W1	**W2**	**W3**	**W4**	**W5**
W1	1.00				
W2	0.63	1.00			
W3	0.46	0.65	1.00		
W4	0.43	0.49	0.56	1.00	
W5	0.50	0.50	0.48	0.60	1.00

Table 3: Correlation of weekly data

This test is designed to observe the stability of centrality over time. There is a tendency for participation levels to be more similar on adjacent weeks than between weeks more temporally distant from each other.

5.3 Comparing the Reference and IRC Lists

A concordance will be established to link all names used in both methods. Lists of users most strongly exhibiting leading characteristics as measured by the respective methods are produced. The two lists are compared in the following sections using a range of tests.

5.3.1 Position and Proportion of Reference List Members on the IRC List

Table 4 lists participants ordered by the centrality score by the IRC method. Those that were also selected by the interview method are highlighted and denoted 'R'. The complete ordered list of participants constitutes the IRC list.

A visual analysis of the presented table with the reference list overlaid on the IRC list demonstrates that that the majority of individuals identified by the reference method are also placed towards the top of the IRC list. The list of the top 30 individuals selected by the IRC method contains greater than fifty percent of those selected by the reference method. Considering that more than 3,500 individuals were active in the community during the study period, this level of selection is considered to be highly successful.

IRC Method	Reference List	Participant	Centrality	Proportion	Magnitude
1	1	26CE	4048	100.0%	R
2	2	26CC	3390	100.0%	R
3		26CD	3337		
4	3	26CA	3249	75.0%	R
5	4	26CB	3114	80.0%	R
6		26A2	3100		
7		26C4	2442		
8		26BF	2113		
9	5	26B9	1996	55.6%	R
10		26B8	1949		
11		26A8	1828		
12		26C7	1808		
13	6	26C3	1688	46.2%	R
14	7	26C9	1585	50.0%	R
15		00C6	1536		
16		26AB	1526		
17	8	26AA	1483	47.1%	R
18		26AE	1420		
19		26AF	1410		
20		046E	1408		
21		26B2	1325		
22		26C0	1269		
23		26AD	1235		
24		269E	1229		
25	9	26B6	1218	36.0%	R
26		26A9	1201		
27		26BD	1170		
28	10	26B0	1166	35.7%	R
29		26B5	1086		
30		26A4	1064		
31	11	26BE	1021	35.5%	R
32		26A5	975		
33	12	2695	943	36.4%	R
48	13	2689	682	27.1%	R
111	14	2658	334	12.6%	R
1128	15	23C9	26	1.3%	R
1134	16	1C75	25	1.4%	R
1560	17	182B	14	1.1%	R
3754				0.45%	

Table 4: Ranking of Participants

Table 4 also shows the proportion of members of the IRC list that is comprised of individuals also on the reference list. The table indicates, for example, that 80% of the top 5 on the IRC list were also identified by the reference method. The fact that the distribution of members of the reference list is strongly skewed towards the upper end of the IRC list is further evidence of the success of this new approach.

The reference list constitutes less than 0.5% of the individuals active in the community, and so the likelihood of selecting any member of the reference list at random is very low.

5.3.2 Comparison of the Reference and IRC Lists

The centrality score achieved on the IRC method by individuals who were also selected by the interview method was 1528.4. The mean score achieved by those not selected by the interview process was 47.5. A single tailed Student's t test for samples with unequal variance applied to the two lists returns a value for p of 0.000088 and so the null hypothesis – that the reference group

will not have a mean higher than that of the non-reference group – is rejected ($p < .01$). Table 5 gives details of the comparison.

	Non-Reference	Reference
n	3737	17
Mean	1528.4	47.7
Std Dev	1258.0	155.9
Student's t	4.853	
p	0.000088	

Table 4: Comparison of Means

6.0 Discussion

6.1 Findings

This application of the reference method produced a list of seventeen individuals who are regarded by their peers as leaders. As might be expected from a process of interviews designed to construct a social network graph, measures of individual's centrality – as shown in table 1 – were quite low, and so it is inappropriate to make any attempt at ordering the reference list before making comparisons. The IRC method, however, produces an ordered list by its very nature, and so it is appropriate to draw comparisons between the reference list and the IRC list on the basis of the positions of reference list members on the ordered IRC list.

Prior to comparison, the nature of the IRC list was analysed by a series of methods designed to better qualify the nature of the data for subsequent tests. A view of periodicity was taken to determine an appropriate epoch size for following tests. Accumulation of centrality scores was observed for a range of individuals, and the example given in Chart 1 is typical. Most individuals participated in the IRC chat on an approximately daily cycle – with some day-to-day variations – consistent with the individual including their participation as a regular part of their daily life. A second indication of periodicity was observed to take place on a weekly cycle, and again the activity recorded in Chart 1 is typical. The magnitude of the weekly cycle tended to be relatively slight however, and no obvious pattern – such as, for example, higher levels of participation on a weekend – that could be generalised to all participants was observed. An epoch of twenty-four hours was set therefore, as the basis for following tests.

In order to gauge the internal reliability of the collected data, a split half approach was used to construct a parallel form (Miner 1992) test of reliability. As is shown by Chart 2, the parallel sets of data became strongly correlated within seven days and so a decision to make practical use of collected data after seven days brings the dual benefits of high internal correlation of data, and absorption of the secondary seven day periodicity observed above.

A matrix correlation of weekly sets of data demonstrated sometimes significant variations in the IRC list from week to week. Whilst adjacent weekly sets were always moderately correlated, sets that were more temporally distant were often significantly less correlated. This issue is further discussed in the Limitations section below.

Having satisfied tests to ascertain the nature of the data collected by the IRC method, comparisons were drawn between the IRC list and the reference list. Immediately apparent by the

data presented in Tables 4 and is that all members of the reference list are placed highly on the IRC list. The top 10 members of the reference list appear inside the top one percent (0.75%) of the IRC list. If participants '23C9', '1C75' and '182B' are excluded as outliers by virtue of their relatively low centrality scores, the remainder of the reference list appears in the top three percent of the IRC list. A comparison of the mean centrality scores of the IRC list members who were and were not selected by the reference method demonstrates that the two lists are very significantly different, and highly unlikely to have been arrived at by chance. The high rate at which centrality scores decay on the IRC list – a reduction of two orders of magnitude in the first 10% of the list, and three orders of magnitude in the first half of the list – is further evidence of the selectivity of this method.

Considering that the IRC method has made no attempt to ascertain the content of the conversations that took place on the IRC channel and that no direct interaction with the community members took place during the IRC method process, this represents a highly successful outcome. The IRC method shows great promise as a tool to improve the lead user selection component of the lead user method.

6.2 Limitations and Suggestions for Further Research

The production of undirected graphs by the IRC method – approach taken to replicate that used by the PieSpy (Mutton 2004b) software and to reduce the technical complexity of the IRC method – attracts the criticism that the individuals identified by the method may in fact not be leading users who rise to the top of the list of users by virtue of their lead status in the community, but simply particularly verbose individuals who do not exhibit leading characteristics at all. This apparent logical flaw may not be important in practise however. Comments made by some interviewees – not reported herein – provided anecdotal evidence that new community members may learn and develop leading characteristics in the course of their participation in the community, and so any initial verbosity on the part of a non-leading user may constitute a pre-indication of an individual who will become a lead user with time. This author's direct experience of activity in other similar communities indicates that highly verbose individuals who do not make contributions are unlikely to be tolerated by the community, and so short term verbosity on the part of individuals may not significantly bias results. Future research should attempt to apply more complex means of textual analysis with a view to producing directed graphs that differentiate between those who ask questions and those who answer questions in problem solving situations. Appropriate analysis would allow observation of the relative status of the two groups and of whether non-leading individuals progress to a leading status over time.

While the lists produced by the IRC method at different times exhibited sometimes significant variations, it should be noted that the reference method may also produce different results at different times. Whilst the lead user method has been demonstrated to be highly successful (Luthje et al. 2004) (Herstatt et al. 2001), little work has been done to measure the repeatability of results produced by the reference networking approach. Because lead users identified by either method will always be further qualified with an interview designed to measure their leading status, it might be considered that highly repeatable results are not a necessity. Research to further examine the repeatability of results produced by the networking approach and the IRC method may serve to further validate the IRC method. Further testing of the existing data might also seek to determine whether the observed variations occur uniformly across the whole IRC list or within some specific segment of the list.

As discussed in section 0 the IRC method relies on specific common characteristics that are exhibited by lead users and by the members of OSS communities. The IRC method may have applications in other communities with similar features, and so an opportunity exists for further research to test the applicability of the method in other communities of practice.

Whilst IRC is a widely used means of communication for OSS and for other communities of practice, many other mediums are also used. It is considered that variations on the IRC method

may be similarly effective in other contexts, such as electronic mail lists and web-based community forums. Work to adapt the method to, and validate on, alternate electronic mediums may allow the method to be more broadly applied.

A key factor of the lead user method is the identification of experts in fields other than the target field, an approach designed to bring skills and technologies from other fields to the new product development process (von Hippel et al. 1999a). Because this study operates within one IRC channel devoted to a particular area, this field-crossing approach could not be carried out automatically. It is considered that individuals identified by the automated pre-screening process can be asked to refer experts in other fields during the post-screening interviews designed to assess their lead user status.

Prior to the execution of this study it was considered that the interactions between human participants and the service bots that exist in many IRC channels may bias results. Interactions of this type were tested for in initial analysis of results and were observed to occur so rarely as to be unlikely to affect outcomes. Service bots are readily identifiable and would be easily excluded from results where their activity was of a sufficiently significant magnitude to cause concern.

6.3 Conclusion

The IRC method was devised of the observation that Raymond's (2001) anthropological discussions of OSS communities and von Hippel's (2004; 1985) work on lead users discussed very similar characteristics in the individuals under study. This application of Mutton's (2004) method to IRC based OSS communities has produced a screening method that identifies leading users without the high resource costs and intrusion into the community that the networking approach suffers, and it has done so with highly encouraging results.

Additionally, the highly selective nature of the IRC method when applied to the IRC based OSS community seems to imply that OSS communities generally may be appropriate venues for a test of this nature and so future research should consider and test a broader application of this method.

References

Angot, J. & Josserand, E. (2001). Analyzing social networks. In *Doing Management Research*. pp. 312-331 Sage: London.

Baudis, P. (2007). *piespy.pl - A perl Adaptation of the PieSKy Social Network Bot*. Retrieved from http://pasky.or.cz/~pasky/dev/irssi/piespy/ - Accessed 10th July, 2007.

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

Churchill, J. (1999). *Interviewing Methods for Lead User Project Teams, Video Presentation*. Retrieved from http://userinnovation.mit.edu/videos/Identifying_users.mpg - Accessed: 28th November 2006.

Dasgupta, P. (1986). Trust as a Commodity. In *Trust: Making and Breaking Cooperative Relations*. University of Cambridge Department of Applied Economics. pp. 49-72.

David, S. & Pinch, T. (2005). Six Degrees of Reputation: The Use and Abuse of Online Review and Recommendation Systems, in *Economic Sociology and Technology Conference.*

Donath, J. (2002) A Semantic Approach to Visualizing Online Conversations. *Communications of the ACM*, *45*(4): 45-49

Garfield, E. (1979). Citation indexing: its theory and application in science, technology, and humanities, Wiley: New York

Goffman, W. (1971) A Mathematical Method for Analyzing the Growth of a Scientific Discipline. *Journal of the ACM*, *18*(2): 173-185.

Golbeck, J. (2005). *Computing and Applying Trust in Web-based Social Networks.* Unpublished doctorial thesis, University of Maryland.

Golbeck, J. & Hendler, J. (2004) Reputation Network Analysis for Email Filtering. *Proceedings of the First Conference on Email and Anti-Spam*, pp30-31.

Harhoff, D., Henkel, J. & von Hippel, E. (2003) Profiting from voluntary information spillovers: How users benefit by freely revealing their innovations. *Research Policy*, *32*(10): 1753-1769.

Herstatt, C. & von Hippel, E. (1992) Developing New Product Concepts via the Lead User Method: A Case Study in a Low Tech Field. *Journal of Product Innovation Management*, *9*,(3): 213-221.

Herstatt, C., Lüthje, C. & Lettl, C. (2001) Innovationsfelder mit Lead Usern erschliessen. *Workingpaper des Lehr-und Forschungsbereichs Technologie-und Innovationsmanagement der Technischen Universität Hamburg-Harburg (forthcoming in Harvard-Manager)*

Intrachooto, S. (2004) Research and Concepts: Lead Users Concept in Building Design: It's Applicability to Member Selection in Technologically Innovative Projects. *The TQM Magazine*, *16*(5): 359

Josserand, E. & Teo, S. (2004). Role Identity and Social Capital Development During HRM Role Change. In , *64th Academy of Management Meetings*, New Orleans: USA.

Kautz, H., Selman, B. & Shah, M. (1997) Referral Web: Combining Social Networks and Collaborative Filtering. *Communications of the ACM*, *40*(3): 63-65.

Lakhani, K. & Wolf, R. (2005). Why Hackers Do What They Do: Understanding Effort and Motivation in Open Source Projects. In *Perspectives on Free and Open Source Software*. MIT Press: Cambridge, MA.

Lettl, C. & Gemünden, H. (2005) The Entrepreneurial Role of Innovative Users. *Journal of Business & Industrial Marketing*, *20*(7): 339-346.

Lettl, C., Herstatt, C. & Gemuenden, H. (2006) Users' Contributions to Radical Innovation: Evidence from Four Cases in the Field of Medical Equipment Technology. *R&D Management*, *36*(3): 251-272.

Luthje, C. & Herstatt, C. (2004) The Lead User Method: An Outline of Empirical Findings and Issues for Future Research. *R & D Management*, *34*(5): 553.

Mantel, S., & Meredith, J. (1986). The Role of Customer Cooperation in the Development, Marketing, and Implementation of Innovations. In *The Art and Science of Innovation Management*. pp. 27-36. Elsevier: Amsterdam.

Marsh, R., Ward, T. & Landau, J. (1999) The Inadvertent Use of Prior knowledge in a Generative Cognitive Task. *Memory & Cognition*, *27*(1): 94-105.

Miner, J. (1992). Industrial Organizational Psychology, McGraw Hill.

Mutton, P. (2004a). *The PieSky Social Network Bot*. Retrieved from http://www.jibble.org/piespy/ - Accessed 10th July, 2007

Mutton, P. (2004b) Inferring and Visualizing Social Networks on Internet Relay Chat. *Eighth International Conference on Information Visualisation,IV. Proceedings.*, pp35-43.

Nortel Networks (2006). Lead Users and Dynamic Information Transfer, Nortel Networks, USA

Oikarinen, J. & Reed, D. (1993) RFC1459: Internet Relay Chat Protocol

Olson, E. L. & Bakke, G. (2001) Implementing the Lead User Method in a High Technology Firm: A Longitudinal Study of Intentions versus Actions. *The Journal of Product Innovation Management*, *18*(6): 388-395

Open Source Initiative (2006). *Open Source Initiative OSI - Certification Mark and Process*. Retrieved from - http://opensource.org/docs/certification_mark.php - Accessed 26th November 2006.

Page, L., Brin, S., Motwani, R. & Winograd, T. (1998). *The PageRank Citation Ranking: Bringing Order to the Web.* Unpublished technical report, Stanford University.

Perens, B. (2001). Open Source Software. In *Revolution OS, 35mm film*. Wonderview Productions: California.

Raymond, E. (2001). The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary, O'Rielly and Associates, Sebastapol, California

Rheingold, H. (1995). The Virtual Community: Finding Connection in a Computerized World, Martin Secker & Warburg Ltd, England

Robertson, T. (1967) The Process of Innovation and the Diffusion of Innovation. *Journal of Marketing*, *31*, 1, pp14-19

Rogers, E. (1962). Diffusion of Innovations, Free Press New York

Rogers, E. (2003). Diffusion of Innovations, 5th Ed, Free Press, New York

Rothenberg, R. & Narramore, J. (1996) Commentary: The Relevance of Social Network Concepts to Sexually Transmitted Disease Control. *Sexually Transmitted Diseases - Journal of the American STD Association*, *23*, , pp24-29

Rothenberg, R., Potterat, J., Woodhouse, D., Muth, S., Darrow, W. & Klovdahl, A. (1998) Social Network Dynamics and HIV Transmission. *AIDS - Official Journal of the International AIDS Society*, *12*, 12, pp1529-1536

Rowntree, D. (1981). Statistics Without Tears. A Primer for Non-Mathematicians, Charles Scribner's Sons. New York

Sabater, J. & Sierra, C. (2002) Reputation and social network analysis in multi-agent systems. *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1*, pp475-482

Schwartz, M. & Wood, D. (1993) Discovering shared interests using graph analysis. *Communications of the ACM*, *36*, 8, pp78-89

Shah, S. (2000) Sources and Patterns of Innovation in a Consumer Products Field: Innovations in Sporting Equipment. *WP-4105, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Mass*

Stallman, R. (2001). *Transcript of Richard M. Stallman's speech, "Free Software: Freedom and Cooperation" New York University in New York, New York on 29 May 2001*. Retrieved from http://www.gnu.org/events/rms-nyu-2001-transcript.txt. Accessed: 26th November 2006

Tuulos, V. & Tirri, H. (2004) Combining Topic Models and Social Networks for Chat Data Mining. *WI'04: Proceedings of the Web Intelligence, IEEE/WIC/ACM International Conference on (WI'04)*, pp206-213

Urbanec, P. (2005). *Personal communication with Peter Urbanec, computer scientist, October 2005*.

von Hippel, E. (1976) The Dominant Role of users in the Scientific Instrument Innovation Process. *Research Policy*, *5*, pp212-239

von Hippel, E. (1986) Lead Users: A Source of Novel Product Concepts. *Management Science*, *32*, 7, p691

von Hippel, E. (1988). The Sources of Innovation, Oxford University Press, USA

von Hippel, E. (2005). Democratizing Innovation, MIT Press, Cambridge, MA

von Hippel, E. & Sonnack, M. (1999a) Breakthroughs to Order at 3M. *MIT Sloan School of Management Working Papers*

von Hippel, E., Thomke, S. & Sonnack, M. (1999b) Creating breakthroughs at 3M. *Harvard Business Review*, *77*, 5, p47