Innovation und Entrepreneurship
Hrsg.: Nikolaus Franke, Dietmar Harhoff
und Joachim Henkel

Celine Schulz

Organising User Communities for Innovation Management

GABLER

RESEARCH

Celine Schulz

Organising User Communities for Innovation Management

GABLER RESEARCH

Innovation und Entrepreneurship

Herausgegeben von
Professor Dr. Nikolaus Franke,
Wirtschaftsuniversität Wien,
Professor Dietmar Harhoff, Ph.D.,
Universität München, und
Professor Dr. Joachim Henkel,
Technische Universität München

Innovative Konzepte und unternehmerische Leistungen sind für Wohlstand und Fortschritt von entscheidender Bedeutung. Diese Schriftenreihe vereint wissenschaftliche Arbeiten zu diesem Themenbereich. Sie beschreiben substanzielle Erkenntnisse auf hohem methodischen Niveau.

Celine Schulz

Organising User Communities for Innovation Management

With a Foreword by Prof. Dietmar Harhoff, Ph. D.

GABLER

RESEARCH

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Dissertation Ludwig-Maximilians-Universität München, 2009

1st Edition 2009

Editorial Office: Claudia Jeske | Jutta Hinrichsen

Gabler is part of the specialist publishing group Springer Science+Business Media.
www.gabler.de

Umschlaggestaltung: KünkelLopka Medienentwicklung, Heidelberg
Printed on acid-free paper
Printed in Germany

ISBN 978-3-8349-2084-3

Preface

With the astounding success of social networking systems such as Facebook, Bebo, LinkedIn, and MySpace, interest in social networks soared amongst researchers and practitioners. Organised groups of individuals can be of considerable value to firms, but a deeper understanding of the phenomenon is still lacking.

User communities are social networks of individuals who group together with the purpose of exchanging ideas, insights and experiences with each other. Although the concept of leveraging user communities for innovation management is not new, most of the literature has assumed the ready existence of user communities and has concentrated on examining how firms can cooperate with them. Research on the emergence and persistence of user communities has mostly been neglected.

Celine Schulz attempts to fill this gap by focussing on why user communities exist, what their goals are and how they function. In her dissertation, Celine Schulz undertakes a study of individuals in user communities of the software firm Computer Associates (CA). Using an online survey, a unique data set of users in 161 different CA user groups was assembled. This dissertation consists of three empirical studies based on this data set. The first two studies analyse why individuals communicate with other members in their user communities, that is why they are willing to share information with each other and how they benefit from the interaction with their user communities. The third study examines how user communities can be leveraged for firms' innovation processes, by specifically investigating the individual- and group-specific attributes of customer collaborators from user communities.

The dissertation presented by Celine Schulz provides important new research results which promote our understanding of how user communities can be organised internally and externally for innovation management. I strongly recommend this book to researchers and practitioners alike.

Munich, September 2009

Prof. Dietmar Harhoff, Ph.D.

Acknowledgements

First and foremost, I would like to thank my doctoral advisor Dietmar Harhoff for his superb research assistance, encouragement and patience. His infectious enthusiasm and unlimited zeal have been major driving forces through my graduate career at the Ludwig-Maximilian University in Munich. I am especially grateful to him for granting me a remarkable degree of freedom and providing me with the resources I needed to pursue my research interests.

Moreover, I would also like to thank Arnold Picot, my thesis referee and scientific advisor for his motivating support and guidance. I also wish to thank Franz Waldenberger for his advice and insightful discussions during my Master of Business studies.

Equally, I would like to thank my colleagues at the Institute for Innovation Research, Technology Management and Entrepreneurship who provided me with a very pleasant work and research environment. In particular, I am deeply grateful to Karin Hoisl who had provided me with countless advice and fruitful discussions on my research. I would also like to thank Joachim Henkel for his inspiration and valuable guidance, which played an essential role in the conceptualisation of this dissertation. In addition, I wish to thank Georg von Graevenitz, Carolin Häussler, Stefan Wagner, and Thomas von Eggelkraut-Gottanka for their helpful support and valuable comments. Finally, thanks to Rosmarie Wilcox for helping me edit the final drafts of this dissertation.

This entire research project would not have been possible without the assistance of numerous Computer Associates user group members. I am indebted to Leo Hauguel for his constant support of my research endevours, and especially for his persuasive skills in convincing other user group presidents to provide me with assistance during the launch of the web-based survey. Special thanks also goes to Ross Markley, Norbert Flügel, Wilhelm Alexy, Hans-Peter Kotowski, Clark Meyer, Thomas Bock, Philip Gineo, Jan Eiken, Othmar Pusch, Johan Peyron, Herman Koester, Chris Poole and Borge Kraft-Pettersen for their valuable comments and assistance. I would also like to thank the following employees of Computer Associates who had provided me with the necessary resources throughout the course of the research project - Giovanni Di Natale, Sibylle Ritthammer, Jan Van Kan, Dennis Sullivan, Wendy Hughes, John Dueckman, Emma Mcgrattan, Danny Sandwell, Dave Clark and Melissa Borza.

I am also grateful to my parents for their support, continuous encouragement and affection. In particular, I would like to thank my father for his assistance with the final editing of this dissertation. Last but not least, I dedicate my thesis to my husband Jan and my daughter Christina. Only their loving support and inexhaustible patience made all of this possible.

Celine Schulz

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

User communities can be defined as social networks that consist of user nodes interconnected by information transfer links which may involve face-to-face, electronic or any other form of communication (von Hippel 2007). These user networks provide members with connection, support, a sense of belonging, and social identity (Wellman et al. 2002). The widespread and rapid adoption of internet technology has not only facilitated systematic interactions between geographically dispersed individuals who share a common purpose and identity (Wellman 1979), but has also led to a rise in countless online user communities, examples of which include the user-generated video sharing community YouTube, the social-networking community Facebook and the virtual reality community Second Life.

Recently, firms have begun to discover the value of user communities. For instance, Google acquired YouTube for US$1.65 billion[1] in November 2006, and in October 2007 Microsoft announced that it would acquire a US$240 million equity stake in Facebook[2]. Having access to communities of users not only provides firms with additional marketing possibilities, but also provides the opportunity to leverage them for innovation management. Specifically, user communities can be perceived as additional sources of external ideas for firms pursuing an open innovation model (Chesbrough 2003).

It is widely accepted that the integration of customers in the innovation process can be extremely beneficial for firms (von Hippel 1986; Salomo et al. 2003; Gruner and Homburg 2000; Lilien et al. 2002). In general, customers can be either passively or actively integrated into the innovation processes. Passive integration involves the acquisition of

[1] www.theage.com.au/news/Busness/Google-closes-A2b-YouTube-deal/2006/11/14/1163266548827.html (accessed on 29th June 2008)
[2] www.broadcastingcable.com/article/CA6494386.html (accessed on 11th November 2007)

customer preferences from the employment of structured inquiry mechanisms such as market surveys or focus groups (Nambisan 2002). User communities have the advantage of assisting firms in such passive integration by organising and prioritising the pool of user information regarding their needs, feedback to existing products as well as ideas for new products. This identification and integration of customer preferences can help reduce uncertainty and risks involved in innovation projects (Enkel et al. 2005).

Additionally, user communities can also be actively integrated into innovation processes, such as when firms provide users with toolkits which would enable them to create their own innovations (von Hippel 2001b; von Hippel and Katz 2002). As user communities facilitate the interaction of individuals to exchange information and to assist each other with the development of their user innovations (Franke and Shah 2003; Shah 2005a; Hienerth 2006; von Hippel 2007), they have the potential to foster the creativity and innovativeness of individual users. Furthermore, as users in user communities create and, under certain conditions, freely reveal[3] their innovations, firms can aim to profit either directly through the integration of the user-generated innovation into their proprietary innovation model (Harhoff and Mayrhofer 2007), or indirectly through the user-generated complements otherwise not available to their proprietary product (Schulz and Wagner 2008). In October 2007 Apple decided to provide iPhone users with a developer's kit, which would allow users to more easily create mobile games, navigation systems, and other tools.[4] Examples of firms that have profited from user-generated innovations stemming from toolkits include the statistical software firm StataCorp (Mayrhofer 2005), the computer game firm EA Games (Prügl 2006) and the computer-controlled music instrument firm Propellerhead (Jeppesen and Frederiksen 2006).

Another example of active customer integration is that of collective customer commitment, whereby users in user communities develop, submit, inspect and approve user-generated new product designs. The manufacturer will then evaluate and refine a selected number of designs, which they then present to potential customers in the user community for solicitation of commitment. The firm will then only manufacture those designs that have garnered the necessary number of pre-orders from interested customers. A major advantage of this integration method is that it greatly reduces the risk of new product development. Examples of firms that have profited from this method are the U.S. fashion T-shirt manufacturer Threadless and the Japanese consumer commodity manufacturer

[3]Free revealing occurs when an innovator voluntarily gives up his intellectual property rights to that innovation and all interested parties are given access to it. See Harhoff et al. (2003) for a discussion of the conditions under which users will freely reveal their innovations. Users will generally freely reveal their innovations when they perceive higher profits than free-riders from disclosing their innovations (von Hippel and von Krogh 2006).

[4]www.nytimes.com/2007/10/18/technology/18apple.html (accessed on 13th November 2007)

Muji (Ogawa and Piller 2006).

Empirical studies that have examined the relationship between user communities and firms have found that there is often a symbiotic relationship from which both users and manufacturers benefit (Jeppesen and Molin 2003; Jeppesen and Frederiksen 2006; Prügl and Schreier 2006). Examining firms that profit from open source software communities, Dahlander and Magnusson (2005) identify two further relationships which they refer to as commensalistic (where the manufacturer gains and the community is indifferent) and parasitic (where the manufacturer gains at the expense of the community) relationships. Despite this widespread view that user communities are beneficial for firms, more recent studies have found that user communities can sometimes be harmful for firms (Schulz and Wagner 2008; Flowers 2008). An example of such a relationship is when innovations stemming from communities aim at bypassing legal or technical safeguards that prevent users from unsolicited usage of the manufacturer's products (Mollick 2004). Hence, before firms can leverage on user communities for innovation management, they need to have a better understanding of the phenomenon of user communities so that they can relate more effectively to it. It is thus important to firstly understand the origins of user communities, why they exist, what their goals are, how they function and how they maintain their existence.

By organising or supporting the organisation of customers into user communities that are self-managed by the users themselves, firms have the opportunity to effectively acquire both passive and active customer input from a large number of customers into the various stages of the innovation process. Thus, user communities may be an effective alternative to more traditional ways of integrating customers (such as through focus groups or telephone interviews) which are often organisationally complex, and are expensive for obtaining customer knowledge (Prandelli et al. 2006).

Although the concept of leveraging user communities for innovation management is not new, much of the literature has concentrated on documenting how firms can cooperate with a large number of customers in virtual communities (Nambisan 2002; Prandelli et al. 2006; Dahan and Hauser 2001; Füller et al. 2006; Jeppesen and Frederiksen 2006; Prügl and Schreier 2006). Little attention has been paid to evaluating the premises for the persistence of user communities per se. To address this issue, this dissertation presents three empirical analyses on the organisation of user communities in the software industry for innovation management. The first two analyses investigate the antecedents of the give-and-take behaviour of individuals in user communities. The third analysis provides a descriptive study of how user communities can be leveraged by firms for innovation management. The empirical analyses are based on survey data collected from users who are members in user groups of the management software firm Computer Associates.

Given that the phenomenon of user community participation in innovation processes is more widely apparent in the software industry than any other industry (Nambisan 2002), software firms offer the most appropriate context for the empirical analyses. In the following section, a short overview of user communities in the information technology industry is presented, followed by an introduction to the study setting of Computer Associates' user groups.

1.2 User Communities in the Information Technology Industry

The Information Technology Association of America (ITAA) defines information technology as the "study, design, development, implementation, support or management of computer-based information systems, particularly software applications and computer hardware."[5] User communities in the information technology industry are often referred to as user groups and have been in existence for over fifty years. User groups began because manufacturers only marginally understood how to support and enrich the products they produced. The need for assistance was obvious and activist communities arose (Ralston and Reilly 1982).

The first user-initiated user group - SHARE - was officially founded in August 1955 by a small group of International Business Machines (IBM) mainframe users in Santa Monica. The initial impulse for the formation of SHARE was that users felt an urgent need to create a united front against a proposed IBM assembler as they were unhappy about the assembler's performance and usefulness. The first meeting was attended by 46 persons representing 17 companies. A few months after the founding of the SHARE user group, a group of IBM users of commercial computers recognised that the user group idea had merit and founded another IBM user group, GUIDE (Guidance for Users of Integrated Data Processing Equipment) in 1956 (Ralston and Reilly 1982).

According to Blau and Scott's (1962) categorisation of organisations, user communities in the information and technology industry can be categorised as "mutual-benefit associations" where the prime beneficiary is the membership. One crucial problem in mutual-benefit associations is that of maintaining internal democratic processes, or in other words that of providing for participation and control by the membership.

In general, user groups can be either user- or manufacturer-initiated; have exclusive or

[5] http://en.wikipedia.org/wiki/Information_technology (accessed 10th July 2008)

"open to the public" memberships; and be manufacturer-dependent or independent groups (Erichsson 1994).

1.2.1 Organisation of User Groups

Organisations can be defined as social entities that persistently pursue specific goals and possess a formal organisation structure which facilitates the activities of its members to fulfill these goals (Kieser and Walgenbach 2007). Based on this definition, the following parameters provide the central dimensions for differentiating between organisations - its goals, structure and members.

1.2.1.1 Goals

In the late 1950s and early 1960s, an era before software was sold, a fundamental purpose of a user group was the swapping of home-grown software. Before manufacturers supplied sub-routines, users had little but their own ingenuity on which to rely for the numerous routines necessary to keep a system running. For example, the purpose of the SHARE user group as stated in its by-laws was "... to foster the development, free exchange and public dissemination of research data pertaining to SHARE computers", which implies that the group existed to generate a climate for the exchange of data (Ralston and Reilly 1982).

In general, one of the central goals of user groups is to support their members by providing them with a platform to exchange information and use experience, and by assisting them to solve general technical problems. User groups are often categorised in the literature as self-help or support groups that consist of like-minded people who gather together to pool resources, troubleshoot problems, learn about new uses, or otherwise discuss the joys and pains associated with ownership of an unfamiliar technology (Gilliam and Sluzenski 1990). Not only do user groups provide support and benefit to new or novice users of the technology, but also to experienced users who can use the contacts they make to assist them in the solving of complex problems and take advantage of seminars and trainings to advance their career prospects (Gilliam and Sluzenski 1987).

An additional central goal that user groups often have is similar to that of pressure groups, that is, the user group may endeavour to collectively influence the manufacturer regarding its particular concerns and priorities. The group achieves this by articulating its requests and demands for future technology products. Suggestions for the future product may be collected by the manufacturer through questionnaires that are distributed during user group meetings. Alternatively, users also have the opportunity to submit suggestions through their user groups, which are first organised and prioritised within the groups before being submitted on behalf of the user groups to the manufacturer. This systematic process of collecting user preferences is intended to increase the bargaining power of the users with regard to their influence on the manufacturer (Kawasaki 1990; Erichsson 1994).

1.2.1.2 Structure

In general, user groups are often structured according to special interests which may be technology product specific or industry specific. Alternatively, user groups may also be structured geographically, whereby support is provided for organisations or individuals who have a common interest and reside in a particular locality (Gonzalez 1982). To embrace the challenges posed by the rapid internationalisation in the information and technology industry, local or national user groups are increasingly working together at the international level (Erichsson 1994).

To illustrate, the German SAP user group (DSAG) is a national user group oriented for German-speaking users of SAP products. The executive committee of the DSAG is responsible for the overall management of the group, coordinates cooperation efforts with other national SAP user groups, initiates information exchanges between SAP managers and DSAG members and manages the finances of the group. As SAP offers a very wide range of software products, the DSAG is structured according to special interest groups (or sub-user groups) which are either industry-specific, SAP product specific or technology-specific. Hence, depending on the type of information users would like to discuss, they have the option to be members in more than one interest group. To assist in the management of the DSAG, each interest group has a spokesperson or leader who is mainly responsible for mediating the communication between the work group and SAP. In general, each work group organises face-to-face meetings two to three times a year. Users are encouraged to use the DSAG online forum for continuous contact with their work groups throughout the year.[6] The organisation structure of the American SAP user group (ASUG) is very similar to that of the DSAG, with the exception that in addition to the special interest groups, the ASUG is also geographically structured according to "chapters", which provide regular opportunities for members to network and share information and best practices at face-to-face meetings. The ASUG currently has thirty-nine "chapters" across the U.S. and Canada.[7]

1.2.1.3 Members

User groups can be considered clubs, whereby membership is applied for either by firms or individuals. As mentioned earlier, one crucial problem of user groups is that of providing

[6] http://www.dsag.de/dsagcorp/ueber-dsag/organisation.html (accessed on 8th July 2008)
[7] http://www.asug.com/Home/WhatWeDo.aspx (accessed on 9th July 2008)

for participation and control by the membership (Blau and Scott 1962). Despite being a formal member of the user group, rational self-interested individuals in a group may seek to maximise their own private interests, which are rarely perfectly aligned with the interests of other individuals or with the groups to which the individuals belong. Individuals may thus be motivated not to act to advance the common interest or goals of the group to which they belong to (Olson 1965; Milgrom and Roberts 1992; Picot et al. 2005). In the case of user groups, this motivation problem may occur when members participate in user group activities with the aim of benefiting from the user group through the acquisition of information from others while at the same time withholding information of their own. When this free riding[8] behaviour is the interest of most or all members in the group, there would be an under-provision or even a non-provision of the collective good (which in the case of user groups is information) - this outcome is often termed the free rider problem.

Since the successful functioning of any organisation must be based on some circumvention of this free rider problem (Frohlich and Oppenheimer 1970), a great deal of research on the issue of collective action in the last half century have focused on various incentives and mechanisms that aim to mitigate this problem (Olson 1965; Hardin 1982; Cornes and Sandler 1996; Ostrom 1998; Oliver et al. 1985; Oliver and Marwell 1988).

This dissertation focusses on this last aspect of the organisation of user communities - that of its members. In particular, it investigates the determinants that influence the give-and-take behaviour of individual members in Computer Associates' user groups. Furthermore, it also examines how Computer Associates selects user community members for integration into its innovation processes. As the study setting of the dissertation is that of user groups of the management software firm Computer Associates, the following will present a brief introduction of the firm and its user group program.[9]

1.2.2 Computer Associates' User Groups

Computer Associates (CA) was founded in 1976 when Charles Wang decided to market a sorting program, CA-SORT for IBM mainframes (Campbell-Kelly 1995). Since then, CA has acquired more than seventy firms, ranging from medium-sized companies valued at a few million U.S. dollars up to larger companies such as Legent, which CA acquired at a cost of US$ 1.8 billion in 1995 (Campbell-Kelly 2000). As a result of the numerous acquisitions, CA today possesses a diverse product portfolio and delivers products and

[8]Free riding can be defined as the failure of individuals to reveal their true preferences for a collective good through their contributions (Cornes and Sandler 1996). Free riders are hence individuals who see incentives to withhold their resources, hoping that the efforts of others will suffice to provide the good (Frohlich and Oppenheimer 1970).

[9]The following information on Computer Associate's user group program was obtained through interviews with various CA employees and CA user group presidents.

services for operations, security, storage, life cycle and service management.[10] CA is currently one of the world's largest independent software vendors.[11]

1.2.2.1 The CA User Group Program

Through their many acquisitions, CA had also inherited a large number of user groups. During a discussion with some customers at a user conference in 1990, Charles Wang recognised the need to have a formal CA user group program to support his customers. The decision to establish this program was to provide CA users with a platform to communicate not only with other users but also with CA. This close communication was expected to bring learning benefits to the users. CA on the other hand was expected to gain not only from the user-input for the development process of their products but also from customer references who were willing to help them sell their products.

Although CA wanted their user groups to be operated independently from them, they wanted to have the possibility to attend these user group meetings to occasionally conduct presentations on updates of their products and to listen in on the types of issues and concerns that the users discussed amongst themselves. Through the years, the user group program became more centralised with CA taking over a large proportion of the organisational planning for the user groups.

In the early nineties, the user group program consisted of around 20 security software groups and roughly 10 systems software groups. To date, there are over 300 user groups worldwide, each supporting any one or more of CA's software product categories.[12]

To be an officially supported CA user group, user groups have to satisfy a number of requirements. Firstly, each group has to have a minimum of 20 members from at least eight companies.[13] Secondly, it is required that members meet physically at least once a year for a full-day meeting. Thirdly, although user groups have the freedom to choose their members, CA specifies that user groups are not allowed to be used for the solicitation for a member's personal or company benefit. Lastly, CA does not support the establishment of a new user group within a one hour commute of an existing CA user group that supports the same CA product/s.

[10] See the Appendix A for a brief overview of the various software products CA offers.

[11] Based on 2003 revenues, the OECD ranked CA fifth in the independent software vendor industry behind Microsoft Corp., Oracle Corp., SAP AG and Softbank Corp. (Houghton 2004). In the worldwide distributed performance and availability management software industry, CA had the fourth largest revenue in 2004 (Grieser 2005).

[12] In 2006, CA restructured its user group program by grouping together various local or national user groups that support the same product/s into (international) product line communities, whereby the existing user groups were then organised as sub-units of the product line community. This new structure is intended to better facilitate the flow of user-specific information from the customers to CA.

[13] User groups established before April 2000 required only a minimum of eight members from at least four companies.

Official CA user groups are entitled to logistical support from CA, such as the assistance in the organisation of user group activities; free refreshments and the use of CA facilities for user group meetings; as well as web administration support for their user group web sites. Furthermore, user group presidents are given free accommodation to attend the Computer Associates Regional Exchange (CARE) conference, which takes place 2 days before CA's official annual user conference, CA World. The CARE conference consists of a 2 day program where user group presidents not only have the opportunity to network with other IT professionals (i.e. other group leaders) and personally communicate with the product and development managers of their respective software products, but to also have the chance to voice out their feedback and opinions to executives on the management board of CA. In addition, group leaders also receive a registration fee waiver to attend CA World.

1.2.2.2 Inter-National User Group for Endevor

To provide an illustration of how CA leverages its user groups, the following will focus on one particular CA product, the life cycle management mainframe software - Endevor. The Inter-National User Group for Endevor (I-NUGE)[14] is a group of Endevor users made up of representatives from individual local user groups from around the world. Since its establishment[15] in 1992, the I-NUGE has provided its members with a platform for the open exchange of knowledge, experiences, resources, and questions to organisations and individuals involved in the design and development of software applications using Endevor tools and related software. Furthermore, the I-NUGE also acts as an intermediary between CA and Endevor users by organising and prioritising enhancement and upgrade requests from its members and by providing a means of presenting a united position on various issues (when requested by the membership) to CA.

Each I-NUGE member not only represents an organisation that is in possession of a CA license of the Endevor software but also one of the 26 local Endevor user groups recognised by CA. Each local group may appoint up to three I-NUGE representatives, although only one vote is allowed per local user group regardless of the number of representatives from that group. Membership of the I-NUGE also includes a CA representative from the CA Endevor product group, whose involvement in the group is only upon request and as a non-voting participant. The executive board of the I-NUGE consists of a president, a vice-president, an administrator, an internet webmaster, a European liaison, a Asia-Pacific

[14]Information of this user group was obtained from the official I-NUGE website, http://www.inuge.com/ and from correspondences with various committee members of the I-NUGE.

[15]In 1992, the NUGE (National User Group for Endevor) was established. To better reflect its growing international membership, the name of the user group was changed to the Inter-National User Group for Endevor in 1996.

liaison, and the past president from the previous year.

One important function of the I-NUGE for CA is its coordination of the Demand Analysis Requests (DAR) process. DARs are change and/or fix requests submitted by users to CA for updating their software product. At the individual local user group level, members suggest and vote for DARs that they would like for CA to implement in their future releases. These locally prioritised DARs are then accumulated by the I-NUGE and further prioritised, this time at the international level. These internationally prioritised DARs are then submitted by the I-NUGE to CA. As compared to requests submitted by individual organisations, these "power in number" requests are significantly more influential. However, it should be noted that CA exercises its discretion in deciding which of the submitted DARs to implement in future releases of their software.

Besides assisting in the transfer of user-specific information from the users to CA, the I-NUGE also disseminates firm-specific information from CA to its members. Examples of such information include updates of the development process of the next Endevor release, announcements of general technology update webcasts, as well as registration requests for beta testers and development buddies.[16]

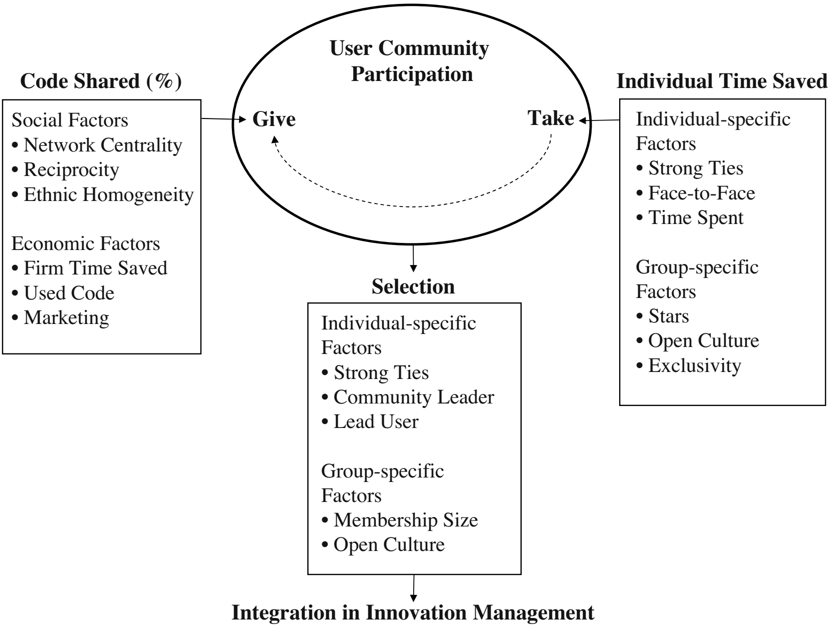
Furthermore, CA also profits from the user-to-user support in the user groups. Instead of contacting the official CA support lines, individual users can browse forum archives of their local Endevor user groups or approach other members for assistance when they encounter difficulties with their software. Additionally, users can also browse and make use of the I-NUGE shareware website to access a compilation of user-generated add-ons, utlities and applications.

1.3 Description of the Dissertation Chapters

This dissertation consists of three empirical investigations which are self-contained as chapters on how user communities can be organised for innovation management. The data used in the three chapters were obtained from a web-based survey on individual users in CA user groups conducted from July 2005 to January 2006. Users were asked questions relating to the CA software they use, the formal user group activities they attended, the personal contacts they have in their group, the software code they wrote, their assessment of their group, the reasons for joining and participating in their group, their position in the group, the general characteristics of the group (these questions were only filled out by committee members of each group), other members in the group, and

[16] The CA Development Buddy Program gives users the opportunity to work closely with CA developers on a specific functionality of the software.

Figure 1.1: Summary of the Dissertation



their personal information.[17] An email with a link to the web-based survey was sent to 398 user group presidents and other committee members from 256 different user groups. These individuals were requested to participate in the survey and to forward the link to their members. Their contact details were obtained from the official CA user group website.[18] A total of 1055 respondents from individuals in 203 different CA user groups participated in the survey. In addition to the survey data, the analyses also draw on results from explorative interviews with twenty-nine CA user group presidents and members as well as thirteen CA employees.

In chapter 2 of the dissertation, socio-economic determinants of an individual's information sharing decision is investigated. Chapter 3 examines individual- and group-specific factors that influence an individual's benefit from participating in user communities. Finally, chapter 4 provides a descriptive study on the individual- and group-specific characteristics of user community members that Computer Associates selects for integration in innovation management.

[17]For more detailed information, please refer to the questionnaire in Appendix B.
[18]http://causergroups.com/UserGroups/ (accessed on 7th July 2005).

Figure 1.1 illustrates how the three chapters are structured within the dissertation. Chapters 2 and 3 are related to user community participation and examine the antecedents of the give-and-take behaviour of individuals. It can be seen from figure 1.1 that the economic benefit (take) from participation in the user community is assumed to influence an individual's information sharing (give) behaviour, but not vice versa. This assumption is substantiated with the theory of collective action which postulates that individuals would be willing to provide or contribute to a collective good (which in the case of user communities is that of information) only if they find it in their private (economic) interest to do so (Olson 1965).[19] This collective community information, which is available to all members in the user community can be considered a collective good which is non-rivalrous in nature - meaning that the cost of its provision does not increase with the number of individuals in the group (Cornes and Sandler 1996). It is thus assumed that the economic benefits which an individual is able to obtain from consumption of the collective good is determined by the information that other individuals contribute to the community, and not by the sharing behaviour of the individual himself. This is because there will be other members in the user community who are highly motivated to contribute to the provision of the public good regardless of the sharing behaviour of other members (Oliver et al. 1985). This dissertation thus assumes that what an individual takes from the user community is not endogenously related to the giving behaviour of the individual.

A brief overview of the contents of this dissertation follows.

Determinants of Sharing Information in User Communities (Chapter 2)

Chapter 2 of this dissertation explicitly investigates the information sharing behaviour of individuals in user communities. As the premise for the existence of user communities is that individuals share and exchange information with each other, the aim of this chapter is to contribute a better understanding of how socio-economic factors influence the information sharing behaviour of an individual. As can be seen from figure 1.1, the factors investigated are network centrality, the reciprocity norm, ethnic homogeneity, the amount of time the user's firm was able to save, if the individual had utilised code from other members in the community, and if the user joined his user community for marketing incentives for his firm. To quantify information shared, the percentage of code written by individuals that are shared with other community members is used. As the share of revealed code is restricted to the interval [0-100%], a Tobit model is selected as the estimation method for the multivariate analysis. The sample used for the empirical analysis is that of survey responses from 519 software code writers in 159 different Computer

[19] Olson (1965) refers to private interest as separate incentives. Separate incentives can be considered extrinsic motivations, which refer to the separable outcome (or indirect reward) that is attained when an activity is done (Ryan and Deci 2000).

Associates' user groups.

To summarise, chapter 2 finds evidence that an individual's information sharing behaviour is not only determined by pecuniary benefits but also by all three dimensions of social embeddedness. These results propose the following implications. Firstly, community leaders should organise activities to actively promote the one-to-one social interaction between members. Secondly, the reciprocity norm should be promoted amongst individuals in the user community. This can be achieved by cultivating trust amongst individuals that every member would contribute and share information with others in the community. Thirdly, it is important that community leaders ensure that users within the same user community have shared collective understandings. One possibility would be to form user communities with members that embrace similar cultures and speak the same language, for instance a German-speaking user community which could include individuals who come from Germany, Austria and Switzerland. Last but not least, community leaders should ensure that the themes of community activities are novel and relevant such that members are able to receive pecuniary benefits from their participation.

Maximising Benefits from Participating in User Communities (Chapter 3)

The objective of chapter 3 of this dissertation is to analyse the impact of both individual- and group-specific factors on the benefits users receive from their user communities. As rational individuals will only participate in and contribute to their user community if they perceive a private (economic) benefit, this knowledge is of particular importance for the management of successful user communities. The amount of time individual members save through having access to their user communities is used to quantify user benefits. As approximately 22% of the respondents indicated the number of hours they saved in the last year as an interval, an interval regression is used for the multivariate analysis. The advantage of the interval regression model is that it is able to take into consideration uncertainty concerning the nature of the exact values within each of these intervals. The empirical analysis is based on survey data from 924 individuals in 161 different Computer Associates' user groups.

Empirical findings demonstrate that the following determinants are relevant factors that influence the benefits users are able to obtain from their user communities: the number of strong intra-group contacts a user has, the amount of time users invest in their user groups, the openness of the communication culture of the user group, the number of star members in the user group and the restriction requirements of membership in the user group. Based on these findings, chapter 3 proposes the following four implications for the management of user communities. Firstly, group leaders are advised to give users more opportunity to socially interact with each other at a more personal level. Secondly, it is essential that members not only have regular contact with one another throughout the course of the year, but that group leaders and forum moderators ensure that the knowledge and information exchanged within the group be of a high quality and relevance for majority of the users. Thirdly, it may be advisable to introduce an incentive for highly experienced and knowledgable users to join and participate in user group meetings, as their presence may increase the potential competence and attractiveness of the group. Lastly, user communities should not only be more open to the general public and have less restrictive membership structures, but also promote an open culture of communication.

Characteristics of User Community Collaborators (Chapter 4)

The final empirical study presented in chapter 4 provides a descriptive analysis of how Computer Associates selects user community members for integration in various phases of the innovation process. Firms often face difficulties in selecting customer collaborators and establishing ties to them (Nambisan 2002; Prügl 2006). In general, it is not efficient to collaborate with every customer due to time, personnel, and monetary restrictions. This paper thus hypothesises that firms are more likely to select user community collaborators who exhibit certain individual- (have a large number of strong ties, are community leaders and lead users[20]), and group-specific (are members in large user groups and in groups with an open culture of communication) characteristics that would enable the firm to effectively acquire the large pool of customer knowledge in user communities. Multivariate binary response and ordered response models are estimated to derive the influence of these characteristics on the customer selection decision for two integration programs that correspond to the idea generation and problem solving phases of the innovation process of the proprietary software firm, Computer Associates (CA). Specifications differ in the design of the dependent variable, which allows for both a detailed analysis of the selection decision for individual collaboration programs as well as a more general analysis of the frequency of selection of individuals in multiple programs. For the case when the dependent variable is a dichotomous (1/0) variable which depicts if a community member participates in a particular customer integration program or not, a Probit estimation method is used. When the dependent variable is a count response - that is the number of different customer collaboration programs an individual was selected for by CA, an Ordered Probit model is estimated.

Empirical results based on data of 808 respondents from 161 CA user groups provide evidence that customers in user communities who are community leaders and lead users are more often selected to cooperate with the firm for both idea generation and problem solv-

[20]Lead users refer to individuals who face needs today that will be general in the marketplace, but face them significantly earlier than the bulk of that marketplace encounters them, and who expect to benefit significantly by obtaining a solution to those needs (von Hippel 1986; Urban and von Hippel 1988).

ing. Additionally, user community collaborators from larger communities have a higher probability of being selected for idea generation; and lead users who are also community leaders are more likely to be chosen for problem solving. These findings suggest that firms should treat their community leaders well, as they are gate keepers who have access to the large pool of customer knowledge in their user community. This is important as such information may be very valuable for the various stages of the innovation process. Secondly, management should also create appropriate incentive mechanisms to motivate lead users in user communities to cooperate with them on various innovation projects. More incentives should be offered to lead users who are also community leaders, especially if they are to be leveraged for the problem solving phase of the innovation process. Finally, firms should be present at activities of large user communities. This would facilitate the selection of active customers for idea generation, who would be better able to filter the large pool of community generated knowledge to the firm.

Conclusion (Chapter 5)

Finally, chapter 5 concludes with a summary of the main findings of chapters 2 to 4, a discussion of the caveats of the empirical studies and suggestions for future research.

Chapter 2

Determinants of Sharing Information in User Communities

2.1 Introduction

Informal information trading is the exchange of know-how, data, and other information between employees working for different companies that is not formally regulated by explicit contractual agreements (von Hippel 1987; Schrader 1991, 1995). A major advantage of such informational exchanges is that firms can save on expenses by avoiding transaction costs related to formal contracts as well as costs incurred by "reinventing the wheel" themselves (Carter 1989). Furthermore, knowledge exchanges between firms can stimulate learning processes and contribute to firms' innovativeness (Allen 1977). Besides these economic incentives to share information, Bouty (2000) suggests that individuals are often socially motivated to engage in information trading. This notion of the embeddedness of economic action was first introduced by Granovetter (1985), who postulates that economic behaviour is closely embedded in structures of social relations.

This paper focusses on the issue of social embeddedness by examining the influence of socio-economic factors on the economic action of sharing information. As the premise for the existence of user communities is that individuals share and exchange information with each other, knowledge of the antecedents of an individual's sharing behaviour is of particular interest for the management of successful user communities. In addition to the influence of pecuniary benefits, this study investigates how structural, relational and cognitive dimensions of social embeddedness affect the amount of information an individual shares with his community.

To empirically examine this phenomenon, survey data from individual users in user groups from the proprietary software producer Computer Associates (CA) will be used. These user groups provide an excellent setting for the analysis because they can be perceived as established informal trading networks that provide a platform for software users to exchange information regarding the usage of their software products. In addition, these user groups consist of individuals who work for different companies and who voluntarily share information with one another in the absence of explicit contractual agreements. Following Henkel (2006), this study will quantify information shared as the percentage of code written by the individual that is revealed to other community members. The empirical analysis is based on 519 survey responses from individuals in 158 Computer Associates' user groups. Results suggest that information sharing behaviour is not only determined by pecuniary benefits but also by all three dimensions of social embeddedness.

Although there are various studies that have examined the information sharing behaviour of individuals working in different firms (von Hippel 1987; Schrader 1991, 1995; Bouty 2000; Henkel 2006), this paper contributes to the information sharing literature by providing the first empirical study that examines this phenomenon using a unique set of survey data from individuals in different informal trading networks. A major advantage of this data set is that it allows for the examination of group- or network-specific determinants on information sharing. Specifically, it is found that a group's ethnic homogeneity (cognitive embeddedness) has a positive influence on the information sharing behaviour of individuals.

The remainder of this paper proceeds as follows. In section 2.2, relevant literature on the embeddedness of economic action and information sharing in scientific and open source software communities will be reviewed. In section 2.3, hypotheses regarding the influence of the three different dimensions of social embeddedness as well as that of pecuniary benefits on information sharing are presented. Section 2.4 describes the data and discusses the key variables used in the multivariate analysis. In section 2.5, the descriptive statistics and the multivariate analysis are presented. The final section provides a discussion of the implications of the results and concludes.

2.2 The Embeddedness of Economic Action

New economic sociology analyses central economic processes with the help of standard sociology, such as network theory, organisational theory, and cultural sociology (Swedberg 1997). The fundamental concept of new economic sociology is that economic actors are not isolated from one another, but are involved in interaction with each other that are of crucial importance for the explanation of economic actions and outcomes (Granovetter 1985, 1990, 1992). This notion of the social embeddedness of economic behaviour was first introduced by Granovetter (1985), who highlighted the fact that the behaviour of rational individuals not only aims at economic goals but also at sociability, approval, status and power.

Economic action is any exercise of an actor's control over resources which is in its main impulse oriented towards economic ends, or more specifically as an individual's behaviour when the satisfaction of a need depends, 1) in the actor's judgement, 2) upon some relatively *scarce* resources and a *limited* number of possible actions, and 3) if this state of affairs evokes specific reactions (Weber [1921] 1978). The embeddedness of economic action thus refers to the fact that economic behaviour is affected by the actor's dyadic (pairwise) relations (relational embeddedness) and the structure of the overall network of relations (structural embeddedness) (Granovetter 1992). In other words, the former conceptualises the relationship at an individual level, whereas structural embeddedness is related to the positions of these relationships in a wider network system.

The concept of social capital will be briefly introduced to assist with explaining the social embeddedness of economic action. Nahapiet and Ghoshal (1998) define social capital as the sum of the actual and potential resources embedded within, available through, and derived from the network of relationships possessed by an individual or social unit. Specifically, they propose that there exists three dimensions of social capital - relational, structural and cognitive. The relational dimension describes the personal relationships individuals have developed with each other through a history of interactions. This concept posits that in ongoing relations, human beings do not start fresh each day, but carry the baggage of previous interactions into each new one (Granovetter 1992). The structural dimension refers to the characteristics of the overall pattern of connections between actors - that is who you reach and how you reach them (Burt 1992). Finally, the cognitive dimension refers to the resources providing shared representations, interpretations, and systems of meaning among parties (Cicourel 1973).

Relational embeddedness will modify economic action through the mere fact of attachment an individual has to other individuals. For instance, an employee may stay in a certain firm despite economic advantages elsewhere because of his attachment to a large number of fellow workers. On the other hand, the structural dimension describes the effect of the efficiency of information spread on the economic action of an individual. Additionally, network structure can also generate normative, symbolic, and cultural structures that can affect economic behaviour (Granovetter 1992). Finally, cognitive embeddedness such as shared collective understandings and shared language is expected to influence economic action by facilitating the perceptive abilities of individuals to evaluate the likely benefits of their actions (Nahapiet and Ghoshal 1998).

As this paper investigates the social embeddedness of the economic action of sharing information in communities, the following will provide a summary of relevant literature investigating this phenomenon in scientific communities and in open source software communities.

In analysing the sociology of science, Merton (1973) postulates that the ethos of science comprises of a set of norms, values and rules which are held to be binding as academic scientists are morally committed to them. These norms, values and rules are examples of what Granovetter (1992) refers to as relational embeddedness. One major concept of this scientific ethos is "communality" - that is, substantive findings of science are a product of social collaboration and are assigned to the community. In particular, Merton argues that scientific knowledge should be freely shared (complete disclosure) so as to be able to contribute to the institutional goal of advancing the boundaries of knowledge. Recognition and esteem are proposed to be the sole property right of the scientist in his discoveries.

Extending the work of Merton (1973), Dasgupta and David (1987, 1994) contrast the prevalence of distinctive norms, customs and institutions that govern university science on the one hand, and industrial Research and Development (R&D) on the other. In particular, they propose that the academic scientific community are mainly concerned with the stock of knowledge and are devoted to furthering its growth, whereas the technological (industrial R&D) community is concerned with the private economic rents that can be earned from that stock. Thus, they suggest that the norm for scientists in the academic community is to disclose their knowledge and for those in the technological community to practise secrecy (Dasgupta and David 1987). Despite this distinction, Dasgupta and David (1994) further propose that the socio-economic structures under which the research takes place makes it possible for a scientist to be part of either, or both communities, within the course of a day. For instance, profit-seeking firms may organise their research facilities in such a way to emulate the open, cooperative environment characteristic of university campuses; and academic scientists may seek to benefit materially by patenting their inventions. Using patent data to analyse networks of inventors in Italy, Balconi et al. (2004) find that academic inventors retain some of their "openness" when moving from the realm of the academic scientific community to the technological community. Specifically, they exchange information with more people and across more organisations as compared to non-academic inventors.

To highlight how economic interests of a firm and employees' social capital are intertwined, Bouty (2000) examined the individual decision process in informal exchanges between industrial scientists (technologists) across organisational boundaries. She finds the following results - 1) scientists exchange non-confidential resources which they have readily at hand, 2) a scientist's willingness to exchange resources depends on the extent of his or her acquaintance, mutual trust and competition with the other partner, and 3) a scientist's willingness to share does not necessarily mean he will actually do it. The interaction logic of weather the exchange will be profitable or equitable is the last step to making the final decision to share.

Open source software communities exhibit characteristics of both academic scientific and technological communities. In the last decade, many researchers have examined the intrinsic and extrinsic motivations as to why individual programmers share codified information in the form of source code with other members in Free/Open Source Software (F/OSS) communities. Intrinsic motivation is defined as the performance of an activity for its inherent satisfaction whereas extrinsic motivations are defined as the separable outcome (or indirect reward) that is attained when an activity is done (Ryan and Deci 2000). Mayrhofer (2005) provides a good overview of the manifestations of the intrinsic and extrinsic motivations of open source software contributors.

Intrinsic motivation can be categorised into obligation-based and enjoyment-based incentives (Lindenberg 2001). Specifically, Lindenberg argues that community or obligation-based motivations exist when individuals act on the basis of principle or when they are socialised into acting in a manner consistent with the norms of a group. This obligation-based motivation is an example of what Granovetter (1992) refers to as relational embeddedness. Examples of obligation-based motivations include reciprocity, altruism, reputation and community identity. The social norm, reciprocity is an important obligation-based intrinsic motivation for F/OSS code contributors. Reciprocity refers to the behaviour of individuals to react to positive actions of others with positive responses and to negative actions of others with negative responses (Ostrom 1998). Empirical studies have found that F/OSS code contributors often feel a personal obligation to contribute because they themselves use F/OSS software (Hars and Ou 2002; Ghosh et al. 2002; Lakhani and Wolf 2005). Altruism refers to the unselfish concern for the well-being of others. Hars and Ou (2002) found that 16.5% of their respondents indicated that they had unconditional altruistic motivations when contributing to the open source software project. Reputation within the community is another example of an obligation-based incentive. Empirical studies analysing the importance of peer recognition for contributors provide differing results. Whereas Hars and Ou (2002) find that peer recognition ranks high in importance for contributors, Ghosh et al. (2002) and Lakhani and Wolf (2005) find that a large majority of the contributors in their samples do not seem to be motivated by the benefits from the gain in reputation within the community. Finally, community identity refers to an individual's self-identification with the hacker community and ethic (Lakhani and Wolf 2005). Similar to peer recognition, studies seem to find differing results regarding the importance of community identification as a motivation for contributors. Hars and Ou (2002) find that community identification to be of low importance, Ghosh et al. (2002) find that community identification is ranked in the middle, and Lakhani and Wolf (2005) and Hertel et al. (2003) find that a large majority of their respondents identify themselves with the hacker community and ethic.

Examples of enjoyment-based motivations include fun and the desire to discover and innovate (creativity). According to Raymond (1999), having fun is part of the hacker[1] (Open Source) identity. For example, Linus Torvalds was largely motivated by fun when creating the successful open source software Linux (Torvalds and Diamond 2001). To be intellectually stimulated or the desire to be creative is another example of an enjoyment-based motivation. Lakhani and Wolf (2005) find that the most important reason as to why individuals contribute to F/OSS projects is because of their desire to be intellectually stimulated. With regard to the importance of fun, Lakhani and Wolf (2005) find that only a moderate percentage (20.3%) of the respondents indicated that they like working with their respective F/OSS development team.

Besides being intrinsically motivated to share software code, individuals are also extrinsically motivated. One of the most important extrinsic incentive for individuals to share their software code is to be able to improve their programming skills through receiving constructive feedback from other developers in the community (peer review) regarding their code contributions (Raymond 1999). Empirical studies by Hars and Ou (2002) and Ghosh et al. (2002) find that a large majority of over 70% of the respondents in both studies contribute their code to F/OSS projects so as to be able to learn and develop new programming skills. This finding is also supported by Lakhani and Wolf (2005). Another extrinsic motive is that of the job signaling incentive proposed by Lerner and Tirole (2001, 2002). They argue that programmers reveal their code to signal their programming capabilities for potential career advancements or future career benefits in the form of better job offers. However, there has been moderate empirical evidence of the importance of this signalling incentive for contributors (Hars and Ou 2002; Ghosh et al. 2002; Lakhani and Wolf 2005). A further extrinsic incentive similar to the job signaling incentive is the "sales signaling incentive", whereby an individual contributes code on behalf of his firm to "advertise" his firm's technological capabilities (Gruber and Henkel 2006) with the desire of increasing the sales of the complementary product/s his firm sells. Hars and Ou (2002) find that only 13.9% of their respondents indicated that they are motivated to contribute code because of their desire to sell related products. However, when analysing the motivations separately for hobbyists and paid professionals, they find that a majority

[1] Software hackers initiated the open source movement in the late 1990s to promote open collaborative projects between software developers (Raymond 1999).

(57.5%) of paid professionals are motivated to contribute software code by the possibility of selling complementary products.

2.3 Determinants of Information Sharing

Informal networking refers to any action that can contribute to the disclosure, dissemination, transmission and communication of knowledge (Pyka 1997). User communities or user groups are examples of such informal trading networks (Schrader 1995) where users who all use a similar product or service, voluntarily communicate with each other on a regular basis to learn more about the product or service and how to use it better (Wenger 2004).

The goal of this paper is to analyse the influence of socio-economic factors on the economic action of sharing information in user communities. Social factors will refer to the three dimensions - structural, relational and cognitive - of social embeddedness as proposed by Nahapiet and Ghoshal (1998). Although Nahapiet and Ghoshal's model analyses social embeddedness at the organisational level, recent research have adapted the model to analyse knowledge sharing at the individual level (Wasko and Faraj 2005; Chiu et al. 2006). Economic factors will refer to the immediate or delayed pecuniary benefits that individuals (and/or the firms they work for) obtain from their actions.

2.3.1 Social Factors

2.3.1.1 Structural Embeddedness

Network structure such as the density (frequency of ties) and connectivity of a network are important predictors of collective action (Marwell et al. 1988; Gould 1993). Collective action refers to the voluntary actions of individuals that contribute to a collective good (Olson 1965). In the case of user communities, the collective good is information such as ideas, insights and experiences, which individuals share with each other. Analyses have shown that close and dense social networks can produce more social capital for maintaining group norms, including norms for cooperation, than atomised networks (Marwell et al. 1988; Coleman 1990; Janky and Takacs 2004). Gould (1995) found positive effects of density on cooperation and Marwell et al. (1988) propose that both the overall density of social ties in a group and the centralisation of network ties have positive effects on the contributions of individuals for a collective good.

Centrality

Individuals who are centrally embedded in a collective have a relatively high proportion of direct ties to other members and are more likely than others to understand and comply with group norms and expectations (Rogers and Kincaid 1981). In this paper, network centrality will refer to the number of direct diadic ties individuals have in their user community. As user communities are examples of informal trading networks whose members are *expected* to communicate with each other to learn more about the product or service and how to use it better, it is expected that

> *H1: Individuals with a higher network centrality in their user community will share more information.*

2.3.1.2 Relational Embeddedness

The relational dimension of social capital refers to assets created and leveraged through relationships and include the following facets - trust, norms and sanctions, obligations and expectations, and identity and identification (Nahapiet and Ghoshal 1998). Of these facets, obligation or the reciprocity norm is suggested to have an important direct influence on the level of cooperation between individuals (Ostrom 1998).[2]

Reciprocity
The norm of reciprocity refers to the behaviour of individuals to react to positive actions of others with positive responses and to negative actions of others with negative responses (Ostrom 1998). According to von Hippel (1987), reciprocity forms the basis for the exchange of information. This is because providing an individual with a favour obliges that individual to reciprocate in order to maintain the stability of the exchange relationship by balancing the benefits and contributions, even in the absence of an explicit agreement.

Thus,

> *H2: Individuals guided by the norm of reciprocity will share more information.*

2.3.1.3 Cognitive Embeddedness

The cognitive dimension of social capital is similar to the principle of *homophily*, which proposes that social interactions tend to take place among individuals with similar lifestyles and socioeconomic characteristics (Lin 2001). Zukin and DiMaggio (1990) suggest that the shared collective understandings amongst individuals play an important role in shaping economic action. An important part of social exchange requires at least some sharing of context between the parties to such exchange (Boland and Tenkasi 1995; Nahapiet and Ghoshal 1998).

Ethnic homogeneity
Shared language and codes influence the perception of individuals by organising sensory

[2] Ostrom (1998) suggests that trust indirectly influences cooperation through its influence on reciprocity.

data into perceptual categories and providing a frame of reference for observing and interpreting the environment, thus encouraging individuals to engage in communication (Nahapiet and Ghoshal 1998; Wasko and Faraj 2005). Empirical studies have shown that communities that are more homogenous in terms of ethnicity, religion and social class tend to exhibit higher levels of cooperation[3] (Banerjee et al. 2004; Miguela and Gugerty 2005; Bandiera et al. 2005). Using individual level panel data on Eastern European workers employed by a UK farm to pick fruit, Bandiera et al. (2005) find that individuals cooperate more when the proportion of co-workers of the same nationality is higher. As the ethnic homogeneity of a community directly influences the extent the community exhibits shared language and codes, it is expected that

> *H3: Individuals who are in user communities with higher ethnic homogeneity will share more information.*

2.3.2 Economic Factors

Cost Savings

Bouty (2000) proposes that profitable exchanges are positively associated with the likelihood of scientists to exchange resources. A major pecuniary benefit from informational exchanges is that firms can save on expenses by avoiding costs incurred by "reinventing the wheel" themselves (Carter 1989). It is thus hypothesised that

> *H4: Individuals who are able to save more costs for their firm through their participation in user communities will share more information.*

Cost savings occur when individuals are able to save time for their firm by utilising the information they obtain from their user communities. One manifestation of this information can take the form of software code that other members in the user community had written. Thus, the following subhypotheses are expected:

> *H4a: Individuals who are able to save more time for their firm through their participation in user communities will share more information.*

> *H4b: Individuals who use code from other members in their user community will share more information.*

[3] See Bandiera et al. (2005) for a review of the theoretical and empirical literature that examine the influence of social heterogeneity on collective action.

Marketing
Pecuniary benefits can also take the form of an increase in expected sales of a firm. Gruber and Henkel (2006) found that active participation of individuals from small firms in F/OSS communities enables firms to be more visible towards potential customers and helps them to build a technical reputation. Hars and Ou (2002) had found that the majority of their respondents who were paid professional programmers are motivated to contribute code because of the possibility to sell complementary products. Thus,

> *H5: Individuals who perceive that participation in user communities will increase the probability of selling complementary products for their firm will share more information.*

2.4 Data Source and Variable Description

2.4.1 Data Source

The empirical investigation concentrates on user groups of a proprietary software firm Computer Associates (CA). As CA's official user group program has been in existence since the early 1990s[4], CA user groups can be perceived as established informal trading networks that provide software users with a platform to exchange information regarding the usage of their software products. Furthermore, members in these user groups work for different companies and voluntarily share information with one another in the absence of explicit contractual agreements.

To prepare for the formulation of the questionnaire, explorative interviews were conducted between October and December 2004 with twenty-nine CA user group presidents and members, as well as a CA employee who was responsible for the user group program. Data for this study were collected from individual users who are members in CA user groups using a web-based survey. The questionnaire was divided into ten sections. In section A, users were asked general information about the number and names of their membership/s in CA user groups, and in section B questions relating to the CA software they use. Section C contained questions regarding their participation and valuation of formal user group activities. Section D asked for information on the personal contacts they have in their group. Section E included questions characterising the software code they wrote and information on their code sharing behaviour. Section F was concerned with their personal evaluation of their user group, and section G with the reasons why they joined and participate in their group. Section H dealt with their position in the group, and included questions regarding general characteristics of the group which were

[4]Refer to section 1.2.2 in chapter 1 for a detailed description of CA user groups.

presented only to committee members of each group. Section I asked for details on other members in the group. Finally, section J included questions on personal information about the users.[5]

A pre-test of the survey was conducted with seven individuals in CA user groups between May and June 2005. Contact details of presidents and committee members of 265 CA user group were obtained from the official CA user group website.[6] On the 19th July 2005, 398 user group presidents and other committee members were then sent an invitation email with a link to the web-based survey, where they were requested to respond to the questionnaire and to forward the link to their members encouraging them to also participate in the questionnaire. 42 emails were returned due to incorrect email addresses. Reminder emails were then sent on the 2nd September 2005, the 11th October 2005, and the 21st November 2005. As of the 11th January 2006, a total of 1055 responses had been received from individuals in 203 different CA user groups.

Although 1055 responses were obtained, only 561 had indicated that they had written code for the CA software they use. As this study will quantify information shared as the percentage of code written by an individual that is revealed to other community members, only software code writers will be included in the empirical analysis. Specifically, users write code in order to use their CA software which is the case when they use a development program; or when they write add-ons, utilities, User Datagram Protocols (UDPs)[7], exits or applications for their CA software. 42 questionnaires were excluded from the analysis because they were not completely answered. Thus the analysis is based on 519 respondents from 158 different CA user groups or 94.4 percent of the relevant sample size. As information on the size of the underlying population of the total number of individual members in all CA user groups is not available, it is not possible to report a precise response rate. However, based on answers by presidents and committee members on the individual membership size of each user group, it can be calculated that approximately 12,149 persons had received the invitation email with the URL link, assuming that presidents and committee members had forwarded the email to all their members. Based on this number, the estimated response rate is approximately 8.7%. A non-response analysis to compare early to late respondents (Armstrong and Overton 1977) yielded no indication of a non-response bias.[8]

[5]See Appendix B for a copy of the questionnaire.

[6]http://causergroups.com/UserGroups/ (accessed on 7th July 2005).

[7]User Datagram Protocol (UDP) is one core protocol of the Internet protocol suite. Using UDP, programs on networked computers can send short messages sometimes known as datagrams to one another. (`http://en.wikipedia.org/wiki/User_Datagram_Protocol`, accessed 2nd June 2008)

[8]In particular, the first 10% was compared with the last 10% of the responses to the survey with regard to the dependent, independent and control variables using T-tests (metric data), Mann-Whitney rank-sum test (ordinal data), and χ^2-tests (nominal data). The observed differences are not significant at the 1% level and do not point to systematic differences between early and late responses.

2.4.2 Variable Description

Dependent Variable

Code shared
According to Henkel (2006), this study will quantify information shared as the percentage of code written by the individual that is revealed to other community members. Individuals who wrote code either for a CA development tool, or for add-ons, utilities, UDPs, exits or applications for their CA software were asked to indicate the percentage of their software code that they revealed in the last year to other members in their CA user group. *Code shared* will be used as a proxy for the amount of information that users reveal to other members in their user community.

Independent Variables

Centrality
As the use of a direct measure of the number of dyadic ties an individual has in his community may introduce a potential endogeneity problem in the econometric estimation models, this study will introduce the use of a proxy measure to identify individuals who have a high level of centrality in their communities. Community leaders are centrally embedded in their user communities. This is because they have direct diadic ties to every member in their community as they are not only responsible for maintaining the member contact list but also for controlling the admittance of new members to the group. Thus, *high centrality* takes the value of one if an individual is a community leader, and zero otherwise; and will be used as a proxy for individuals who have a high level of centrality in the community.

Reciprocity
This variable takes the value one if individuals indicated that they help other members because they received help from their CA user group in the past, and/or because they help other members because they anticipate needing help from their CA user group in the future, and zero otherwise. The items for this variable were adapted from existing empirical open source software literature (Hars and Ou 2002; Ghosh et al. 2002; Lakhani and Wolf 2005).

Ethnic homogeneity
Users were asked to indicate the country in which they work and live.[9] *Ethnic homogeneity* is measured by the number of different countries respondents from each user group come from. A lower number of countries per user group is a proxy of a higher ethnic

[9] Although the author acknowledges the fact that there may exist instances where individuals do not work in the country of their nationality, it is assumed that individuals who work in a particular country are aware of and embrace the national cultures of that country.

homogeneity. *Ethnic homogeneity* is a group-specific variable, which takes the same value for all individuals in the same user group. This measure of ethnic homogeneity was conceptualised by the author.

Time saved
The perceived work time individuals were able to save for their firm over the course of the last year will be used as a measure of the firm's cost savings. This measure is similar that of Lakhani and von Hippel (2003) who had found that one benefit of having access to a user network and community documents is that individuals can save valuable time. Users were asked to estimate their own personal time saved as well as that of other colleagues by making use of all the information, help and advice they had received from and through their CA user group over the course of the last year. If users were uncertain of the approximate number of hours they saved, they were given the alternative of indicating a range of hours.

If users indicated the number of hours saved as an interval, *time saved lower* refers to the lower bound and *time saved upper* the upper bound of their indicated interval. The variable *firm time saved* refers to the sum of the geometric means[10] of the time saved by the individual as well as that of his colleagues. To take into account users who had indicated intervals where the lower bound was zero, the following formula is used to calculate *firm time saved*,

$$firm\,time\,saved = \sum_{i=1}^{2} \sqrt{(time\,saved\,lower_i + 1)(time\,saved\,upper_i)},$$

where i refers to the individual or his colleagues.

To ensure consistency, it is also necessary to perform a similar transformation for users who responded with a point estimate, *time saved point* with the following formula,

$$firm\,time\,saved = \sum_{i=1}^{2} \sqrt{(time\,saved\,point_i + 1)(time\,saved\,point_i)}.$$

Used code
The usage of software code that other individuals wrote can be considered a pecuniary benefit related to cost savings. The *used code* variable takes the value one if the individual indicated that he used code of other members in the community and zero otherwise.

[10]The geometric mean is smaller than or equal to the arithmetic mean, as it gives more weight to the lower as compared to the upper bound. This measure is more appropriate than the arithmetic mean, because the amount of time users saved is positively skewed.

Marketing
The item for this variable was adapted from Gruber and Henkel (2006). Users were asked to indicate the extent to which they agreed or disagreed to the following statement as to why they joined their CA user group, "Because it increases the opportunity of selling my complementary product and/or service". The *marketing* variable takes the value one if the individual indicated that they strongly or somewhat agreed to the statement and zero otherwise.

Control Variables

Competitors
This variable takes the value of one if users indicated that they are users who work for competitors of their firm in the user community, and zero otherwise. As this information may have an influence on the information sharing behaviour of individuals, it is necessary to control for this in the multivariate analysis.

Development tool
Users who use a CA development tool may write different types and a larger amount of code as compared to users of other CA software. As a result, development tool users may reveal a smaller percentage of the code they write. To control for this, this variable takes the value one if users use a CA development tool and zero otherwise.

IP policy
This set of dummy variables control for the different types of intellectual property (IP) policies of users' firms. This is important as the IP policy of the firm is expected to have a large influence on the information sharing behaviour of the individual. *IP restrict* takes the value one if individuals are not allowed by their firms to reveal code, and zero otherwise. The second variable, *IP open* takes the value one when users are allowed to reveal software code that they write to the user community, and zero otherwise. Finally, *IP uncertain* takes the value one when users indicated they are unsure of their firm's IP policy, that is if they are allowed to reveal code or not, and zero otherwise. This measurement of this variable is very similar to that from Henkel (2006).

Use experience
As the level of experience of users may affect their sharing behaviour, users were asked to indicate the year in which they started using the most important CA software that they most frequently use at their present employment. The *use experience* variable is constructed by subtracting their indicated year from 2006.

Low quality code
Mayrhofer (2005) measured the quality of a programmer's code by asking respondents to

Table 2.1: Descriptive Statistics (N = 519)

Variable	Mean	S.D.	Min.	Max.
code shared (%)	9.85	22.75	0	100
high centrality (0/1)	0.21		0	1
reciprocity (0/1)	0.69		0	1
ethnic homogeneity	2.96	3.78	1	15
firm time saved	75.47	158.24	0	1,436
used code (0/1)	0.28		0	1
marketing (0/1)	0.22		0	1
competitors (0/1)	0.36		0	1
DevTool (0/1)	0.54		0	1
IP restrict (0/1)	0.23		0	1
IP open (0/1)	0.46		0	1
IP uncertain (0/1)	0.31		0	1
use experience	11.92	6.28	1	34
low quality code (0/1)	0.11		0	1

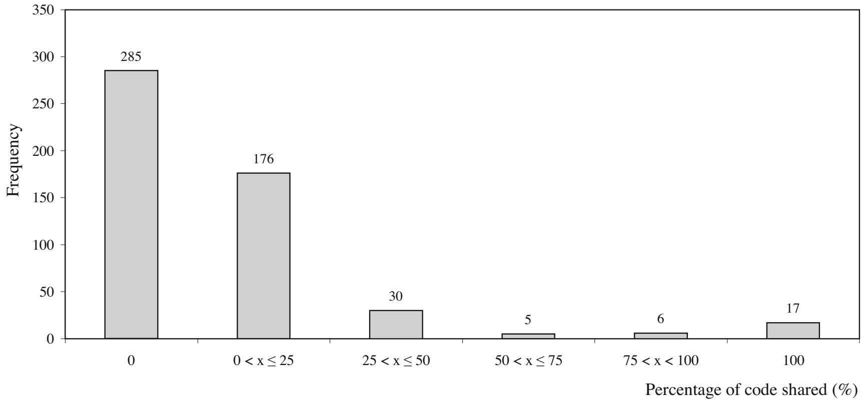
indicate how they perceive the quality of the code they write in comparison to other code writers. Results of the pre-test of the survey indicated that this item should be rephrased. Hence, users were asked to indicate the extent to which they agreed with the following statement "I didn't think my code was of a high quality.". The *low quality code* variable takes the value one if the quality of the code respondents write is of a low quality, and zero otherwise.

2.5 Empirical Analysis

2.5.1 Descriptive Statistics

Table 2.1 presents descriptive statistics of the above-mentioned variables. The percentage of software code that individuals reveal to other members in their community ranges from 0% to 100% at an average of 9.85%. Taking a closer look at the distribution of *code shared* reveals that its median (second quartile) is zero and its third quartile is 5%. These findings indicate that *code shared* has a positively skewed distribution. As illustrated in Figure 2.1, it can be seen that the percentage of code shared decreases monotonically till the interval (50%, 75%] where it then increases slightly in the last two intervals. One reason for the marginal increase in the last category of individuals who revealed 100% of their code is the fact that percentages are naturally bounded by 100. As the percentage of code shared is a measure of an individual's sharing behaviour, this finding (17 individuals who revealed 100% of the code they wrote) indicates that there may be individuals who would have revealed more if they had had more to reveal. This bounded nature of the

Figure 2.1: Percentage of Code Shared (N = 519)



dependent variable will be taken into consideration in section 2.5.2 of the multivariate analysis.

With regard to an individual's level of centrality, it can be observed that 21% of the respondents have a high level of connectivity. This relatively high percentage is due to the fact that community leaders were contacted directly to participate in the survey, whereas regular user group members were informed of the survey indirectly through their community leaders. Approximately 69% of the respondents are guided by the reciprocity norm. This result is in accordance with von Hippel (1987), who suggests that reciprocity forms the basis for the exchange of information. As user communities are supposed to be informal exchange networks, it is expected that the adherence to this reciprocity norm is an important attribute for the sustenance of these communities. The next variable is a user group-specific characteristic. As displayed in Table 2.1, the mean of the number of different countries each user group (at the individual level) represents is approximately 3, with a minimum of 1 and a maximum of 15. At the group level of analysis, the average number of countries per user group is 1.5. This result indicates that majority of the CA user groups tend to be national or country-specific user communities. One explanation for this is that to be an officially sponsored CA user group, members are required to meet physically at least once a year for a full-day meeting. Thus most user groups tend to be localised to a certain country or state in the case of the U.S.A..[11]

On average, users were able to save 75.47 hours for their firm annually by participating

[11]Please refer to section 1.2.2.1 in chapter 1 for more details.

in user groups. *Firm time saved* has a minimum of 0 and a maximum of 1,436 number of hours. The first (5 hours), second (29 hours) and third quartile (80 hours) suggest that *firm time saved* has a right-skewed distribution. Interestingly, only 28% of the respondents used software code in the last year from other members in their user community. Approximately 44% of the respondents revealed code to their user community. This finding suggests that there appears to be more "code givers" than there are "code takers". Examining the contingency table for *used code* and individuals who had shared code reveals that approximately 75% of those who used code also shared code, whereas only 46% of those who shared code had used code of others. Next, 22% of the respondents indicated that they participate in their user group for marketing purposes. Specifically, they had joined their user group to increase the opportunity of selling a complementary product and/or service.

36% of the respondents indicated that there are other individuals in their user group who work for competitors of the firm they are employed in. Approximately 54% of the respondents write code for a CA development tool that they use, whereas the other 46% had written add-ons, utilities, User Datagram Protocols (UDPs), exits or applications for their CA software. 23% indicated that they work for firms with restrictive IP policies that restrict them from revealing any code they write. 46% indicated that their firms have an open IP policy where they are allowed to reveal their code and 31% were uncertain of the existence of a restrictive or open policy. These results suggest that a large percentage of firms are either aware of the cost advantages of trading information or do not perceive software code that is written by their employees to be valuable competitive information for the firm. The average number of years that users have had experience with their CA software is 11.92, with a standard deviation of 6.28. A closer inspection of this variable at the user group level reveals that majority of the user groups have users with a relatively heterogeneous level of use experience, thus suggesting that user groups provide newer users with the possibility to learn from more experienced users. Finally, approximately 11% of the respondents indicated that they did not think that the code they wrote was of a high quality.

An overview of the bivariate relationships between the independent and control variables[12] are presented in Table 2.2. Firstly, individuals who used code of other members appear to exhibit the reciprocity norm more than other individuals who did not use code ($\rho = 0.26$). This relationship is reasonably conceivable as it can be expected that individuals who have used code of others will want to reward the fair intentions of others (Rabin 1993). Secondly, it is not surprising that individuals who used code of other members

[12]The variable *IP open* is omitted in the table, as this variable will not be included in the multivariate analysis in section 2.5.2, as it will be used as the reference group.

Table 2.2: Correlation Matrix of the Independent and Control Variables (N = 519)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**
(1) high centrality (0/1)	1.00										
(2) reciprocity (0/1)	0.08	1.00									
(3) ethnic homogeneity	−0.18◊	0.04	1.00								
(4) firm time saved	0.16◊	0.16	0.14♠	1.00							
(5) used code (0/1)	0.06	0.26	0.16◊	0.25◊	1.00						
(6) marketing (0/1)	0.03	0.13	0.06◊	0.11◊	0.14	1.00					
(7) competitors (0/1)	0.14	0.15	−0.07◊	0.11◊	0.02	0.12	1.00				
(8) DevTool (0/1)	−0.06	−0.03	0.31◊	0.07◊	0.02	0.13	−0.01	1.00			
(9) IP restrict (0/1)	−0.03	−0.08	0.02◊	−0.01◊	−0.17	−0.06	0.02	0.08	1.00		
(10) IP uncertain (0/1)	−0.03	−0.06	−0.02◊	−0.09◊	−0.05	−0.09	−0.08	−0.14	−0.36	1.00	
(11) use experience	0.05◊	0.02◊	0.00♠	0.02♠	−0.03◊	0.08◊	0.12◊	0.12◊	0.01◊	−0.08◊	
(12) low quality code (0/1)	0.00	−0.05	−0.01◊	0.00◊	0.11	0.01	−0.04	−0.01	−0.06	0.06	−0.10◊

Note: Phi-coefficient values unless otherwise stated; ◊ Point biserial correlation coefficient,
♠ Pearson product-moment correlation coefficient

seem to be able to save more time for their firm (ρ = 0.25). This finding is relatively plausible, as individuals do not have to spend time duplicating software code which exists and is available to them for use.[13]

Thirdly, individuals who use a CA development tool tend to come from user groups that have a higher level of ethnic heterogeneity or in other words more international user groups (ρ = 0.31). This positive relationship is a result of the fact that there are two relatively large international user groups in the sample that support two different CA development tools. Responses from these two user groups comprise approximately 12.5% of the sample. Finally, there appears to be a negative relationship (ρ = -0.36) between respondents who work for firms that have a restrictive IP policy and those who were uncertain of their firms' IP policy. As these 2 variables have dichotomous values that represent nominal data, this negative relationship is a result of the approximately equal number of individuals who work in firms with a restrictive IP policy (N=119) and those who were uncertain of their firms' IP policy (N=159).

2.5.2 Multivariate Analysis

2.5.2.1 Model Specification

Tobit Model

As the dependent variable, percentage of code shared is naturally bounded between 0 and 100, a Tobit model (Tobin 1956) is selected as the estimation method for the multivariate analysis. In contrast to an Ordinary Least Squares (OLS) method, a Tobit model allows for the censoring of the dependent variable.

Let the OLS model be defined as

$$y = \boldsymbol{x}\beta + \varepsilon, \tag{2.1}$$

where y represents a $N \times 1$ vector of users' percentage of code shared with other members in their user community, $\boldsymbol{x}$ represents the $N \times K$ matrix of independent and control variables and β represents a $K \times 1$ vector of coefficients.[14] The model assumes $\varepsilon \sim N(0, \sigma^2 I)$. To estimate β, the OLS method minimises the sum of squared errors ($\varepsilon'\varepsilon$), where

$$\varepsilon'\varepsilon = (y - \boldsymbol{x}\beta)'(y - \boldsymbol{x}\beta). \tag{2.2}$$

Unfortunately, the OLS method does not take appropriate consideration of the nature

[13]It should be noted that CA does not certify the software code that individual members reveal to one another.

[14]N represents the total number of observations and K the total number of independent and control variables plus 1 for the constant.

of the underlying dependent variable: sharing behaviour is a latent variable which is not observable. However, what can be observed is a proxy of the latent variable - the percentage of code that users reveal to other community members. It is important to note that the observed variable is a function of the unbounded underlying latent variable but is - as a percentage - naturally bounded between 0 and 100. Denoting a user's unobserved inclination to share or not share information (sharing behaviour) as y^*, the observed model can be formally written as

$$y = \begin{cases} 0 & \text{if} \quad y^* < 0 \\ y^* & \text{if} \quad 0 \leq y^* \leq 100 \\ 100 & \text{if} \quad y^* > 100. \end{cases} \tag{2.3}$$

The underlying relation between the unobserved sharing behaviour y^* is defined by

$$y^* = \boldsymbol{x}\beta + \varepsilon \tag{2.4}$$

where the distribution of ε is $\varepsilon \sim N(0, \sigma^2 I)$. In situations where the dependent variable is censored as described by equations 2.3 and 2.4, an OLS estimation of equation 2.1 will in general be biased towards zero. The Tobit model has the advantage that it takes into account the censored nature of the observed proxy variable, thus providing unbiased β coefficient estimates. Amemiya (1984) contains a survey of Tobit models.[15]

As the analysis is based on responses from individuals in different user groups, there is a high possibility that responses from individuals within the same user group may be correlated, possibly through unobserved cluster effects (Wooldridge 2003). This intra-cluster correlation is often referred to as over-dispersion or extra variation in an estimated statistic beyond what would be expected under independence. Hence, analyses that assume independence of the observations will tend to underestimate the true variance, which would lead to test statistics with inflated Type I errors (Williams 2000). To avoid this over-dispersion problem, unbiased variance estimators will be computed to control for intra-cluster correlation (Rogers 1993).

Ordered Probit Model

To examine the robustness of the results from the Tobit model, this study will estimate two additional models - an Ordered Probit model and an interval regression model. In the Ordered Probit model, the dependent variable *code shared* takes on discrete values from 1 to 6, where '1' indicates that the share is 0%, '2' codes the interval (0%, 25%],

[15]The model specified by equations 2.3 and 2.4 can be considered as a type 1 Tobit model according to the classification proposed in Amemiya (1984).

‘3’ codes the interval (25%, 50%], ‘4’ codes the interval (50%, 75%], ‘5’ codes the interval (75%, 100%), and finally ‘6’ indicates that the share is 100%. The following provides a brief overview of the Ordered Probit estimation method.

Let y now be an ordered response taking on values 0, 1, 2, ..., J for some known integer J (Wooldridge 2002). Assume that, the latent variable y^* is determined by

$$y^* = \boldsymbol{x}\beta + e, \;\; e|\boldsymbol{x} \sim N(0,1), \tag{2.5}$$

Let $\alpha_1 < \alpha_2 < \alpha_3 < \cdots < \alpha_J$ be unknown cut points or threshold parameters, and define

$$\begin{aligned} y &= 0 \quad \text{if } y^* \le \alpha_1 \\ y &= 1 \quad \text{if } \alpha_1 < y^* \le \alpha_2 \\ &\;\;. \\ &\;\;. \\ &\;\;. \\ y &= J \quad \text{if } y^* \ge \alpha_J \end{aligned} \tag{2.6}$$

For instance, if y takes on the values 0, 1, and 2, then there will be two cut points, α_1 and α_2. Given the standard normal assumption for e, each response probability is computed as:

$$\begin{aligned} &P(y=0|\boldsymbol{x}) = P(y^* \le \alpha_1|\boldsymbol{x}) = P(\boldsymbol{x}\beta + e \le \alpha_1|\boldsymbol{x}) = \Phi(\alpha_1 - \boldsymbol{x}\beta) \\ &P(y=1|\boldsymbol{x}) = P(\alpha_1 < y^* \le \alpha_2|\boldsymbol{x}) = \Phi(\alpha_2 - \boldsymbol{x}\beta) - \Phi(\alpha_1 - \boldsymbol{x}\beta) \\ &. \\ &. \\ &. \\ &P(y=J|\boldsymbol{x}) = P(y^* > \alpha_J|\boldsymbol{x}) = 1 - \Phi(\alpha_J - \boldsymbol{x}\beta) \end{aligned}$$

The parameters α and β can be estimated by maximum likelihood. Hence, in comparison to the Tobit model, the Ordered Probit model has the advantage that it allows for a non-linear dependence of the dependent variable on the explanatory variables within the range of [0%, 100%].

Interval Regression Model

The discrete values of 1 to 6 of the dependent variable computed for the Ordered Probit model can be termed interval coded data, as they can take on interval or censored values (Wooldridge 2002). For instance, the discrete value '1' takes on the left censored value 0%, '2' takes on the interval (0%, 25%], '3' the interval (25%, 50%], '4' the interval (50%, 75%], '5' the interval (75%, 100%), and finally '6' takes on the right censored value 100%. The following provides a short overview of how the Ordered Probit Model can be applied to interval coded data.

Let $a_1 < a_2 < \cdots < a_J$ denote the known cell limits, and define y as in equation 2.6, but with a_j replacing the unknown parameter α_j. As the interest is now in the linear model for $E(y^*|\boldsymbol{x})$, the standard normal assumption in equation 2.5 is replaced with the assumption $y^*|\boldsymbol{x} \sim N(\boldsymbol{x}\beta, \sigma^2)$, where $\sigma^2 = \text{Var}(y^*|\boldsymbol{x})$ is assumed not to depend on $\boldsymbol{x}$. The parameters of β and σ^2 can be estimated by maximum likelihood. This estimation method of the Ordered Probit with the cut points fixed and with β and σ^2 estimated by maximum likelihood is called the interval regression model (Wooldridge 2002).

2.5.2.2 Empirical Results

In this section, the influence of the socio-economic factors on information sharing will be analysed using the above-mentioned Tobit estimation method.[16] As can be observed from Model 1 in Table 2.3, users who have a high level of centrality have a positive significant influence on code shared at the 1% level. Specifically, individuals who have dyadic ties to everyone in the community share approximately 16.6% more code with their community. It can hence be concluded that users who are more centrally embedded in their user communities share more information with other members.[17] As a result, hypothesis 1 can be supported by the data.

From Model 1 in Table 2.3, it can also be seen that users who embrace the reciprocity norm share on average 15.2% more code as compared to someone who does not exhibit this reciprocity characteristic. This finding is significant at the 1% level. Accordingly, it is concluded that hypothesis 2 is supported. Hypothesis 3 proposes that users from user

[16] The logarithmical forms of the number of hours firms saved will be used in the analysis. To accommodate for 0 values, the logarithm of 1 plus the values of each variable is used.

[17] Although this paper acknowledges the fact that network centrality and the amount of information that an individual shares may be endogenous, it is expected that the effect of centrality on information shared has a much larger causal relationship as compared to the effect of information shared on centrality. This is because an individual's network position is largely determined by enduring personal characteristics such as demographic characteristics (education), values (activity preference, hedonism and tradition), and personality (conscientiousness, extraversion, neuroticism, agreeableness and openness to experience) (Klein et al. 2004). This potential endogeneity problem is minimised in the empirical analysis as the measure of network centrality used (community leader) is exogenously given and is not related to information shared.

Table 2.3: Tobit Regression Models with Heteroscedasticity-Robust Standard Errors in the Presence of Intra-Cluster Correlation

	Model 1	**Model 2**
	code shared (%)	
high centrality (0/1)	16.638***	11.571*
	[5.371]	[5.895]
reciprocity (0/1)	15.171***	16.446***
	[5.048]	[5.442]
ethnic homogeneity	−0.899**	−0.764**
	[0.367]	[0.381]
ln (firm time saved + 1)	2.858**	3.286**
	[1.191]	[1.278]
used code (0/1)	18.866***	17.562***
	[5.221]	[5.657]
marketing (0/1)	2.424	3.353
	[4.298]	[4.491]
competitors (0/1)	0.158	1.100
	[3.643]	[3.667]
DevTool (0/1)	−7.372*	−9.433**
	[3.844]	[3.971]
IP restrict (0/1)	−33.187***	−28.634***
	[6.336]	[6.920]
IP uncertain (0/1)	−15.897***	−15.610***
	[4.854]	[5.221]
ln (use experience)	0.283	−0.093
	[3.387]	[3.620]
low quality code	4.962	1.615
	[6.004]	[7.153]
Constant	−22.926***	−23.705**
	[8.100]	[9.222]
Observations	519	438
No. of left-censored observations	285	247
No. of right-censored observations	17	12
No. of uncensored observations	217	179
No. of user groups	158	97
Log pseudolikelihood	−1246.28	−1024.02

Robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

groups with higher ethnic homogeneity would share more code. As ethnic homogeneity is measured using the number of different countries user groups represent, it can be observed that users in groups with a lower ethnic homogeneity (or who represent a larger number of countries) share significantly less code at the 5% level. In particular, if the number of countries user groups represent increases by one, users will reveal approximately 0.9% less code. To check for consistency of the results, it is necessary to exclude user groups for which only one member responded to the survey as the *ethnic homogeneity* variable would be automatically one. Model 2 is estimated to include only user groups, whereby at least four members responded to the survey.[18] As can be seen, the coefficient of *ethnic homogeneity* in Model 2 is still significantly negative at the 5% level. Therefore, it can be concluded that the data supports hypothesis 3.

As predicted, individuals who were able to save more time for their firms share more information. Specifically, when the amount of time individuals save for their firms doubles, they would reveal on average 2.9% (significant at the 5% level) more code to members in their user groups. Hence, hypothesis 4a is supported by the data. Next, users who used code of other members reveal approximately 18.9% more code (significant at the 1% level). Thus, hypothesis 4b is also supported by the data. A joint-hypothesis test of the significance of both these variables further concludes that they are jointly significant. The chi-squared statistic of this joint-test is 12.42 with 2 degrees of freedom, and a p-value of 0. These results provide evidence that at the 1% level of significance, individuals who are able to save more costs for their firm through their participation in user communities will share more information, thus supporting hypothesis 4. Hypothesis 5 predicts that users may be motivated for marketing reasons to reveal code. As observed from Table 2.3, the variable *marketing* is not significant. Accordingly, hypothesis 5 is not supported. This finding is consistent with Henkel (2006) who also found no significant effect for the marketing motive. Henkel's interpretation is that firms for which marketing is an important reason to reveal information may contribute code in a more visible way, but not necessarily in higher quantity.

Controlling for other factors that may have an influence on code sharing behaviour reveals that the existence of other individuals working for competing firms has no effect on the percentage of code shared. This finding is in accordance with von Hippel (1987) who found that individuals working in competing firms also share and trade information with each other. Users who use a CA development tool appear to reveal less code as compared to others who use other CA software, although this effect is not significant at the 5% level. Next, in comparison to individuals who work in firms with an open IP policy

[18] This means that the *ethnic homogeneity* variable is constructed for each user group using responses from at least four members.

Table 2.4: Dependent Variable for the Ordered Probit and Interval Regression

Dependent Variable	
code shared (rank)	**code shared (interval)**
Ordered Probit	Interval Regression
1	0%
2	(0%, 25%]
3	(25%, 50%]
4	(50%, 75%]
5	(75%, 100%)
6	100%

who are allowed to reveal code, those working in firms with a restrictive IP policy reveal significantly less code. Similarly, those who were uncertain of their firm's IP policy also revealed significantly less code. Finally, it can be observed that the number of years of experience an individual has with the software and the quality of code he writes have no effect on his information sharing behaviour. A joint test revealed that all six control variables are significant at the 1% level.

To test the robustness of the results of the Tobit model in Table 2.3, two additional models are estimated - an Ordered Probit and an interval regression. Table 2.4 displays the dependent variables in each these models. As can be observed from Table 2.5, *ethnic homogeneity* is not significant in both the Ordered Probit and the interval regression estimations. However, it should be noted that the coefficient of *ethnic homogeneity* has the hypothesised negative sign.[19] This finding suggests that the variability of the dependent variable *code shared* that can be explained by ethnic homogeneity is greatly reduced when *code shared* is coded as ordinal data or categorised as intervals. This is because categorising numeric data to ordinal data or into intervals has the disadvantage that the amount of information on the dependent variable is reduced. Nevertheless, with the exception of hypothesis 3, the results from the Ordered Probit[20] and interval regression produce very similar results as the Tobit model with regard to the significance of the tests of the hypotheses. Furthermore, the interpretation of significant coefficients in the interval regression are very close to those of the Tobit estimations in Table 2.3. Hence, it can be concluded that the results of the Tobit model are relatively robust and reliable.

[19]Marginal effects for Model 3 were estimated for each specific parameter value of the dependent variable - *code shared (rank)*. The marginal effects for the variable *ethnic homogeneity* all indicate the hypothesised negative sign, although they are not significant.

[20]Marginal effects for Model 3 were estimated for each specific parameter value of the dependent variable - *code shared (rank)*. All marginal effects of each independent variable have the same sign (interpretation) as that of their coefficients.

Table 2.5: Ordered Probit and Interval Regression Models with Heteroscedasticity-Robust Standard Errors in the Presence of Intra-Cluster Correlation

	Model 3	**Model 4**
	Ordered Probit	**Interval Regression**
	code shared (rank)	code shared (interval)
high centrality (0/1)	0.491***	16.552***
	[0.142]	[5.154]
reciprocity (0/1)	0.442***	13.844***
	[0.132]	[4.837]
ethnic homogeneity	−0.011	−0.563
	[0.011]	[0.397]
ln (firm time saved + 1)	0.094***	2.908**
	[0.033]	[1.139]
used code (0/1)	0.571***	18.677***
	[0.129]	[5.075]
marketing (0/1)	0.100	2.174
	[0.122]	[4.047]
competitors (0/1)	0.039	0.022
	[0.098]	[3.470]
DevTool (0/1)	−0.114	−3.522
	[0.108]	[3.737]
IP restrict (0/1)	−1.026***	−34.865***
	[0.170]	[5.984]
IP uncertain (0/1)	−0.445***	−15.899***
	[0.121]	[4.515]
ln (use experience)	−0.006	0.300
	[0.097]	[3.295]
low quality code	0.185	4.31
	[0.168]	[5.923]
Constant		−22.355***
		[7.994]
Observations	519	519
No. of left-censored observations		285
No. of right-censored observations		17
No. of interval observations		217
Log pseudolikelihood	−471.73	−548.15

Robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

2.6 Discussion of Findings

An essential prerequisite for the existence and sustenance of user communities is that individuals share information with each other. Thus, the aim of this paper is to examine socio-economic determinants of the information sharing decision of individuals in user communities. Results from the empirical findings suggest that economic action is indeed closely embedded in structures of social relations.

In particular, this study finds evidence that all three dimensions of social embeddedness play a significant role in the information sharing decision of individuals. Firstly, with regard to the structural dimension, the empirical results show that individuals who are more centrally embedded in their user communities share more information. This finding is consistent with research by Wasko and Faraj (2005) and Chiu et al. (2006) who find that a higher network centrality increases the likelihood of contributing more information in online communities. Secondly, the influence of the norm of reciprocity (relational embeddedness) is found to be significantly positive. This result is in accordance with von Hippel (1987) and Schrader (1991, 1995), who propose that the premise for the exchange of information is reciprocity. In the absence of explicit contract agreements, adherence to the reciprocity norm maintains the stability of the exchange relationships and hence the existence of user communities. The third dimension of social embeddedness examined is that of the cognitive dimension. Findings indicate that individuals in user communities with a higher racial homogeneity share more information with each other. Therefore individuals who are more cognitively embedded in their user communities will share more information with each other.

Besides the influence of social embeddedness on information sharing, this paper also examined the role of economic factors. Results provide evidence that pecuniary benefits motivate individuals to share information with each other. Specifically, it is observed that individuals who are able to save more time for their firms and who use code of other individuals share more information.

These findings provide guidance for the management of user communities. Specifically, it suggests that attempts must be made to create an environment conducive for the exchange of information amongst individuals. Firstly, individuals should be given more opportunities to come into direct contact with other individuals in the user community. This could be achieved by organising regular user group meetings either physically or virtually and by promoting discussions in forums. Secondly, it is important to cultivate the reciprocity norm amongst individuals in the user community, as it was shown that individuals who embrace this norm reveal more information. As trust is suggested to directly influence reciprocity (Ostrom 1998), the promotion of trust amongst individuals could encourage individuals to reveal information because they believe that others would also feel obligated to do so. Thirdly, special attention should be paid to ensure that individuals within the same user community have shared collective understandings, as results show that individuals in more racially homogenous user communities are willing to reveal more information. For instance, it may be important for user communities to be country-specific as members would embrace similar cultures and speak the same language. Last but not least, it is essential that individuals receive pecuniary benefits for sharing information, as firms would only allow their employees to take part in user community activities if they perceive potential cost savings from their participation.

One limitation of this study is that it focuses only on the sharing of codified information (software code). However, information sharing can also take the form of tacit information.[21] For instance, instead of sharing the actual code, individuals may share tacit information of how they wrote their code. Future studies could try to analyse if there are any differences between the influence of social embeddedness on the sharing of codified and tacit information. A further drawback of this study is in the measurement of centrality. This paper makes the assumption that community leaders are the most central figures within their user communities and ignores the existence of other individuals who may also have as many dyadic ties to all other individuals in the community as the community leader. Despite this drawback, it should be noted that this measure of centrality avoids potential endogeneity problems, as community leaders are exogenously determined.[22] Future research could attempt to better model the causality between centrality and the information sharing decision using alternative instrumental variables.

Overall, this study provides an important contribution to the literature on information sharing (von Hippel 1987; Schrader 1991, 1995; Bouty 2000; Henkel 2006) by being the first empirical study to investigate the influence of a group-specific social factor (the group's ethnic homogeneity) on the information sharing behaviour of individuals. Further studies should extend on the work of this paper by examining the influence of other group-specific factors on an individual's sharing behaviour.

[21] In contrast to codified knowledge, tacit knowledge can be neither expressed in a particular language nor recorded on a particular medium, and as a result is not readily transferable (Cowan et al. 2000; Foray 2004).

[22] Community leaders are selected by other members in the CA user group. Based on interviews conducted between October and December 2004, community leaders tend to be selected largely based on personal characteristics such as their personality (extraversion, agreeableness and openness to experience) and their capabilities in managing a user group.

Chapter 3

Maximising Benefits from Participating in User Communities

3.1 Introduction

Groups of customers in the form of user communities can exercise a powerful influence on the market, because of the speed with which they can disseminate information (Prahalad and Ramaswamy 2000). For instance, closely-knit, networked groups of customers can play a vital role for the adoption and diffusion of innovations (Chakravorti 2004). Furthermore, user communities can also be seen as an additional source of external ideas for firms pursuing an open innovation model (Chesbrough 2003). In particular, these communities have the potential to assist firms in organising and prioritising the pool of user information regarding their needs, feedback on existing products as well as ideas for new products. The access to and integration of such community-generated information have been demonstrated by various studies to contribute significantly to product development (Jeppesen and Molin 2003; Jeppesen and Frederiksen 2006; Prügl and Schreier 2006). But before firms can leverage user communities for innovation management, it is important to firstly understand why these user communities exist, what their goals are and how they function. To address these issues, this paper examines user groups from a proprietary software producer Computer Associates (CA), with the objective of identifying determinants of successful user communities.

An important premise for the existence of user communities is that the transmission and acquisition of (tacit and codified) knowledge can be both difficult and costly. In contrast to codified knowledge, tacit knowledge can be neither expressed in a particular language nor recorded on a particular medium, and as a result is not readily transferable (Cowan et al. 2000; Foray 2004). The term 'tacit' was first introduced by Polyani (1966), who proposed the notion that individuals can know more than they can tell, because they are often not conscious of the knowledge they possess. Frequent interactions with their user communities can assist users to exchange both tacit and codified knowledge. This is because these regular exchanges foster the establishment of common cognitive platforms and common social norms (Cohendet 2006), which in turn increase the willingness of individuals to share such knowledge with each other. Furthermore, the easy access to potential knowledge pools within user communities can significantly reduce costs users incur when searching for solutions to their problems; or for innovative ideas as to how they can better optimise the use of their products (Lakhani and von Hippel 2003; Franke and Shah 2003).

To understand how user communities function, the theory of collective action (Olson 1965) is used to examine the incentives and mechanisms that motivate individuals to act in support of their group interests. With the phenomenal success of open source software in the last decade, interest in the theory of collective action and its application in explaining the microfoundations of open source communities soared (von Hippel and von Krogh 2003; von Krogh et al. 2003a). The reasons as to why these horizontal innovation networks are successful, have been widely discussed (von Hippel 2007). Apart from having analysed the extrinsic and intrinsic motivations - such as future career benefits, skill improvement, fun, reciprocity and altruism - of individual programmers to voluntarily contribute to a public good (Lerner and Tirole 2002; Lakhani and Wolf 2005), studies have also examined the organisational structures responsible for the internal functioning of these communities (Weber 2004; Brügge et al. 2004).

The objective of this paper is to analyse the impact of both individual- and group-specific factors on the benefits users receive from their user communities. As rational individuals will only participate in and contribute to their user community if they perceive a net benefit, this knowledge is of particular importance for the management of successful user communities. This paper hence contributes to the literature on the management of user communities by providing a discussion of how user communities should be organised in order that members' benefits are maximised. To quantify user benefits, the amount of time individual members save through having access to their user communities is used. The empirical analysis is based on 924 survey responses from individuals in 161 Computer Associates' user groups.

Hypotheses for the individual- and group-specific factors are derived from a diversity of disciplines ranging from sociology and psychology to economics and business. The empirical results propose that the following determinants have an impact on the benefits individuals are able to receive from their user communities: the number of intra-group informal ties, the amount of time users invest in their groups, the communication culture of the group, the number of star members in the group and the groups' membership requirements.

The rest of this paper is structured as follows. Relevant literature will be reviewed in section 3.2 of this paper. Next, the hypotheses of the determinants of successful user communities will be derived in section 3.3. This will be followed by the research methods and data description in section 3.4. Section 3.5 presents the descriptive statistics and the results of the multivariate analysis. Finally, this paper concludes by presenting a set of implications for the successful management of user communities in section 3.6.

3.2 Literature Review

Olson (1965) was the first scholar to formalise the notion of collective action and the free rider problem[1] in his book, *The Logic of Collective Action.* According to Olson, the characteristic and primary function of organisations is to advance the common interests of groups of individuals. Hence, it can be considered that when any one common interest of the group has been satisfied, then a collective good has been provided for that group. Olson particularly addressed the issue of the conflict between individual and collective interests and contended that in the absence of coercion and some separate individual incentive, rational self-interested individuals in a group will seek to maximise their personal welfare and not act to advance the interest of the group to which they belong to. His central thesis was that larger groups are less likely to succeed in achieving their goals as compared to smaller groups.

Olson supported his thesis with the argument that individuals will tend to free ride[2] on the contributions of others in large groups. This is because an increase in the size of the group decreases the likelihood that the contribution and non-contribution of any one individual will be perceptible. In other words, if one member does or does not help to provide the collective good, no other one member will be significantly affected. Hence, individuals have little or no incentive to contribute to the group. When this free riding behaviour is in the interest of most or all members in the group, there would be an under-provision or even a non-provision of the collective good - this outcome is often termed the

[1]Modern political theorists in the eighteenth century such as David Hume, Alexander Hamilton, James Madison, and John Stuart Mill had developed theories of the free riding behaviour way before Olson's work. However, most of these theories were hidden in written works that were totally disregarded in the mid-twentieth century until Olson gave them relevance again (Dougherty 2003).

[2]Free riding can also be defined as the failure of individuals to reveal their true preferences for a collective good through their contributions (Cornes and Sandler 1996). Free riders are hence individuals who see incentives to withhold their resources, hoping that the efforts of others will suffice to provide the good (Frohlich and Oppenheimer 1970).

free rider problem.

Since the successful functioning of any organisation must be based on some circumvention of this free rider problem (Frohlich and Oppenheimer 1970), a great deal of research on the issue of collective action in the last half century has focused on various incentives and mechanisms that mitigate this problem. In the following, these incentives and mechanisms will be illustrated using examples from the conventional collective action literature and the more recent open source software literature.

Although Olson (1965) first introduced the idea that the presence of separate incentives can motivate individuals to advance the interest of the group they belong to, it was Hardin (1982) who focused his work on the internal incentives that motivate groups to act. Separate incentives can be considered extrinsic motivations, which refer to the separable outcome (or indirect reward) that is attained when an activity is done (Ryan and Deci 2000). One such separate incentive Hardin (1982) addressed, was the *political entrepreneurship motivation*, which occurs when individuals, who for their own career reasons, find it in their private interest to work to provide collective benefits to relevant groups. Lerner and Tirole (2001, 2002) propose a similar separate incentive, which they refer to as the signaling motive for career advancements and/or future career benefits. Additional extrinsic motivations from the open source software literature include: monetary rewards, skill improvement and active peer review (Ghosh et al. 2002; Hertel et al. 2003; Lakhani and Wolf 2005).

Besides extrinsic motivations, Hardin (1982) also acknowledged the existence of intrinsic motivations, which refer to the inherent satisfaction of performing an activity (Ryan and Deci 2000). Hardin's example of an intrinsic motivation was his concept of *extrarational motivations*, which occurs when the unselfish and social motives of members bring about the provision of a collective good. For example, fervent nationalism might inspire a group of individuals to organise a political party. The social norm, reciprocity is also an example of an intrinsic motivation. Reciprocity refers to the behaviour of individuals to react to positive actions of others with positive responses and to negative actions of others with negative responses (Ostrom 1998). Hence individuals may be willing to contribute to the group because they have benefited from the contributions of other group members in the past. Additional intrinsic motivations from the open source software literature include: fun (Torvalds and Diamond 2001; Lakhani and Wolf 2005; Bitzer et al. 2004), reputation (Ghosh et al. 2002; Lakhani and Wolf 2005), and altruism (Zeitlyn 2003; Bitzer et al. 2004).

Without the presence of a formal organisation or an informal group agreement, individuals may have insufficient intrinsic and extrinsic motivations to voluntarily contribute to the group (Olson 1965; Hardin 1982). Hence, it is sometimes necessary for a central authority, in the form of an external institution or a group leader, to establish control mechanisms in the group that encourage members to participate and discourage free-riding (von Hippel and von Krogh 2003). Examples of such control mechanisms include: facilitating the coordination between members, monitoring their contributions, or enforcing penalties against free riding (Oliver 1980; Swanson 1992; Taylor and Singleton 1993). Additionally, the central authority can also promote a collective identity (Friedman and McAdam 1992) and social norms such as reciprocity, reputation and trust (Ostrom 1998) between members in the group. As mentioned above, such norms can be referred to as intrinsic motivations that can help solve social dilemmas[3] such as the free rider problem (Ostrom 1998).

In the following, the importance of group leadership to coordinate and monitor members' contributions of two successful open source software projects are presented. It is a common misconception that open source communities are simply 'networked hordes of programmer/anarchists' (Raymond 1999). Although individuals are intrinsically and extrinsically motivated to contribute to the open source community, it is nevertheless necessary for an individual or a group of individuals to take the lead in coordinating and organising community activities (Weber 2004). One of the most successful open source software, Linux was created in 1991 by Linus Torvalds. In its early development stages, the Linux project was chiefly led and controlled by Torvalds. As the size of the community of developers increased, Torvalds reorganised his community and delegated responsibilities for subsystems and components of the kernel to a core group of developers. As a result of his high involvement and control of the Linux project, Torvalds was also sometimes known as its 'Benevolent Dictator' (Torvalds and Diamond 2001; Weber 2004). In contrast to Linux, the Apache server software project did not originate from a single leader, but rather a core group of seven developers in 1994 (von Hippel 2001a; Lerner and Tirole 2002). The group devised a voting system based on a minimal quorum consensus rule to collectively involve all developers who contributed to Apache in the group's decision-making processes (Mockus et al. 2000). This voting system is used not only to make code changes but also for the inclusion of new members (Fielding 1999).

To summarise, the theoretical and empirical literature on the conventional theory of collective action and that of open source software suggest that well managed groups can assist in promoting the intrinsic and/or extrinsic motivations of individuals to participate in and contribute to the group to which they belong to.

[3] Social dilemmas are situations when individuals face choices in which the maximisation of short-term self-interest yields outcomes that leave all participants worse off than feasible alternatives (Ostrom 1998). In other words, each individual will receive a higher payoff for defecting than for cooperating, but all are better off if all cooperate than if all defect.

3.3 Determinants of Successful User Communities

This study will base its analysis on a specific type of organisation of individuals, that of user communities. User communities are defined as groups of individuals who all use a similar product and who learn how to use it better as they interact on a regular basis [4] (Wenger 2004). Through the sharing of ideas, insights and experiences with each other, user communities provide benefits for their members. For instance, access to the user network and community documents can save individuals valuable time (McDermott 2002; Lakhani and von Hippel 2003). This paper hence defines successful user communities as groups whose members have benefited significantly from their acquisition of knowledge and information from their user community. To quantify users' benefits, the amount of time individual members save through having access to their user networks is used. In the following, hypotheses of the impact of individual- and group-specific factors on users' benefits will be derived.

3.3.1 Individual-Specific Determinants

Number of Strong Ties

A tie is a relationship between two individuals. The strength of a tie is the combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterise the tie (Granovetter 1973). These relational ties between individuals can be considered channels for the transfer or the flow of resources, such as information and knowledge (Wasserman and Faust 1994).

Although Granovetter (1973, 1983) proposes that weak ties distribute information more efficiently because they provide individuals with access to information and resources beyond those available in their own social circles, he acknowledges that strong ties have greater motivation to be of assistance and are generally more easily available. Applying Granovetter's strong tie hypothesis to the study of entrepreneurship, Jack (2005) finds that strong ties are instrumental for business activity and are used extensively to provide knowledge and information, as well as to maintain, extend and enhance business and personal reputations. Carpenter et al. (2003) also find that in inter-organisational contact-making in U.S. healthcare politics, individuals tend to convey new information to their strong ties.

As individuals with a greater number of close contacts may be more accessible to receive new information and potential assistance for difficulties they encounter, it is expected that:

[4] Franke and Shah (2003) used a similar "communities of practice" definition for their notion of user communities.

H1: Users who have a larger number of strong ties to other members benefit more from their user communities.

Learning Ability

An individual's experience on a series of related or similar tasks may improve his ability to learn new tasks. This progressive improvement in the performance of learning tasks is a form of knowledge transfer known as *learning to learn* (Ellis 1965; Estes 1970). To illustrate this, Ward (1937) conducted an experiment where participants were required to learn successive lists of twelve syllables, one list a day, which were all about equal in difficulty. He found that participants required approximately thirty-eight trials to completely master the first list; twenty trials to reach mastery after six lists; and only fourteen trials to reach mastery after fifteen lists. Improvement in participants' learning rates were most rapid during the first six lists and were more gradual afterwards. Similar results were obtained in a study by Melton and von Lackum (1941).

As the investment in an individual's learning ability is expected to positively impact his ability to identify, assimilate and use new knowledge (Cohen and Levinthal 1990), it can be assumed that:

H2: Users who invest more time in improving their learning abilities benefit more from their user communities.

Face-to-Face Communication

It has been shown that face-to-face communication promotes higher levels of cooperation[5] (Ostrom 1998; Ostrom et al. 1994). In particular, Sally (1995) finds that when participants communicate face-to-face in one-shot prisoners' dilemma and public good experimental games, their cooperation rate rises, on average, by more than 45 percentage points.

Additionally, face-to-face settings help promote the transfer of tacit knowledge. As tacit knowledge can be neither expressed in a particular language nor recorded on a particular medium (Cowan et al. 2000; Foray 2004), face-to-face dialogues or interactions can facilitate the articulation of tacit knowledge into codified knowledge (McKenny et al. 1992; von Krogh 1998; Foray 2004). Polyani (1958) contends that such tacit knowledge can only be passed on by example from master to apprentice. Hence,

H3: Users benefit more from face-to-face interactions as compared to other forms of communication with their user communities.

[5] See Ostrom (1998) for an overview of experimental studies that find a positive effect of face-to-face communication on cooperation.

3.3.2 Group-Specific Determinants

Open Communication

An open exchange of information within groups can limit conflict between members (Lau and Murninghan 1998). Furthermore, a culture that encourages a free flow of knowledge would encourage individuals to share more of their knowledge with each other (von Krogh 1998), thus increasing the quantity of the potential knowledge pool in the community.

Shah (2005b) finds that open communication amongst users in innovation communities can increase the diversity of expertise in problem solving and allow the results of trial-and-error experimentation by multiple parties to be exchanged. Furthermore, open communication not only increases the likelihood that users find effective solutions to their problems, but also reduces the time required to find or create such solutions.

Therefore, the following relationship is expected:

> *H4: Users benefit more from user communities that have an open culture of communication.*

Number of Star Members

Star members in open source software projects refer to the core group of developers who contribute most of the code and oversee the design and evolution of the project (Crowston et al. 2006; von Krogh et al. 2003b). In the case of the biotechnology industry, Zucker and Darby (1996) found that the number of collaborative links to *star scientists* had significant positive effects on the success[6] of new biotechnology enterprises. These *star scientists* were highly productive researchers who accounted for only 0.8% of all scientists listed in GenBank[7] through to 1990, but for 17.3% of all published articles.

Similar for open source software communities, *star members* in user communities are characterised as having a high level of experience, skill and knowledge of the product they use. The participation of these *star members* can help stimulate constructive discussions and informational exchanges between the users. Therefore, a greater number of these *star members* can result in a higher quality of the available pool of knowledge in the community. Thus,

> *H5: Users benefit more from user communities that have a larger number of star members.*

[6]Three measures of success were analysed - products in development, products on the market and employment growth.

[7]GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. (http://www.psc.edu/general/software/packages/genbank/genbank.html, accessed on 14th February 2006.)

Exclusivity

In the theory of clubs, individuals group together to share the cost of a public good. One of the underlying premises of club theory is that group size must be controlled in order to ensure optimal sharing (Buchanan 1965). Hence, exclusion mechanisms are frequently used to monitor users' rates of utilisation and to bar nonmembers and/or nonpayers. It is considered that such exclusion mechanisms provide incentives for members to join, to contribute to the group and to pay dues (Cornes and Sandler 1996). It is therefore, proposed that:

> *H6: Users benefit more from user communities that have exclusion mechanisms.*

As exclusivity can take more than one form, the following subhypotheses are expected:

> *H6a: Users benefit less from user communities with memberships that are open to the general public.*

> *H6b: Users benefit more from user communities that require their members to pay a membership fee.*

3.4 Research Methods and Data Description

3.4.1 Methods

Computer Associates' user groups were selected as the specific research context, to quantitatively analyse the above-mentioned hypotheses on the determinants of successful user communities. To assist in the formulation of the questionnaire, explorative interviews with twenty-nine user group presidents and members, as well as a CA employee who was responsible for the user group program, were conducted between October and December 2004. Specifically, these interviews served to identify similarities between the numerous CA user groups that support different types of software and provided valuable insight to the general goals and functioning of the groups. From the interviews, it was decided that a web-based questionnaire would be the ideal data-collection method.

Between May and June 2005, a pre-test of the questionnaire was conducted with seven individuals in CA user groups that use various CA software. On the 19th July 2005, an email with a link to the web-based survey was sent to 398 user group presidents and other committee members from 265 different CA user groups. These individuals were requested to participate in the survey and to forward the link to their members. Their contact details were obtained from the official CA user group website.[8] 42 emails were returned due to incorrect email addresses. Three additional follow-up emails were sent to remind individuals to complete the survey and to also encourage their members to participate. The collection of data from 19th July 2005 to 11th January 2006 yielded 1055 responses from individual users in 203 different CA user groups.[9]

The empirical analysis uses answers from 924 respondents who are members in 161 different CA user groups. This reduction in the sample size is due to incomplete questionnaires from 131 respondents. Due to lack of information on the actual size of the basic population of the total number of individual members in all CA user groups, a precise response rate cannot be computed. However, a non-response bias analysis is conducted. Although information from non-respondents is not available, it has been argued that late respondents who participate only after an appreciable follow-up effort can be regarded as "almost non-respondents" and, hence, as closely similar to those who do not reply (Ferber 1948). Thus, answers to the dependent, independent and control variables from the first 10% of respondents were compared with the last 10% of respondents using T-tests (metric data) and χ^2-tests (nominal data). At the 1% level of significance, results suggest a sample selection bias (Heckman 1979) in favour of motivated individuals who are active participants[10] in their user group and who are in user groups with an open culture of communication. Such self-selection could result in a disproportionately high rate of active and "openly communicative" participants in the sample. Although a self-selection bias could undermine the external validity of the results (Ellis et al. 1970), this study had made several attempts to reduce this bias by sending out at least three follow-up emails in approximate monthly intervals to remind individuals to participate in the survey. Additionally, presidents were personally encouraged to participate at the CARE conference during CA World 2005, where approximately 200 user group presidents were present. Nevertheless, as these two variables are included as independent variables in the multivariate analysis, empirical results are unlikely to be greatly affected.

3.4.2 Variable Description

Before the determinants of successful CA user groups can be identified, it is necessary to firstly provide a brief description of the dependent, independent and control variables to be used in the multivariate analysis.

[8]http://causergroups.com/UserGroups/ (accessed on the 7th July 2005).
[9]See Appendix B for a copy of the questionnaire.
[10]Active participants are individuals who had participated in formal user group activities in the course of the last year.

Dependent Variable

Time Saved

Lakhani and von Hippel (2003) had found that having access to a user network and community documents can save individuals valuable time. Thus, the benefits users received from their user communities will be measured by the perceived number of hours of personal work time they were able to save over the course of the last year. If users were uncertain of the approximate number of hours they saved, they had the possibility of indicating a range of hours.

If users indicated the number of hours they saved as an interval, *time saved lower* refers to the lower bound and *time saved upper* the upper bound of their indicated interval.

The variable *time saved mean* refers to the geometric mean[11] of *time saved upper* and *time saved lower*. To take into account users who had indicated intervals with a zero lower bound, the following formula is used to calculate *time saved mean*,

$$Time\, saved\, mean = \sqrt{(time\, saved\, lower + 1)(time\, saved\, upper)}.$$

To ensure consistency, it is also necessary to perform a similar transformation for users who responded with a point estimate, *time saved point* with the following formula,

$$Time\, saved\, mean = \sqrt{(time\, saved\, point + 1)(time\, saved\, point)}.$$

Independent Variables

Strong Ties

The conceptualisation of this question closely follows Granovetter's (1973) definition of strong ties. The number of strong ties is the number of close contacts (excluding colleagues) users have in their CA user group. Close contacts are defined as individuals with whom they communicate frequently, whom they can approach at any time for help and whom they would willingly help at any time.

Time Spent

The number of days individuals spent participating in formal user group activities in the course of the last year will be used as a proxy for an individual's investment in his learning ability. Ten dummy variables are generated. The first dummy *no participation* takes the value one if individuals did not participate in any formal user group activities in the last

[11]The geometric mean is smaller than or equal to the arithmetic mean, as it gives more weight to the lower as compared to the upper bound. This measure is more appropriate than the arithmetic mean, because the amount of time users saved is positively skewed. See Figure 1 for an illustration of this skewed distribution.

year, and zero otherwise. Similarly, nine more time dummies for the following amounts of time spent are generated - less than 1 day, 1 day, 2 days, 3 to 4 days, 5 to 6 days, 7 to 10 days, 11 to 14 days, 15 to 20 days, and 21 days or more. This variable is not measured using an open-ended question because results from the pre-test suggested that the provision of categories would greatly increase the simplicity of answering the question.

Face-to-Face Communication
In the course of the last year, users could formally communicate with other members in the group through physical meetings; web, video and telephone meetings; as well as through the user group's forum. Two dummy variables are constructed. The first dummy *face-to-face* takes the value one if users interacted face-to-face with their user group in the last year and zero otherwise. The second dummy *no face-to-face* takes the value one if users only interacted with their user groups through web, video and telephone meetings, and/or through the user group's forum; and zero otherwise.[12]

Stars
Users were asked to indicate their estimate of the number of highly qualified individuals in their CA user group, from whom they had learnt a lot. Highly qualified individuals are defined as experienced and skilled users whom have a good knowledge of the CA software. To proxy for the number of stars in each user group, the average number of respondents' estimates per user group is used. The measurement of this variable was conceptualised by the author using interview results with CA user group members.

Open Communication
Users were asked to indicate the extent to which they agreed or disagreed with the following statement: "This CA user group has been extremely successful in promoting an open exchange of information between its members.". An open exchange of information is defined as the willingness of members to share and exchange information with one another. The *open communication* variable takes the value one when at least 50% of the members in each user group who responded to the survey indicated that they either strongly agreed or somewhat agreed with the statement, and zero otherwise. The formulation of the item for this variable was inspired by Shah (2005b).

Exclusivity
Users who are on the management committee of their user group were asked to indicate how individuals can become members in their user groups. Two dummy variables are generated - one to control for the openness of a group's membership and the other for the existence of a membership fee. The first dummy *open membership* takes the value

[12] These two dummy variables need not be perfectly multicollinear with each other, as there may exist users who did not participate in any formal user group activities. See the description of *time spent* for more details.

one if membership is open to the general public, and zero otherwise[13]. The other dummy *membership fee* takes the value one if members are required to pay a membership fee, and zero otherwise. If multiple differing answers were provided for each group, the answer of the group leader (i.e. president, co-president and/or vice-president) is used. The measurement of the above-mentioned variables was motivated from interview results with CA user group presidents and CA employees.

Control Variables

Female

To control for gender effects, this variable takes the value one for females and zero otherwise.

Learn

Users who joined the user group with the intention to learn may have different expectations from the user community as compared to users who joined for other reasons. The measurement of this variable was adapted from empirical open source software literature (Hars and Ou 2002; Ghosh et al. 2002), which propose that one important reason for individuals to participate in open source software projects is to learn and to increase their programming skills. As not all user group members write software code, it was necessary to rephrase the item to suit the research context of this paper. Thus, users were asked to indicate the extent to which they agreed or disagreed with the following statement on why they joined their user group: "For the opportunity to learn from experienced users". *Learn* takes the value one if users indicated that they either strongly agreed or somewhat agreed with the statement, and zero otherwise.

Used code

To control for individuals who made use of source code from other members in their user community this variable takes the value one if respondents indicated that they used source code of other members in their user groups, and zero otherwise.

Reciprocity

Reciprocity is a social norm that may affect the benefits users are able to receive from their user communities. To control for this effect, *reciprocity* takes the value one if respondents indicated that they help other members because they received help from their CA user group in the past, and/or because they anticipate needing help from CA user group members in the future, and zero otherwise. The items for this variable were adapted from existing empirical open source software literature (Hars and Ou 2002; Ghosh et al. 2002; Lakhani and Wolf 2005).

[13] A more restrictive form of membership is to allow *only* individuals of firms who have a license of the CA product to be members.

Committee
User group presidents and other committee members invest additional time managing and organising activities for their user groups. To control for this extra time, which is not accounted for in the *time spent* variable, the *committee* variable takes the value one if the user is a committee member of his user group and zero otherwise.

Contribute
This variable was conceptualised using interview results with CA user group members. Individuals can contribute to the user group by sharing information on their use experience with the CA software. They can do this by making a presentation, by participating in discussion-sessions or by making posts on the user group's forum. Users were asked to indicate the extent to which they agreed or disagreed with the following statement: "I reveal information on how my company utilises the CA software because I benefit from the resulting discussions.". This variable takes the value one if respondents indicated that they either strongly agreed or somewhat agreed with the statement, and zero otherwise.

Use Experience
To control for the level of experience of users, users were asked to indicate the year in which they started using the most important CA software that they most frequently used at their present employment. The *use experience* variable is constructed by subtracting their indicated year from 2006.

3.5 Empirical Analysis

3.5.1 Descriptive Statistics

Table 3.1 presents the descriptive statistics of the above mentioned dependent, independent and control variables. The final sample consists of 924 responses from individuals in 161 different user groups. The number of responses per user group ranges from 1 to 40, with at least 87.12% of the respondents belonging to user groups from which at least 4 responses were received. It can be seen that users are able to save an average of 22.29 number of hours per year.[14] The minimum number of hours they saved is 0 and the maximum[15] is 600. 26.62% of the respondents did not save any time in the last year. An observation of the first to fourth percentiles of *time saved mean* suggests that the variable has a positively skewed distribution. Figure 3.1 illustrates this skewed distribution.[16]

CA user groups also appear to provide a platform for users to network and to make valuable social contacts within their user communities. Specifically, users have on average

[14]Approximately 22% of the respondents indicated the amount of hours they saved as an interval.
[15]The maximum reported here is the *time saved point* estimate of 600 hours.
[16]The number of hours presented here refer to the *time saved mean* variable as described in section 3.4.2.

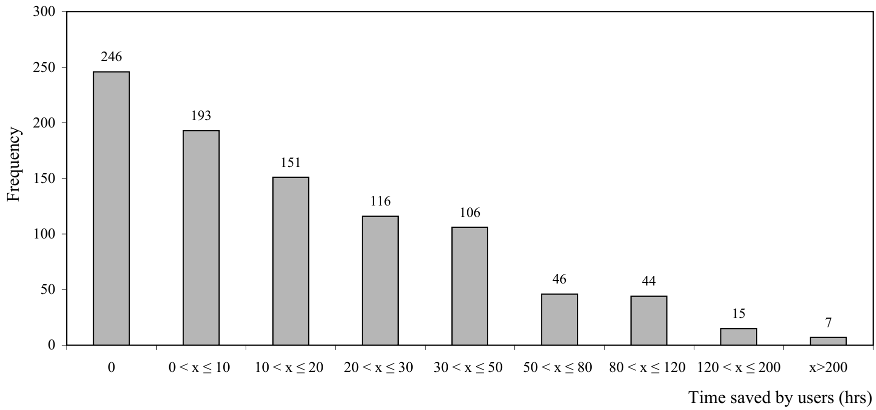
Table 3.1: Descriptive Statistics I - Across Individuals (N = 924)

Variable	Mean	S.D.	Min.	Max.
time saved mean	22.29	38.53	0	600.5
time saved lower*	20.49	37.65	0	600
time saved upper*	25.31	51.60	0	1,000
strong ties	3.49	8.81	0	200
time spent				
no participation (0/1)	0.22		0	1
less than 1 day (0/1)	0.06		0	1
1 day (0/1)	0.11		0	1
2 days (0/1)	0.21		0	1
3 - 4 days (0/1)	0.21		0	1
5 - 6 days (0/1)	0.08		0	1
7 - 10 days (0/1)	0.05		0	1
11 - 14 days (0/1)	0.02		0	1
15 - 20 days (0/1)	0.01		0	1
21 days or more (0/1)	0.02		0	1
face-to-face (0/1)	0.73		0	1
no face-to-face (0/1)	0.05		0	1
open communication (0/1)	0.87		0	1
stars	6.78	4.07	0	25
open membership (0/1)	0.37		0	1
membership fee (0/1)	0.17		0	1
female (0/1)	0.22		0	1
learn (0/1)	0.93		0	1
used code (0/1)	0.21		0	1
reciprocity (0/1)	0.64		0	1
committee (0/1)	0.28		0	1
contribute (0/1)	0.74		0	1
use experience	10.57	6.40	1	34

* For point estimates, time saved lower has the same value as time saved upper.

3.49 number of strong ties in their user groups. This number ranges between 0 to 200. Approximately 40% of the respondents do not have any close contacts in their user groups. Similar to the *time saved mean* variable, the first to fourth percentiles of the number of strong ties also indicate a positively skewed distribution. Of the 924 respondents, 22% did not attend any formal user group activities in the last year. 67% of the respondents (excluding those who did not participate) spent less than a week with their user communities, with the majority having spent 2 days (21%) and 3 to 4 days (21%). Only 10% had invested 7 days or more in their user groups, whereby 2% indicated that they spent 3 weeks or more of their time annually participating in user group activities. Face-to-face interaction seems to be the most popular means of communication between users and

Figure 3.1: Number of hours saved per year (N = 924)



their user groups as 73% of respondents[17] had attended physical user group meetings. This finding is not surprising as CA requires its user groups to meet at least once a year for a full-day meeting. 5% of the respondents participated *only* in the user group's forum and/or in web, video and telephone meetings in the last year.

The next four variables are user group-specific characteristics. 87% of the respondents belong to CA user groups that have an open culture of communication. As displayed in Table 3.2, 84% of the 161 user groups have members who are willing to share and exchange information with one another. Table 3.2 also shows that the average number of star members in each user group is 6.38 with values ranging from 0 to 25. Similarly, the mean of the number of stars of users' user groups (from Table 3.1) is 6.78. 37% of the respondents are in user groups that have an open membership structure that allows anyone interested (i.e. from the general public) to become a member. From Table 3.2, it can be seen that 35% of the 161 user groups have such an open membership structure. Finally, only 13% of the user groups require their members to pay a membership fee; 17% of the respondents are in such user groups.

22% of the respondents are female. Consistent with the definition of user communities as introduced in section 3.3, 93% of the respondents indicated that they joined their user groups with the aim to learn. Users were also asked if they had utilised source code written by other users in their user group. 21% indicated that they make use of add-ons, utilities,

[17] 93% of all participants who participated in formal user group activities (N=719) interacted with their user groups face-to-face.

Table 3.2: Descriptive Statistics II - Across User Groups (N = 161)

Variable	Mean	S.D.	Min.	Max.
open communication (0/1)	0.84		0	1
stars	6.38	4.38	0	25
open membership (0/1)	0.35		0	1
membership fee (0/1)	0.13		0	1

applications and/or exits from their user community. The majority of the respondents seem to exhibit the social norm reciprocity, as 64% of the respondents indicated that they were willing to help other users because they had either received assistance from their user group in the past or expect help from their user group in the future. 28% are committee members in their user groups. This finding is not surprising as these individuals are more committed and interested in participating in a survey regarding their user group. Another reason for this relatively high percentage could also be because committee members were contacted directly to participate in the survey, whereas regular user group members were informed of the survey indirectly through their committee members. As 74% of the respondents indicated that they contribute to their user communities, the free-rider problem (introduced in section 3.2) does not seem to be a substantial problem in majority of the CA user groups surveyed. Finally, the average number of years that users have had experience with their CA software is 10.57 and there appears to be a relatively diverse level of use experience amongst the users as can be seen from the first percentile (6 years), the second percentile (9 years), the third percentile (15 years) and the 95% percentile (23 years). This result suggests that CA user groups provide a platform for newer users to learn from more experienced individuals. A closer inspection of this variable at the user group level reveals that the majority of the user groups have users with a relatively heterogeneous level of use experience.

Tables 3.3 and 3.4 present an overview of bivariate relationships between the independent and control variables.[18] It can be seen from Table 3.3 that with the exception of three correlation coefficients, all the rest of the coefficients are below 0.25. These stronger interrelationships will be elaborated in the following. Firstly, committee members appear to have a larger number of strong ties to other members in their user groups ($\rho = 0.28$). This relationship is reasonably conceivable as committee members play a more central role in the user group.[19] They are more accessible to the full member contact list and have more interactions with individual members when organising user group activities.

[18] The variables *no participation* and *face-to-face* are omitted in the bivariate analyses as these variables will not be included in the multivariate analysis in section 3.5.2.

[19] See Wasserman and Faust (1994) for a more detailed description of centrality in social networks.

Table 3.3: Correlation Matrix of the Independent and Control Variables I (N = 924)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) strong ties	1.00											
(2) no face-to-face (0/1)	−0.04♢	1.00										
(3) open communication (0/1)	0.05♢	0.02	1.00									
(4) stars	0.22♠	0.07♢	0.26♢	1.00								
(5) open membership (0/1)	−0.04♢	0.07	0.05	0.03	1.00							
(6) fee (0/1)	0.01♢	−0.03	−0.11	0.04	0.12	1.00						
(7) female (0/1)	0.00♢	−0.01	0.13	−0.02	0.04	−0.10	1.00					
(8) learn (0/1)	0.00♢	−0.03	0.10	0.08	−0.06	−0.01	0.08	1.00				
(9) used code (0/1)	0.13♢	−0.01	0.04	0.09	0.01	0.00	−0.03	0.08	1.00			
(10) reciprocity (0/1)	0.17♢	0.02	0.16	0.09	0.06	−0.08	0.05	0.14	0.22	1.00		
(11) committee (0/1)	0.28♢	−0.10	−0.01	0.01	0.02	0.00	0.05	0.07	0.08	0.19	1.00	
(12) contribute (0/1)	0.12♢	−0.03	0.08	−0.02	0.00	−0.04	0.08	0.15	0.14	0.39	0.20	1.00
(13) use experience	0.19♠	−0.00♢	−0.04♢	0.13♠	0.05♢	0.22♢	−0.08♢	−0.08♢	0.05♢	0.10♢	0.16♢	0.04♢

Note: Phi-coefficient values unless otherwise stated; ♢ Point biserial correlation coefficient,
♠ Pearson product-moment correlation coefficient

Table 3.4: Correlation Matrix of the Independent and Control Variables II (N = 924)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) less than 1 day (0/1)	1.00								
(2) 1 day (0/1)	−0.09	1.00							
(3) 2 days (0/1)	−0.13	−0.19	1.00						
(4) 3 - 4 days (0/1)	−0.13	−0.19	−0.27	1.00					
(5) 5 - 6 days (0/1)	−0.08	−0.11	−0.16	−0.16	1.00				
(6) 7 - 10 days (0/1)	−0.06	−0.08	−0.12	−0.12	−0.07	1.00			
(7) 11 - 14 days (0/1)	−0.03	−0.05	−0.07	−0.07	−0.04	−0.03	1.00		
(8) 15 - 20 days (0/1)	−0.02	−0.03	−0.05	−0.05	−0.03	−0.02	−0.01	1.00	
(9) 21 days or more (0/1)	−0.04	−0.06	−0.08	−0.08	−0.05	−0.03	−0.02	−0.01	1.00
(10) strong ties	−0.06◊	−0.05◊	−0.03◊	0.00◊	0.11◊	0.20◊	0.05◊	0.00◊	0.10◊
(11) no face-to-face (0/1)	0.27	0.04	0.00	−0.10	−0.05	−0.05	−0.03	0.04	0.25
(12) open communication (0/1)	0.04	−0.14	0.03	0.12	0.09	0.09	0.00	0.00	0.06
(13) stars	−0.01◊	−0.01◊	−0.08◊	−0.04◊	0.08◊	0.06◊	0.07◊	0.02◊	0.20◊
(14) open membership (0/1)	0.01	−0.05	−0.02	0.00	−0.04	−0.01	0.02	0.04	0.06
(15) fee (0/1)	−0.05	−0.02	0.00	−0.03	0.04	−0.01	−0.02	−0.04	0.01
(16) female (0/1)	0.06	−0.08	0.01	0.04	0.03	0.09	−0.01	−0.02	0.00
(17) learn (0/1)	−0.05	−0.04	0.00	0.03	0.04	0.04	−0.03	0.02	0.02
(18) used code (0/1)	−0.11	−0.08	0.01	0.00	0.07	0.09	0.07	0.02	0.13
(10) reciprocity (0/1)	−0.04	−0.03	−0.01	0.08	0.08	0.09	0.01	0.08	0.09
(20) committee (0/1)	−0.12	−0.03	0.01	0.09	0.10	0.19	0.06	0.06	0.06
(21) contribute (0/1)	−0.09	−0.03	0.02	0.14	0.10	0.05	0.02	0.05	0.04
(22) use experience	−0.04◊	0.03◊	−0.02◊	−0.04◊	0.09◊	0.05◊	−0.02◊	−0.02◊	0.01◊

Note: Phi-coefficient values unless otherwise stated; ◊ Point biserial correlation coefficient

Secondly, it is not surprising that respondents who exhibit the social norm reciprocity are also those who contribute to the user group (ρ = 0.39), as individuals who have benefitted from the user group in the past or expect to benefit in the future would be more willing to contribute to their user group by giving oral presentations, assisting members in solving their difficulties, participating in discussion-sessions or in answering posts on the user group's forum. Thirdly, the number of star members and the openness of the communication culture of each users' user group are positively related (ρ = 0.26). To interpret this positive relationship, it is also necessary to see how these 2 variables are correlated at the user group level. The Phi-coefficient at the user group level (ρ = 0.23) is fairly similar to that at the individual level. These findings suggest that the participation of more star members may help promote an open culture of communication; and/or that as star members may be interested in interacting with other users, they are motivated to join user groups whose members are more open to share and exchange information with one another.

From Table 3.4, it can be observed that two correlation coefficients are greater than 0.25 in magnitude. Firstly, it appears that there is a negative correlation between individuals who participated in their user groups for 2 days and those who participated for 3 to 4 days (ρ = -0.27). As these two variables have dichotomous values that represent nominal data and are mutually exclusive from each other, this negative correlation simply reflects the approximately equal numbers of individuals who spent 2 days (N=195) and those who spent between 3 to 4 days (N=197) in formal user group activities. Finally, there appears to be a positive relationship between individuals who had spent less than one day with their user groups and those who did not participate in face-to-face activities (ρ = 0.27). This finding is relatively plausible as face-to-face meetings are generally organised for at least one day.[20]

3.5.2 Multivariate Analysis

To empirically test the hypotheses introduced in section 3.3, this paper will apply a pooled interval estimation method with robust standard errors in the presence of intra-cluster correlation. First, a short overview of the estimation method is presented. This is followed by the empirical analysis of the estimated models.

3.5.2.1 Model Specification

As the analysis uses data from individuals in different user groups, there is a high probability that responses from individuals within the same user group may be correlated, possibly through unobserved cluster effects (Wooldridge 2002). This intra-cluster correlation is of-

[20]This information is obtained from interviews with CA user group presidents.

ten referred to as over-dispersion or extra variation in an estimated statistic beyond what would be expected under independence. Hence, analyses that assume independence of the observations will tend to underestimate the true variance, which would lead to test statistics with inflated Type I errors (Williams 2000). To avoid this over-dispersion problem in ordinary least squares (OLS) regressions, unbiased variance estimators can be computed (Rogers 1993).

Let the model be

$$y = \boldsymbol{x}\beta + \varepsilon, \tag{3.1}$$

where y represents a $N \times 1$ vector of *time saved*, $\boldsymbol{x}$ represents the $N \times K$ matrix of independent and control variables and β represents a $K \times 1$ vector of coefficients.[21] The model assumes $\varepsilon \sim N(0, \sigma^2 I)$. To estimate β, the OLS method minimises the sum of squared errors ($\varepsilon'\varepsilon$), where

$$\varepsilon'\varepsilon = (y - \boldsymbol{x}\beta)'(y - \boldsymbol{x}\beta). \tag{3.2}$$

The OLS estimate of β is

$$b = (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'y, \tag{3.3}$$

and its variance is

$$var(b) = E(\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'(y - Ey)(y - Ey)'\boldsymbol{x}(\boldsymbol{x}'\boldsymbol{x})^{-1}. \tag{3.4}$$

Under the assumption that the observations are independently and identically distributed, this variance reduces to $(\boldsymbol{x}'\boldsymbol{x})^{-1}\sigma^2$. In the presence of intra-cluster correlation, this assumption is relaxed, which allows for non-identical diagonal terms as well as nonzero off-diagonal terms. However, as observations from different clusters are assumed to be independent, these off-diagonal terms can only take on nonzero values if observations come from the same cluster. These nonzero terms are then represented by the appropriate products of the residuals (Rogers 1993).

Approximately 22% of the respondents indicated the number of hours they saved in the last year as an interval. One possibility to analyse such interval data is to estimate a pooled OLS regression on the arithmetic or geometric mean of these intervals. The drawback of this OLS method is that the analysis would not reflect any uncertainty concerning the nature of the exact values within each of these intervals[22]. Hence analysing the data with

[21] N represents the total number of observations and K the total number of independent and control variables plus one for the constant term.

[22] www.ats.ucla.edu/stat/stata/dae/intreg.htm (accessed on 4th April 2006)

an interval regression - a slightly modified version of the standard censored Tobit model (Tobin 1956) - would provide more accuracy for the estimation results. An advantage of the interval regression over the OLS estimation method is that it is able to estimate the β coefficients even if the dependent variable y is non-observable, provided the interval in which y falls is observable (Wooldridge 2002).

Assuming the same model as in equation (3.1), y is now defined as continuous outcomes which are either observed or unobserved. For observations $j \in \zeta$, y_j is observed as point outcomes. Observations $j \in \iota$ are left-censored outcomes; that is it is known that y_j is less than or equal to $y_{\iota j}$. Similarly, observations $j \in \omega$ are right-censored. Observations $j \in \theta$ are intervals, such that y_j is in the interval $[y_{1j}, y_{2j}]$. The interval regression is estimated using the maximum likelihood method. The log likelihood to be maximised is

$$
\begin{aligned}
ln\, L = & -\frac{1}{2}\sum_{j\in\zeta}\{(\frac{y_j - \boldsymbol{x}\beta}{\sigma})^2 + log\, 2\pi\sigma^2\} && (3.5)\\
& + \sum_{j\in\iota} log\,\{\Phi(\frac{y_{\iota j} - \boldsymbol{x}\beta}{\sigma})\} && (3.6)\\
& + \sum_{j\in\omega} log\,\{1 - \Phi(\frac{y_{\omega j} - \boldsymbol{x}\beta}{\sigma})\} && (3.7)\\
& + \sum_{j\in\theta} log\,\{\Phi(\frac{y_{2j} - \boldsymbol{x}\beta}{\sigma}) - \Phi(\frac{y_{1j} - \boldsymbol{x}\beta}{\sigma})\},
\end{aligned}
$$

where $\Phi()$ is the standard cumulative normal.[23] Similar to the OLS regression method, unbiased variance estimates can also be computed for the interval regression to control for intra-cluster correlation.

3.5.2.2 Empirical Results

Table 3.5 presents the results of the above-mentioned estimation method. The logarithmical forms of the number of hours saved, the number of strong ties and the number of star members will be used in the analysis. To accommodate for 0 values, the logarithm of 1 plus the values of each variable is used. As time saved has a natural lower bound of 0, 263 of the 924 observations in Model 1 are left-censored observations representing individuals who did not save any time in the past year; 473 are uncensored observations; and 188 are interval observations. There are no right-censored observations as time saved does not have an upper bound.

[23]STATA 9 Reference A-J, 2005. Note that the right- and left-censored outcomes have been omitted for simplicity.

Table 3.5: Interval Regression Models with Heteroscedasticity-Robust Standard Errors in the Presence of Intra-Cluster Correlation

	Model 1	**Model 2**
	ln (time saved lower + 1) ln (time saved upper + 1)	ln (time saved lower + 1) ln (time saved upper + 1)
ln (strongties + 1)	0.356***	0.398***
	[0.081]	[0.087]
less than 1 day (0/1)	−0.099	−0.092
	[0.343]	[0.351]
1 day (0/1)	0.740***	0.719***
	[0.233]	[0.249]
2 days (0/1)	0.820***	0.835***
	[0.199]	[0.207]
3 - 4 days (0/1)	1.058***	0.984***
	[0.188]	[0.194]
5 - 6 days (0/1)	0.815***	0.781***
	[0.242]	[0.252]
7 - 10 days (0/1)	1.260***	1.150***
	[0.309]	[0.340]
11 - 14 days (0/1)	1.468***	1.380***
	[0.479]	[0.518]
15 - 20 days (0/1)	1.942***	1.907***
	[0.321]	[0.400]
21 days or more (0/1)	1.146**	0.645
	[0.514]	[0.402]
no face-to-face (0/1)	0.119	0.156
	[0.237]	[0.243]
open communication (0/1)	0.524***	0.400**
	[0.166]	[0.173]
ln (stars + 1)	0.314**	0.421***
	[0.142]	[0.154]
open membership (0/1)	0.215**	0.274**
	[0.107]	[0.114]
Observations	924	805
Min. obs per group	1	4
No. of user groups	161	92
Log pseudolikelihood	−1625.826	−1397.425

Continued on Next Page…

Table 3.5: Interval Regression Models with Heteroscedasticity-Robust Standard Errors in the Presence of Intra-Cluster Correlation (continued)

	Model 1	**Model 2**
fee (0/1)	-0.325^{**}	-0.293^{**}
	[0.132]	[0.132]
female (0/1)	-0.486^{***}	-0.539^{***}
	[0.130]	[0.142]
learn (0/1)	1.148^{***}	1.012^{***}
	[0.310]	[0.329]
used code (0/1)	0.990^{***}	1.028^{***}
	[0.117]	[0.129]
reciprocity (0/1)	0.592^{***}	0.598^{***}
	[0.153]	[0.169]
committee (0/1)	-0.257^{**}	-0.325^{**}
	[0.129]	[0.142]
contribute (0/1)	0.405^{***}	0.446^{**}
	[0.170]	[0.181]
ln (use experience)	−0.004	0.024
	[0.090]	[0.097]
constant	-2.067^{***}	-2.185^{***}
Observations	924	805
Min. obs per group	1	4
No. of user groups	161	92
No. of left-censored obs.	263	239
No. of uncensored obs.	473	408
No. of interval obs.	188	158
Log pseudolikelihood	−1625.826	−1397.425

Robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

To test for the presence of heteroscedasticity in Model 1, the σ term was expressed in the log-likelihood function of equation 3.5 as a set of all 22 independent and control variables. A likelihood-ratio test is then conducted with this unrestricted model and the restricted model where σ is kept constant (i.e. homoscedasticity). The likelihood-ratio test statistic is 130.39 with 22 degrees of freedom and a p-value of zero. This result indicates that the null hypothesis of homoscedasticity can be rejected at the 1% level of significance. Hence, heteroscedasticity-robust standard errors in the presence of intra-cluster correlation are estimated to control for both heteroscedasticity and cluster effects.

As can be seen from Model 1 in Table 3.5, the number of strong ties users have in their user groups is positively significant at the 1% level.[24] In particular, when a user's number of strong ties doubles, the number of hours he can save increases by approximately 35.6%. Hence, the results support hypothesis 1 that users benefit more from a greater number of strong relational ties as this increases the number of channels from which they can receive relevant information and knowledge.[25]

As a proxy for an individual's investment in his learning abilities, nine dummy variables on the amount of time the user invests in participating in user group activities are included. The reference group are users who did not participate in any formal user group activities in the last year. As can be seen, results show that the more time individuals invest in their user groups, the more they benefit.[26] For instance, individuals who spend 1 day (2 days) per annum attending user group activities receive 74% (82%) more benefits from their user groups as compared to individuals who do not attend any activities per year. With the exception of individuals who spend between 5 to 6 days, those that invest 3 or more days participating in user group activities are able to save more than twice as much time as compared to individuals who did not spend any time. Except for the two extreme categories - those that spend 1 day (negative and insignificant) and those that spend 3 weeks or more (positive and significant at the 5% level) - the coefficients of all the time spent variables are positively significant at the 1% level. In tendency[27], it can be seen from Table 3.5 that larger coefficients are observed for the *time spent* dummy variables that reflect more time spent in user group activities.

To test if an individual's benefits increases when he invests more time to improve his learning abilities (hypothesis 2), it is necessary to conduct a joint-test with the null hypotheses that the coefficients of all the time spent variables are equal to zero. The chi-

[24] Models 1 and 2 are also estimated with the pooled OLS method, with *time saved mean* as the dependent variable. Coefficient estimates and their significance levels were very similar to those from the interval estimation method.

[25] The author acknowledges the argumentation that an individual who wants to benefit more from his user community could actively forge and maintain a larger number of strong ties, thus introducing a potential endogeneity problem. However, as individuals have limited time and emotional resources to invest in social relationships, it is not expected that they would be able to sustain a large number of strong ties at will as this would require a high investment in time and energy (Granovetter 1973). Thus, the author argues that the number of weak ties (and not strong ties) is potentially endogenous to the benefits an individual obtains from his user community, as the maintenance of weak ties does not require much personal investment (Granovetter 1973; Seibert et al. 2001).

[26] There could exist an endogeneity problem here as individuals who want to benefit more from their user community would logically participate more in user group activities. However, it is argued that the amount of time an individual spends attending CA user group activities is largely determined by his employer (who grants him time off regular work to participate), and thus exogenous to the amount of personal benefits he actually obtains.

[27] It should be noted that the coefficient for the 5-6 days dummy is smaller than that of the 3-4 days dummy. However, the coefficient of the 7-10 days dummy is larger than that of the 3-4 days dummy. Furthermore, the coefficient of the 21 days or more dummy is smaller than the coefficients of the 7-10 days, 11-14 days and 15-20 days dummies.

squared statistic of this joint-test is 66.73 with 9 degrees of freedom, and a p-value of 0. This means that the joint null hypotheses can be rejected at the 1% level of significance. It can hence be concluded that users who spend more time with their user groups, i.e., invest more in their learning abilities, benefit more from their user communities. As a result, hypothesis 2 can also be supported by the data.

The coefficient of the *no face-to-face* variable is not significant. This indicates that individuals who only interact with their user groups either through their user group's forum and/or through web, video and telephone meetings do not benefit significantly less from their user groups as compared to individuals who participate in physical meetings[28]. It should however be noted that as only 6.54% of users who did engage in user group activities (i.e. only 47 respondents) did not have face-to-face interaction with other members in their user communities, the power of the above-mentioned joint-test is relatively small. Nevertheless, hypothesis 3 cannot be supported by the data.

Users in user groups with an open culture of communication are able to save 52.4% (significant at the 1% level) more hours as compared to users in user groups whose members are hesitant to share and exchange information with one another. When the number of star members in a user group doubles, users are able to increase the number of hours they save by approximately 31.4%. This group-specific determinant is significant at the 5% level. Hence, the data provides evidence that hypothesis 4 and 5 can be supported.

Individuals who are in user groups with an open membership save 21.5% (significant at the 5% level) more hours per year as compared to individuals who are in user groups whose memberships are not open to the public. Hence, hypothesis 6a cannot be confirmed by the data. Next, users who have to pay a membership fee save 32.5% (significant at the 5% level) less hours as compared to users in user groups that do not charge membership fees. Thus, hypothesis 6b is also not supported by the data. A joint-hypothesis test of the significance of both these variables further concludes that they are jointly significant. The chi-squared statistic of this joint-test is 9.25 with 2 degrees of freedom, and a p-value of 0.0098. These results provide evidence that at the 1% level of significance, individuals in user groups with less restrictive membership requirements are able to benefit more from their user groups. Hypothesis 6 is thus not supported by the data.

In addition to the dependent variables, 7 additional control variables are also included in the estimation. Results show that females save significantly less hours as compared to males. Individuals who had joined their user groups with the intention to learn are able to save more hours than individuals who had joined their communities for other

[28] 78.7% of respondents who participated in physical meetings also participated in their user group's forum and/or in web, video and telephone meetings.

reasons. In comparison to users who do not utilise source code from other members in their community, users who do are able to save significantly more hours. Users who exhibit reciprocal behaviour are also able to save more hours than those who do not have this reciprocity characteristic. Committee members appear to save less time as compared to regular members. One reason for this finding could be that committee members may be more intrinsically than extrinsically motivated to participate in user group activities. Additionally, users who contribute to their user groups are also able to benefit more from them. Finally, more experienced users do not significantly benefit more (or less) from their user communities. A joint-hypothesis test is conducted to test for the joint-significance of all the control variables. The chi-squared statistic of this test is 185.98 with 7 degrees of freedom, and a p-value of 0. Hence, all 7 control variables are jointly significant.

To check the reliability and robustness of the results from Model 1, the sample is reduced to include only user groups from which at least 4 responses were received. This reduction in the sample also seeks to avoid potential endogeneity problems that may arise, as group specific variables such as the number of stars in each group and the openness of the group will now be based on at least 4 members of each group. Model 2 provides the results of the pooled interval regressions with heteroscedasticity-robust standard errors in the presence of intra-cluster correlation using only responses from individuals in user groups, whereby at least 4 members from each CA user group responded to the survey. These results are based on 805 responses from individuals in 92 different user groups. It can be seen that the results are relatively similar to those of Model 1. Hypothesis 1 is still significantly supported by the data at the 1% level. A joint-hypothesis test[29] provides evidence to support hypothesis 2 at the 1% level. There is still no evidence from the data that hypothesis 3 is supported. Hypothesis 4 can be supported at the 5% level and hypothesis 5 at the 1% level. Lastly, hypothesis 6a, 6b and 6 are again not supported by the data as the coefficients have a significant opposite sign as compared to that predicted.[30]

3.6 Conclusion

By comparing the amount of time individuals in different user groups were able to save over the course of the last year, the aim of this study is to identify the determinants of successful user communities. Empirical findings demonstrate that the following determinants are relevant individual-specific factors: the number of strong intra-group contacts a user has and the amount of time users invest in their user groups. Furthermore, results also show that user communities with an open communication culture, a larger number of star members and less restrictive membership requirements are more successful.

[29]The chi-squared statistic of this joint-test is 53.73 with 9 degrees of freedom, and a p-value of 0.
[30]The chi-squared statistic of this joint-test is 10.14 with 2 degrees of freedom, and a p-value of 0.

As user communities are voluntary in nature, rational individuals will not participate in and contribute to these groups unless they perceive a net benefit. Hence, the successful organisation of these communities is essential for members to be able to constantly leverage their user network for their own benefit. This paper proposes the following four implications for the management of user communities.

Social Bonding Activities

An essential characteristic of a successful user community is the existence of a social network. It has been shown that users who have a greater number of close contacts (i.e. strong ties) within their communities and who exhibit reciprocal behaviour benefit more from their user groups. User group meetings are a very good means of stimulating social interactions between members. For instance, members can get to know each other better through social events or functions that may take place before, between or after formal group meetings. Alternatively, more user-to-user contact can be stimulated through longer or more frequent intermission breaks between presentations at physical user group meetings. Such social bonding activities may not only promote social norms such as reciprocity and trust but can also facilitate the establishment of close contacts between group members.

Quality and Frequency of Informational Exchanges

User group meeting programs should consist of topics and issues that a majority of the members are interested in and that are of a high quality as this would motivate more members to participate. Moreover, an active user group forum that has a regular exchange of knowledge between members will also motivate users to invest time to read and make posts on their group's forum (Lazar and Preece 2002). This is important because the empirical results indicate that users who spend more time engaging in group activities are able to benefit more from their groups.

Incentive Scheme for Stars

Results provide evidence that users in user groups with a larger number of star members are able to benefit more from their communities. This suggests that experienced stars contribute valuable input into the knowledge pool of user communities from which other members are able to benefit from. Hence, one could provide incentives to attract and motivate highly experienced individuals (stars) to join and actively participate in user communities. There are various ways in which such a scheme can be implemented, either user- or firm-initiated. For instance, star members (as voted by the group) may receive a complementary voucher for a meal or a drink, sponsored by the rest of the members. Firms can also introduce incentives for star members, such as to provide more customised support or a reduced fee of the product to companies of star members. Alternatively, firms can recognise a star's participation in the community by inviting the star to be part of an executive user committee that is involved in the decision making process of future development strategies for the product[31].

Open Membership and Communication
Results suggest that user communities should be open to the general public and have less restrictive membership structures. Individuals in user groups that do not impose a membership fee and/or allow interested individuals from the public to become members, benefited more from their communities. A less restrictive membership structure may increase the potential knowledge pool available by allowing a wider diversity of individuals to participate and contribute to the community. Additionally, the empirical results also suggest that there should be a free flow of knowledge and information within user communities. Hence, users should be encouraged to share their ideas, insights and experiences with each other.

Although these implications have been based on user groups associated to a proprietary software company, there exists sufficient group heterogeneity (due to the diversity of products CA provides and the large number of acquisitions CA made in the past) to be able to generalise these results for similar forms of user communities where users voluntarily group together to learn from each other. However, more industry-specific information will be needed before these implications can be applied to industries outside the software industry.

A further drawback of this study is that users are mainly extrinsically motivated to participate in their user communities. The extrinsic measure made use in this study was that of the amount time individuals were able to save from utilising information they had obtained through their user communities. This measure ignores other forms of benefits such as intrinsic incentives or social benefits. However, it should be noted that in the case of user communities in the management (proprietary) software industry, whereby individuals consider participation in these user communities as part of their job, it is expected that extrinsic incentives are the main driving force for their participation.

Nevertheless, the above-mentioned implications for the management of user communities provides a first insight into how groups of users can be organised. This knowledge is of particular importance for firms that intend to leverage user communities for innovation management. A well managed user community can be used to efficiently organise and prioritise the pool of user information regarding their needs, feedback to existing products as well as ideas for new products. Furthermore, an open exchange of information and

[31] An example of such a committee is CA's Product Advisory Councils, which consist of product-specific groups of experienced, prominent IT and business professionals that help guide CA in its future offerings.

knowledge within these user communities can help enhance the creativity and innovative potential of individual users, which may result in user innovations (Shah 2005b). Under certain conditions, firms may even be able to harness these user innovations and integrate them into their own proprietary solutions (Mayrhofer 2005).

A symbiotic relationship between firms and their user communities can provide conducive conditions for firms to constantly leverage their communities for innovation management (Dahlander and Magnusson 2005). This study has focused its analysis on the benefits users can obtain from their user communities. Future research should be directed towards investigating the integration of user communities into firms' research and development teams and examining the benefits firms are able to obtain from this integration. More research is still necessary before a framework for the optimal integration of user communities into firms' product development processes can be developed. One possibility could be to use user communities to identify leading-edge users, who have been shown to be essential sources of innovation (Urban and von Hippel 1988). In particular, user communities may be able to help reduce the search costs involved in finding these leading-edge users as well as in reducing the uncertainty of their dedication and commitment to cooperate with the firm.

Further research could also examine potential endogeneity issues between the dependent and independent variables. For instance, individuals who want to benefit more from their communities may be motivated to forge and maintain more strong ties. Although existing literature argues that the maintenance of a large number of strong ties is not feasible due to time constraints (Granovetter 1973; Seibert et al. 2001), there does not exist any empirical proof of this argument. An additional endogenous variable could be the amount of time individuals spend participating in user community activities, as individuals who want to benefit more may logically invest more of their time. Although it is argued in this paper that the amount of time members spend in CA user groups is largely determined by their employers, this may not be the case for user communities in other industries such as that of consumer products. Thus, future work should suggest instrumental variables for these potentially endogenous variables and apply appropriate estimation methods.

Chapter 4

Selecting User Community Collaborators for Innovation Management

4.1 Introduction

An organisation can be viewed as a bounded, open system effecting transactions with its external environment (Adams 1976). As innovation processes are information processes in which knowledge is acquired, processed, and transferred (Hauschildt 1992), a critical factor for successful performance is the ability of the innovating organisation to effectively gather information from and transmit information to several external domains (Keller and Holland 1975; Tushman 1977). By focussing on the external customer domain, this paper analyses how firms can gain access to a large pool of customer knowledge by integrating customers from user communities (user community collaborators) into the innovation process.

Studies on the integration of customers in the innovation process have come to the general consensus that cooperation with customers can be extremely beneficial. These benefits can take several forms, such as in enhancing the innovation process by contributing to idea generation and evaluation, concept development and design, development, and prototype testing (von Hippel 1986; Souder et al. 1998; Salomo et al. 2003; Veryzer and Borja de Mozota 2003); increasing new product success (Gruner and Homburg 2000; Lilien et al. 2002; Hsieh and Chen 2005; Ogawa and Piller 2006); and having positive impacts on the new product's effectiveness and efficiency, that is in optimising the fit of the product with firm competencies and market needs (Brown and Eisenhardt 1995; Koufteros et al. 2005).

One major challenge in leveraging customers for innovation management is the selection of customer collaborators and the establishment of ties with them (Nambisan 2002; Prügl 2006). Due to restrictions in time, personnel, and money, it is generally not economical for firms to collaborate with every customer. Using literature from a diversity of disciplines ranging from sociology and communication studies to economics and innovation management, this paper hypothesises that firms are more likely to select user community collaborators who exhibit certain individual- and group-specific characteristics that would enable the firm to effectively acquire a large pool of customer knowledge in user communities. User communities can be defined as horizontal user networks that consist of user nodes interconnected by information transfer links which may involve face-to-face, electronic or any other form of communication that provide members sociability, support, a sense of belonging, and social identity (Wellman et al. 2002; von Hippel 2007). This user-to-user interaction provides community members with access to a large pool of user-specific information as well as to general information regarding market and technological developments.

The purpose of this study is to investigate the extent to which user community collaborators exhibit the hypothesised individual- and group-specific characteristics. Empirical evidence is presented on the influence these characteristics have on the customer selection decision for two integration programs that correspond to the idea generation and problem solving phases of the innovation process of a proprietary software firm, Computer Associates (CA). Empirical results based on survey data of 808 respondents from 161 different CA user groups indicate that customers in user communities who are community leaders and lead users[1] are more often selected to cooperate with the firm for both idea generation and problem solving. Additionally, user community collaborators from larger communities have a higher probability of being selected for idea generation; and lead users who are also community leaders are more likely to be chosen for problem solving. Furthermore, results also show that individuals who are community leaders and lead users tend to be integrated in more collaboration programs.

While the idea of integrating communities of customers in the various phases of the innovation process is not new, much of the literature has focused on how firms can cooperate virtually with a large number of customers in online communities (Nambisan 2002; Prandelli et al. 2006; Dahan and Hauser 2001; Füller et al. 2006; Jeppesen and Frederiksen 2006; Prügl and Schreier 2006). This paper contributes to the literature by providing a descriptive study on how the integration of a small number of user community collaborators into the innovation process may provide firms with an efficient way of indirectly gaining access to input from a much larger number of customers in user communities.

In section 4.2 of this paper, literature on the integration of customers in the innovation

[1] Lead users refer to individuals who face needs today that will be general in the marketplace, but face them significantly earlier than the bulk of that marketplace encounters them, and who expect to benefit significantly by obtaining a solution to those needs (von Hippel 1986; Urban and von Hippel 1988).

process will be summarised. Next, the hypotheses of the characteristics of user collaborators will be derived in section 4.3. This will be followed by the presentation of the research setting and method in section 4.4. Section 4.5 describes the variables to be used in the empirical analysis and section 4.6 presents the descriptive statistics and the results of the multivariate binary response and ordered response analyses. Section 4.2 concludes with a discussion of the findings.

4.2 Customer Integration in the Innovation Process

The innovation process can be defined as occurring in three overlapping phases, which are 1) idea generation: the development of a design concept or technical proposal by integrating market needs with technical information; 2) problem solving: research, development, and engineering to develop a technical solution to the proposal; and 3) implementation and diffusion: market introduction, and the communication and use by which the innovation comes to have a wider social and economic impact (Myers and Marquis 1969; Utterback 1971). The following sub-sections will present the importance of customer integration in each of these phases.

4.2.1 Idea Generation

Studies have shown that customers can provide stimulating inputs for the idea generation phase of the innovation process (von Hippel 1986; Gruner and Homburg 2000; Prandelli et al. 2006). In this phase, customers can play two roles - as a resource to supply customer-need information and to evaluate ideas for new products, and as a co-creator where customers actively participate in the design of new concepts (Lengnick-Hall 1996; Nambisan 2002). The main motivation behind integrating customers in this phase of the innovation process is to reduce uncertainty by identifying customer preferences and interacting directly with customers to acquire new knowledge (Prandelli et al. 2006; Ogawa and Piller 2006). As user community members have access to a large pool of information regarding the product (Lakhani and von Hippel 2003; Shah 2005a; von Hippel 2007), interacting with customers in user communities can provide firms with the additional advantage of being better able to organise and prioritise need and feedback information from their customers.

Customers who are lead users have been shown to be particularly valuable for the idea generation phase as co-creators (Herstatt and von Hippel 1992; von Hippel et al. 1999; Olson and Bakke 2001). Lead users refer to individuals who face needs today that will be general in the marketplace, but face them significantly earlier than the bulk of that marketplace encounters them, and who expect to benefit significantly by obtaining a solution to those needs (von Hippel 1986; Urban and von Hippel 1988). The integration of lead users for idea generation has been shown to produce concepts that are truly novel and of value to the market as a whole in the future (Urban and von Hippel 1988; Lilien et al. 2002; Olson and Bakke 2001; Franke et al. 2006). Studies have shown that lead users can be found in user communities and that they create innovations either by themselves or with the assistance of other community members (Franke and Shah 2003; Jeppesen and Frederiksen 2006; Prügl and Schreier 2006; Füller et al. 2006). Hence, by cooperating with lead users from user communities, firms not only profit from their innate level of expertise but also from the additional information and "creativity boost" lead users obtain from their user communities.

4.2.2 Problem Solving

The primary contribution of customers in the problem solving phase is in product testing (Nambisan 2002). Product testing is often done by customers to determine if the product does what it is designed to do either at the manufacturer's facilities or in the customers' actual environments. Usability testing is user testing at the manufacturer's facility, where the manufacturer is able to control the conditions, directly observe prototype use, and obtain real-time customer feedback, whereas beta testing assesses the performance of the product in the users' actual environments for a real-world test (Dolan and Matthews 1993). Customer involvement for product testing enables firms to validate the product concept, eliminate product flaws prior to market introduction and minimise costly redesign and rework (Dolan and Matthews 1993; Nambisan 2002). The advantage of conducting product tests in user communities is the acceleration of the debugging process, since "given enough eyeballs, all bugs are shallow" (Raymond 1999).

Customers can also be co-creators in this phase of the innovation process when they actively create their own innovations. One way firms can integrate customers as co-creators is to introduce user-friendly toolkits that allow customers to create a preliminary design, stimulate or prototype it, evaluate its functioning in their own use environment and then iteratively improve it until satisfied (von Hippel 2001b; Thomke and von Hippel 2002). User toolkits allow manufacturers to abandon the attempt to understand user needs in detail in favor of transferring need-related aspects of product and service development to users along with an appropriate toolkit (von Hippel 2001b). User toolkits for innovation have been successfully implemented in a wide range of industries such as the food and beverage, manufacturing, semiconductor and software industries (von Hippel 2001b; Thomke and von Hippel 2002; von Hippel and Franke 2003). As users in user communities create and under certain conditions freely reveal[2] their innovations, firms can profit from these user-generated innovations (Mayrhofer 2005; Jeppesen and Frederiksen 2006; Prügl and Schreier 2006).

4.2.3 Implementation and Diffusion

Ebadi and Utterback (1984) propose that contact with external sources such as with potential users can assist the firm in the implementation phase of the innovation process. Urban and von Hippel (1988) propose that lead users who have been working together with the firm on new concepts can be useful after the product has been launched as opinion leaders to positively influence others' attitudes towards the new product. Schreier et al. (2007) find evidence that lead users demonstrate stronger opinion leadership as compared to ordinary users. As they are regarded as credible and trustworthy sources, opinion leaders are able to informally influence others' attitudes or overt behaviour in a desired way with relative frequency (Katz and Lazarsfeld 1955; Rogers and Agarwala-Rogers 1976; Adams 1976). Hence, they are individuals who can lead in influencing others' opinions about innovations (Rogers 1962; Czepiel 1975; Valente and Davis 1999).

Lead users in user communities who have a lot of intra-community contacts are often perceived as influential and can thus be used by firms as seeding agents to speed up the innovation diffusion process[3] (Morrison et al. 2000). Harrison and Waluszewski (2007) propose that user networks can be leveraged for the successful re-launch of an unwanted product. Specifically, they find that lead users play a pivotal role in the diffusion process by sharing their use experience to new and potential adopters of the product.

As the technical process of innovation ends once the product or process has been introduced in the market, the technical uncertainties involved in the first two phases of the innovation process are thought to be much higher as compared to the implementation and diffusion phase (Utterback 1971). Given that various studies in innovation management have shown that customer integration is particularly valuable for the technical process (von Hippel 1986; Souder et al. 1998; Salomo et al. 2003; Veryzer and Borja de Mozota 2003), this paper will focus its analysis on the integration of customers in the idea generation and problem solving phases of the innovation process.

[2]Free revealing occurs when an innovator voluntarily gives up his intellectual property rights to that innovation and all interested parties are given access to it. Conditions under which free-revealing may occur is, when perceived competition among users is low, when the user innovation is general in nature such that other users can also benefit, when the value to the innovating user that free revealing brings is high as it induces the manufacturer of the innovation to make and distribute to all users equally, and when the cost to each user of adopting the improved commercial product is low. (Harhoff et al. 2003). In general, users will freely reveal their innovations when they perceive higher profits than free-riders from disclosing their innovations (von Hippel and von Krogh 2006).

[3]The diffusion of an innovation is the process by which the innovation is communicated through certain channels over time among members in a social system (Rogers 1983).

4.3 Characteristics of User Community Collaborators

Section 4.2 has shown the importance of acquiring information from customers for the technical innovation process, that is for idea generation and problem solving. As it is generally not economical for firms to collaborate with every customer, firms are more likely to select user community collaborators who exhibit certain individual- and group-specific characteristics that would enable the firm to effectively acquire the large pool of customer knowledge in user communities. In the following, these individual- and group-specific hypotheses will be derived.

4.3.1 Individual-Specific Characteristics

Strong Ties

Relational ties or personal contacts between individuals can be considered channels for the transfer or the flow of resources, such as information and knowledge (Wasserman and Faust 1994). Information that individuals receive from their personal contacts can have a big impact on their decision making process because this information is considered more credible than information from non-personal sources (Lazarsfeld et al. 1944; Parthasarathy et al. 1995).

The strength of a tie is the combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterise the tie (Granovetter 1973). Strong ties can be considered trusted informational contacts (Carpenter et al. 2003). Hence, due to the high frequency of communication and the fact that information they receive from their strong ties is considered credible, it is expected that

> *H1: Users who have a larger number of strong ties within their communities are more likely to be selected for collaboration in innovation projects.*

Community Leader

Yukl (1981) defines leadership as "... the process of influencing others to understand and agree about what needs to be done and how it can be done effectively, and the process of facilitating individual and collective efforts to accomplish the shared objectives.". Due to their position in the community, community leaders enjoy a certain level of recognition from and hence possess a certain level of influence and authority over other members in the community. Thus, the community leader can be viewed as a formal opinion leader.

Community leaders can also be perceived as gate keepers who control the flow of information that flows in and out of his community to the firm. The term "gate keeper" was introduced by Lewin (1947) to describe a mother as a person who decided which foods end up on the family's dinner table. White (1950) had first applied this gating process to mass communication research, where the "gate keeper" was depicted as an individual who controlled information focal points. The community leader can perform this gate keeping function by assisting the firm in organising and prioritising the pool of user information regarding user needs, feedback to existing products as well as ideas for new products from his community.

Due to the above mentioned reasons and the fact that firms tend to look for persons in "key positions" in the community (Lewin 1947), it is expected that

> *H2: Community leaders have a higher likelihood of being leveraged for innovation management.*

Lead Users

Lead users are an important external source of information for the innovation process. Due to their high level of expertise, lead users who are members in user communities may be considered informal opinion leaders (Urban and von Hippel 1988; Morrison et al. 2000; Schreier et al. 2007), whom other users consult for advice and are hence in a good position to filter and interpret relevant technological and market information from their user community to the firm during collaboration projects. Thus,

> *H3: Firms are more likely to cooperate with lead users from user communities in innovation projects.*

Lead users who are also community leaders are similar to technological gatekeepers. The technological gatekeeper as defined by Allen and Cohen (1969), is a boundary spanner who occupies a key position in the communication network of an R&D laboratory; that is those to whom others in a laboratory most frequently turn to for technical advice and consultation, and who show more contact with technical activity outside of the laboratory. Tushman and Scanlan (1981) find that such boundary spanners are often strongly linked internally and externally, so as to be in a position to efficiently gather and disseminate information from outside their subunits.

Lead users who are also community leaders would be especially valuable for the firm as they can also be leveraged as gate keepers. Hence it can be assumed that:

> *H3a: Firms are more likely to cooperate with lead users who are also community leaders in innovation projects.*

4.3.2 Group-Specific Characteristics

Membership Size

Literature on the theory of collective action propose that when the costs of providing the collective good vary little with group size, larger groups exhibit more collective action as they have a higher probability of having more highly interested and resourceful individuals who are willing and able to provide the collective good for everyone (Oliver et al. 1985; Oliver and Marwell 1988). The collective good in the case of user communities is that of information and knowledge which are non-rivalrous in nature, meaning that the cost of its provision does not increase with the number of individuals in the group (Cornes and Sandler 1996). Hence, members in larger groups have access to a larger pool of community resources.

User collaborators in larger groups could be considered more valuable to the firm since they would be in a position to filter larger pools of information to the firm and to represent the opinions of a large number of customers. It is thus expected that:

> *H4: Firms are more likely to choose user collaborators from larger user communities.*

Open Communication

Gladstein (1984) proposes that open communication within groups will increase the group's information processing capacity. An open culture of communication that encourages the free flow of knowledge would encourage individuals to share more of their knowledge with each other (von Krogh 1998), thus increasing the quantity of the potential knowledge pool in the community. Furthermore, open communication can also lead to more objective discussions regarding the product. Therefore, as users from user communities with an open culture of communication would have access to more user-generated information as compared to groups with a reserved communication culture, the following relationship is expected:

> *H5: Firms are more likely to choose user collaborators from user communities with an open culture of communication.*

4.4 Study Setting and Method

User communities from the proprietary software firm Computer Associates (CA) were selected for analysis in this study. CA supports over 300 official CA user groups that provide a platform for user-to-user contact to exchange information relating to any one of their management software product lines. Official CA user groups receive assistance from

CA in the organisation of their activities, free refreshments and the use of CA facilities for user group meetings, as well as web administration support for their user group web sites. CA is expected to gain not only from the user-input for the development process of their products but also from customer references who are willing to help them sell their products. Independent to their CA user group program, CA has two additional customer integration programs, the CA Product Advisory Council (PAC) program, and the Development Buddy (DevBud) program.[4] Detailed information on these two programs were obtained from explorative interviews with twelve CA employees in November 2005.

The aim of the CA PAC program is to assist CA to design effective development strategies for future product releases that correspond to real-world needs. CA sponsors more than 20 PACs which focus on a variety of CA product lines as well as specific platforms and technologies. Each PAC consists of between 12 to 15 experienced, prominent IT and business professionals. PAC members discuss topics that include future product functionality, competitive concerns, industry standards, government regulation, and other issues and trends. Although PAC members can be nominated by CA's field marketing group, sales area managers, customer advocates and other internal resources, PAC leaders (vice-president of product and/or development of the respective product line) have the final say on the selection of their members, who may also include users whom they know through user group activities. Tenure within the council is at the discretion of PAC leaders. Users would voluntarily accept invitation to the program if they want to play a strategic role in deciding the direction of CA's products and to ensure that new releases meet their firms' needs thereby enhancing the return on their investments. Furthermore PAC members often receive early access to the latest CA technologies, copies of alpha and beta products, regularly scheduled development briefings and direct access to high-level CA development and marketing managers. The PAC program is an example of customer collaboration at the idea generation phase of the innovation process.

The CA DevBud program is designed to provide customers with access to CA's development resources and to provide CA with a feedback mechanism to help ensure that the products developed meet the needs of individual clients and the overall market. DevBud participants are expected (but not required) to participate in appropriate alpha and beta programs and provide timely feedback; regularly interact with CA's developers during scheduled webcasts, conference calls, and CA World conferences; maintain a test environment representative of their production environment in which new technology can be deployed and exercised; discuss beta products with a limited number of press and analysts during beta testing; and participate in the CA reference program after products

[4]Information on the Development Buddy program is also obtained from CA's website: www.ca.com/cadevelopmentbuddy/faq.htm (accessed on 22nd November 2005).

become generally available. Participants are selected from a list of users who apply for the program. Selection is determined by the senior vice-president of development of the respective product line and is based on the type of environment, CA products used, other products used and the willingness of users to invest the necessary time and associated expenses to participate. Customers sign up to participate in the DevBud program because they want early access to the latest CA technologies, on-site team training and implementation assistance for alpha and beta products, access to scheduled conference calls and webcasts between CA developers and other DevBuds and the possibility to offer insight and input on CA development roadmaps. The DevBud program corresponds to the problem solving phase of the innovation process.

A web-based survey was used to gather data from individual users in CA user groups. In preparation of the construction of the questionnaire used in the survey, explorative interviews were conducted with twenty-nine user group members as well as a CA employee who was responsible for the user group program between October and December 2004. In the questionnaire, users were asked questions relating to the CA software they use, the formal user group activities they attended, the personal contacts they have in their group, the software code they wrote, their assessment of their group, the reasons for joining and participating in their group, their position in the group, the general characteristics of the group (these questions were only filled out by committee members of each group), other members in the group, and their personal information which includes information on if they participate in the PAC and/or DevBud programs.[5]

Seven user group members had participated in the pre-test of the survey between May and June 2005. An email with a link to the web-based survey was sent to 398 user group presidents and other committee members from 265 different user groups. These individuals were requested to participate in the survey and to forward the link to their members. Their contact details were obtained from the official CA user group website[6]. All emails were sent out on the 19th July 2005 and information from user group members were collected until the 11th January 2006. In total, 1055 individuals in 203 different CA user groups responded to the survey. However, some users had filled out the survey more than once because they were members in more than one CA user group. For the purpose of this study, email addresses were used to exclude 66 multiple responses as the focus here is on the characteristics of the individual user. The first response of multiple respondents were kept as this response related to the most important CA user group they were in. A further 181 questionnaires were not included in the analysis because they were not properly answered. Thus the analysis is based on 808 users from 161 different CA

[5]See Appendix B for a copy of the questionnaire.
[6]http://causergroups.com/UserGroups/ (accessed on 7th July 2005).

user groups or 77 percent of the collected sample.

It is not possible to calculate an exact response rate as data on the size of the underlying population was not available. However, a non-response analysis comparing the responses of early and late respondents was carried out (Armstrong and Overton 1977).[7] Assuming that late respondents are similar in characteristics to non-respondents (Ferber 1948), results of the test indicate that at the 1% level of significance, there could be a potential self-selection bias favouring individuals in user groups with an open communication culture. The implication of a self-selection bias is that when respondents differ from non-respondents, statistics based on responses alone may not validly depict the population investigated and could result in inaccurate and misleading results (Ellis et al. 1970). Thus, "openly communicative" individuals may be over-represented in the sample. However, this bias is not considered to be problematic as this variable is included as an independent variable in the multivariate analysis presented in section 4.6.2.

4.5 Data Description

Before the characteristics of customer collaborators are analysed, it is necessary to firstly provide a brief description of the dependent, independent and control variables to be used in the multivariate analysis.

Dependent Variable

This study will employ two different dependent variables for the multivariate analysis. For one analysis, the dependent variable is a binary response, that is if an individual is a user collaborator who has been selected to participate in either the PAC or DevBud programs. For another analysis, the dependent variable is a count response, that is the number of different customer collaboration programs an individual was selected for by CA.

PAC member
Members were asked if they are members in a Product Advisory Council. This variable takes the value one if they are a member and zero otherwise.

DevBud
Members were asked if they participate in the CA Development Buddy program. This variable *DevBud* takes the value one if they are Development Buddies and zero otherwise.

[7] Specifically, the first 10% was compared with the last 10% of the responses to the survey with regard to the dependent, independent and control variables using T-tests (metric data) and χ^2-tests (nominal data).

Num Programs

This variable is simply a sum of *PAC member* and *DevBud*, and reflects the number of different customer collaboration programs an individual was selected for by CA.

Independent Variables

Strong Ties

The number of strong ties is the number of close contacts (excluding colleagues) users have who are members in CA user groups. Close contacts are defined as individuals with whom they communicate frequently, whom they can approach at any time for help and whom they would willingly help at any time. This definition follows Granovetter's (1973) definition of strong ties very closely.

Lead User

Based on existing lead user literature (Urban and von Hippel 1988; Herstatt and von Hippel 1992; Morrison et al. 2004; Jeppesen and Frederiksen 2006), lead users in CA user groups are identified as satisfying three conditions - 1) users must be using the most recent version of their software, 2) users must have submitted at least one "need" suggestion (DAR) for new product releases in the course of the last year, and 3) users must have written software code that they would consider "new to the world". Two variables measuring lead userness are constructed. *Lead user-2* is a dummy variable that takes the value one if an individual satisfied all the above-mentioned conditions and if his suggestions were forwarded by CA for development into new releases or slated as "in the process for a future release", and zero otherwise. *Lead user-1* is a dummy variable that takes the value one if an individual satisfied all the above-mentioned conditions but whose suggestions were not acknowledged by CA for future product releases, and zero otherwise.

Community Leader

The *community leader* variable takes the value one if the user is a president, co-president or vice-president of his user group and zero otherwise.

Group Size

Users who are on the management committee of their user group were asked to indicate the number of individual members in their user group. If multiple differing answers were provided for each group, the answer of the group leader (i.e. president, co-president and/or vice-president) was taken.

Open Communication

Users were asked to indicate the extent to which they agreed or disagreed that their CA user group has been extremely successful in promoting an open exchange of information between its members. This *open communication* variable takes the value one when at least 50% of the members in each user group who responded to the survey indicated that they either strongly agreed or somewhat agreed with the statement, and zero otherwise. Individuals in CA user groups with an open exchange of information are expected to willingly share and exchange information with one another. The formulation of the item for this variable was inspired by Shah (2005b).

Control Variables

Company Size
Individuals who work for large companies may have a higher probability of being nominated by the sales department or other internal sources of CA to be a member of the PAC. To control for this effect, *company size* measures the total number of employees working for the user's company.

Perceived Fairness
Users who feel fairly treated by CA for their participation in their user group may be willing to participate in innovation collaboration projects with CA. Users were asked to indicate the extent to which they agreed or disagreed with the following statement: "CA treats me fairly by acknowledging my involvement in this CA user group". *Perceived fairness* takes the value one if users indicated that they either strongly agreed or somewhat agreed with the statement, and zero otherwise. This variable represents one of four items that Mayrhofer (2005) used to measure perceived fairness.

Multiple Member
Users who are members in more than one user group may be enthusiastic about the product and hence be more willing to agree to participate in collaboration innovation projects. This variable takes the value one if users are members in more than one user group, and zero otherwise.

Feedback
Users who joined the user group with the aim of getting their feedback acknowledged by CA by participating in the group may be interested to cooperate with CA in collaboration programs. Users were asked to indicate the extent to which they agreed or disagreed with the following statement on why they joined their user group: "For the "power in numbers" approach in getting my feedback acknowledged by CA". *Feedback* takes the value one if users indicated that they either strongly agreed or somewhat agreed with the statement, and zero otherwise. This variable was conceptualised using interview results with CA user group members.

Table 4.1: Descriptive Statistics (N = 808)

Variable	Mean	S.D.	Min.	Max.
PAC member (0/1)	0.08		0	1
DevBud (0/1)	0.13		0	1
Num Programs	0.21	0.47	0	2
strong ties	3.39	9.02	0	200
community leader (0/1)	0.20		0	1
lead user-2 (0/1)	0.04		0	1
lead user-1 (0/1)	0.02		0	1
group size	119.13	136.68	7	503
open communication (0/1)	0.86		0	1
company size	19,354.17	46,272.72	1	550,000
perceived fairness (0/1)	0.50		0	1
multiple member (0/1)	0.18		0	1
feedback (0/1)	0.71		0	1
female (0/1)	0.21		0	1
use experience	10.39	6.33	1	34

Female
To control for gender effects, this variable takes the value one for females and zero for males.

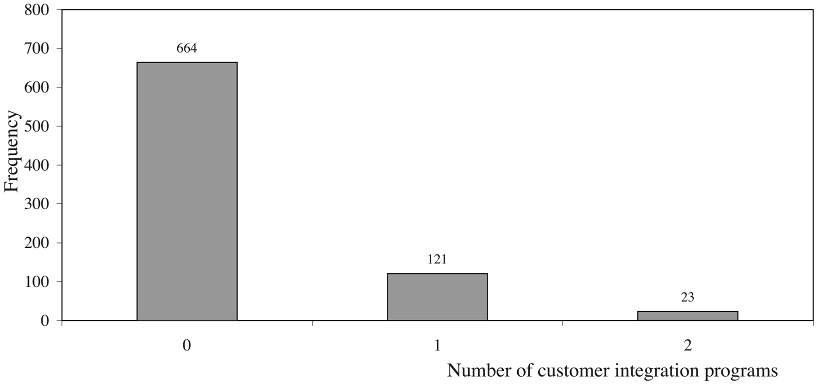
Use Experience
More experienced individuals may have a higher likelihood of being selected for collaboration programs. To control for this, the number of years which users have used their most important CA software will be used to proxy a user's level of experience.

4.6 Empirical Analysis

4.6.1 Descriptive Statistics

Descriptive statistics are displayed in Table 4.1. The final sample consists of 808 responses from individuals in 161 different user groups. Of the 808 respondents, 8% are members in a PAC and 13% participate in DevBud programs. The 65 PAC members originate from 45, and the 102 Dev-Buds from 63 different CA user groups. This is an interesting result as it suggests that there may be more "valuable" user groups with certain group-specific characteristics which CA favours when searching for user community collaborators. Next, it can be seen that the mean number of programs respondents participate in is 0.21. *Num Programs* has a minimum of 0 and a maximum of 2. From Figure 4.1, it can be seen that 664 respondents (82.2%) do not participate in any innovation collaboration programs, 121 respondents (15%) participate in only one program and only 23 respondents (2.8%) were

Figure 4.1: Number of Customer Integration Programs per Individual (N = 808)



selected by CA to participate in both programs. These results indicate that 17.8% of the sample are user community collaborators. These 144 individuals come from 94 different user groups.

It can be observed that on average, respondents have approximately 3 strong ties to members in their user community. *Strong ties* ranges from 0 to 200 and has a positively skewed distribution, where approximately 41.1% of the respondents do not have any close contacts in their community. Due to the skewed nature of *strong ties* and to accommodate for zero values, the logarithm of one plus *strong ties* will be used in the multivariate analysis.

20% of the respondents are community leaders. This high percentage is due to the fact that community leaders were contacted directly to participate in the survey, whereas regular user group members were informed of the survey indirectly through their community leaders. Lead users consist of only 6% of the sample. Specifically, lead users (*lead user-2*) whose need suggestions were recognised as corresponding to that of future needs of the market comprise 4% of the respondents. Lead users (*lead user-1*) whose needs were not recognised by CA comprise 2% of the sample.

The following two variables are user group-specific characteristics. As displayed in Table 4.1, the average number of individual members in user groups across the 808 respondents is 119.13 with a standard deviation of 136.85. *Group size* ranges from 7 to 503 members. Analysing the descriptive statistics across the 161 user groups shows that the mean of *group size* is 75.24 with a standard deviation of 90.82. This finding is not surprising as it indicates that there are more respondents from larger user groups as compared to smaller ones in the sample. 86% of the respondents belong to CA user groups that have an open culture of communication. Similarly, there are approximately 84% of the 161 user groups that have members who are willing to share and exchange information with one another.

The average number of employees working for the respondents' firms is 19,354. Respondents ranged from independent free lance consultants to individuals that work for very large companies consisting of 550,000 employees. As *company size* has a positively skewed distribution, the logarithmic form of *company size* will be utilised in the multivariate analysis. Although CA selects which customers they want to cooperate with, individuals make the final decision to cooperate or not. The next three variables control for the motivation or willingness of individuals to collaborate with CA. As can be seen, half of the respondents felt that CA treated them fairly for their participation in their user groups and may hence be willing to collaborate with CA in innovation projects. It is observed that 18% of the respondents are individuals who are so enthusiastic about their CA product that they join more than one CA user group to be able to access more information and contacts. 71% had indicated that they had joined their user group with the aim of getting their feedback acknowledged by CA and may thus be motivated to participate in collaboration programs with CA to have an influence in the future direction of the CA product. Next, 21% of the respondents are female. Finally, the average number of years that users have had experience with their CA software is 10.39. This finding indicates that in general, members in CA user groups are relatively loyal customers who have continued using their CA product over many years.

Table 4.2 displays the differences[8] in characteristics between PAC and non-PAC members. It can be observed that PAC members are significantly different from non-PAC members in the following aspects: PAC members have more than twice the number of strong ties in their community as compared to non-PAC numbers, are more often community leaders, tend to be lead users whose suggestions were acknowledged by CA, work in larger companies, feel that CA acknowledges their participation in their user group, are members in more than one user group, and had joined their user group with the aim of getting their feedback acknowledged by CA. Similar results are found for the differences between DevBuds and non-DevBuds as shown in Table 4.3 with the exception that DevBuds also tend to be lead users whose needs are not recognised as corresponding to that of the future market, do not tend to work for large companies, did not necessarily join the user community to get their feedback acknowledged by CA, and tend to be males.

[8] T-tests were conducted for interval-scaled variables (connectedness, group size and company size) and Pearson χ^2 tests for all other nominal-scaled variables for both Tables 4.2 and 4.3.

Table 4.2: Differences between PAC and Non-PAC Members (N = 808)

Variable	Non-PAC members Mean (743 obs)	PAC members Mean (65 obs)
strong ties ***	2.97	8.17
community leader (0/1) ***	0.18	0.46
lead user-2 (0/1) ***	0.03	0.15
lead user-1 (0/1)	0.02	0.05
group size	117.54	137.37
open communication (0/1)	0.85	0.89
company size ***	18,085.65	33,854.35
perceived fairness (0/1) ***	0.48	0.66
multiple member (0/1) ***	0.16	0.37
feedback (0/1) **	0.70	0.85
female (0/1)	0.21	0.17
use experience	10.35	10.89

** 5% and *** 1% significant difference between PAC and non-PAC members

Table 4.3: Differences between DevBud and Non-DevBuds (N = 808)

Variable	Non-DevBuds Mean (706 obs)	DevBuds Mean (102 obs)
strong ties ***	2.90	6.80
community leader (0/1) ***	0.17	0.44
lead user-2 (0/1) ***	0.03	0.10
lead user-1 (0/1) ***	0.01	0.07
group size	119.60	115.87
open communication (0/1)	0.86	0.84
company size	19,681.53	17,088.32
perceived fairness (0/1) ***	0.47	0.69
multiple member (0/1) ***	0.15	0.32
feedback (0/1)	0.70	0.75
female (0/1) **	0.22	0.13
use experience	10.32	10.89

** 5% and *** 1% significant difference between DevBuds and non-DevBuds

Table 4.4: Differences between Individuals who Participate in neither Program, in only one Program, and in both Programs (N = 808)

Variable	**Non-Participant Mean (664 obs)**	**PAC only or DevBud only Mean (121 obs)**	**PAC and DevBud Mean (23 obs)**
strong ties	2.78	4.64	14.43
community leader (0/1)	0.15	0.37	0.65
lead user-2 (0/1)	0.03	0.05	0.30
lead user-1 (0/1)	0.01	0.07	0.04
group size	117.43	130.73	107.17
open communication (0/1)	0.86	0.84	0.91
company size	19,167.09	15,528.08	44,883.57
fairness (0/1)	0.46	0.65	0.74
multiple member (0/1)	0.14	0.29	0.48
feedback (0/1)	0.70	0.76	0.87
female (0/1)	0.22	0.10	0.26
use experience	10.33	10.39	12.22

Table 4.4 displays the differences in characteristics of individuals between user community members who did not participate in either program, who are members in only one of the programs and who are members in both programs respectively. It can be seen that the results are very similar to those from Tables 4.2 and 4.3. The results from Tables 4.2 and 4.3 suggest that there are significant differences between the characteristics of user collaborators and ordinary community members, whereas those from Table 4.4 suggest that individuals with certain characteristics are more often chosen for customer integration programs. These differences will be analysed in more detail in the multivariate analysis where cross-correlations between all independent and control variables would be accounted for.

Table 4.5 presents an overview of the bivariate relationships between the independent and control variables. It can be seen that there is a positive correlation ($\rho = 0.29$) between community leaders and the number of strong ties. This is a reasonable finding as community leaders play a central role in the organisation and management of the user community and hence have more opportunity to make close contacts in their communities as compared to regular user group members. Furthermore, it can also be observed that community leaders feel fairly treated by CA for managing their user groups and for participating in user group activities ($\rho = 0.28$). This result is not surprising as user group presidents are provided with incentives from CA such as free accommodation to attend the Computer Associates Regional Exchange (CARE) conference, which takes place 2 days before CA's official annual user conference, CA World. In addition, group leaders

Table 4.5: Correlation Matrix of the Independent and Control Variables (N = 808)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) strong ties	1.00										
(2) community leader (0/1)	0.29◊	1.00									
(3) lead user-2 (0/1)	0.08◊	0.11	1.00								
(4) lead user-1 (0/1)	0.07◊	0.06	−0.03	1.00							
(5) group size	0.11♠	−0.14◊	0.04◊	0.10◊	1.00						
(6) open communication (0/1)	0.04◊	−0.04	0.01	0.04	−0.01	1.00					
(7) company size	0.07♠	0.00◊	0.01◊	−0.01◊	0.01♠	0.11◊	1.00				
(8) fairness (0/1)	0.09◊	0.28	0.01	0.01	−0.13	0.14	0.02	1.00			
(9) multiple member (0/1)	0.15◊	0.20	0.03	0.09	0.06	0.09	−0.02	0.07	1.00		
(10) feedback (0/1)	0.06◊	0.13	0.04	0.02	−0.03	0.11	0.10	0.23	0.11	1.00	
(11) female (0/1)	−0.01◊	0.03	0.00	−0.03	−0.07	0.11	0.05	0.08	0.04	0.05	1.00
(12) use experience	0.20♠	0.12◊	0.06◊	0.08◊	0.10♠	−0.04◊	−0.04♠	0.01◊	0.06◊	0.04◊	−0.09◊

Note: Phi-coefficient values unless otherwise stated; ◊ Point biserial correlation coefficient,
♠ Pearson product-moment correlation coefficient

also receive a registration fee waiver to attend CA World.

4.6.2 Multivariate Analysis

The multivariate analysis will estimate two different econometric models. A probit model will be estimated when the dependent variable is a binary response - that is, if the user is a user community collaborator or not. When the dependent variable is a count response - that is the number of different customer collaboration programs CA selects an individual for, an ordered probit model will be estimated. This section will first present the model specifications of the two different estimation methods, followed by the empirical results of the multivariate analysis.

4.6.2.1 Model Specifications

Probit Model

When the dependent variable is a binary response - that is, if an individual is a user community collaborator or not, a linear probability model (LPM) can be estimated to test the hypotheses (Wooldridge 2003).

Let the linear probability model be defined as

$$P(y = 1|\boldsymbol{x}) = \boldsymbol{x}\beta + \epsilon, \tag{4.1}$$

where y represents a $N \times 1$ vector of the binary variable indicating if an individual is a user community collaborator or not, $\boldsymbol{x}$ represents the $N \times K$ matrix of independent and control variables and β represents a $K \times 1$ vector of coefficients.[9] The model assumes $\epsilon|\boldsymbol{x} \sim N(0, \sigma^2 I)$. The β_js are estimated using ordinary least squares (OLS) and interpreted as fixed marginal effects, that is the change in probability of success when a variable x_k changes, holding other factors fixed:

$$\Delta P(y = 1|\boldsymbol{x}) = \beta_k \Delta x_k \tag{4.2}$$

Although the linear probability model is easy to estimate and interpret, it has a couple of shortcomings. First, with certain combinations of values of the independent and control variables, it is possible to get predicted values either less than zero or greater than one, which do not make any sense considering probabilities are bounded between zero and one. A further limitation with the linear probability model is that a probability cannot be linearly related to the independent and control variables for all their possible values,

[9] N represents the total number of observations and K the total number of independent and control variables plus one for the constant term.

that is the marginal effects of a change in one variable is always constant regardless of the original value of the variable. (Wooldridge 2003)

To avoid these above-mentioned LPM limitations, this study will use a more sophisticated binary response model - the Probit Model. Probit models are generally derived from an underlying latent variable model (Wooldridge 2003). The underlying unobserved, latent variable in this study is CA's cooperative behaviour with user community collaborators. However, what can be observed is a proxy of the latent variable - CA's selection decision of user community collaborators. This observed variable is a function of the unbounded underlying latent variable and is bounded between 0 and 1. Let y^* be the unobserved, latent variable

$$y^* = \boldsymbol{x}\beta + e, \;\; y = 1[y* > 0], \tag{4.3}$$

The notation 1[.] is called the indicator function, which takes on the value one if the event in brackets is true., and zero otherwise. Hence, y is one if $y^* > 0$ and y is zero if $y^* \leq 0$. e is assumed to be independent of x and has the standard normal distribution. From equation 4.3, the following response probability for y can be derived;

$$\begin{aligned} P(y = 1|\boldsymbol{x}) =& P(y^* > 0|\boldsymbol{x}) \\ =& P(e > -\boldsymbol{x}\beta|\boldsymbol{x}) \\ =& 1 - G[-\boldsymbol{x}\beta] \\ =& G(\boldsymbol{x}\beta), \end{aligned}$$

where G is the standard normal cumulative distribution function. In comparison to the LPM, the β_js in the probit model are estimated by maximum likelihood. The partial effect of a roughly continuous variable x_j on $p(\boldsymbol{x}) = P(y = 1|\boldsymbol{x})$ is obtained from the partial derivative:

$$\frac{\delta p(\boldsymbol{x})}{\delta x_j} = g(\boldsymbol{x}\beta)\beta_j, \tag{4.4}$$

where $g(z) \equiv \frac{dG}{dz}(z)$. As G(.) is a strictly increasing function, the partial or marginal effect always has the same sign as β_j. In the following estimation results for the probit model, marginal effects of the various independent variables are reported.

Ordered Probit Model

As the alternative dependent variable - that is the number of different customer collaboration programs an individual was selected for by CA - is count data, a poisson model may at first glance appear to be a suitable estimation method (Wooldridge 2002). However, an ordered probit model will be applied in this study, because the dependent variable *Num Programs* has only values from 0 to 2. Hence, it would be inappropriate to impose a poisson distribution on the data.

Let y now be an ordered response taking on values 0,1,2,...,J for some known integer J (Wooldridge 2002). Assume now that, the latent variable y^* is determined by

$$y^* = \boldsymbol{x}\beta + e, \;\; e|\boldsymbol{x} \sim \text{Normal}(0,1), \tag{4.5}$$

Let $\alpha_1 < \alpha_2 < \alpha_3 < \cdots < \alpha_J$ be unknown cut points or threshold parameters, and define

$$\begin{aligned} y &= 0 \quad \text{if } y^* \leq \alpha_1 \\ y &= 1 \quad \text{if } \alpha_1 < y^* \leq \alpha_2 \\ &. \\ &. \\ &. \\ y &= J \quad \text{if } y^* \geq \alpha_J \end{aligned}$$

For instance, if y takes on the values 0, 1, and 2, then there will be two cut points, α_1 and α_2. Given the standard normal assumption for e, each response probability is computed as:

$$\begin{aligned} &P(y = 0|\boldsymbol{x}) = P(y^* \leq \alpha_1|\boldsymbol{x}) = P(\boldsymbol{x}\beta + e \leq \alpha_1|\boldsymbol{x}) = \Phi(\alpha_1 - \boldsymbol{x}\beta) \\ &P(y = 1|\boldsymbol{x}) = P(\alpha_1 < y^* \leq \alpha_2|\boldsymbol{x}) = \Phi(\alpha_2 - \boldsymbol{x}\beta) - \Phi(\alpha_1 - \boldsymbol{x}\beta) \\ &. \\ &. \\ &. \\ &P(y = J|\boldsymbol{x}) = P(y^* > \alpha_J|\boldsymbol{x}) = 1 - \Phi(\alpha_J - \boldsymbol{x}\beta) \end{aligned}$$

The parameters α and β can be estimated by maximum likelihood. In comparison to the standard probit model, the direction of the partial effect of x_j on the response probabilities does not always correspond to the sign of β_j. Although the partial effect of x_j on the probabilities $\mathrm{P}(y = 0|\boldsymbol{x})$ and $\mathrm{P}(y = J|\boldsymbol{x})$ are unambiguously determined by the sign of β_k, the sign of β_k does not always determine the direction of the effect for the intermediate outcomes, 1,2,...,J-1. This is because the standard normal probability density function is symmetric about zero, reaches a maximum at zero, and declines monotonically as its argument increases in absolute value. In the following estimation results for the ordered probit model, estimations of β_js are reported. Marginal effects of the independent variables are also estimated and reported separately.

4.6.2.2 Estimation Results

In the following, characteristics of user collaborators from user communities are analysed in a multivariate framework. Results of the hypothesis tests for individuals who are PAC members and DevBuds are examined separately in Tables 4.6 and 4.7 using a probit model. Both tables present not only the estimated coefficients but also the marginal effects of independent and control variables calculated at their means.[10] Table 4.8 presents the coefficients of Ordered Probit estimations for the number of collaboration programs individuals are selected by CA for integration. As the sample size is relatively large (N=808), only results with significance at the 5% are elaborated on.

Models 1 and 2 in Table 4.6 present the results for PAC members with and without the interaction terms to test for hypothesis 3a. Unless otherwise stated, the findings of Model 1 without the interaction terms will be elaborated on. As hypothesised, an individual's number of strong ties appears to have a positive influence on the probability that an individual is a user collaborator in the idea generation phase.[11] Unfortunately, this marginal effect is not significant. One possible explanation for this is that the effect of the number of strong ties can be explained through the community leader variable, as it was shown in section 4.6.1 that *community leader* is correlated with the number of

[10] For dummy variables, the marginal effect presented is for a discrete change of the dummy variable from 0 to 1.

[11] It is acknowledged that there could be a potential endogeneity problem as individuals who want to be selected for customer integration programs may actively forge and maintain a larger number of strong ties. However, this paper argues that due to limited time and emotional resources, it is not expected that individuals would be able to sustain a large number of strong ties at will as this would require a high investment in time and energy (Granovetter 1973). It is thus argued that the number of weak ties (and not strong ties) is potentially endogenous to an individual being selected for customer integration programs, as the maintenance of weak ties does not require much personal investment (Granovetter 1973; Seibert et al. 2001).

Table 4.6: Probit Regression Models for PAC members with Heteroscedasticity-Robust Standard Errors in the Presence of Intra-Cluster Correlation

	Model 1		**Model 2**	
	PAC member (0/1)		PAC member (0/1)	
	Coefficients	Marginal Effects (dF/dx)	Coefficients	Marginal Effects (dF/dx)
ln (strong ties + 1)	0.103	0.012	0.102	0.012
	[0.078]	[0.009]	[0.078]	[0.009]
community leader (0/1)	0.472***	0.067**	0.459***	0.064**
	[0.161]	[0.026]	[0.166]	[0.026]
lead user-2 (0/1)	0.769***	0.144**	0.763***	0.143*
	[0.230]	[0.064]	[0.287]	[0.075]
lead user-1 (0/1)	0.237	0.032	0.111	0.014
	[0.430]	[0.068]	[0.605]	[0.080]
lead user-2 (0/1) * community leader (0/1)			0.020	0.002
			[0.475]	[0.055]
lead user-1 * community leader			0.270	0.037
			[0.831]	[0.138]
ln (group size)	0.132**	0.015**	0.130**	0.015**
	[0.061]	[0.006]	[0.061]	[0.006]
open communication (0/1)	0.041	0.004	0.042	0.005
	[0.208]	[0.022]	[0.208]	[0.022]
company size	0.087***	0.010***	0.086***	0.010***
	[0.025]	[0.003]	[0.025]	[0.003]
fairness (0/1)	0.236	0.027	0.239*	0.027
	[0.144]	[0.017]	[0.144]	[0.017]
multiple member (0/1)	0.468***	0.067**	0.469***	0.067**
	[0.163]	[0.029]	[0.161]	[0.029]
feedback (0/1)	0.172	0.018	0.175	0.019
	[0.164]	[0.016]	[0.167]	[0.016]
female (0/1)	−0.369**	−0.035**	−0.371**	−0.035**
	[0.181]	[0.015]	[0.180]	[0.015]
ln (use experience)	−0.042	−0.005	−0.043	−0.005
	[0.094]	[0.011]	[0.094]	[0.011]
Constant	−3.215***		−3.200***	
	[0.456]		[0.462]	
Observations	808		808	
Log pseudolikelihood	−192.887		−192.829	

Robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

strong ties.[12] Nevertheless, this means that hypothesis 1 is not supported.

Hypothesis 2 proposes that community leaders have a higher likelihood of being leveraged for innovation management.[13] As expected, the marginal effect of being a community leader is positively significant at the 5% level. Specifically, community leaders have a 6.7% higher probability of being selected for collaboration in the idea generation phase.

Two dummy variables are used to proxy the degree of lead userness of an individual. The reference group are users who do not satisfy *all* three lead user criteria. As can be observed in Table 4.6, the coefficient of the lead user-2 in Model 1 is significant at the 1% level. Specifically, lead users whose suggestions were acknowledged by CA have a 14.4% higher probability of being selected for the PAC program. This marginal effect is significant at the 5% level. The marginal effect of lead users whose suggestions were not acknowledged by CA is positive but does not have a significant effect on the selection decision for customer cooperation. To test if firms want to cooperate with lead users using the two lead userness variables, a joint test on the coefficients (not marginal effects) of the lead user variables with the following null hypothesis is conducted:

$$H_0 : \begin{bmatrix} lead\,user\ 1 = 0 \\ lead\,user\ 2 = 0 \end{bmatrix}$$

The chi-squared statistic of this joint-test is 11.35 with 2 degrees of freedom and a p-value of 0.003. Thus, the null hypothesis can be rejected at the 1% level of significance. This finding indicates support for hypothesis 3 that lead users have a higher likelihood of being selected for collaboration in the idea generation phase, and corresponds to the lead user literature that propose the importance of integrating lead users in the idea generation phase of the innovation process. As shown in Model 2, it can be seen that the estimated coefficients and the marginal effects of the two interaction terms are positive, suggesting that lead users who are community leaders may have a higher probability of being selected for the PAC program. However, no significant effect can be reported. A joint test on the coefficients of the two interaction terms produces a chi-squared statistic of 0.11 and a p-value of 0.948. Thus, the data provide no evidence that hypothesis 3a is supported for the idea generation phase.

As predicted, the coefficient of *group size* is positive. Specifically, when the number of

[12]Model 1 was re-estimated without the *community leader* variable and *strong ties* is found to be positively significant at the 1% level. Specifically, when the number of strong ties doubles from 1.6 (which is approximately equal to the mean of *ln (strong ties + 1)*) to 3.2, the probability of being selected to be on the PAC increases by 2.3%.

[13]It should be noted that as community leaders are selected by community members based on leadership quality and management skills, it is not expected that this variable introduces an endogeneity problem to the empirical results.

Table 4.7: Probit Regression Models for DevBuds with Heteroscedasticity-Robust Standard Errors in the Presence of Intra-Cluster Correlation

	Model 3		**Model 4**	
	DevBud member (0/1)		DevBud member (0/1)	
	Coefficients	Marginal Effects (dF/dx)	Coefficients	Marginal Effects (dF/dx)
ln (strong ties + 1)	0.093	0.016	0.103	0.018
	[0.081]	[0.014]	[0.082]	[0.014]
community leader (0/1)	0.504***	0.107***	0.613***	0.133***
	[0.149]	[0.037]	[0.156]	[0.041]
lead user-2 (0/1)	0.541*	0.127	0.856**	0.227*
	[0.278]	[0.082]	[0.354]	[0.125]
lead user-1 (0/1)	0.821**	0.220*	1.197***	0.359***
	[0.334]	[0.122]	[0.340]	[0.136]
lead user-2 (0/1) * community leader (0/1)			−0.809	−0.082***
			[0.535]	[0.029]
lead user-1 * community leader			−1.089*	−0.091***
			[0.641]	[0.019]
ln (group size)	0.037	0.006	0.043	0.007
	[0.092]	[0.016]	[0.091]	[0.016]
open communication (0/1)	−0.182	−0.035	−0.188	−0.036
	[0.223]	[0.046]	[0.223]	[0.045]
company size	−0.006	−0.001	−0.003	−0.001
	[0.022]	[0.004]	[0.021]	[0.004]
fairness (0/1)	0.396***	0.070***	0.367***	0.064***
	[0.140]	[0.025]	[0.140]	[0.025]
multiple member (0/1)	0.371**	0.076**	0.373**	0.075**
	[0.166]	[0.038]	[0.161]	[0.037]
feedback (0/1)	−0.070	−0.013	−0.075	−0.013
	[0.143]	[0.026]	[0.142]	[0.026]
female (0/1)	−0.456***	−0.067***	−0.450***	−0.066***
	[0.166]	[0.022]	[0.166]	[0.022]
ln (use experience)	−0.003	−0.001	0.001	0.000
	[0.085]	[0.015]	[0.085]	[0.015]
Constant	−1.579***		−1.653***	
	[0.499]		[0.493]	
Observations	808		808	
Log pseudolikelihood	−269.269		−266.677	

Robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

individuals in the group doubles from 72 (which is approximately equivalent to the mean of *ln (group size)*) to 144, the probability that an individual would be selected increases by 1.5%, or in other words, the percentage of PAC members in the sample would increase from 8% to approximately 9.5%. This effect is significant at the 5% level. Accordingly, it is concluded that hypothesis 4 is supported. Hypothesis 5 proposes that users from user groups with an open culture of communication would have a higher probability of being selected to participate in the PAC program. As can be observed, the coefficient and marginal effect are not significant.

The marginal effect of the control variable, *company size* in Model 1 shows that the size of the company that an individual is working for has a positive significant influence on the probability of being selected to be on the PAC. This finding is in line with CA's selection criteria of PAC members, as CA's field marketing group, sales area managers, customer advocates and other internal resources would tend to nominate more prominent, larger firms as they are often perceived as important customers. Additionally, PAC members tend to be individuals in more than one CA user group, and tend to be males. A joint hypothesis-test for the significance of all 6 control variables in Model 1 generates a chi-squared statistic of 33.72 with 6 degrees of freedom and a p-value of 0. Thus all control variables are jointly significant.

Table 4.7 presents results of the probit estimations for the DevBud program without (Model 3) and with (Model 4) the interaction terms. Unless otherwise stated, the results of Model 3 will be presented in the following. Similar to results from Model 1, the marginal effect of the number of strong ties has no significant effect on the likelihood of being a DevBud.[14] Thus, the data do not support hypothesis 1.

Corresponding with hypothesis 2, community leaders are found to have a higher likelihood of being leveraged for innovation management for the problem solving phase. In particular, it can be observed that community leaders have a 10.7% higher probability of being a DevBud. This marginal effect is significant at the 1% level.

As Model 3 indicates, the marginal effects of both lead user variables are positive but not significant at the 5% level. However, it should be noted that the coefficient of *lead user-1* is significant at the 5%. Similar to that of PAC members in Model 1, a joint-test on the coefficients of the lead user variables is conducted to test for hypothesis 3. The chi-squared statistic of this joint-test is 7.55 with 2 degrees of freedom and a p-value of 0.02. Hence, the null hypothesis can be rejected at the 5% level of significance. Accordingly, it

[14]Model 3 was re-estimated without the *community leader* variable and the *strong ties* variable is found to be positively significant at the 1% level. Specifically, when the number of strong ties doubles from 1.6 (which is approximately equal to the mean of *ln (strong ties + 1)*) to 3.2, the probability of being selected as a DevBud increases by 3.3%.

is concluded that hypothesis 3 is supported for the problem solving phase.

From Model 4, it can be seen that the marginal effects of the two interaction terms of lead users and community leaders are both significantly negative at the 1% level. To calculate the marginal effect of lead users who are community leaders, it is necessary to add the marginal effects of each interaction term with that of the respective lead user variable and that of *community leader*. Thus, individuals who belong to the lead user-2 category and who are community leaders have a 27.8% higher probability of being selected as a DevBud. Similarly, individuals who belong to the lead user-1 category and who are also community leaders have a 40.1% higher probability of being a DevBud. Furthermore, it can be observed that the marginal effects of lead users who are not community leaders are respectively smaller than that of lead users who are also community leaders. A joint test on the coefficients of the two interaction terms produces a chi-squared statistic of 4.65 and a p-value of 0.098. The result of this joint-test suggests that the coefficients of the two interaction terms are not jointly significant at the 5% level. It can thus be concluded that hypothesis 3a is partially supported by the data for the problem solving phase.

Next, it can be observed that *group size* does not have a significant effect on the selection decision of DevBuds. Similarly, the communication culture of a user's user group also does not have an influence on the probability that the user is a DevBud. Hence, the data do not provide evidence that the two group-specific characteristics are supported. Model 3 further indicates that DevBuds tend to feel fairly treated by CA for their participation in user group activities, tend to be members in more than 1 CA user group and tend to be males. A joint hypothesis-test for the significance of all 6 control variables in Model 1 generates a chi-squared statistic of 22.39 with a p-value of 0. Thus all control variables are jointly significant at the 1% level.

Table 4.8 presents estimation results from ordered probit regressions examining the determinants that influence the number of different CA customer collaboration programs an individual is selected for. Table 4.9 presents the marginal effects calculated at the means of the independent variables in Model 5 for each specific parameter value of the dependent variable - *Num Programs*. First, similar to the Probit models in Tables 4.6 and 4.7, hypothesis 1 cannot be supported by the data. Next, it can be seen from the positive significant coefficient (at the 1% level) in Model 5 that community leaders are more often integrated in collaboration programs. Specifically, community leaders have a 15.1% lower probability of *not* being selected for any program; a 12.1% higher probability of being chosen for one program; and a 3.0% higher probability of being selected to participate in two programs. These results provide evidence that hypothesis 2 is supported at the 1% level.

Table 4.8: Coefficients of Ordered Probit Regression Models for the Number of Customer Collaboration Programs Individuals Participate in with Heteroscedasticity-Robust Standard Errors in the Presence of Intra-Cluster Correlation

	Model 5	**Model 6**
	Num Programs	
ln (strong ties + 1)	0.110	0.116*
	[0.067]	[0.068]
community leader (0/1)	0.556***	0.619***
	[0.138]	[0.142]
lead user-2 (0/1)	0.710***	0.880***
	[0.261]	[0.326]
lead user-1 (0/1)	0.725**	0.960***
	[0.285]	[0.276]
lead user-2 * community leader		−0.422
		[0.533]
lead user-1 (0/1) * community leader (0/1)		−0.687
		[0.623]
ln (group size)	0.089	0.094
	[0.065]	[0.064]
open communication (0/1)	−0.107	−0.110
	[0.182]	[0.180]
company size	0.030*	0.032**
	[0.016]	[0.016]
fairness (0/1)	0.359***	0.342***
	[0.127]	[0.127]
multiple member (0/1)	0.463***	0.462***
	[0.141]	[0.140]
feedback (0/1)	0.023	0.023
	[0.130]	[0.130]
female (0/1)	−0.491***	−0.489***
	[0.154]	[0.154]
ln (use experience)	−0.020	−0.018
	[0.076]	[0.076]
Observations	808	808
Log pseudolikelihood	−387.603	−386.542

Robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4.9: Marginal Effects of Independent Variables in Model 5 of the Ordered Probit Regression

	Marginal Effects (dF/dx) on the following parameter values of Num Programs		
Variable	0	1	2
ln (strong ties + 1)	−0.026	0.022	0.004
	[0.016]	[0.013]	[0.003]
community leader (0/1)	−0.151***	0.121***	0.030***
	[0.042]	[0.033]	[0.011]
lead user-2 (0/1)	−0.217**	0.165**	0.052
	[0.098]	[0.066]	[0.034]
lead user-1 (0/1)	−0.225**	0.169**	0.055
	[0.108]	[0.072]	[0.039]
ln (group size)	−0.021	0.018	0.0023
	[0.015]	[0.013]	[0.002]
open culture (0/1)	0.026	−0.022	−0.004
	[0.045]	[0.038]	[0.008]

Standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

† Marginal effect is calculated for a discrete change of the dummy variable from 0 to 1.

Furthermore, individuals who are lead users participate in more programs. In particular, individuals in the *lead user-2* and *lead user-1* category have a 21.7% and a 22.5% lower probability of *not* being selected for any program respectively; and a 16.5% and 16.9% higher probability of being selected for one program respectively. These marginal effects are all significant at the 5% level. A joint hypothesis test of the coefficients of the lead user variables in Model 5 to test for hypothesis 2 produces a chi-squared statistic of 10.5 with 2 degrees of freedom and a p-value of 0. Thus, as the null hypothesis can be rejected, these results indicate that hypothesis 2 is supported by the data. Results for the interaction terms *lead user-1 * community leader* and *lead user-2 * community leader* in Model 6 as well as their estimated marginal effects indicate that hypothesis 3a is not supported. A joint test on the coefficients of the two interaction terms in Model 6 further confirms the non-significance of the two variables.[15] Furthermore, it can also be observed that hypothesis 4 and 5 are not supported by the data. A joint hypothesis-test for the significance of all 6 control variables in Model 5 computes a chi-squared statistic of 36.94 with a p-value of 0. Thus all control variables are jointly significant at the 1% level.

[15]The joint-test produces a chi-squared statistic of 1.57 and a p-value of 0.457.

4.7 Discussion of the Findings

Based on data from members of user groups supported by the proprietary software firm Computer Associates, the aim of this study was to examine the characteristics of customer collaborators from user communities. The following summarises the findings of the hypothesised relationships.

Firstly, empirical results from Table 4.6 and Table 4.7 show that for both the idea generation and problem solving phases of the innovation process hypotheses 2 is supported by the data. This finding is also supported with results from Table 4.8. These findings suggest that community leaders have a higher likelihood and are more often selected for cooperation in innovation projects. A closer examination of the centrality of community leaders shows that approximately 55% of the 654 respondents who are not community leaders indicated that majority of their communication with their user group is with the community leader. Thus, community leaders can be leveraged for their gate keeper role in controlling the flow of information from and to his user community. As community leaders appear to have a larger number of strong ties, they can also be leveraged to acquire credible information from members in their user community.

Consistent with earlier research (Herstatt and von Hippel 1992; von Hippel et al. 1999; Olson and Bakke 2001), it is found that lead users are more often selected for the idea generation phase. Furthermore, results in this study show that lead users also appear to participate in the problem solving phase and are integrated in more collaboration programs as compared to non-lead users. These results are interesting as they suggest that CA was successfully able to identify lead users considering that only 6% of the sample are individuals who exhibited lead user characteristics.

Additionally, it is found that lead users who are also community leaders have a higher probability of being selected for the problem solving but not the idea generation phase. One possible explanation for this finding is that in the problem solving phase, lead users with gate keeping characteristics may be especially valuable because they would be better able to source for information from their user communities to assist CA in the development of solutions. This result is similar to that from Allen et al. (1979) and Tushman and Katz (1980), who found that technological gatekeepers are especially important for development projects but not for research projects.

Findings also indicate that an individual's group size has an influence on the probability of the individual being selected to participate in the idea generation phase but not in the problem solving phase. One reason for this finding could be that users from larger groups would be able to provide feedback on existing products and ideas for new products representative of a much larger number of customers as compared to users from smaller groups. Such information is especially important for the idea generation phase when a design concept or technical proposal is developed by integrating market with technical information, as compared to the problem solving phase where the main aim is to develop a technical solution to the proposal developed in the idea generation phase.

The findings of this study suggest several management implications. First, firms must provide community leaders with appropriate incentives for managing their communities, as this would lay the foundation for a successful cooperation should the community leaders be selected as user collaborators. This is important as community leaders are gate keepers and firms want to be able to leverage them optimally to acquire the large pool of customer knowledge in their user community. Furthermore, it is also vital that community leaders treat sensitive information which they receive during innovation collaboration activities discretely and not broadcast it to the rest of their community. Next, management should also create appropriate incentives for lead users in user communities to cooperate with them. Special attention should be paid to lead users who are also community leaders when they are to be leveraged for innovation projects for problem solving. Finally, firms should make attempts to participate in activities of large user communities. This would better enable the firm to select active customers for idea generation, who would be in a position to filter the large pool of community generated knowledge to the firm.

The above-mentioned implications have been based on the selection of user community collaborators of a proprietary software company. It should be noted that the nature of software development makes it easy to integrate customers into the various phases of the innovation process. Further studies of a similar nature should be conducted in other industries to test if these implications can be generalised.

One limitation of the study is that it is not known how these individual- and group-specific characteristics influence the actual collaboration between customers and the firm and the extent to which they contribute to new product success. Further studies could also examine the influence of these characteristics on the implementation and diffusion phase, which was not analysed in this study.

A further drawback of this study is that it does not explicitly address any potential endogeneity problems between the dependent and independent variables. For instance individuals who want to be selected as user collaborators may be highly motivated to have a greater number of strong ties and/or be willing to volunteer as community leaders. Further research should attempt to identify possible instrumental variables for these independent variables and to apply appropriate estimation methods such as the two-stage least squares method.

Chapter 5

Conclusion

5.1 Summary

This dissertation consists of three empirical investigations that provide new insights as to how groups of users can be organised and leveraged by firms for their innovation processes. By drawing on various theories in economics, business, communication studies, sociology and psychology, a theoretical framework for the analysis was developed. This framework is structured into two parts. The first part deals with the give-and-take behaviour of individuals in user communities. In particular, it identifies factors that affect individuals' decisions to share information with each other (give) and factors that influence the amount of benefits individuals are able to receive from their user communities (take). The second part of the framework addresses the issue of how user communities can be leveraged for innovation management. Specifically, it analyses the extent that attributes of the individual member and that of his community play a role in the selection of community members for integration in the various phases of the innovation process of the firm.

This theoretical framework is then empirically investigated using a unique data set obtained from a web-based survey on individual users in user groups from the proprietary software firm Computer Associates. A series of econometric models are applied to the data in the empirical analyses; these include the Tobit model, the Probit model, the Ordered Probit model and the interval regression model. As the analyses are based on responses from individuals in different user groups, there is a strong possibility that responses from individuals within the same user group would be correlated, possibly through unobserved cluster effects (Wooldridge 2003). Analyses that assume independence of the observations will tend to underestimate the true variance, which would lead to test statistics with inflated Type I errors (Williams 2000). To avoid this over-dispersion problem, unbiased variance estimators are computed to control for intra-cluster correlation for all econometric models in this dissertation.

Figure 5.1: Overview of Empirical Results

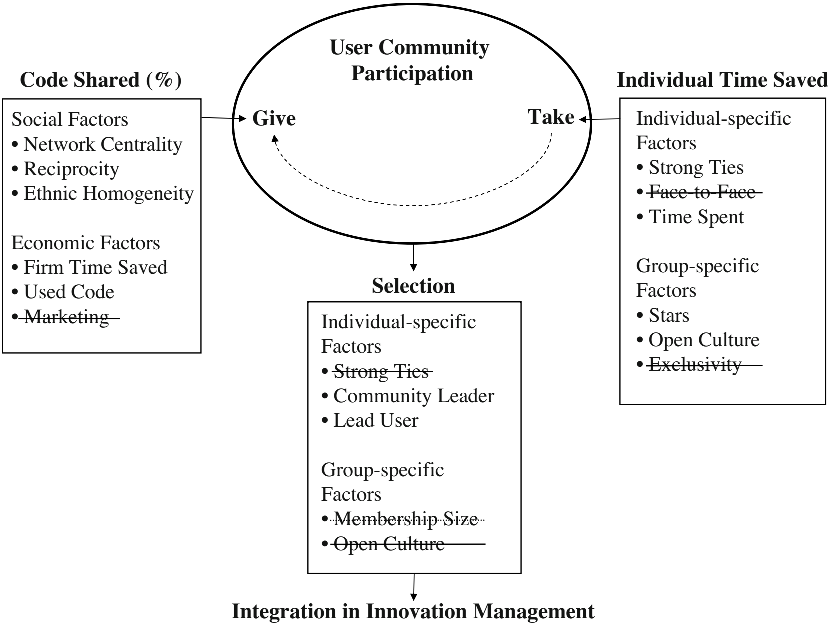


Figure 5.1 summarises the main findings from chapters 2 to 4.

The first empirical study (chapter 2) addressed the issue of information sharing in user communities. In particular, it analysed how socio-economic factors affect the information sharing behaviour of individuals in user communities. The following factors were hypothesised to be determinants of information sharing - network centrality, the reciprocity norm, ethnic homogeneity, the amount of time the user's firm was able to save, if the individual had utilised code from other members in the community, and if the user uses his user community as a marketing tool for his firm. Following Henkel (2006), information shared was quantified using the percentage of code written by individuals that they shared with other community members. To take into account the natural boundaries of the percentage of code revealed [0-100%], a Tobit model was estimated. An Ordered Probit model and an interval regression model were also estimated to check the robustness of the results from the Tobit model. Empirical results from this study were based on individual responses from 519 software code writers in 159 different Computer Associates' user groups.

The second empirical study (chapter 3) analysed the extent of how individual- (number of strong ties, amount of time spent attending community activities, and face-to-face communication) and group-specific (number of star members in each group, communication culture, and membership restrictions of the group) factors influence the benefits users receive from their user communities. To quantify users' benefits, the amount of time individual members save through having access to their user communities was used. Approximately a fifth of the respondents indicated the number of hours they saved as an interval. The interval regression model was identified as being an appropriate estimation method to model the effects of these individual- and group-specific factors on users' benefits, as this method specifically takes into consideration the uncertainty concerning the nature of the exact values within each of these intervals. This empirical analysis made use of survey data from 924 individuals in 161 different Computer Associates' user groups.

From the empirical results in chapters 2 and 3, the following summarises the antecedents of the give-and-take behaviour of individuals in user communities. Firstly, findings from chapter 2 suggest that both social and economic factors have an influence on the information sharing (give) decision of individuals. Specifically, the economic decision of sharing information is embedded structurally (higher network centrality), relationally (reciprocity norm) and cognitively (higher racial homogeneity). Furthermore, pecuniary benefits (more firm time saved and the usage of code from other members) also motivate individuals to share information with their community. Secondly, results from chapter 3 suggest that a greater number of strong ties, the more time members spend participating in user group activities, the greater number of experienced star members in the user community, an open culture of communication as well as less restrictive membership requirements all have a positive influence on the amount of time individuals can save (take) from having access to their user communities.

Based on the above-mentioned results, the following implications for the internal organisation of user communities - that motivate individuals to voluntarily contribute to and participate in their communities - can be derived. In general, user communities should be organised in such a way that they provide a conducive environment for the exchange of information amongst members. Firstly, an important finding is that social factors are critical for both the contribution to and participation in user communities. To promote sociability between members, community leaders can organise social events or functions for members to interact with each other. Such social bonding activities not only allow individuals to establish more contacts (higher network centrality and number of strong ties) within their community, but also promote social norms such as reciprocity and trust. Additionally, the arrangement of such social activities before formal meetings would create a more relaxed ambience during the meetings themselves, which could in turn have a positive influence on the willingness of members to share information. Furthermore, community leaders should also ensure that members have shared collective understandings (racial homogeneity) as this would create a good foundation for individuals to want to contribute to the community.

Next, community leaders should organise activities that are of high informational quality with respect to the novelty and relevance of the information provided and exchanged. This would not only motivate members to participate more (spend more time) in their communities but also to share more information with others as they would be able to receive more pecuniary benefits (save more time for their firm) from their participation. This is an important point as firms may only permit their employees to participate in community activities provided the firms are able to profit from their participation. Thirdly, it is advisable that incentive mechanisms are available for highly experienced and knowledgable users (stars) to join and participate in user group meetings, as the presence of such individuals increase the attractiveness of community activities and create an incentive for other members to participate. Finally, a further implication is that user communities be more open in general. Specifically, community leaders should not only open membership to the general public and have less restrictive membership structures (be less exclusive), but also promote an open culture of communication; as openness is suggested to have a positive influence on the benefits individuals are able to receive from their user communities.

Additionally, this dissertation also explored how firms can leverage their user communities for innovation management. In chapter 4, it was proposed that firms can use user communities to select user community collaborators that would enable them to effectively acquire the large pool of customer knowledge in user communities. Specifically, it was investigated if the following individual and group attributes have an effect on the decision of a firm to select user community members for collaboration projects - the number of strong ties of the individual, if an individual is a community leader, if an individual is a lead user, the membership size of the group and the communication culture of the group. To test the hypotheses, two different dependent variables were used. A Probit estimation method was applied when the dependent variable was a dichotomous (1/0) variable which depicted if a community member participated in a particular customer integration program or not; and an Ordered Probit model was estimated when the dependent variable was a count response - that is the number of different customer collaboration programs an individual was selected for by CA. A total of 808 observations were used from individuals in 161 different CA user groups.

In general, individuals who are community leaders and lead users have a higher likelihood of being selected for integration in the innovation process.[1] These results suggest the following implications for the external organisation of user communities, that is how firms can interact with their user community and leverage them for innovation management. Firstly, firms should have a healthy symbiotic relationship with their community leaders, as these individuals have access to customer feedback and suggestions that may be very valuable for the innovation process. For instance, firms could give community leaders products for free or at a discounted rate, or officially acknowledge their contributions by presenting them with plaques and certificates. Secondly, firms should create appropriate incentives for lead users in user communities to cooperate with them. Possible incentives may include giving lead users early access to alpha and beta versions of the product to experiment and tinker around with, or in allowing lead users direct access to technical personnel, who may be able to assist them in their endeavours.

To summarise, this dissertation has firstly explored how groups of individuals in the form of user communities can be organised and self-managed by the users themselves. An important finding is that social relations play a vital role for the give-and-take behaviour of individuals in the community, and hence for the existence and sustenance of the community. Secondly, it has demonstrated how firms can leverage these user communities to acquire customer input into the various stages of the innovation process. In particular, individuals who are community leaders and lead users tend to be selected for integration. On the whole, it can be drawn from the empirical results of chapters 2 to 4 that the theoretical framework developed for this study (as illustrated in figure 5.1) is largely supported.

5.2 Limitations and Future Research

As this dissertation empirically analysed user communities in the software industry, it would be premature to generally apply the implications of this study for user communities in other industries. In particular, more industry-specific information is needed before these implications can be directly applied to other industries. For instance, in the case of sports communities (Franke and Shah 2003) or brand communities (Muniz and O'Guinn 2001) which are often categorised as "hobbyist" communities, it would be necessary to firstly perform explorative work to identify and assess the importance of pecuniary benefits members are able to obtain from their communities. However, as the theoretical framework developed in this dissertation is not specific to the software industry, it is suggested that the framework can be applied to analyse any user community where users

[1] Individuals in larger user groups have a higher probability of being selected for the idea generation phase of the innovation process, but not for the problem solving phase; and not for being selected for a larger number of collaboration programs.

voluntarily group together to learn from each other.

One drawback of this study is that endogenous concerns between dependent and independent variables were not explicitly taken into consideration in chapters 2 to 4. In particular, there appears to be causality issues between "network" variables such as the number of strong ties or the centrality of individuals and certain dependent variables in chapters 2 to 4. An important contribution of future research would be to address these endogeneity issues by providing a deeper understanding into how changes in the network occur as a consequence of motivated actors driving those changes. Specifically, it is suggested that this endogeneity is modeled with appropriate econometric methods so as to be able to obtain more accurate estimates of various network effects on actors' give-and-take behaviour and the selection decision of firms for user collaborators.

A further limitation of this dissertation is the negligence of a potential endogeneity between the give-and-take behaviour of members. To address this issue, one possibility would be to perform a netnography analysis[2] of an online user community, whereby the posts of every member in the community are systematically sorted according to information seeking posts and information giving posts. The advantage of such an analysis is that it allows for the examination of the give-and-take behaviour over time, as one would be able to construct a panel data set using the dates of the posts. As an important factor for causal inference is that the cause has to precede the effect in time (Cook and Campbell 1979), a panel data set would be ideal to examine if there is a two-way causality between the give-and-take behaviour of individuals. Interesting research questions include how free-riders can be converted to contributors and if contributors (due to their higher visibility or reputation) benefit more as their information seeking requests may have a higher probability of being answered by other members.

Although chapter 4 analyses how firms can harness the innovative potential of user communities for innovation management, the theoretical framework for the external management of user communities by firms is not entirely satisfying. Further work needs to be done to explicitly examine how firms should react and relate to their user communities. This is important as user communities are similar to pressure or interest groups which may not always share the same interests as that of the firm. For instance, outlaw communities are examples of user communities whose members create innovations that aim at bypassing legal or technical safeguards that prevent users from unsolicited usage of the manufacturer's products (Mollick 2004; Flowers 2008; Schulz and Wagner 2008). Specifically, more work should be done on how the needs of the user communities can be aligned with that of firms.

[2]Netnography is the systematic analysis of information in online forums (Kozinets 2002).

In addition, future research should suggest a quantifiable measure of the benefits firms can obtain by leveraging user communities for innovation management. It would be especially interesting to investigate the determinants of individual- and group-specific characteristics on the benefits firms are able to obtain from user community collaborators. In combination with empirical results of this dissertation, this information would provide firms with an optimal framework for managing user communities. It is advisable to take into consideration the life-cycle phase of the user community as well as the maturity of the firm's product/s in the analysis.

In conclusion, this study has shown that the phenomenon of user communities presents major opportunities for both individuals who are members in them as well as firms who intend to profit from them. In particular, there often exists an information asymmetry between firms and their (potential) customers which would result in inefficient outcomes (Picot et al. 2003). User communities have the potential to make markets more efficient as the information provided to transaction participants improves - firms benefit from the good information on the perceptions and preferences of potential buyers; and community members (potential buyers) benefit from good information on the characteristics of the various offerings in the market (von Hippel 2005). However, more research is still needed to understand more about how user communities can be optimally organised for innovation management.

Appendix A

Due to the wide variety of software products CA offers, it is not possible to provide detailed descriptions of each product. However, a brief introduction of CA's six main product categories will be presented in the following[3]:

- **Operations Management**: CA's Unicenter operations management solutions provide network, systems, database, applications, and web infrastructure management. These solutions map IT infrastructure to business processes and automatically detect, diagnose, repair and recover complex problems across the entire technology stack supporting critical business applications and services. Additionally, they also streamline the delivery of critical software updates and patches to internal customers; as well as automate and optimize jobs and business workloads.

- **Storage Management**: BrightStor intelligent storage management solutions enable the proactive simplification, secure management and protection of information and storage assets.

- **Data Management and Application Development**: CA's Advantage[4] software provide solutions for application development and generation, application code validation, testing and fault management; enterprise database and data management; and data access, transformation and transportation, for both the mainframe and distributed environment.

- **Security Management**: The eTrust security management solutions manages companies' entire security environment by integrating three key components - identity and access management, threat management and security information management.

- **Life Cycle Management**: AllFusion life cycle management solutions empowers companies to centrally store, administer, secure, version, and track the entire software inventory of their enterprise. These solutions provide for the integration and synchronisation of all software running on any platform, including mainframe, distributed, and web applications.

- **Service Management**: The CleverPath solutions from CA enable businesses to centralise information access; refine, analyse and sort data; create and distribute informative reports; and develop decision support systems that mirror their business processes.

[3]See CA's home page for detailed information on each product category (http://www3.ca.com/products/, accessed on 19th December 2004 and 22nd August 2005).

[4]CA has for marketing purposes, recently shifted a number of application development software from the Advantage suite to the Allfusion suite.

Appendix B
Questionnaire

Survey on CA User Groups

A. Membership in CA user groups

A.1. Of how many CA user groups are you **currently** a **member**?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 or more

A.2 What are the **names** of these CA user groups?

The first user group you joined ______________________________

The second user group you joined ______________________________

The third user group you joined ______________________________

Other user groups ______________________________

____

____

A.3 Is this your **first** time participating in this survey?

☐ Yes → **proceed to A.3a**
☐ No → **proceed to A.3b**

*Please answer **A.3a** if your answer to **A.3** was **yes**, otherwise please proceed to **A.3b**.*

A.3a The questions in this survey are user group specific and deal mostly with **informational exchanges** regarding your CA software.

Please indicate the **name of the CA user group** with which you **most frequently** exchanged information regarding your CA software in the **last year**. Please note that your **answers for the rest of this survey** will be **based on the CA user group you indicate here.**

A.3b The questions in this survey are user group specific. **Please indicate the name of another** of your CA user groups that you would like to base your answers on for the rest of this survey.

A.4 **In which year** did you join this CA user group that you would like to base your answers on for the rest of this survey? (*If this CA user group is a result of a merger between two or more user groups, please enter the year you joined your original user group.*)

In all the following questions, the term "this CA user group" will always refer to the above named CA user group that you chose to base your answers on.

B. CA software

B.1 Please indicate **the most important** CA software that you **frequently use** at your present employment that is **supported by this CA user group**.

B.2 **In which year** did you start using the above-mentioned CA software?

B.3 Please indicate **in order of importance 2 other** CA software, which you frequently use that are supported by this CA user group.

B.4 Did you or your company **upgrade** any of the **above mentioned CA software** in the last year?

☐ Yes ☐ No

C. Formal user group activities

Formal user group activities refer to **physical, web, video and telephone meetings** as well as **forums of user groups**.

C.1 Please estimate the **number of days** you **spent** in the last year **attending or participating** in all of the above-mentioned activities of **this CA user group**.

☐ None → **proceed to Section D**
☐ Less than 1 day
☐ 1 day
☐ 2 days
☐ 3-4 days
☐ 5-6 days
☐ 7-10 days
☐ 11-14 days
☐ 15-20 days
☐ 21 days or more

Please answer questions C.2 and C.3 if you participated in any of the above-mentioned formal user groups activities of this CA user group in the last year.
*If you answered "**None**" to C.1, please proceed to **Section D**.*

C.2 Please rank the following formal user group activities **in descending order** according to the **amount of time**, you **spent** in the last year participating in or attending. *(Please **rank** only activities **you use**.)*
Please number each of the following from 1 (most amount of time spent) to 3 (least amount of time spent).

☐ Physical meetings
☐ Web, video and/or telephone meetings
☐ This user group's forum

C.3 To what extent were the following **types of information** you received from the above-mentioned activities in the last year **relevant or irrelevant** for your work? (*Relevant refers to how applicable and useful the information was for your work.*)

	extremely relevant		neither relevant nor irrelevant		extremely irrelevant	I didn't receive this
	1	2	3	4	5	
Updates, news and/or talks from CA	☐	☐	☐	☐	☐	☐
Advice to **solve my difficulties** with the CA software	☐	☐	☐	☐	☐	☐
Tips on how I can **optimise the usage** of the CA software	☐	☐	☐	☐	☐	☐
Information on **related products or services** to the CA software	☐	☐	☐	☐	☐	☐
Information on **potential problems** that could arise with the CA software	☐	☐	☐	☐	☐	☐

D. Close contacts in this CA user group

Close contacts are individuals with whom you **communicate frequently**, whom you **can approach at any time for help** and whom you **would willingly help at any time**.

D.1 Excluding your colleagues, how many **close contacts** do you have **in this CA user group**?

Please answer question D.2 if your answer to D.1 was not zero, otherwise proceed to Section D1.

D.2 To what extent do you agree or disagree with the following statements on your **one-to-one communication** with your **close contacts** (excluding your colleagues) in the last year?

	strongly agree		neither agree nor disagree		strongly disagree	I don't know
	1	2	3	4	5	
I **frequently approached** these close contacts **for help** last year.	☐	☐	☐	☐	☐	☐
The **information** I **received from** these **close contacts** in the last year were **extremely relevant in solving my difficulties.**	☐	☐	☐	☐	☐	☐
I **frequently helped** these close contacts last year.	☐	☐	☐	☐	☐	☐

D1. Close contacts outside this CA user group

D1.1 **How many close contacts** do you have **in other CA user groups** that support the same CA software as this one?

E. Software code

E.1 Do you **use** a CA **development tool** that is supported by this CA user group? (*Development tools assist in the creation of new software (i.e. using programming languages like Java, COBOL, Pascal, etc.).*)

☐ Yes → **proceed to E.2**
☐ No

Please proceed to ***E.1.1*** *if your answer to E.1 was* ***No****, otherwise proceed to E.2.*

E.1.1 Have you written **source code for add-ons, utilities, UDPs, exits or applications** for the CA software supported by this CA user group?

☐ Yes → **proceed to E.2**
☐ No

Please proceed to ***E.1.2*** *if your answer to E.1.1 was* ***No****, otherwise proceed to E.2.*

E.1.2 Did you **use** in the **last year** source code of add-ons, utilities, UDPs, exits or applications written by other members in this CA user group?

☐ Yes
☐ No

Please answer questions E.2-E.5 if your answer to E.1 or E.1.1 was ***Yes****, otherwise proceed to* ***Section F****.*

E.2 The following questions relate to the source code that is written by yourself or other members with the **CA development tool** or for **add-ons, utilities, UDPs, exits or applications** for the CA software supported by this CA user group.

To what extent did you willingly or unwillingly **reveal information** in the **last year**, on the **source code you wrote to members** in this CA user group?

	extremely willing		neither willing nor unwilling		extremely unwilling	I don't know
	1	2	3	4	5	
The code itself	☐	☐	☐	☐	☐	☐
Information on how I wrote the code	☐	☐	☐	☐	☐	☐

E.3 What **percentage of your source code** that **you wrote** did you **reveal** in the **last year** to other members in this CA user group? *(Please indicate the approximate percentage you reveal.)*

E.4 To what extent do you agree or disagree with the following statements about the **source code** that was **written** by yourself or other members? *("New to the world" refers to the uniqueness of your code.)*

	strongly agree		neither agree nor disagree		strongly disagree	I don't know
	1	2	3	4	5	
I revealed in the **last year** mostly **generic code** that could be utilised by any user that uses the same CA software.	☐	☐	☐	☐	☐	☐
I **didn't think** my code was of a **high quality**.	☐	☐	☐	☐	☐	☐
I **used** code in the **last year** from other members in this CA user group.	☐	☐	☐	☐	☐	☐
I received **assistance from CA** to write my code.	☐	☐	☐	☐	☐	☐
Majority of my code were **new to the world** at the time I wrote them.	☐	☐	☐	☐	☐	☐

E.5 Are you **allowed** by your company to **reveal code** that you wrote to members in this CA user group?

☐ Yes
☐ No
☐ I don't know

F. Your assessment of this CA user group

We are trying to estimate **the number of hours you and your company saved** through your access to this CA user group.

Please take some time to think of **all the information, help and advice** you have received **from and through this CA user group** over the course of the **last year**.

F.1 Please estimate: **how many hours** of your **personal work time** did you save by making use of this information, help and advice in the last year?
(Do not take into account time spent organising, attending and/or participating in activities of this CA user group. If you are uncertain of the approximate number of hours saved, you may also indicate a range. For instance: X-Y)

F.2 Please estimate: **how many hours** of work time did **other employees in your company** save by making use of this information, help and advice in the last year?
(Do not take into account time spent organising, attending and/or participating in activities of this CA user group. If you are uncertain of the approximate number of hours saved, you may also indicate a range. For instance: X-Y)

G. Reasons for joining and participating in this CA user group

G.1 To what extent do you agree or disagree with the following statements on **why you joined** this CA user group?

	strongly agree		neither agree nor disagree		strongly disagree	I don't know
	1	2	3	4	5	
For the opportunity to **learn from experienced users**.	☐	☐	☐	☐	☐	☐
Because it **increases the value** of the complementary product and/or service I sell.	☐	☐	☐	☐	☐	☐
Because it **increases the opportunity of selling** my complementary product and/or service.	☐	☐	☐	☐	☐	☐
For the "**power in numbers**" approach in getting my feedback acknowledged by CA.	☐	☐	☐	☐	☐	☐

G.2 To what extent do you agree or disagree with the following statement on the **openness of members** in this CA user group? *(An open exchange of information refers to the willingness of members to share and exchange information with one another.)*

	strongly agree		neither agree nor disagree		strongly disagree	I don't know
	1	2	3	4	5	
This CA user group has been **extremely successful** in promoting an **open exchange of information** between it's members.	☐	☐	☐	☐	☐	☐

G.3 To what extent do you agree or disagree with the following statements on **your participation** in this CA user group?

	strongly agree		neither agree nor disagree		strongly disagree	I don't know
	1	2	3	4	5	
I help other members because I **received** help from this CA user group in the past.	☐	☐	☐	☐	☐	☐
I help other members because I **anticipate** needing help from this CA user group in the future.	☐	☐	☐	☐	☐	☐
I **reveal** information on how my company utilises the CA software because I **benefit** from the resulting **discussions**.	☐	☐	☐	☐	☐	☐
I contribute to this CA user group because I benefit from the **gain in reputation**.	☐	☐	☐	☐	☐	☐
I participate because it is **fun**.	☐	☐	☐	☐	☐	☐
CA **treats me fairly** by **acknowledging** my involvement in this CA user group.	☐	☐	☐	☐	☐	☐

G.4 **How many** suggestions (DARs) for the **Suggestion Box** for the CA software supported by this CA user group **did you make** in **2004**?

Please answer G.4.1 if you submitted any suggestions (DARS) in 2004.

G.4.1 Of the suggestions submitted, **how many** were **forwarded by CA for development** into new releases or **slated as** "in the process for a future release"?

H. Position in this CA user group

H.1 What is your **position** in this CA user group?

- ☐ President, co-president or vice-president → **proceed to H.2**
- ☐ Other committee member → **proceed to H.1a**
- ☐ User group member → **proceed to H.1a**

Please answer H.1a if you are a ***committee member*** *other than the president or vice-president, or a* ***user group member****.*

H.1a To what extent do you agree or disagree with the following statement?

	strongly agree		neither agree nor disagree		strongly disagree	I don't know
	1	2	3	4	5	
Majority of my communication with this CA user group was with the **president, co-president and/or vice-president**.	☐	☐	☐	☐	☐	☐
I am **extremely satisfied** with the **leadership performance** of my president, co-president and/or vice-president.	☐	☐	☐	☐	☐	☐

Please proceed to Section H1 if you are a ***committee member*** *other than the president or vice-president of this CA user group. If you are a* ***user group member****, please proceed to Section* ***I****.*

Please answer H.2 if you are the president, co-president or vice-president of this user group.

H.2 The following statements relate to **leadership behaviour**. Using the following scale, decide the extent to which these statements **apply to you**. *(For best results, please answer as truthfully as possible.*

	never		sometimes		always
	1	2	3	4	5
I **enjoy managing** this CA user group.	☐	☐	☐	☐	☐
I **encourage** members in this CA user group to **participate** when it comes to **decision-making time** and I try to **implement** their ideas and suggestions.	☐	☐	☐	☐	☐
It is **extremely important** that this CA user group **achieves** its aims and goals.	☐	☐	☐	☐	☐
I am **willing** to **invest considerable time to organise** activities for this CA user group.	☐	☐	☐	☐	☐
I **enjoy advising** members in this CA user group on how they can **better optimise** their CA software.	☐	☐	☐	☐	☐
It is **extremely important** that there is **rapport** between members in this CA user group.	☐	☐	☐	☐	☐

*Please proceed to Section H1 if you are the **president, co-president or vice-president** or **a committee member** of this CA user group. If you are a **user group member**, please proceed to Section I.*

H1. General characteristics of this CA user group

H1.1 **Who** can become a member of this CA user group?

- ☐ Anyone interested, i.e. the general public
- ☐ Users who have a license of the CA software
- ☐ Users who pay a membership fee
- ☐ CA partners
- ☐ Other ____

H1.2a Please estimate the **total number of individual persons** who are members in this CA user group.

____

H1.2b Please estimate the **total number of companies** that are members in this CA user group.

____

H1.2c Please estimate the **number of active individuals** in this CA user group (include yourself if you are active). *(Active individuals are very involved in the user group - be it in the form of* ***attending or presenting*** *at user group meetings and/or* ***participating*** *in this user group's forum.)*

____

H1.2d Please estimate the **number of individual persons** in this CA user group who work for **consultants, contractors, or for software or hardware companies**. *(How many members work in companies that sell a complementary product or service to the CA software?)*

____

H1.3 **How many members** are on the **management committee** of this CA user group?

____

H1.4 Who **initiated** the **founding** of this CA user group?

- ☐ CA
- ☐ Users
- ☐ Other ____

H1.5 **In which year** was this CA user group **founded**?

____

H1.6 Is this CA user group a result of a **merger** between two or more user groups?

☐ Yes → **proceed to H1.6a**
☐ No
☐ I don't know

Please answer H1.6a if you answered ***yes*** *to H1.6, otherwise proceed to H1.7.*

H1.6a In which year was your **original user group founded** (i.e. before the merger occurred)?

H1.7 Please estimate the number of **hours** you **spent organising** activities of this CA user group in the last year? *(If you are uncertain of the approximate number of hours, you may also indicate a range. For instance: X-Y)*

H1.8 Are the members in this CA user group provided with a **contact list** of all other members?

☐ Yes
☐ No

I. Other members in this CA user group

I.1 **How many** of your **work colleagues** are also members in this CA user group?

I.2 Please estimate the number of **highly qualified individuals** in this CA user group of whom you have learnt a lot from. (*Highly qualified individuals are experienced and skilled users whom have a high knowledge of the CA software.*)

I.3 Are there **individual persons** in this CA user group who work for **competitors of your company**?

☐ Yes
☐ No → **proceed to Section J**
☐ I don't know → **proceed to Section J**

Please proceed to ***Section I1*** *if you answered* ***yes*** *to I.3, otherwise proceed to Section J.*

I1. Competitors in this CA user group

I1.1 Please **estimate the number of individual persons in this CA user group** who work for **competitors** of your company.

I1.2 Please **estimate the number of different companies** that these individuals come from.

I1.3 To what extent do you agree or disagree with the following statement?

	strongly agree		neither agree nor disagree		strongly disagree	I don't know
	1	2	3	4	5	
I have **exchanged information** regarding the use of my CA software with those members who work for **competitors** of my company.	☐	☐	☐	☐	☐	☐

J. User specific information

J.1 In which **country** do you currently **work and live**?

____

J.1.1 In which **state** do you currently **work**?

____

J.1.a In which **city** do you currently **work**?

____

J.2 In which **industry** do you currently work?

- ☐ Automobile
- ☐ Aviation
- ☐ Financial
- ☐ Government or non-profit organisations
- ☐ IT (software, hardware or consulting/contracting services) → **proceed to J.2a**
- ☐ Semiconductor
- ☐ Telecommunication
- ☐ Other ____

*Please answer J.2a if you answered **IT** to J.2, otherwise proceed to J.3.*

J.2a Are you currently an employee of **Computer Associates**?

- ☐ Yes
- ☐ No

J.3 What is your **age**?

____

J.4 What is your **gender**?

☐ Female
☐ Male

J.5 What is your highest level of **education**?

☐ High school diploma or lower
☐ College diploma or equivalent
☐ Bachelor degree or equivalent
☐ Masters degree or equivalent
☐ Doctorate or higher

J.6 What was your annual **total net (after tax) income in US$** from all sources in 2004? Please select one of the following broad categories.

☐ US$ 20,000 or less
☐ US$ 20,001 to 40,000
☐ US$ 40,001 to 60,000
☐ US$ 60,001 to 80,000
☐ US$ 80,001 to 100,000
☐ US$ 100,001 to 150,000
☐ US$ 150,001 to 200,000
☐ US$ 200,000 or more
☐ No answer

J.7 Which of the following **jobs** best describes your current employment?

☐ Administrator
☐ Consultant
☐ Developer
☐ Manager
☐ Other ____

J.8 Do you have your **own business** that sells a complementary product or service to the CA software supported by this CA user group?

☐ Yes → **proceed to J.8a**
☐ No
☐ I don't know

*Please answer **J.8a** if you answered **Yes** to J.8, otherwise proceed to J.9.*

J.8a In which **year** was your business founded?

J.9 How many **people** work for **your company**?

J.10 Please answer the following questions.

	Yes	No	I don't know
Does your company have access to **maintenance support** from **CA**?	☐	☐	☐
Does your company have access to **maintenance support** from an IT **contractor or consultant**?	☐	☐	☐
Are you a member of the **Product Advisory Council** for the CA software supported by this CA user group?	☐	☐	☐
Have you ever taken part in the **Buddy Development Program** for the CA software supported by this CA user group?	☐	☐	☐

J.11 To what extent do you agree or disagree with the following statement?

	strongly agree		neither agree nor disagree		strongly disagree	I don't know
	1	2	3	4	5	
I am **encouraged** by my company **to participate** in activities of this CA user group.	☐	☐	☐	☐	☐	☐

K. Your contact details

K.1 Would you like to receive a summary report of the survey?

☐ Yes
☐ No

K.2 You will automatically be entered into the INNO-tec Lottery drawing of the official Munich Oktoberfest beer mug. Please enter your email address so we will know where to contact you.

____

K.2 Do you have any comments to the survey?

____
____
____
____
____
____
____

Bibliography

Adams, S. (1976). The Structure and Dynamics of Behavior in Organizational Boundary Roles. In Dunette, M., editor, *Handbook of Organizational and Industrial Psychology*, pages 1175–1199. Rand McNally, Chicago.

Allen, T. J. (1977). *Managing the Flow of Technology.* MIT Press, Cambridge, third edition.

Allen, T. J. and Cohen, S. I. (1969). Information Flow in Research and Development Laboratories. *Administrative Science Quarterly*, 14(1):12–19.

Allen, T. J., Tushman, M. L., and Lee, D. M. S. (1979). Technology Transfer as a Function of Position in the Spectrum from Research through Development to Technical Services. *Academy of Management Journal*, 22(4):694–708.

Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics*, 24:3–61.

Armstrong, J. S. and Overton, T. S. (1977). Estimating nonresponse bias in mail surveys. *Journal of Marketing Research*, 14:396–402.

Balconi, M., Breschi, S., and Lissoni, F. (2004). Networks of inventors and the role of academia: an exploration of italian patent data. *Research Policy*, 33:127–145.

Bandiera, O., Barankay, I., and Rasul, I. (2005). Cooperation in collective action. *The Economics of Transition*, 13(3):473–498.

Banerjee, A., Iyer, L., and Somanathan, R. (2004). History, Social Divisions and Public Goods in Rural India. unpublished manuscript, Massachusetts Institute of Technology.

Bitzer, J., Schrettl, W., and Schröder, P. J. H. (2004). Intrinsic Motivation in Open Source Software Development. Nr. 2004/19, Diskussionsbeiträge des Fachbereichs Wissenschaft der Freien Universität Berlin.

Blau, P. M. and Scott, W. R. (1962). *Formal Organisations: A comparative Approach.* Chandler, San Francisco.

Boland, R. J. and Tenkasi, J. R. V. (1995). Perspective Making and Perspective Taking in Communities of Knowing. *Organization Science*, 6(4):350–372.

Bouty, I. (2000). Interpersonal and Interation Influences on Informal Resource Exchanges between R&D Researchers across Organizational Boundaries. *Academy of Management Journal*, 43(1):50–65.

Brügge, B., Harhoff, D., Picot, A., Creighton, O., Fiedler, M., and Henkel, J. (2004). *Open Source Software – eine ökonomische und technische Analyse*. Springer, Berlin.

Brown, S. and Eisenhardt, K. (1995). Product Development: Past Research, Present Findings, and Future Directions. *Academy of Management Review*, 20(2):343–378.

Buchanan, J. M. (1965). An Economic Theory of Clubs. *Economica*, 32(125):1–14.

Burt, R. (1992). The Social Structure of Competition. In Nohria, N. and Eccles, R. G., editors, *Networks and Organizations: Structure, Form, and Action*. Harvard Business School Press, Boston.

Campbell-Kelly, M. (1995). Development and Structure of the International Software Industry, 1950-1990. *Business and Economic History*, 24(2):73–110.

Campbell-Kelly, M. (2000). *From Airline Reservations to Sonic the Hedgehog. A History of the Software Industry*. MIT Press, Cambridge.

Carpenter, D., Esterling, K., and Lazer, D. (2003). The Strength of Strong Ties. A Model of Contact-Making in Policy Networks with evidence from U.S. Health Politics. *Rationality and Society*, 15(4):411–440.

Carter, A. P. (1989). Knowhow trading as economic exchange. *Research Policy*, 18(3):155–163.

Chakravorti, B. (2004). The New Rules for Bringing Innovations to Market. *Harvard Business Review*, 82(3):59–67.

Chesbrough, H. W. (2003). The Era of Open Innovation. *MIT Sloan Management Review*, 44(3):35–41.

Chiu, C.-M., Hsub, M.-H., and Wang, E. T. (2006). Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision Support Systems*, 42(3):1872–1888.

Cicourel, A. V. (1973). *Cognitive Sociology*. Penguin Books, England.

Cohen, W. M. and Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1):128–152.

Cohendet, P. (2006). Knowing Communities in Organizations. In Kahin, B. and Foray, D., editors, *Advancing Knowledge and The Knowledge Economy*, pages 91–109. MIT Press, Boston.

Coleman, J. S. (1990). *Foundations of Social Theory*. Belknap Press, Cambridge, MA.

Cook, T. D. and Campbell, D. T. (1979). *Quasi-Experimentation Design & Analysis Issues for Field Settings*. Houghton Mifflin Company, Boston.

Cornes, R. and Sandler, T. (1996). *The Theory of Externalities, Public Goods and Club Goods*. Cambridge University Press, Cambridge UK, second edition.

Cowan, R., Davis, P. A., and Foray, D. (2000). The Explicit Economics of Knowledge Codification and Tacitness. *Industrial and Corporate Change*, 9(2):211–253.

Crowston, K., Wei, K., Li, Q., and Howison, J. (2006). Core and Periphery in Free/Libre and Open Source Software Team Communications. In *HICSS*, Los Alamitos, CA, USA. IEEE Computer Society.

Czepiel, J. A. (1975). Patterns of Interorganizational Communications and the Diffusion of a Major Technological Innovation in a Competitive Industrial Community. *Academy of Management Journal*, 18(1):6–24.

Dahan, E. and Hauser, J. R. (2001). The virtual customer. *Journal of Product Innovation Management*, 19(5):332—-353.

Dahlander, L. and Magnusson, M. G. (2005). Relationships between open source software companies and communities: Observations from Nordic firms. *Research Policy*, 34(4):481—-493.

Dasgupta, P. and David, P. (1987). Information Disclosure and the Economics of Science and Technology. In Feiwel, G. R., editor, *Arrow and the Ascent of Modern Economic Theory*. MacMillan, London.

Dasgupta, P. and David, P. (1994). Toward a new economics of science. *Research Policy*, 23(5):487–521.

Dolan, R. J. and Matthews, J. M. (1993). Maximising the utility of customer product testing: Beta test design and management. *Journal of Product Innovation Management*, 10(4):318—-330.

Dougherty, K. L. (2003). Public goods theory from eighteenth century political philosophy to twentieth century economics. *Public Choice*, 117(2-3):239–253.

Ebadi, Y. M. and Utterback, J. M. (1984). The Effects of Communication on Technological Innovation. *Management Science*, 30(5):572–585.

Ellis, H. C. (1965). *The Transfer of Learning*. MacMillan, New York.

Ellis, R. A., Endo, C. M., and Armer, J. M. (1970). The problem of bias in mail returns: A solution. *Pacific Sociological Review*, 13(2):103–109.

Enkel, E., Kausch, C., and Gassmann, O. (2005). Managing the Risk of Customer Integration. *European Management Journal*, 23(2):203–213.

Erichsson, S. K. (1994). *User Groups im Systemgeschäft. Ansatzpunkte für das Systemmarketing.* Gabler Verlag, Deutscher Universitäts-Verlag, Wiesbaden.

Estes, W. K. (1970). *Learning Theory and Mental Development.* Academic Press, New York.

Ferber, R. (1948). The problem of bias in mail returns: A solution. *Public Opinion Quarterly*, 12(4):669–672.

Fielding, R. T. (1999). Shared Leadership in the Apache Project. *Communications of the ACM*, 42(4):42–43.

Flowers, S. (2008). Harnessing the hackers: The emergence and exploitation of outlaw innovation. *Research Policy*, 37(2):177–193.

Foray, D. (2004). *The Economics of Knowledge.* MIT Press, Cambridge.

Füller, J., Bartl, M., Ernst, H., and Mühlbacher, H. (2006). Community based innovation: How to integrate members of virtual communities into new product development. *Electronic Commerce Research Journal*, 6(1):57–73.

Franke, N. and Shah, S. (2003). How Communities Support Innovative Activities: An Exploration of Assistance and Sharing Among End-Users. *Research Policy*, 32(1):157–178.

Franke, N., von Hippel, E., and Schreier, M. (2006). Finding Commercially Attractive User Innovations. *Journal of Product Innovation Management*, 23(4):301–315.

Friedman, D. and McAdam, D. (1992). Collective Identity and Activism. In Morris, A. D. and Mueller, C. M., editors, *Frontiers in Social Movement Theory*, pages 156–173. Yale University Press, New Haven.

Frohlich, N. and Oppenheimer, J. A. (1970). I Get by with a Little Help from my Friends. *World Politics*, 23(1):105–120.

Ghosh, R. A., Glott, R., Krieger, B., and Robles, G. (2002). Free/Libre and Open Source Software: Survey and Study. Technical report, International Institute of Infonomics, University of Maastricht and Berlecon Research GmbH.

Gilliam, E. and Sluzenski, S. (1987). User groups and their vendors: the ins and outs of a partnership. *EDN*, 13(August 6):269–271.

Gilliam, E. and Sluzenski, S. (1990). CD ROM user groups–the experiences of Digital Equipment Corporation's Digital Library Network. *Database*, 13(6):105–108.

Gladstein, D. L. (1984). Groups in Contect: A Model of Task Group Effectiveness. *Administrative Science Quarterly*, 29(4):499–517.

Gonzalez, R. A. (1982). The user group phenomenon. *Online*, 6(4):80–82.

Gould, R. V. (1993). Collective Action and Network Structure. *American Sociological Review*, 58(2):182–196.

Gould, R. V. (1995). *Insurgent Identities: Class, Community, and Protest in Paris from 1848 to the Commune*. University of Chicago Press, Chicago.

Granovetter, M. (1985). Economic Action and Social Structure: The Porblem of Embeddedness. *American Journal of Sociology*, 91(3):481–510.

Granovetter, M. (1990). The Old and the New Economic Sociology: A History and an Agenda. In Friedland, R. and Robertson, A. F., editors, *Beyond the Market Place: Rethinking Economy and Society*, pages 89–112. Aldine de Gruyter, New York.

Granovetter, M. (1992). Problems of Explanation in Economic Sociology. In Nohria, N. and Eccles, R. G., editors, *Networks and Organizations: Structure, Form, and Action*, pages 25–56. Harvard Business School Press, Boston.

Granovetter, M. S. (1973). The Strength of Weak Ties. *The American Journal of Sociology*, 78(6):1360–1380.

Granovetter, M. S. (1983). The Strength of Weak Ties: A Network Theory revisited. *Sociological Theory*, 1:201–233.

Grieser, T. (2005). Market Analysis: Worldwide Distributed Performance and Availability Management Software 2005-2009 Forecast Summary and 2004 Vendor Shares. IDC No. 33752, IDC.

Gruber, M. and Henkel, J. (2006). New ventures based on open innovation an empirical analysis of start-up firms in embedded Linux. *International Journal of Technology Management*, 33(4):356–372.

Gruner, K. E. and Homburg, C. (2000). Does Customer Interaction Enhance New Product Success? *Journal of Business Research*, 49:1–14.

Hardin, R. (1982). *Collective Action*. John Hopkins Press, Baltimore, MD.

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

Harhoff, D. and Mayrhofer, P. (2007). User Communities and Hybrid Innovation Processes - Theoretical Foundations and Implications for Policy and Research. unpublished manuscript, University of Munich.

Harrison, D. and Waluszewski, A. (2007). The development of a user network as a way to re-launch an unwanted product. *Research Policy*, 37(1):115–130.

Hars, A. and Ou, S. (2002). Working for free? motivations for participating in open-source projects. *International Journal of Electronic Commerce*, 6(3):25–39.

Hauschildt, p. (1992). External acquisition of knowledge for innovations - A research agenda. *R&D Management*, 22(2):105–110.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–61.

Henkel, J. (2006). Selective Revealing in open innovation processes: The case of embedded Linux. *Research Policy*, 35(7):953–969.

Herstatt, C. and von Hippel, E. (1992). From Experience: 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.

Hertel, G., Nieder, S., and Herrmann, S. (2003). Motivation of Software Developers in Open Source projects: an Internet-based Survey of Contributors to the Linux Kernel. *Research Policy*, 32(7):1159–1177.

Hienerth, C. (2006). The commercialization of user innovations: the development of the rodeo kayak industry. *R&D Management*, 36(3):273–294.

Houghton, J. W. (2004). Digital Delivery of Business Services. DSTI/ICCP/IE(2003)2/FINAL, OECD.

Hsieh, L.-F. and Chen, S. K. (2005). Incorporating voice of the consumer: does it really work? *Industrial Management and Data Systems*, 105(6):769–785.

Jack, S. L. (2005). The Role, Use and Activation of Strong and Weak Network Ties: A Qualitative Analysis. *Journal of Management Studies*, 42(6):1233–1259.

Janky, B. and Takacs, K. (2004). Social Control, Network Structure and Participation in Collective Action. unpublished manuscript, Budapest University.

Jeppesen, L. B. and Frederiksen, L. (2006). Why Do Users Contribute to Firm-Hosted User Communities? The Case of Computer-Controlled Music Instruments. *Organization Science*, 17(1):45–63.

Jeppesen, L. B. and Molin, M. J. (2003). Consumers as Co-developers: Learning and Innovation Outside the Firm. *Technology Analysis & Strategic Management*, 15(3):363–383.

Katz, E. and Lazarsfeld, P. F. (1955). *Personal Influence*. Free Press, New York.

Kawasaki, A. (1990). *The Machintosh Way*. Scott, Foresman, Glenview, IL.

Keller, P. and Holland, W. (1975). Boundary Scanning Roles in a Research and Development Organization. *Academy of Management Journal*, 18(2):388–393.

Kieser, A. and Walgenbach, P. (2007). *Organisation*. Schäffer-Poeschel Verlag, Stuttgart, 5 edition.

Klein, K. J., Lim, B.-C., Saltz, J. L., and Mayer, D. M. (2004). How do they get there? An examination of the antecedents of centrality in team networks. *Academy of Management Journal*, 47(6):952—963.

Koufteros, X., Vonderembse, M., and Jayaram, J. (2005). Internal and External Integration for Product Development: The Contingency Effects of Uncertainty, Equivocality, and Platform Strategy. *Decision Sciences*, 36(1):97–133.

Kozinets, R. (2002). The Field Behind the Screen: Using Netnography for Marketing Research in Online Communities. *Journal of Marketing Research*, 39(1):61–72.

Lakhani, K. R. and von Hippel, E. (2003). How Open Source Software works: "Free" User-to-User Assistance. *Research Policy*, 32(6):923–943.

Lakhani, K. R. and Wolf, R. G. (2005). Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. In Feller, J., Fitzgerald, B., Hissam, S., and Lakhani, K. R., editors, *Perspectives on Free and Open Source Software*, pages 3–22. The MIT Press, Cambridge.

Lau, D. C. and Murninghan, J. K. (1998). Demographic Diversity and Faultlines: The Compositional Dynamics of Organizational Groups. *Academy of Management Review*, 23(2):325–340.

Lazar, J. and Preece, J. (2002). Social considerations in online communities : usability, sociability, and success factors. In Oostendorp, H. V., editor, *Cognition in a Digital World*, pages 127–151. Lawrence Erlbaum Associates Inc. Publishers, Mahwah: NJ.

Lazarsfeld, P. F., Berelson, B., and Gaudet, H. (1944). *The people's choice; how the voter makes up his mind in a presidential campaign.* New York, Columbia University Press.

Lengnick-Hall, C. A. (1996). Customer contributions to quality: A different view of the customer-oriented firm. *Academy of Management Review*, 21(3):791–824.

Lerner, J. and Tirole, J. (2001). The Open Source Movement: Key Research Questions. *European Economic Review*, 45(4-6):819–826.

Lerner, J. and Tirole, J. (2002). Some Simple Economics of Open Source. *The Journal of Industrial Economics*, 50(2):197–243.

Lewin, K. (1947). Frontiers in Group Dynamics. *Human Relations*, 1(2):145.

Lilien, G. L., Morrison, P. D., Searls, K., Sonnack, M., and von Hippel, E. (2002). Performance Assessment of the Lead User Idea–Generation Process for New Product Development. *Management Science*, 48(8):1042–1059.

Lin, N. (2001). *Social Capital: A Theory of Social Structure and Action.* Cambridge University Press, Cambridge.

Lindenberg, S. (2001). Intrinsic Motivation in a new light. *Kyklos*, 54:317–342.

Marwell, G., Oliver, P. E., and Prahl, R. (1988). Social Networks and Collective Action: A Theory of the Critical Mass. III. *American Journal of Sociology*, 94(3):502–534.

Mayrhofer, P. (2005). Does Commercial Interest Hinder Innovation in User Communities? - A Theoretical and Empirical Study. unpublished Masters thesis, University of Munich.

McDermott, R. (2002). Measuring the impact of communities – how to draw meaning from measures of communities of practice. *Knowledge Management Review*, 5(2):26–29.

McKenny, J. L., Zack, M. H., and Doherty, V. S. (1992). Complementary Communication Media: A Comparison of Electronic Mail and Face-to-Face Communication in a Programming Team. In Nohria, N. and Eccles, R. G., editors, *Networks and Organizations: Structure, Form, and Action*, pages 262–287. Boston: Harvard Business School Press.

Melton, A. W. and von Lackum, W. J. (1941). Retroactive and proactive inhibition in retention: Evidence for a two-factor theory of retroactive inhibition. *American Journal of Psychology*, 54(2):157–173.

Merton, R. K. (1973). *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press, Chicago.

Miguela, E. and Gugerty, M. K. (2005). Cooperation in collective action. *Journal of Public Economics*, 89(11-12):2325–2368.

Milgrom, P. and Roberts, J. (1992). *Econonomics, Organizations and Management*. Prentice Hall, New Jersey.

Mockus, A., Fielding, R. T., and Herbsleb, J. (2000). A Case Study of Open Source Software Development: The Apache Server. In *Proceedings of of the 22nd International Conference on Software Engineering*. ACM Press.

Mollick, E. (2004). Innovations from the Underground: Towards a Theory of Parasitic Innovation. unpublished masters thesis, MIT Library, Cambridge, MA.

Morrison, P. D., Roberts, J. H., and Midgley, D. F. (2000). Opinion Leadership Amongst Leading Edge Users. *Australasian Marketing Journal*, 8(1):5–14.

Morrison, P. D., Roberts, J. H., and Midgley, D. F. (2004). The nature of lead users and measurement of leading edge status. *Research Policy*, 33:351–362.

Muniz, A. M. and O'Guinn, T. C. (2001). Brand Community. *Journal of Consumer Research*, 27:412–432.

Myers, S. and Marquis, D. (1969). *Successful Commercial Innovations*. National Science Foundation, Washington D.C.

Nahapiet, J. and Ghoshal, S. (1998). Social Capital, Intellectual Capital, and the Organizational Advantage. *Academy of Management Review*, 23(2):242–266.

Nambisan, S. (2002). Designing Virtual Customer Environments for New Product Development: Toward a Theory. *Academy of Management Review*, 27(3):392–413.

Ogawa, S. and Piller, F. T. (2006). Reducing the Risks of New Product Development. *MIT Sloan Management Review*, 47(2):65–71.

Oliver, P. (1980). Rewards and Punishments as Selective Incentives for Collective Action: Theoretical Investigations. *American Journal of Sociology*, 85(6):1356–1375.

Oliver, P. E. and Marwell, G. (1988). The Paradox of Group Size in Collective Action: A Theory of the Critical Mass. II. *American Sociological Review*, 53(1):1–8.

Oliver, P. E., Marwell, G., and Teixeira, R. (1985). A Theory of Critical Mass. I. Interdependence, Group Heterogeneity, and the Production of Collective Action. *American Sociological Review*, 53(1):1–8.

Olson, E. L. and Bakke, G. (2001). Implementing the lead user method in a high technology firm: a longitudinal study of intentions versus actions. *Journal of Product Innovation Management*, 18:388–395.

Olson, M. (1965). *The Logic of Collective Action – Public Goods and the Theory of Groups*. Harvard University Press, Cambridge, MA.

Ostrom, E. (1998). A Behavioral Approach to the Rational Choice Theory of Collective Action. *American Political Science Review*, 92(1):1–22.

Ostrom, E., Gardner, R., and Walker, J. (1994). *Rules, Games, and Common–Pool Resources*. University of Michigan Press, Ann Arbor.

Parthasarathy, M., Rittenburg, T. L., and Ball, A. D. (1995). A re-evaluation of the product innovation-decision process: the implications for product management. *Journal of Product & Brand Management*, 4(4):35–47.

Picot, A., Dietl, H., and Franck, E. (2005). *Organisation - Eine ökonomische Perspektive*. Schäffer-Poeschel, Stuttgart, fourth edition.

Picot, A., Reichwald, R., and Wigand, R. (2003). *Die Grenzenlose Unternehmung - Information, Organisation und Management: Lehrbuch zur Unternehmensführung im Informationszeitalter*. Gabler, Wiesbaden, fifth edition.

Polyani, M. (1958). *Personal Knowledge: Towards a Post-critical Philosophy*. Routledge & Kegen Paul, London.

Polyani, M. (1966). *The Tacit Dimension*. Doubleday and Co., New York.

Prahalad, C. K. and Ramaswamy, V. (2000). Co-opting Customer Competence. *Harvard Business Review*, 78(1):79–87.

Prandelli, E., Verona, G., and Raccagni, D. (2006). Diffusion of web-based product innovation. *California Management Review*, 48(4):109—-135.

Prügl, R. W. (2006). Die Identifikation von Personen mit besonderen Merkmalen : Eine empirische Analyse zur Effizienz der Suchmethode Pyramiding. Doctoral thesis, Vienna University of Economics and Business Administration.

Prügl, R. and Schreier, M. (2006). Learning from leading-edge customers at The Sims: Opening up the innovation process using toolkits. *R&D Management*, 36(3):237–250.

Pyka, A. (1997). Informal Networking. *Technovation*, 17(4):207–220.

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *American Economic Review*, 83(5):1281–1302.

Ralston, A. and Reilly, E. D. J. (1982). *Encyclopedia of Computer Science and Engineering*, pages 375–377, 670–671, 1319–1320. Van Nostrand Reinhold Company, New York, second edition.

Raymond, E. S. (1999). *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary.* O'Reily & Associates, Inc., Cambridge.

Rogers, E. and Kincaid, D. (1981). *Communication Networks: Toward a New Paradigm for Research.* The Free Press, New York.

Rogers, E. M. (1962). *Diffusion of Innovations.* Free Press, New York, first edition.

Rogers, E. M. (1983). *Diffusion of Innovations.* Free Press, New York, third edition.

Rogers, E. M. and Agarwala-Rogers, R. (1976). *Communication in Organizations.* Free Press, New York.

Rogers, W. H. (1993). Regression standard errors in clustered samples. Technical report, Stata Technical Bulletin 13: 19–23. Reprinted in Stata Technical Bulletin Reprints, vol. 3, 88–94.

Ryan, R. M. and Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25(1):54–67.

Sally, D. (1995). Conservation and Cooperation in Social Dilemmas. A Meta-Analysis of Experiments from 1958 to 1992. *Rationality and Society*, 7:58–92.

Salomo, S., Steinhoff, F., and Trommsdorff, V. (2003). Customer orientation in innovation projects and new product development success - the moderating effect of product innovativeness. *International Journal of Technology Management*, 26(5-6):442 – 463.

Schrader, S. (1991). Informal technology transfer between firms: Cooperation through information trading. *Research Policy*, 20:153–170.

Schrader, S. (1995). Gaining Advantage by 'Leaking' Information: Informal Information Trading. *European Management Journal*, 13(2):156–163.

Schreier, M., Oberhauser, S., and Prugl, R. (2007). Lead users and the adoption and diffusion of new products: Insights from two extreme sports communities. *Marketing Letters*, 18(1):15–30.

Schulz, C. and Wagner, S. (2008). Outlaw community innovations. *Internatinal Journal of Innovation Management*, 12(3):399–418.

Seibert, S. E., Kraimer, M. L., and Liden, R. C. (2001). A Social Capital Theory of Career Success. *Academy of Management Journal*, 44(2):219–237.

Shah, S. (2005a). From Innovation to Firm Formation in the Windsurfing, Skateboarding, and Snowboarding Industries. Working Paper 05-0107, University of Illinois.

Shah, S. K. (2005b). Open Beyond Software. In Cooper, D., DiBona, C., and Stone, M., editors, *Open Sources 2.0: The Continuing Evolution*. O'Reilly Media, Sebastopol, CA.

Souder, W. E., Sherman, J. D., and Davies-Cooper, R. (1998). Environmental Uncertainty, Organizational Integration, and New Product Development Effectiveness: A Test of Contingency Theory. *Journal of Product Innovation Management*, 15:520–533.

Swanson, G. E. (1992). Doing Things Together: Some Basic Forms of Agency and Structure in Collective Action and Some Explanations. *Social Psychology Quarterly*, 55(2):94–117.

Swedberg, R. (1997). New Economic Sociology: What Has Been Accomplished, What Is Ahead? *Acta Sociologica*, 40(2):161–182.

Taylor, M. and Singleton, S. (1993). The Communal Resource: Transaction Costs and the Solution of Collective Action Problems. *Politics and Society*, 21(2):195–214.

Thomke, S. and von Hippel, E. (2002). Customers as Innovators: A New Way to Create Value. *Harvard Business Review*, 80(4):74–81.

Tobin, J. (1956). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, 26:24–36.

Torvalds, L. and Diamond, D. (2001). *Just for Fun: The Story of an Accidental Revolutionary*. Harper Business, New York.

Tushman, M. L. (1977). Special Coundary Roles in the Innovation Process. *Administrative Science Quarterly*, 22:587–605.

Tushman, M. L. and Katz, R. (1980). External Communication and Project Performance: An Investigation into the Role of Gatekeepers . *Management Science*, 26(11):1071–1085.

Tushman, M. L. and Scanlan, T. J. (1981). Boundary Spanning Individuals: Their Role in Information Transfer and Their Antecedents. *Academy of Management Journal*, 24(2):289–305.

Urban, G. L. and von Hippel, E. (1988). Lead User Analyses for the Development of New Industrial Products. *Management Science*, 34(5):569–582.

Utterback, J. M. (1971). The Process of Innovation: A Study of the Origination and Development of Ideas for New Scientific Instruments. *IEEE Transactions on Engineering Management*, 18(4):124–131.

Valente, T. W. and Davis, R. L. (1999). Accelerating the Diffusion of Innovations using Opinion Leaders. *Annals of the American Academy of Political and Social Science*, 566(1):55–67.

Veryzer, R. W. and Borja de Mozota, B. (2003). The Impact of User-Oriented Design on New Product Development: An Examination of Fundamental Relationships. *Journal of Product Innovation Management*, 22(2):128–143.

von Hippel, E. (1986). Lead Users. A Source of novel product concepts. *Management Science*, 32(7):791–805.

von Hippel, E. (1987). Cooperation between rivals: Informal know-how trading. *Research Policy*, 16(6):291–302.

von Hippel, E. (2001a). Innovation by user communities: Learning from open-source software. *MIT Sloan Management Review*, 42(4):82–86.

von Hippel, E. (2001b). User toolkits for innovation . *Journal of Product Innovation Management*, 18(4):247–257.

von Hippel, E. (2005). *Democratizing Innovation*. The MIT Press, Cambridge–London.

von Hippel, E. (2007). Horizontal innovation networks by and for users. *Industrial and Corporate Change*, 16(2):1—-23.

von Hippel, E. and Franke, N. (2003). Satisfying heterogeneous user needs via innovation toolkits: the case of Apache security software. *Research Policy*, 32(7):1199–1215.

von Hippel, E. and Katz, R. (2002). Shifting Innovation to Users via Toolkits. *Management Science*, 48(7):821–833.

von Hippel, E., Thomke, S., and Sonnack, M. (1999). Creating Breakthroughs at 3M. *Harvard Business Review*, 77:47–57.

von Hippel, E. and von Krogh, G. (2003). Open Source Software and the "Private-Collective" Innovation Model: Issues for Organization Science. *Organization Science*, 14(2):209–223.

von Hippel, E. and von Krogh, G. (2006). Free Revealing and the Private-Collective Model for Innovation. *R&D Management*, 36(3):295–306.

von Krogh, G. (1998). Care in Knowledge Creation. *California Management Review*, 40(3):133–153.

von Krogh, G., Haefliger, S., and Spaeth, S. (2003a). Collective Action and Communal Resources in Open Source Software Development: The Case of Freenet. unpublished working paper, Institute of Manangement, University of St. Gallen, Switzerland.

von Krogh, G., Lakhani, K., and Spaeth, S. (2003b). Community Joining, and Specialization in Open Source Software Innovation. *Research Policy*, 32(7):1217–1241.

Ward, L. B. (1937). Reminiscence and rote learning. *Psychological Monographs*, 49(4):64.

Wasko, M. and Faraj, S. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, 29(1):35–57.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge.

Weber, M. ([1921] 1978). *Economy and Society: An Outline of Interpretive Sociology*. Berkeley: University of California Press. (Original work published 1921).

Weber, S. (2004). *The Success of Open Source*. Harvard University Press, Cambridge–London.

Wellman, B. (1979). The community question: The intimate networks of east yorkers. *American Journal of Sociology*, 84(5):1201–1231.

Wellman, B., Boase, J., and Chen, W. (2002). The Networked Nature of Community: Online and Offline. *IT & Society*, 1(1):151–165.

Wenger, E. (2004). Knowledge Management as a doughnut: Shaping your knowledge strategy through communities of practice. *Ivey Business Journal*, 68(3):1–8.

White, D. M. (1950). The "Gate Keeper": A Case Study In the Selection of News. *Journalism Quarterly*, 27:383–390.

Williams, R. L. (2000). A Note on Robust Variance Estimation for Cluster-Correlated Data. Technical report, Biometrics Volume 56 Page 645.

Wooldridge, J. M. (2002). *Econometrics Analysis of Cross Section and Panel Data*. MIT Press, Cambridge.

Wooldridge, J. M. (2003). *Introductory Econometrics - A Modern Approach*. Thomson Learning, South-Western College Publishing, Stamford, CT, 2 edition.

Yukl, G. (1981). *Leadership in Organizations*. New Jersy: Prentice Hall.

Zeitlyn, D. (2003). Gift Economies in the Development of Open Source Software: Anthropological Reflections. *Research Policy*, 32(7):1287–1291.

Zucker, L. G. and Darby, M. R. (1996). Star Scientists and Institutional Transformation: Patterns of Invention and Innovation in the Formation of the Biotechnology Industry. *Proceedings of the National Academy of Sciences*, 93(23):12709–12716.

Zukin, S. and DiMaggio, P. (1990). *Introduction to Social Structures of Capital*. Cambridge University Press, Cambridge.