Knowledge Creation in Health IT Online Communities

Hani Safadi
Desautels Faculty of Management
McGill University, Montreal
hani.safadi@mail.mcgill.ca

September 16, 2014

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy.

DESAUTELS | McGill

Abstract

Research on knowledge and digital production in Communities of Innovation is still in its early stages. The most sustained research has focused on open-source software development as it has emerged as a serious competitor to the traditional proprietary software. The Linux operating system, the Apache web server, and the Firefox browser are all successful open-source products that achieved wide commercial success over their proprietary competitors. It has been argued that open-source software combines aspects of private investment and common action production models. This new combination requires fundamental revisions to theories of innovation. With the objective of gaining a deeper understanding of open innovation in healthcare, this research focuses on how online communities focused on innovation create, share, and evolve knowledge artifacts.

Specifically, I explore the development, evolution, and knowledge creation in a community of innovation centered on an open-source Electronic Medical Record (EMR) named OSCAR. This community is primarily Canadian, has been in existence for a decade, and has developed an EMR that is rapidly diffusing (currently used by over 1,500 Canadian doctors to follow over a million patients). The OSCAR EMR is freely available open-source software and is gaining market share against commercial products typically costing $25,000 per year. Given the complexity of such software, the mission-critical nature of patient records for solving Canadian and world health issues, and the fact that the vast majority of users (family doctors) are not computer savvy, this success is unusual and significant. Contrary to other open-source projects, this software is embedded in a community of doctors rather than a community of programmers. The community is active both face to face via user meetings and online via discussion forums.

Focusing on knowledge creation and software evolution, I take on three interrelated studies of this healthcare community of innovation to answer the following research questions: (1) What determines individual members' contributions in the community? (2) How does the community integrate individual contributions into the digital artifact? (3) How does the community grow and evolve overtime creating and sustaining its collaborative structure? Taken together, the studies contribute theoretically by exploring the knowledge exchanges and innovation dynamics in a community of innovation that involves diverse groups of participants (e.g., doctors, programmers, nurses, administrators). They will also help shed light on an important class of innovation communities, one where Canada is a leader and where the innovation outcomes are helping computerize healthcare in Canada.

Resumé

Recherche sur le savoir et la production numérique dans les communautés de l'innovation en est encore à ses premiers stades. La recherche la plus soutenue a mis l'accent sur le développement de logiciels libres et open-source qui devinent un concurrent sérieux pour les logiciels propriétaires. Le système d'exploitation Linux, le serveur Web Apache et le navigateur Firefox sont tous des produits open-source qui ont obtenu un grand succès commercial contre leurs concurrents propriétaires. Il a été avancé que le logiciel open-source combine des aspects de l'investissement privé et des modèles de production d'action commune. Cette nouvelle combinaison nécessite des révisions fondamentales aux théories de l'innovation. Avec l'objectif d'obtenir une meilleure compréhension de l'innovation ouverte dans la domaine de soins de santé informatique, cette recherche se concentre sur la façon dont les communautés d'innovation en ligne créent, partagent et évoluent des artefacts de connaissance.

Plus précisément, j'explore le développement, l'évolution et la création de connaissances dans une communauté de l'innovation centrée sur un dossier médical informatisé open-source nommé OSCAR. Cette communauté est canadienne, a été en existence depuis une décennie, et a développé un dossier médical informatisé qui a diffusé rapidement (actuellement utilisé par plus de 1500 médecins canadiens suivant plus d'un million de patients). OSCAR est disponible gratuitement comme un logiciel open-source et a gagné des parts de marché contre des produits commerciaux coûte généralement $25000 par an. Ce succès est inhabituel et significatif. Contrairement à d'autres projets open-source, ce logiciel est intégré dans une communauté de médecins plutôt que d'une communauté de programmeurs. La communauté est active à la fois face à face par des réunions d'utilisateurs et en ligne par l'intermédiaire de forums de discussion.

En se concentrant sur la création du logiciel et l'évolution de connaissance, je prends sur trois études connexes de cette communauté d'innovation pour répondre aux questions de recherche suivantes: (1) Qu'est ce qui détermine les contributions des membres de la communauté? (2) Comment peut la communauté intégrer les contributions individuelles dans l'artefact numérique? (3) Comment peut la communauté se grandir, s'évoluer et maintenir sa structure de collaboration? Ces études contribuent théoriquement en explorant les échanges de connaissances et de la dynamique de l'innovation dans une communauté d'innovation qui implique divers groupes de participants (par exemple, des médecins, des programmeurs, des infirmières, des administrateurs). Ils aideront aussi à faire la lumière sur une classe importante de communautés de l'innovation, celle où le Canada est un leader et où les résultats de l'innovation aident informatiser les soins de santé au Canada.

Acknowledgements

I would like to acknowledge several people who helped me during my journey throughout the doctoral program. First, I would like to acknowledge my supervisor Prof. Samer Faraj who used to say that a PHD is an apprenticeship and an opportunity to learn from fellow scholar and academicians. I have learned a great deal from my "master" and I am indebt for his help and support. He has been an excellent role model as a researcher, instructor and mentor. I would also like to thank my committee members Prof. Alain Pinsonneault and Prof. Christiane Demers who oversaw the development of this thesis and were always generous to provide their advice and guidance.

I want to thank the professors at McGill University and HEC Montreal whom I benefited from collaborating with and also attending their courses: Prof. Animesh Animesh, Prof. Genevieve Bassellier, Prof. Kunsoo Han, Prof. Yeona Jang, Prof. Ann Langley, Prof. Liette Lapointe, Prof. Jui Ramaprasad, Prof. Suzanne Rivard, and Prof. Emmanuelle Vaast. I have also benefited from interacting with my fellow doctoral students at McGill, HEC, Concordia and UQAM. I cherish our interaction and look forward to continuing our journey in academia.

I would like to thank members of the OSCAR community who in addition to creating this amazing EMR also greatly facilitated my research. In particular I am grateful to Dr. David Chan and Dr. Mark Roper for their help obtaining access to sites and informants.

During my PHD studies I received financial support from the Social Sciences and Humanities Research Council, les Fonds de recherche sur la société et la culture, the Rathlyn Foundation, and Mr. Hian Siang Chan. I am indebted to all of you. Thank you for your generosity and support.

Finally, I dedicate this work to my family, my brothers Sami and Hassan, and my parents Housam and Sahar who were always standing up for me even during their difficult times. Last but not least, I dedicate this work to sweetheart Ghalia her support and unconditional confidence in me and my son Housam Junior who came as a joy and a challenge in my last year of studies. I am very grateful to all of you.

Contents

List of Figures

List of Tables

1 Introduction

New forms of organization and work such as crowdsourcing, open-source development and open user innovation have become increasingly recognized as alternatives to traditional within-firm forms of organizing (Benkler, 2006; Tapscott and Williams, 2008). Those forms of collaboration, enabled by information and communication technology (ICT), extend knowledge production, traditionally performed within firm boundaries, to include a larger collectivity of users and participants outside traditional organizational boundaries. This extensive change in the way innovations, products, and services are initialized and developed led researchers to call to theorize the phenomenon (Von Krogh et al., 2003; Von Hippel, 2005; Von Krogh, 2012). Multiples names are used to refer to this new phenomenon with slight variations in their definition. They include: communities of innovation, online communities of practice, and open-source communities. As the majority of activity in these new forms of organization happens online, we refer to them as online communities of innovation or simply as online communities (OC).

Online communities (OC) are groups of people who work together to achieve a common goal not because of institutional mandates (Coakes and Smith, 2007), but because they share similar interests, have common motivations and are convinced of their common cause (Sproull and Arriaga, 2007). Organizational design scholars are keenly focused on how community-based innovation is transforming our understanding of firm boundaries and of market mechanisms in product development (Argote and Miron-Spektor, 2011; Tushman et al., 2012). It has been argued that open-source software combines aspects of private investment and common action production models (Von Hippel and Von Krogh, 2003). This new combination requires fundamental revisions to theories of innovation.

Open-source has been identified as an alternative production model for creating software and IT services. Open-source development relies on distributed talent of many interested developers to produce a common good that is available to the public. At the same time, it offers several material and non-material incentives to members in the project (Lakhani and Wolf, 2005; Roberts et al., 2006). The open-source software development movement has gained momentum in the last years and has emerged as a serious competitor to the traditional proprietary software. For example, the Linux operating system, the Apache web server, and Mozilla Firefox browser are all successful open-source products that gained market share and attained commercial success over their proprietary competitors.

Success stories of open-source (e.g., Wikipedia, Linux) have been the focus of much attention but we still know little about how these communities innovate in an ecosystem that includes firms, users, and developers (Tushman et al., 2012). Given that open-source communities often represent an ecosystem of interested firms and individual developers, it is no surprise that some 40 percent of developers in some key open-source software projects are paid to participate in the community (Lakhani and Wolf, 2005), a factor that leads many to challenge the traditional shared interest and communitarian understanding of these communities (West and Lakhani, 2008).

On the surface, online communities appear as anarchic collections of individuals largely devoid of a formal authority. Yet comparable and sometimes superior to traditional organizations, online communities develop strong group norms and leadership structure (O'Mahony and Ferraro, 2007), successfully generate information goods, and satisfy members' needs (Tapscott and Williams, 2008). The key to understand this discrepancy between the apparent disorder and the effective order is to look at the internal dynamics of organization in the community that harmonize a loosely coupled heterogeneous group of individuals into an effective and productive effort (Ahuja et al., 2012). Community members act, react and interact and out of this continuous interaction the community is continuously being reshaped and common goods are produced.

1.1 Open Innovation in Healthcare

The current crisis of healthcare in North America demonstrates the limits of conventional organizational forms to solve problems of healthcare and healthcare IT in particular. The use of technology in healthcare is associated with many advantages such as cutting down healthcare costs and improving the quality of care (Walker et al., 2005). Health IT has been recognized as "one of the keys to modernizing the health system and improving access and outcomes for Canadians" (Romanow, 2002). Despite such advantages, the rate of adoption of Health IT in North America is much lower compared to other developed countries (DesRoches et al., 2008). The USA and Canada also contribute more of their GDP toward healthcare than other OECD countries (OECD, 2011). Solutions to problems surrounding health IT and healthcare require rethinking the basic assumptions about those problems. The combination of a community of innovation that develops open-source software in healthcare provides an opportunity to study an alternative form of organization using an alternative production model to solve an enduring problem in society that established institutions could not resolve.

1.2 Knowledge Creation in Healthcare Online Communities of Innovation

Open innovation communities rely on tapping knowledge from diverse members of different backgrounds. Of a particular importance are those members who utilize the innovation (the software in the case of open-source communities) in their daily profession. Those professional members and users are most often the primary source of innovation because of their field expertise and vested interests in the community (Von Hippel, 2005). In an open-source health IT community such as the one studied here, it is medical practitioners such as doctors and nurses who use the open-source software and benefit in their work from other community-provided resources such as discussion threads, plugins and documentation. There are of course other members who have other interests in the community but do not necessarily benefit from its production. In return, many members including professional users and other members contribute back to the community, engage in conversation, and jointly produce knowledge and innovation. Online communities focused on healthcare and healthcare IT are unique in many aspects:

- Communities focused on developing and maintaining health software involve besides programmers and developers professional members such as doctors and other medical practitioners more than is typical in an open-source project. This different membership composition may lead to different dynamics and different distribution of roles than what typically exist in other open-source communities.

- Given their reliance on the core product, professional users (e.g. doctors in this study) are unlikely to leave the development and evolution of the product to the whims of the techies. However, at the same time, professional users lack the time, experience, and perhaps the incentives to contribute to the community and have their voice heard by the members in charge of writing code and developing the software.

- Different than general software packages such as browsers and productivity suites that are designed to offer broad functions that cover the needs of different users, the success of specialized software such as health IT depends on its ability to fit the particularities and needs of professional users who are the sole users of the software.

- Healthcare is a complex industry that is interwoven with cultural, institutional and cultural aspects in the society (Safadi and Faraj, 2011; Berg and Bowker, 1997). Given the complexity of this industry and the uncertainty of its inputs and outputs (Arrow, 1963), it is no surprise that adoption of technology in healthcare is lagging behind other industries (DesRoches et al., 2008; Jha et al., 2009).

Research on community driven software development and open-source software has been too much focused on system software developed by hackers, developers and programmers. Literature of open-source development most often relies on case studies of well-known projects such as the GNU/Linux operating system, the Apache webserver, MySQL database and Firefox browser (Feller et al., 2005). This is understood given the pioneering role and the success of these projects. However, such projects are stereotypical and do not necessarily represent the software landscape. End users of such software are typically a different group of people than developers. The outcome of such communities is software that does not correspond to a real need in the society but serve rather a niche market (e.g. a web server, or a database system). In addition, the membership of such open-source communities is idiosyncratic which limits the generalizability of results found studying them.

We aim to fill this gap in open-source research by studying software developed by a community whose members include in addition to programmers and developers other members who have professional affiliation such as doctors and nurses. Moreover, health IT corresponds to a visible and well-studied sector in the society. This setting allows for examining the role of technology in shaping institutional and social change.

1.3 Thesis Summary

In a recent article in Organization Science, Ahuja et al. (2012) call for a reconsideration of theoretical and methodological issues for studying organizational networks and argued for a perspective that emphasizes the drivers and key dimensions of network change as well as the role of time in this process. Cross-sectional research that studies static snapshots of an organization fails to capture the continuous dynamics of change (Ahuja et al., 2012). Previous research on online communities of innovation and open-source has been too focused on the convergence of collaboration online. Recently, this view has been challenged by research that considered the divergence of interests, passions and time as the mode of operation in online communities (Faraj et al., 2011). Under this perspective, online communities are nexus of tensions and conflicts rather than agreement and harmony.

Contrary to conventional organizational forms, online communities are characterized as fluid because they morph and change their boundaries, norms, interactions, and foci yet retain their basic shape and characteristics (Faraj et al., 2011). Membership in online communities is open. New members join and bring fresh ideas that established actors cannot conceive (Jeppesen and Lakhani, 2010). Innovation communities evolve practices that allow for their sustainability, governance and growth (Kudaravalli and Faraj, 2008; Ridings and Wasko, 2010; O'Mahony and Ferraro, 2007). This diversity in online communities leads to divergence of goals, processes and solutions. The alteration between convergence and divergence allow communities of innovation to transform contestation into collaboration (O'Mahony and Bechky, 2008).

A fundamental question about communities of innovation is to understand the participation dynamics in the community and how online communities focused on innovation evolve overtime creating and sharing knowledge artifacts. We propose three interrelated studies that ask the following research questions:

1. What determines individual members' contributions in the community?
2. How does the community integrate individual contributions into the digital artifact?
3. How does the community grow and evolve overtime creating and sustaining its collaborative structure?

In order to answer the research questions, we explore the development, evolution, and participation dynamics in a Community of Innovation centered on an open-source Electronic Medical Record (EMR) named OSCAR (OSCAR, 2012). This community is primarily Canadian, has been in existence for a decade, and has developed an EMR that is rapidly diffusing (currently used by over 1,500 Canadian doctors to follow over a million patients). The OSCAR EMR is freely available open-source software and is gaining market share against commercial products typically costing $25,000 per year per user. Given the complexity of such software, the mission-critical nature of patient records for solving Canadian and world health issues, and the fact that the vast majority of users (family doctors) are not computer savvy, this success is unusual and significant. Contrary to other open-source projects, this software is embedded in a community of doctors rather than a community of programmers. We focus our attention on online communication among members in the community forums, knowledge contribution, and software evolution. A more detailed outline of the thesis is:

- Chapter 2 reviews relevant literature on open innovation communities and open-source software with a particular focus on the context of healthcare. The chapter discusses open-source development model and the role of community in the development process.

In addition, the chapter examines past research on communities of innovation including the motivation to contribute, the mechanisms of collaboration and the governance of communities.

- Chapter 3 outlines the research design, describes the data collection procedures and presents the conceptual framework for the three empirical studies.
- Chapter 4 presents an in-depth case study of the OSCAR open-source Electronic Medical Record (EMR) including its history and growth, hardware and software architecture, functionality, and also the ecosystem and the activity of the community. It concludes with an analysis of how OSCAR stacks against commercial EMRs and how open-source.
- Chapter 5 presents the first empirical study that focuses on the determinants of individual members' contributions to the community and the role their structural position within the community shapes their contribution.
- Chapter 6 presents the second study that considers how the community integrates individual contributions into the digital artifact (i.e. the EMR). It focuses on the dynamics of the community and the roles that various stakeholders including developers and medical practitioners play in shaping the production of the community.
- Chapter 7 presents the third study that examines the growth of the OSCAR community and how new members establish relationships that evolve the community overtime.
- Chapter 8 outlines the expected contributions of the thesis and the future research plan.

Taken together, the three studies contribute theoretically by exploring the knowledge exchanges and innovation dynamics in a community of innovation that involves diverse groups of participants (e.g., doctors, programmers, nurses, administrators). They will also help shed light on an important class of innovation communities, one where Canada is a leader and where the innovation outcomes are helping computerize healthcare in Canada.

2 Literature Review

The thesis focuses on Online Communities of Innovation (OC) of Open-Source Software (OSS) in healthcare. There are commonalities between open-source software and online communities of innovation. First, they both rely on the effort of members to produce a public good. Whereas this public good is knowledge in online communities, it is more materialized as software in open-source communities. Second, membership of both online communities and open-source software is usually open and unrestricted. They both rely on volunteers who are distributed organizationally, spatially and temporally and who have different motivations to join the project or the community (Lakhani and Wolf, 2005; Sproull et al., 2005; Wasko and Faraj, 2005). Finally, the output of online communities and open-source is usually available to the public with little or no restrictions.

With all of those similar aspects at hand, it is not surprising to see OSS and OC coexist in many situations. For example, Wikipedia is based on MediaWiki, an open-source platform that is also empowering other wikis on the Internet. The most common situation where OSS and OC coexist where open-source software employ a form of online community as a medium of communication among members and developers. For example, most open-source development platforms such as Sourceforge, Launcpad, and Github offer various kinds of online community support such as threaded forums, mail-lists, and wikis. Members of such communities discuss several aspects of the software in order to learn, share experiences, and solve problems and issues arising from using the software.

2.1 Knowledge Production of Online Communities

The literature on open-source software and online communities of innovation is very vast and continuously expanding (Boyd and Ellison, 2007). Dedicated review papers and books cover this literature in depth (Flowers and Henwood, 2010; Kilduff and Brass, 2010; Lerner and Schankerman, 2010). For the purpose of this research, we focus our review on the process of getting people to collaborate online and produce common goods: why people contribute, how they contribute, and how they organize themselves and structure their contribution.

2.1.1 Contribution in Online Communities

The main challenge facing most OC is to elicit members both as contributors and administrators who oversee the function of the group. Research has shown different kinds of members with different degrees of involvement in the community including active members (Butler et al., 2013), lurkers (Berdou, 2011), administrators and leaders (O'Mahony and Ferraro, 2007). Interestingly, research has also found that important contributions comes often from peripheral members in the community indicating the importance of casual contributors (Jeppesen and Lakhani, 2010).

Altruistic behavior of people in OC has been the focus of many studies. The question of why people contribute is an important one because it defines the online community and distinguishes it from traditional organizations. Community members participate primarily out of their interest in the community, generalized reciprocity and prosocial behavior (Wasko and Faraj, 2000). The altruistic behavior online has often described as extension to other forms of prosocial behaviors such as helping bystanders (Sproull et al., 2005, p. 141). However, such characterization does not fit the complex milieu in which online collaboration happens. Other factors such as collective identity and social learning may affect altruism online and differentiate it from offline behavior (Sproull et al., 2005). Self-interested motivations of participating in online groups include building and maintaining social ties with people already known offline, and gaining personal benefits such as becoming visible, knowledgeable and skilled at a certain matter (Butler et al., 2013). Finally, professional motivations such as self-esteem at work and perception of peers at work are important motivators for contributing online regardless of expectations of reciprocity (Wasko and Faraj, 2005).

Research has looked at the contingencies leading to members' contributions to online communities. For example, a strong relationship between two members is not a prerequisite for them to interact and collaborate online (Constant et al., 1996). On the other hand, strong ties are beneficial for the output of the community because they along with network structural characteristics lead to a higher level of knowledge creation (McFadyen et al., 2009). Specific IT features lead to better participation and contribution in OC (Levina and Arriaga, 2015). Technological characteristics that facilitate interaction online by providing a reliable and easy to use platform are a precursor to members' information seeking and providing behavior in the online community (Phang et al., 2009). Specific feature such as virtual co-presence support, persistent labeling, self-presentation, and deep profiling are enablers of online knowledge contribution is (Ma and Agarwal, 2007). Finally, barriers to contribution such as fear of criticism and evaluation are documented (Ardichvili et al., 2003).

The location of members within the community is a determinant to various types of contribution. Dahlander and Frederiksen (2012) found that users' innovation in OC depends on their position both inside and outside the community. Boundary spanning between communities and a core position within the community characterize innovators in OC. Jeppesen and Lakhani (2010) present a complementary view where periphery members play an important role in providing solutions to problems that core members could not solve by bringing expertise from different fields than the core field of the problem and by coming from a different social stratum than core members.

2.1.2 Mechanisms of Collaboration

Beyond looking at the determinants of contribution in OC, research also looked at the mechanisms that enable this contribution. Olivera et al. (2008) uncovered three mediating mechanisms between the technological and social context of the community and the contribution behaviors of its members. The three mechanisms are: awareness, searching and matching, and formulation and delivery. Faraj and Johnson (2011) took a broader perspective and considered the dyadic exchange patterns in online communities of technological software. They found that tie formation in those communities tend to follow the norms direct reciprocity and indirect reciprocity, and to deviate away from preferential attachment found generally in social networks.

Research has also looked at the knowledge production of OC. Members of OC contribute different types of knowledge including solutions, referrals, problem reformulation, validation, and legitimation (Cross and Sproull, 2004). The usefulness of knowledge created online vis-à-vis traditional organization way is also considered. OC is found to be a source of knowledge and solutions to problems that traditional organizations could not solve (Jeppesen and Lakhani, 2010). The effectiveness of OC comes from the heterogeneity and diversity of its members and also the network effectiveness in OC by relying on social mechanisms, rather than legal mechanisms that traditional organizations rely on (Feller et al., 2008).

2.1.3 Governance of Online Communities

Because OC lack a formal organizational structure providing leadership and governance, research has looked into the emerging structure within communities that on the surface appear to be anarchic. Looking at the evolution of policies and control mechanisms in Wikipedia, Butler et al. (2008) found that members in Wikipedia were able to evolve a wide variety of organizational structures using the affordances of the platform. O'Mahony and Ferraro (2007) carried a longitudinal study of the Debian Linux operating system user community focusing on governance and leadership. They have found that the governance structure has passed several phases shifting from an autocratic leadership community with de facto leaders to a democratic community with elected leaders. The characteristics of leaders are important because of its effects on members' interest and continued participation in the community (Johnson, 2010).

Along with governance, studies have looked at the sustainability of online groups. In the absent of direct monetary reward, OC need mechanisms to motivate members and sustain their existence. The role of systematic feedback in inducing better contribution for longer duration is documented (Moon and Sproull, 2008). Sustainability is affected by a variety of structural and social factors such as message volume, content, and contributors' characteristics (Ridings and Wasko, 2010). Online communities are also sustained by a critical mass of active members who develop strong ties with the community as a whole rather than develop interpersonal relationships (Wasko et al., 2009).

2.1.4 Dynamics of Change in Online Communities

Research in organizational change theories proposes that quantum change may better describe organizational change than incremental change (Miller and Friesen, 1982). Inertial forces build momentum which drags the organization into its current configuration allowing only for small incremental changes. Radical change occurs infrequently and results into a quantum change of organizational configurations. The idea of structural inertia (Miller and Friesen, 1982) suggests that organizations stick with what they know because change is detrimental for them. Indeed, structural inertia inhibits large organization from making change and reduces their strategic choices (Hannan and Freeman, 1984), on the other hand, young organizations such as startups and communities of innovation usually have more agility and flexibility in taking strategic decisions that outperform decisions made by well-established organizations.

Not all researchers agree on the radical nature and revolutionary characterization of organizational change. Some organizational change theories propose that change is gradual and evolutionary (Demers, 2007, Ch. 7). In the 1990s the evolutionary school of organizational change gained tractions arguing that organizations do indeed change and adapt. Orlikowski (1996) advocates "a perspective that posits change rather than stability as a way of organizational life.” Researchers have posited different mechanisms of this change including planned management, external environment pressures and internal growth (Demers, 2007). Under this perspective, the organization is always under continuous change because of the various internal and external forces affecting it. Ultimately, combining different perspectives on organizational change provides a richer understanding the phenomenon than any one perspective provides by itself (de Ven, 2005).

One recent perspective to study organizational change is to focus on the discourse and language use in speech and written communication among organizational actors. This approach construes organizations as texts, conversations and discursive performances where change is considered a linguistic process (Demers, 2007, Ch. 8). The analysis of language posits that the linguistic artifacts such as text and speech are not merely representational but rather constituent of social and organizational reality (Alvesson and Kärreman, 2000). As such, the dynamics of discourse dictate the process of change in the organization (Young and Fitzgerald, 2006, Ch. 5). The communication-as-constitutive perspective is instrumental in describing how people get organized and how organizations come to be reenacted and reproduced through analysis of organizational communication (Cooren et al., 2011). For example, the choice of vocabulary is related with social categorization. The different vocabularies used to refer to men and women in media assibilates people to pre-existing gender discrimination (Fairclough, 1995). A discursive approach suits well studying change in online communities where communication is written and archived.

2.1.5 Economics of Collaboration

Notwithstanding the previous points, diversity comes at cost. While it is true that more members bring knowledge, experiences and extra resources to the community, processing such resources is an additional cost that other community members will incur. For example, in the context of collaborative content generation there is an inherent transaction cost of reading what other members posts in addition to rule out low quality content, fight spam and correct mistakes. In the context of mailing list and message exchanging communities, there is the extra cost of reading more threads and more messages per thread in addition to getting to know the extra members.

Employing a resource-based perspective, Butler (2001) studied how sustainable social structures form in online communities (listserv). Despite the fact that members contribute time, energy, and other resources, such resource comes with an associative cost. Resources need to be processed by members in order to create benefits. Social structure in online communities forms as a balance between membership size and communication activity. These limitations of online collaboration are confirmed by other research in different contexts. For example, in peer-to-peer file sharing, the negative externalities associated with community growth in addition to free riding limits the size of the network (Asvanund et al., 2004). In virtual investment-relating communities, there is a diminishing marginal cost of additional information posted by members. The cost incurred by a user increases with the total number of postings in the community and is convex with regard to the total number of postings (Gu et al., 2007).

However, as many organizations have discovered when trying to embrace knowledge management, the creation of an online social space is no guarantee that knowledge creation and sharing will actually take place (Alavi and Leidner, 2001; Levina and Arriaga, 2015). One core issue is that communities of innovation produce a public good, and given a digital public good, the costs of copying and reproducing it tend to zero. All members of the community are free to enjoy the benefits of the collective good, irrespective of the extent of their involvement in its development. Thus, a rational participant could enjoy the public good for free, without contributing in return, a problem known as the free riding problem (Olson, 1965). However, the economics of digital goods can play a countervailing role: if the cost of production is fixed regardless of the number of users benefiting from the good, then a small group of active participants can generate the good without the necessity of equal participation. As is evidenced in projects such as Wikipedia or Linux, a core group can sustain the community activity but with many millions benefiting or only peripherally contributing (Oliver and Marwell, 2001; Wasko et al., 2009).

Finally, the concept of OC and online communities is not limited to collections of individuals. Organizational level research of OC looks at communities or networks whose members are organizations. For example, Venkatraman and Lee (2004) studied the relationships between developers, publishers, and platform enablers in the video game industry. They found that network characteristics (density overlap and embeddedness) formed by relationships among the various entities and technology characteristics (dominance and newness) predetermine the probability of launching new games in different platforms.

2.1.6 Is Healthcare Different?

Doubts about community-based development new models of production are not new. For example, "Steven Ballmer of Microsoft denounced the shared production of software as communism. Robert McHenry, a former editor in chief of Encyclopedia Britannica, likened Wikipedia to public rest room. Andrew Keen, author of The Cult of the Amateur, compared bloggers to Monkeys" (Shirky, 2010, p. 162). Such early skepticism has, however, faded away with the success of many projects adopting new modes of production (Von Hippel, 2005).

Nevertheless arguments against a community model in healthcare are still prevalent. First, there is a widespread misconception that because community-based development and open-source software is free, nothing of a quality can be produced. Second, there is the issue of trusting a system that is not centrally organized nor managed. Innovation communities are like an organism where business has minute control over their evolution and development (Gu et al., 2007; Hof, 2005). Third, the open-source development model is not properly understood by many stakeholders including medical practitioners who are typically profession-oriented and far from the geek culture of open-source. Finally, the significant power of lobby groups that represent commercial vendors contributes to weak government support for open-source health IT initiative (Reynolds and Wyatt, 2011).

The healthcare is different argument is not new. In fact, the healthcare market has unique characteristics that distinguish it from other traditional markets. Those characteristics include the nature of supply and demand, the uncertainty of the product, and the expected behavior of the provider (Arrow, 1963). In addition, the consumption (and the lack of) of healthcare is associated with many externalities that do not exist in other products (Arrow, 1963; Rice and Unruh, 1998). Poor health status of low-income people can trickle to other via contamination and external behavior such as drinking, smoking and perhaps crimes (Cutler, 2005). It is argued that traditional laissez-faire economic model does not work for healthcare because it ignores that the private healthcare markets do not behave like traditional competitive markets (Rothschild and Stiglitz, 1976).

The characteristics of healthcare require new thinking of economics and social production and delivery models. But do these arguments hold true for healthcare IT? Are new production models required to develop and support healthcare technology? There is still lack of evidence of the viability of community development and open-source in healthcare beyond the cost benefit analysis and the technical comparison between HIT systems (Faus and Sujansky, 2008; Reynolds and Wyatt, 2011). It is unknown whether the success of open-source in infrastructure and backend software can translate to business applications in healthcare.

2.2 Open-Source Development

Open-source software is a software development and distribution model that releases source and binary code for use and modification by any person or party interested in the project as long as it maintains the original license under which the software is released. While this is in theory the case, in practice few people are interested in developing and modifying the software. However, contributions are not limited to code modification. Users can contribute by providing help in supporting the software, testing it, reporting bugs, writing documents and communicating with other users and the developers. Contributors benefit from information and communication technology that is capable of effectively supporting rich communication medium (Carlson and Zmud, 1999; Dennis et al., 2008) in spite of the spatial and temporal fragmentation (O'Leary and Cummings, 2007; Pinsonneault and Caya, 2005).

During the last two decades, open-source software has spread and gained a wide acceptance especially in government, educational, and non-for-profit institutions (Foster, 2010). Open-source software has become mainstream in infrastructure and server environments. For example, the Linux operating system, the Apache web server, and Mozilla Firefox browser are all successful open-source products that gained market share and attained commercial success over their proprietary competitors.

Open-source development is a phenomenon that extends beyond freely distributing and manipulating source code (O'Reilly et al., 2005). Indeed, recent research advocates new naming of the phenomenon (Von Krogh, 2012). Open-source has been identified as an alternative production model for creating software and IT services. Moreover, broader community driven development models, also called crowdsourcing, have grown to include a wide range of products other than software including encyclopedia articles (e.g. Wikipedia.org), venture funding (e.g. Kickstarter.com), and car design and manufacturing (e.g. Localmotors.com).

Success stories of open-source (e.g. Wikipedia, Linux) have been the focus of much attention but we still know little about how these communities innovate in an ecosystem that includes firms, users, and developers (Tushman et al., 2012). At the center of open-source development and crowdsourcing in general is the economics argument. Why do programmers and developers share code and work together? Given that open-source communities of innovation often represent an ecosystem of interested firms and individual developers, it is no surprise that some 40 percent of developers in some key open-source software projects are paid to participate in the community (Lakhani and Wolf, 2005), a factor that leads many to challenge the traditional shared interest and communitarian understanding of these communities (West and Lakhani, 2008).

Economists have long realized that knowledge and innovation are not exclusive to the elite (Hayek, 1945). Information and communication technology has made it possible to tap talent, experience and knowledge from diverse and dispersed members by providing platforms of communication, sharing and collaboration while offering at the same time various material and non-material incentives to contribute (Lakhani and Wolf, 2005; Roberts et al., 2006). Recent work focused on studying the economics of open-source within classical frameworks such as labor economics and industrial organization (Lerner and Tirole, 2002). Early licenses such as the GPL provided the legal framework of operation with the exclusive goal of producing a public good (Laurent, 2004). Later, new licenses were created to add more flexibility and make OSS development more appealing for business (Valimaki, 2003; Singh and Phelps, 2013). This allowed firms to mix the development of open-source code and proprietary code in order to specialize and diversify its license types, business models, markets, and marketing strategies (Lerner and Schankerman, 2010; Mehra et al., 2011; Wen et al., 2013). Research on OSS is very vast and continuously expanding (Feller et al., 2005; Berdou, 2011). We focus here on the community dynamics of open-source development versus the traditional software development model.

2.2.1 Traditional Software Development Cycle

The software engineering literature identifies six main stages in the classic waterfall model of software development: Requirement analysis, system design, coding and programming, integrating and testing the system, and finally using and maintaining it (Sommerville, 2007). For example, the iterative model (Figure 1) repeats different stages of development iterations. In each major phase the developers have the chance of interacting with stakeholders and users of the software.

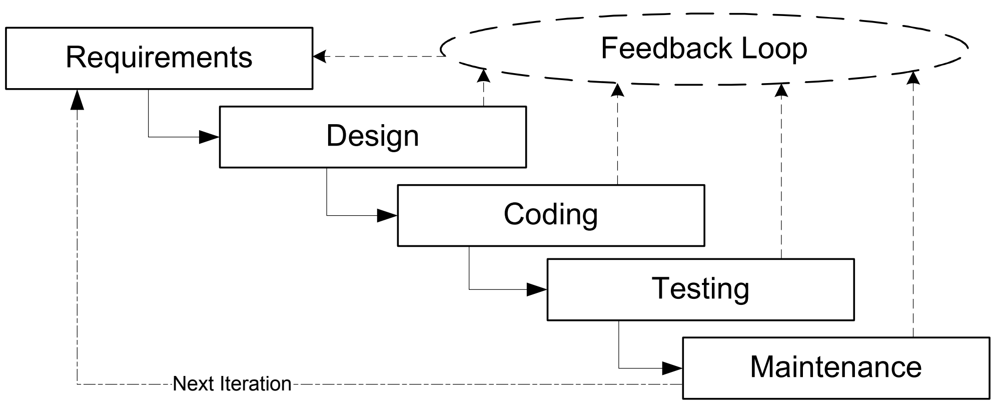


Figure 1: Iterative development model

This traditional development method is seldom used in open-source projects because a central authority does not typically plan them. Instead, open-source projects are incrementally developed using rapid prototyping, iterative development and agile development methods that starts with coding then reviewing and debugging the code before releasing it (Mockus et al., 2000). Open-source development: is parallel rather than linear, involves large communities of distributed developers, utilizes independent peer review, provides feedback to user and developer contributions, includes the participation of talented and motivated developers, includes increased levels of user involvement, and makes use of rapid release schedules (Feller and Fitzgerald, 2002, p. 84).

Based on the activity of the developers, the contributions of the users, and the continuous feedback from the community, the software evolves and matures overtime. System maturity represents the degree of sophistication that the system achieved by incorporating different features to satisfy the needs the users. Software maturity is thus identified by the number of modified or added features to the system. In the test documentation IEEE 829-1998, the Institute of Electrical and Electronics Engineers (IEEE) defines a software feature as a "distinguishing characteristic of a software item (e.g., performance, portability, or functionality)" (Kurbel, 2008, p. 397). Using this definition, a software feature is a tangible aspect of the system that relates to the system usability.

2.2.2 Community Development Model

Open-source software requires the availability of talented and motivated developers to write the code and build the software. However, this is not enough. Users of OSS play an important role in shaping the development of open-source software far more than they do in proprietary software. Users elaborate requirements, test the functionality, report bugs, and write documentation (Feller and Fitzgerald, 2002, p. 92).

Open-source development relies heavily on the online community and the activities that are carried within it (Von Krogh et al., 2003). The activeness of an open-source community is determined by the frequency, intensity and quality of the interaction of its members to collaboratively develop and support the software. The community provides feedback, system improvement requests, missing features wishes and bug fixes. As a result many open-source software projects evolve faster than competing proprietary system because of the activities of the communities. In addition to the actual feedback, the symbolic role of feedback is important in the development process. Many developers in the open-source paradigm are volunteers who are often working for noneconomic incentives (Von Hippel and Von Krogh, 2003). Volunteers value discretionary feedback and peer review from other participants because it shows the value of their work and allow them to express their expertise and values (Constant et al., 1996; Leonardi, 2007).

Communication is the activity through which users' participation and contribution reach the developers and evolve the software. Communication is an important aspect of user participation in software development. By communicating with developers users can express their opinion and concerns about the system. Hartwick and Barki (2001) relate communication to user participation. They define user participation as the extent to which users carries out tasks and perform various activities during the system development process. In addition, communication allows for exchanging facts and experiences with other users. Although the study of Hartwick and Barki (2001) is limited to traditional inside organization system development, the concepts are still relevant to open-source development. Open-source development relies on users and developers engaging in fruitful communication and collaborative efforts in order to develop the system. In fact the development of open-source systems relies heavily on the online community and the activities that are carried within it. Communication and interaction among users and developers in the online community reflects its activeness, which is an important factor in the success of the software (Von Krogh et al., 2003).

2.3 Open Innovation in Healthcare IT

Health Information Technology (HIT) is a diverse set of technologies that serve in storing, managing and transmitting health information for the use of health consumers (patients), health providers (physicians & nurses), payers, insurers and all other groups that have an interest in health and health care (Agarwal and Angst, 2006; Blumenthal and Glaser, 2007). Implementing nationwide HIT systems is a major undertaking because of the associated financial, technical, and institutional challenges (DesRoches et al., 2008; Robertson et al., 2010). At the same time, the use of technology in healthcare is associated with many advantages such as cutting down healthcare costs and improving the quality of care. It is estimated that, by improving health care efficiency and safety, the widespread adoption of electronic medical record (EMR) in the United States can save more than $81 billion annually (Hillestad et al., 2005). Other research estimates that the value to consumers could be more than $77 billion annually (Walker et al., 2005). In addition, the use of IT in hospitals could save over 65,000 lives, prevent over 907,000 serious medication errors, and save approximately $9.7 billion annually (Cutler, 2005; Hillman and Given, 2005; LeapfrogGroup, 2004).

Open-source software (OSS) development has gained momentum in the last years and has emerged as a serious competitor to the traditional proprietary software. Adoption of OSS gained a strong foothold in public sector and academia (Foster, 2010). In healthcare, open-source has yet to gain wide acceptance. Commercial adoption of OSS is still not fully understood because of the complex and novel issues surrounding open-source (Glynn et al., 2005). For example, several open-source HIT systems were developed in North America but are still not widely adopted (Faus and Sujansky, 2008). There is an increasing interest in the applications of OSS in healthcare. An early indicator of such interest is a 2004 Journal of the American Medical Informatics Association (JAMIA) letter to editor calling for considering the unique characteristics of OSS as potential answers for hurdles associated with adoption of EMR systems (Kantor et al., 2003).

Whether the success of OSS can translate to healthcare is still an open question. However, the unique features of OSS may alleviate some of the obstacles associated with proprietary systems. Such features include: low cost (no software cost, no licensing cost) of acquisition and maintenance, greater possibility of customization, and lower exposure to vendor failure or product termination. At the same time, there are exclusive challenges that face open-source health IT. They include the lack of familiarity with the open-source development model, the lack of integration with existing vendor-based hospital systems, a fragmented development effort, a limited number of firms that support installation and training, and an absence of clear software development roadmap, and the lobbying and marketing power of commercial vendors of healthcare software (Reynolds and Wyatt, 2011).

2.3.1 Challenges of Wide Adoption of Health IT

Despite all of the above mentioned advantages of health IT systems, their adoption in North America is still low compared to the adoption of other organizational information systems. For example, in a 2008 national survey, only 4% of physicians in the United States reported having a fully functional EMR system and 13% reported a basic system (DesRoches et al., 2008). In hospitals the situation is not much better, in a 2009 survey; only 1.5% of US hospitals reported having a comprehensive EMR system and 7.6% had a basic system. Computerized physician order entry (CPOE) systems were implemented in 17% of hospitals (Jha et al., 2009).

There are many challenges that face the widespread adoption of IS in healthcare. The healthcare market has unique characteristics that distinguish it from other traditional markets. Those characteristics include the nature of supply and demand, the uncertainty of the product, and the expected behavior of the provider (i.e. the physician) (Arrow, 1963). Those characteristics along with interconnections with other facets of the society such as economics and politics make introducing a change difficult in healthcare (Romanow et al., 2012). Barriers against wide adoption of HIT exist from both providers and consumers' perspectives.

Providers of healthcare include both large institutions such as hospitals and health centers and small providers such as independent physicians and small clinics. Five broad issues must be addressed to promote successful implementation of electronic health records in a small office: financing; interpretability, standardization, and connectivity of clinical information systems; help with redesign of work flow; technical support and training; and help with change management (Baron et al., 2005). Those issues are similar for hospitals but on a more complex and larger scale where various stakeholders are involved in the process (Mintzberg, 2002).

Cost

Cost is the primary barrier of adopting HIT systems. A recent estimate of the cost of purchasing and implementing a clinic EMR system in the US is $15,000 to $50,000 per physician (Blumenthal and Glaser, 2007). Operating costs may reach $20,000 per year (Miller and Page, 2007). Cost soars when the system is implemented in a larger organization. For example, a computerized physician order entry (CPOE) project took three years to complete and cost $18 million dollars in 1995 money (Davidson and Chismar, 2007). Nationwide implementations require huge investment to be realized. For example of, the Quebec Health Record had an initial budget of $562 million (Castonguay et al., 2008).

Complexity

HIT systems should be user-friendly and intuitive in order to reach wide acceptance and save on learning time which an important factor for adoption. As an example, the transition to EMR system in one clinic slowed down physicians and resulted in a 10% to 20% reduction in productivity for a period of months or more (Wang et al., 2003). Another research reports major slow down at the initial stages of adoption (Baron et al., 2005). Customizability is also required because of the richness and diversity of medical practices. Unfortunately, this customization requirement pressures the already soaring price of the software (Safadi and Faraj, 2010).

Interoperability

Because patients need to interact with multiple health workers and insurers in multiple health units, patients' data should be portable and compatible with the different systems used in different sites. Some countries such as UK and Taiwan chose to implement a nationwide EMR system to be used nationwide (Robertson et al., 2010; Liu et al., 2006). In North America, however, the trend is toward allowing different competing systems but insuring interoperability of data at the same time. One solution is to store some data centrally and allow access to it by different systems. For example, the Quebec Health Record stores records of patients in Quebec centrally (Castonguay et al., 2008). Along with storing some data in shared repositories, different systems by different vendors should use a common protocol for communication such as HL7 (Kush et al., 2008).

Social Acceptance

The challenges of health IT goes beyond the characteristics of the technology artifact and how best to design an IS that replaces the existing system (Sujansky, 1998). Adopting information systems in healthcare is not only a question of cost-benefit analysis but rather a complicated decision that depends on the unique social factors that are embedded in healthcare (Anderson, 2007). Time required to learn something novel, fear of potential lawsuits, risk of data openings, fear of automation and deskilling, and poor track record of health IT are among many obstacle that may arise when computerization of patients' records is introduced (Kaplan and Harris-Salamone, 2009).

Consumers' Acceptance

Most HIT issues discussed in the literature are provider-side issues because health data including patients' records is traditionally stored at the provider-side and accessed and processed solely by providers. Computerization of health data removes this restriction as data can be stored in different locations and shared among providers, insurance and government agencies, and even patients themselves. One big issue is privacy as patients may have concerns when their health data is collected and processed by HIT. Concerns of privacy may impede the diffusion of an EMR system that has been demonstrated to reduce medical errors (Angst and Agarwal, 2006). Other issues include data ownership, possession, and disclosure of health data. New legislations and legal structures are required to deal with those issues (Hodge Jr et al., 1999).

Those recurring challenges provide an opportunity for new forms of organization and new models of innovation to prove their viability (Tapscott and Williams, 2008; Jeppesen and Lakhani, 2010). Research has already established that alternative models of innovation and technology production such as open innovation and open-source development can solve problems that traditional models could not (Von Hippel, 2005; Tushman et al., 2012).

2.3.2 Advantages and Disadvantages of Open-Source in Healthcare

The main benefit of open-source HIT is its lower acquisition and maintenance cost compared to proprietary systems. Open-source systems subsidize the initial big acquisition cost over yearly maintenance and support. Moreover, the client is free in choosing support providers if any. This eliminates the risk of vendor locks-in and results in a stronger position for the purchaser and a lower cost of service (Kantor et al., 2003). The bulk of the initial cost comes from hardware cost, installation, training and support cost. Support cost for open-source HIT systems is significantly lower than that of proprietary HITs. If a local IT expertise is available in the site, it may be able to carry on those supporting tasks without being tied to a supporting company. Moreover, the cost of developing and supporting open-source HIT is shared by the different sites and actors participating in this effort.

OSS development relies on pooling distributed talent and skill and organizing it over the Internet. The decentralized development and support model decreases the risk of vendor failure or product termination. It also favors open standards and protocols because the different parties want to insure that their efforts can benefit from others'. The free distribution of source code, in addition to the availability of expertise eliminates the lock-in situation that happens with proprietary systems. This *openness* in term of exposing the underlying data structure and the free distribution of code has its internal and external advantages. Internally, the free and unrestricted access to the code allows for more flexibility in changing the system and customizing it. Externally, open-source systems tend to comply with standard programming interfaces and open architectures such as the Health Level 7 (HL7) protocol in the medical context. This insures higher interoperability and easier connection to other health systems and infrastructure.

The community of users also plays a vital role in providing information and support. In contrast to proprietary systems where the vendor is in charge of providing support, users of open-source HIT rely on each other to acquire and share knowledge about the system. Because of their size matured open-source communities can devote more time and resources to solve arising issues and problems than do traditional organizations (Raymond, 1999). Arguments for the distributed community-supported model of open-source in healthcare include: economic factors where the economy of scale works in favor of large open-source projects with large communities (Reynolds and Wyatt, 2011), psychological factors because members of open-source communities have personal motivation such as peer recognition to be involved in the community (Lerner and Tirole, 2002; Wasko and Faraj, 2005), and social factors because open-source communities tend to be fluid favoring creativity and innovation (Faraj et al., 2011).

Notwithstanding these benefits, open-source in healthcare faces unique challenges that limit its potential advantages. First, open-source HIT may lack the availability of coded data, including databases of diseases, drugs, procedures, and billing codes. Such coded data is necessary to comply with standards and to limit the amount and the variation of free text entry in the system. While some of these databases are standardized and available (e.g.

ICD-9 disease database), some of them require extensive effort to compile and maintain (e.g. drug databases) and some depend on the local context (e.g. billing and procedure codes). Fortunately, there is a recent movement towards open access platforms and databases (Natter et al., 2013; Kapur et al., 2012). Second, Open-source HIT systems may not integrate well with other health systems deployed at medical institutions such as hospitals and pharmacies. This is because large-scale systems are typically developed by big vendors and designed to integrate well with complementary products for the same vendor. Finally, some argue that being open, OSS are less secure because an attacker can access the source code and learn about the inner working of the system in order to exploit vulnerabilities. However, this argument is controversial because vulnerabilities can be quickly fixed since the source code is accessible by many people.

In addition, there are some environmental obstacles that may impede open-source HIT from reaching wide adoption. Many stakeholders including medical practitioners do not properly understand the open-source development model. The lack of understanding along with the significant power of lobby groups that represent commercial vendors contribute to weak government support for open-source health IT initiative (Kantor et al., 2003). There is also still a widespread misconception that because of open-source HITs are cheaper than their commercial counterparts, they must be of lower quality. In addition, there is the issue of trusting a system that is not sponsored by a big vendor. Enterprise open-source systems usually have well-known companies and organizations supporting them (e.g. RedHat supporting Linux). However, there is a lack of big companies that back open-source HIT. The reliance on small companies to support open-source HIT is an issue for clinics and health institutions that are looking for a durable and trusty source of support. Because of fragmented development, some argue that the total cost of ownership is higher in OSS than that of proprietary software (Kantor et al., 2003).

2.3.3 Status of Open-Source Software in Healthcare

Several open-source HIT systems were developed in North America but are still not widely adopted. Faus and Sujansky (2008) enumerate twelve active open-source Electronic Medical Record (EMR) projects in USA and Canada: ClearHealth, IndivoHealth, FreeMed, GNUmed, OpenVista, OpenEMR, OSCAR, PrimaCare, Res Medicinae, Tolven Healthcare, Untimate EMR, and WorldVistA EHR. Among these projects, two have achieved significant visibility and adoption in Canada and the USA.

The first software is a Canadian system called OSCAR (Open Source Clinical Application Resource). OSCAR is developed by the Department of Family Medicine at McMaster University and deployed in multiple clinics in Canada. OSCAR has an active community of users and developers who interact dominantly online. OSCAR gained substantial market share in major Canadian provinces reaching 12.5% and third place in Ontario in 2012 (OntarioMD, 2012b). The second software is OpenVista , an American EMR developed by Medsphere Corp. OpenVista is an open-source re-implementation of VistA which is a seven billion dollar costing EMR developed by U.S. Veterans Affairs and deployed in their affiliated hospitals (Byrne et al., 2010). OpenVisa meets the *meaningful use* standards and provides an affordable and usable solution that is community supported.

OSCAR and OpenVista have different business models. While OSCAR is mainly sponsored by educational institutions and non-for-profit organizations, OpenVista is developed by MedSphere a company that leverages it for providing integrated solutions to clinics and hospitals. Nonetheless, both systems rely on communities of interested users, developers, sponsors and commercial partners to provide an ecosystem that drives the development of the software and also provides help and support for its users and operators. Such successful large-scale collaboration between heterogeneous parties has attracted a lot of research from different disciplines. The puzzle of community driven software development is not resolved yet. Our goal is to uncover an important piece in the puzzle of knowledge creation and digital production of healthcare communities.

2.3.4 The Importance of Professional Users of Healthcare IT

Implementing IT in the hospital or the clinic brings change into a long established profession. Medicine is perhaps the oldest profession on earth with practices rooted in a long human tradition. In their study of the sociological aspect of medical records, Berg and Bowker (1997) argue that the implementation of electronic records is not merely a technical problem of designing and implementing the appropriate interface but rather: “When it is acknowledged that the medical record is interwoven with the structure of medical work in fundamental ways, and that different medical record systems embody different notions of how work is organized, different modes of configuring patient bodies and so forth, we are in a position better to understand and intervene upon the issues at stake” (Berg and Bowker, 1997, p. 532).

Although the implementation of health IT touches the interests of several groups such as medical practitioners, management, staff, insurers, patients and even politicians (Mintzberg, 2002), the importance of satisfying the needs of physicians who are the key users of the technology has been confirmed by research in different disciplines. For example, Lapointe and Rivard (2005) studied the implementation of three EMR systems in three hospitals. In one case, they found that physicians' resistance to the system was triggered by its slowing down their process of carrying out medical procedures. Physicians' resistance was passive at the beginning but started to become active when the distribution of power between physicians and nurses was affected-physicians felt that the system is making them do more clerical tasks. Baron et al. (2005) highlighted several key decisions they made when implementing an EMR system in their clinic. One key decision was not to impose any additional requirement on the physicians because of the new system. "We operate under the assumption that the physician is the most skilled, and most expensive, person in the office and should only do what no one other than a physician could do" Baron et al. (2005, p. 223).

Professional users (mainly doctors) of open-source health IT face a dilemma: at one hand they are highly trained and highly paid professionals who are in high demand. They lack the material incentives to spend time on off-job activities. Moreover, most doctors lack the technical knowledge and expertise that is necessary to contribute to the development of the software. On the other hand, the involvement of doctors in the community is necessary to ensure that the open-source software meets their needs and work requirements. Adopting an open-source health IT mandates different responsibilities that acquiring a commercial system, mainly because of the lack of single party in charge of maintaining and supporting the system. We build on this dilemma in this thesis and aim to uncover the role that professional users of open-source health IT play in contributing to the community.

Much of the research on communities of innovation and open-source focused on the final production of the collaborative effort of community members such as source-code in open-source projects and articles in wikis. However, it is important to note that the successful production of public good is a consequence of the capability of communities to enact and sustain organizational structures that allow members to engage in multiple activities leading finally to the production of common goods and knowledge. A fundamental question about communities of innovation is to understand the participation dynamics in the community. Toward answering this question we tackle three interdependent studies:

1. How does knowledge emerge and evolve from online interactions? What is the structure that enables collaboration in online settings? There are different interaction patterns in online communities. For example, few members can act like gurus providing the most help and advice for others. Alternatively, interaction can be more democratic where many members collaborate toward providing a solution or solving a problem. In the context of health IT OC, what remains unsettled is how requests for help will be answered in a community where knowledge is divided between professional users (with deep knowledge of medical work) and the techies (with deep knowledge of

programming aspects). We examine the interaction patterns among members along with the microstructures enacted by these interactions.

2. How does the community put together individual members' contributions into the final digital artifact? Recent research emphasized the importance of considering the changing, morphing and fluid nature of online communities (Faraj et al., 2011). In such entities, roles are dynamics and independent from professional affiliation. For example, doctors can assume programmers' responsibilities and programmers can engage in medically oriented discussions. Leadership is seen as emergent and more of a role-in-the-making (O'Mahony and Ferraro, 2007). Those unique dynamics are one main differentiator of online communities from traditional organizational settings and are the motor of production in online settings.

3. How does the community grow and evolve overtime creating and sustaining its collaborative structure? How do online communities grow overtime? How do members establish and sustain mutual relationships online? There are various incentives for members to join online communities. Those incentives include both utilitarian and communitarian motives (Preece and Shneiderman, 2009; Wasko and Faraj, 2005; Butler et al., 2013). However, we do not know much about how members establish relationships within the community once they join. Given the diverse membership that is spatially and temporally fragmented (O'Leary and Cummings, 2007). This is an important question to ask because although online communities appear as anarchic collections of individuals largely devoid of a formal authority, their members develop strong group norms and leadership structure (O'Mahony and Ferraro, 2007), successfully generate information goods, and outperform traditional organizations in many domains (Tapscott and Williams, 2008).

3 Research Design

Research in organization science has established that knowledge creation is a central aspect of organizations (Grant, 1996a) and of communities (Brown and Duguid, 2001). Knowledge production is considered to be the main outcome of communities of innovation. Members edit articles in wikis (Ransbotham et al., 2012), answer posts in forums (Johnson, 2010), and write code in open-source communities (Lerner and Schankerman, 2010). We take a knowledge perspective to understand the patterns collaboration leading to knowledge creation and integration in open-source health IT communities of innovation and to understand the role of various members in the process.

One advantage of community-driven development is the diversity of community membership and the ability to integrate knowledge from members who have different backgrounds and experiences, skills and expertise (Raymond, 1999). If Hayek (1945) is right about the equal distribution of knowledge among members of society then the more diverse a community of innovation is, the better it will be in creating knowledge and producing innovation. This premise has been validated in the context of problem solving via contest and prizes where a solution is more likely to be found when the problem is broadcasted to search for more participants (Jeppesen and Lakhani, 2010). However, contrary to the structure of broadcast search where members have equal incentives to participate (same prize), members in many other communities may have different motives to contribute to the community (Ardichvili et al., 2003; Lakhani and Wolf, 2005; Roberts et al., 2006). In particular members of health IT communities belong to different professional affiliations (e.g. programmers, developers, doctors, nurses, administrators) and have different roles in the community.

The previously stated points in the previous chapters entail that the characteristics of an online community of innovation are situated in the context in which the community operates. We take the advantage of studying such a community in the very particular context of Health IT. We propose to study a community of innovation in healthcare in charge of producing an open-source health IT. The software is shaped by the interests of different members in the community including health practitioners and other members. Furthermore, the community is distributed and the software is used in various Canadian provinces with different language, regulatory, and institutional settings, yet the project is kept unified.

Medical practitioners (mainly doctors and nurses) are the main users of healthcare IT. As such, they are also important members in health IT communities of innovation. Those users are highly paid professionals with busy work schedules who do not have the technical skills to participate in software building. Yet doctors are the most important stakeholders because they use the software in their daily work. This is a unique setting that allows for extending research on open-source beyond the idiosyncrasy of developer-based communities and for examining the implications of open innovation communities in transforming the profession of delivering healthcare.

We divide our exploration of the dynamics of interaction within the community that lead to knowledge creation and the production of open-source software into three interrelated research questions. First, we take a structural perspective and examine the effect of the structural characteristic of members in the community along with other determinants on their knowledge contribution. Second, we consider the community as a whole and examine the changes of discourse overtime and try to link it to software evolution and artifact creation. Third, we put the two previous questions together and explain how the community is able to evolve and grow the structure that facilitates knowledge creation and digital production. We outline here the data collection procedures and highlight the methodology. Figure 2 illustrates the conceptual framework of the thesis.

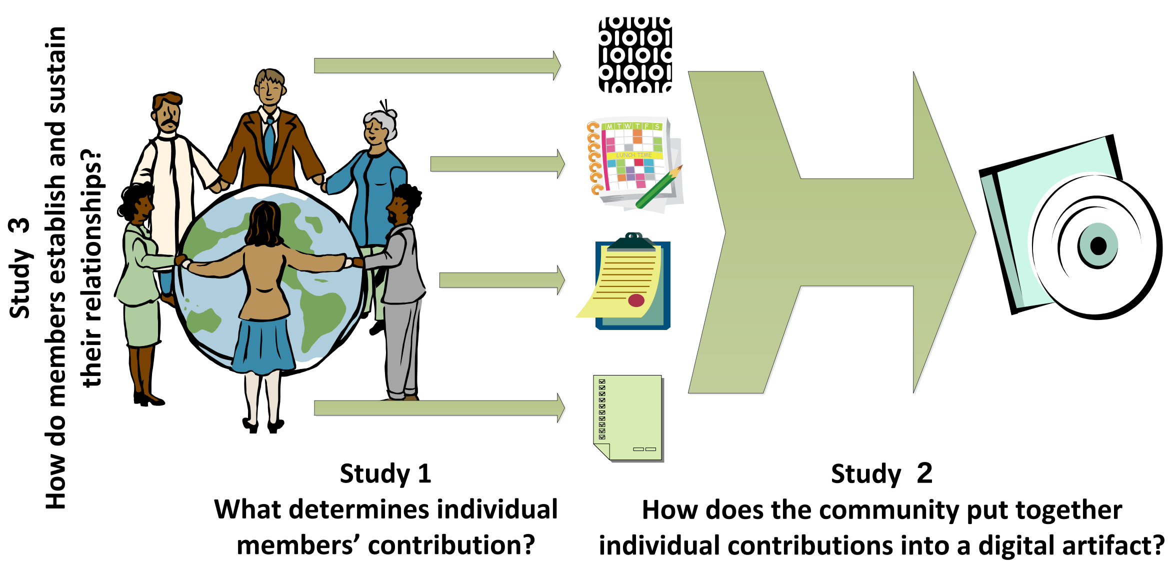


Figure 2: Conceptual Framework

3.1 Research Sites and Data Collection

We focus our attention on the Canadian context by studying the community of innovation centered on an open-source Electronic Medical Record (EMR) named OSCAR This community is primarily Canadian, has been in existence for a decade, and has developed an EMR that is rapidly diffusing (currently used by over 1,500 Canadian doctors to follow over a million patients). The OSCAR EMR is freely available open-source software and is gaining market share against commercial products typically costing $25,000 per year. Given the complexity of such software, the mission-critical nature of patient records for solving Canadian and world health issues, and the fact that the vast majority of users (family doctors) are not computer savvy, this success is unusual and significant. Contrary to other open-source projects, this software is embedded in a community of doctors rather than a community of programmers. The community is active both face-to-face via user meetings and online via discussion forums.

The goal of this research is to obtain a comprehensive in-depth understanding of community-driven healthcare IT innovation. We focus primarily on the OSCAR's online community and members' interactions within it. One advantage of open-source online communities from a research perspective is that community data is self-archived. Members' interactions and contributions are documented as well as software changes and evolution. We have relied on other data sources and informants that shaped our understanding of the phenomenon including becoming members of the OSCAR community and participating in both online and offline activities. Before describing in more detail how data is collected and will be analyzed, we provide a confessional account on how we got involved in the project.

3.1.1 Confessional Account

I have immersed myself[1] in the OSCAR project and its community from the start of my doctoral studies granting me an in-depth understanding of the OSCAR software and the community driven development model. I have also built personal relationships with clinicians, developers and other community members which helped me gain access to various sites and also get more reliable information from members.

I did not have a predefined theoretical lens when I was first engaged in the study. With time however, the theoretical perspective of community-driven knowledge creation and the role of that process in shaping the evolution of the software crystalized in my thinking during the first two years (2009 & 2010). The later encounters and interactions with community

[1] I am using the third person pronoun (we) across the thesis. Only in this section which is much more personal, I use the first person pronoun

members (2011 & 2012) were used to sharpen my understanding of the phenomenon under the knowledge perspective and to frame the three studies on community dynamics. During the course of my studies and before I engaged in my doctoral data collection, I did the following:

- Visiting the development site: I have visited the main development site at McMaster University in Hamilton twice. I had the chance to talk to lead developers and other interns working in the core development team in charge of main development and maintenance of the OSCAR project including Dr. David Chan the visionary of the project and Jay Gallagher the lead developer.

- Contributing to the community: I have been a member of the OSCAR community since 2009 receiving and following community conversations. I also made several contributions to OSCAR including: writing a tutorial on hosting OSCAR in the cloud on Amazon Web Services, compiling a bootable live disturbing of OSCAR on a USB drive, and writing a page in the OSCAR manual. These contributions allowed me to get a hands-on experience on community-driven development and gave me the chance to interact with many members of the community.

- Attending user group meetings: I have also traveled and attended a few OSCAR user group meetings: in Montreal, Hamilton, and Vancouver in the years of 2011 and 2012. During those meetings I documented the discussed topics and followed the interests and contributions of different groups of users. I had the chance to get to know community members in person which gave me a better understanding of the various groups, key persons, and also the process of evolving the OSCAR software.

- Establishing relationships with OSCAR Service Providers (OSPs): OSPs are software companies that make money by providing OSCAR installation, support and training services to their clients. I established a good relationship with one OSP based in Montreal that provided me with information about OSCAR growth in the city and the various clients' cases, common issues and experiences with OSCAR. I also had the chance to talk to four other OSPs in several occasions.

- Observing a clinic adopting OSCAR: over the last four years, I followed the implementation of OSCAR and the switch from paper records to electronic records at a medium-size walk-in clinic in Montreal. The clinic is affiliated with McGill University and serves approximately 30,000 patients per year. Its staff complement includes 38 physicians, 15 nurses, and 10 clerks. Embedded in the health network of the university, the clinic receives and refers patients to other health centers affiliated with the university health network.

3.2 Data Collection

3.2.1 OSCAR Community

OSCAR has a vibrant community of users. Interaction among members is done mostly online via five different mail-lists on SourceForge[2]. Figure 3 shows an example thread from the community. Those mail-lists are accessible either from a web browser or via email clients for people who subscribe. We have collected archival data from the online mail-lists of the community including all posts and threads from all mail-lists from January 2006 until January 2013. Collected data includes network data, textual content of conversation, and members' characteristics. Figure depicts the data structure used to model the community. The collected data includes timestamps for all events happening in the community including thread initiation, message postings and replies. The threads are parsed to their essential components (message content, timestamps, senders, receivers, attachments, forums, etc ...) and stored in a database for further queries. Table 9 summarizes the volume of data collected in each mail-list.

Mail-list	**Members**	**Threads**	**Messages**
Advanced users	91	680	2823
BC Medical Office Assistants	99	321	715
BC Users	604	3497	14944
CAISI Users	25	40	69
Developers	699	2293	8366
Ontario Users	9	8	11
Other Users	55	86	18

Table 1: Data collection from online mail-lists from January 2006 to January 2013

3.2.2 Software Metrics

Software metrics are quantitative measurements of computer software that has been proposed as programming evolved into a profession involving large investments and critical stakes (Lincke et al., 2008). Extensive metrics has been proposed over the last three decades including metrics to measure performance, evolution, quality and cost (Fenton and Pfleeger, 1998).

[2] `http://sourceforge.net/mail/?group_id=66701`

[Oscarmcmaster-bc-users] Anticoagulation Bridging Therapy. 5 messages

John Robertson <johngmrobertson@shaw.ca> Sat, Dec 17, 2011 at 3:05 PM
Reply-To: The OSCAR UserGroup list <oscarmcmaster-bc-users@lists.sourceforge.net>
To: A list for advanced OSCAR user discussions <oscarmcmaster-advanced-users@lists.sourceforge.net>, oscarmcmaster-bc-users <oscarmcmaster-bc-users@lists.sourceforge.net>

This is a Fraser Health Form put out by the pharmacy. It is supposed to be for both surgeries and invasive procedures (so likely interventional radiology, endoscopy, and the like). It is on "FormFast" which means there is no way to get an electronic copy, I have Used CS4 to render a good quality image with a low file size.

Back on September 24th Peter shared some code with us for using calendars in eforms. I have had a chance to use it 9 times in this form. I know there is a better way to do this form as the pregnancy calculator shows the way. When I have time I will try to understand that form better, and I may improve this one. As it stands, here it is with all 9 calendars. JohnR.

You can find it at:

http://oscarcanada.org/oscar-users/emr-resource/eform/eforms-in-development-beta-testing/anticoagulation-bridging-therapy

---

John Yap <john-yap@shaw.ca> Sat, Dec 17, 2011 at 3:35 PM
Reply-To: The OSCAR UserGroup list <oscarmcmaster-bc-users@lists.sourceforge.net>
To: The OSCAR UserGroup list <oscarmcmaster-bc-users@lists.sourceforge.net>

JohnR,
I had tried the calendar feature previously, without success. Seems I need all the calendar-related *.js files that go with it. Where can we find those?

JohnY

Peter Hutten-Czapski <phc@srpc.ca> Sat, Dec 17, 2011 at 4:02 PM
Reply-To: The OSCAR UserGroup list <oscarmcmaster-bc-users@lists.sourceforge.net>
To: The OSCAR UserGroup list <oscarmcmaster-bc-users@lists.sourceforge.net>

the calander.js ships with Oscar

John Robertson <johngmrobertson@shaw.ca> Fri, Dec 23, 2011 at 12:20 PM
Reply-To: The OSCAR UserGroup list <oscarmcmaster-bc-users@lists.sourceforge.net>
To: A list for advanced OSCAR user discussions <oscarmcmaster-advanced-users@lists.sourceforge.net>, oscarmcmaster-bc-users <oscarmcmaster-bc-users@lists.sourceforge.net>

I love this community! David P has taken my form and added the functionality I was envisioning. Now when you put in the procedure date, you can click a button and it calculates all the other dates for starting and stopping anti-coagulation therapy. I have tweaked the image slightly to avoid overlap between the date box and some text. I have also changed some formatting slightly. The new html and png have replaced the old one at the same address (see below). If you have uploaded the old image, it would still work, or you could delete it and replace it with the new one. JohnR.

Figure 3: One example thread from the community as seen in an email client

In the context of open-source software, an important metric is community contribution that leads to software evolution. Because open-source development is decentralized, several developers can contribute code to the project that is typically hosted in a platform that allows concurrent changes by different parties. One metric we look at here is the number of code commits to the software repository. That metric can be examined in any time unity

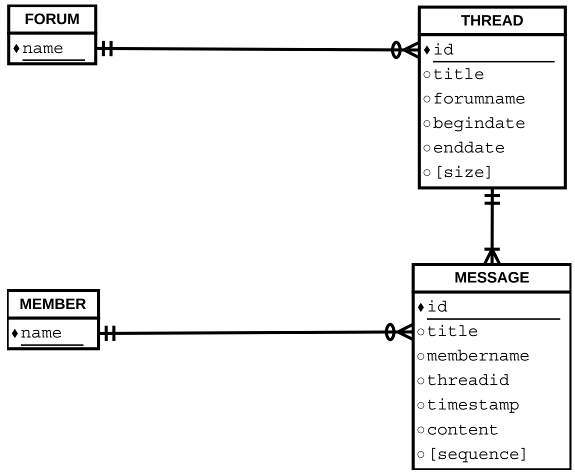


Figure 4: Data structure used to model online interactions in the community

including daily, weekly, and monthly units. The number of commits is a metric that measures the activity within the community to develop and evolve the software. This activity may or may not be accompanied by general interest in using the software. Figure 5 show monthly commits to the OSCAR project from January 2006 to January 2011[3]. Another metric we look at is the number of open and closed tickets in the bug and feature requests tracker of the project[4].

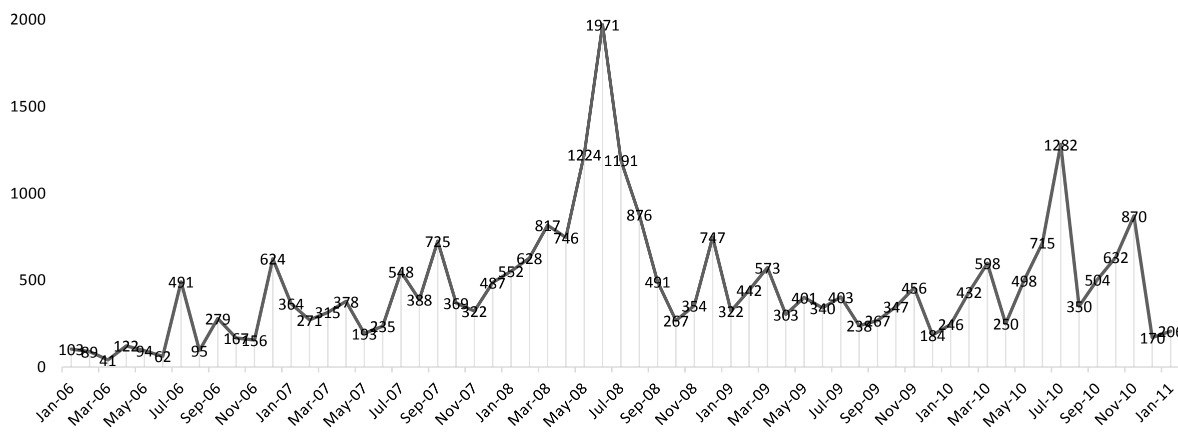


Figure 5: Monthly code commits to the software repository

[3]In 2011, the OSCAR project has migrated from SourceForge to a private GIT repository that is more tightly controlled by the core team. Unfortunately, we were unable to obtain software metrics data beyond 2011

[4]`http://sourceforge.net/p/oscarmcmaster/bugs/stats/`

3.3 Methodology

Each of three studies on community dynamics is self-contained and use different parts of the dataset with different analytical methods. In this section we highlight the commonalties among the studies in terms of methodology and focus on the big picture of how the community dataset is analyzed to study dynamics. In particular, this thesis combines three analytical methods: social network analysis of online interactions, computational content analysis of members' communication, and time series analysis of software evolution.

In communities of innovation relationships are formed via participation in collaborative activities that happen primarily in the cyberspace. One predominant method is to study the communication networks formed by these interactions using social network analysis techniques (Knoke and Yang, 2008; Monge and Contractor, 2003). Traces of members' engagement in activities can be found in different places. Analyzing the social networks of various community activities sheds light on the participation structure in the community. In particular the various roles of members can be examined in light of their structural characteristics.

Research on online communities has been criticized as content agnostic (Hansen, 1999) as it focuses mostly on relationships among social actors while ignoring the content conveyed via these relationships. Content used to be hard to collect and to analyze. Fortunately, this is changing with the advent of online communication in which content is always archived and the development of computerized text analysis methods (Lebart et al., 1998). Recent research has already confirmed the importance of considering content in addition to social relationships in online communities and virtual settings (Yoo and Alavi, 2004; Danescu-Niculescu-Mizil et al., 2013).

Another data component that has been often ignored in social and organization science research is *time*. Discounting time leads to overlooking the dynamics of change in social relationships and organizational networks (Ahuja et al., 2012; South, 1995). Recent advances in statistical methods have made the incorporation of the temporal dimension into the research model possible. Longitudinal analysis techniques include: growth modeling, multilevel modeling, survival analysis, and time-series analysis (Singer and Willett, 2003).

Research in organization science has established that knowledge creation is a central aspect of organizations (Grant, 1996a) and of communities (Brown and Duguid, 2001). Knowledge production is considered to be the main outcome of communities of innovation. Knowledge can be tacit in form of exchanged discussions or codified in form of a digital artifact. Regardless of the form, knowledge creation in online communities is embedded within social interactions that are dynamic and changing and conveyed via written communication that is rich in content. Understanding the dynamics of change in such settings require being able to evaluate those three components.

Before delving into the details of the three studies, the next Chapter provides a detailed description of the OSCAR project. Understanding the history and evolution of the OSCAR project is important to put the rest of the thesis in context.

4 The OSCAR Project

OSCAR[5] (Open Source Clinical Application Resource is an Electronic Medical Record (EMR) system developed at the Department of Family Medicine at McMaster University and deployed in multiple clinics in Canada and worldwide. OSCAR EMR features a web-based interface that allows multiple users to interface with the system simultaneously through different browsers and platforms. The web interface also allows accessing the system over the Internet through a secure connection. The number of sites adopting OSCAR for their practices has steadily increased since its release in 2002. Today OSCAR is used by more than a thousand sites that serve more than one million patient records. OSCAR gained substantial market share in major Canadian provinces reaching third place in Ontario in 2012 (OntarioMD, 2012a). In addition to independent clinics, OSCAR is the EMR of choice of many teaching faculties in Canada (OSCARCanada, 2012a). Finally, OSCAR has been adopted by communities of users if different countries including a strong presence in Argentina, Brazil, and Poland (OSCARCanada, 2012e).

4.1 Hardware and Software Architecture

As open-source software, OSCAR is capable to run on different hardware platforms and configurations. The typical hardware configuration usually consists of a server with data duplication mechanisms such as RAID-1 or a secondary server that is synced with the main one. Clients many be any number of terminals such as laptops and desktops are connected to the server via local network or the Internet if the server has Internet connection (Figure 6). Many other configurations are possible as well. The choice of a platform and a configuration depends on the requirements of the site.

Individual users who only need a local installation of OSCAR can benefit from a prepackaged distribution for Microsoft Windows that can be installed on a PC or a laptop via a single click installer. This method, however, does not support multi-site access nor does it include

[5] `http://www.oscarcanada.org/`

backup and scalability capabilities. Other options are to host OSCAR in a small independent server and access it via local network. OSCAR appliances are available for small-sites and users who need a prepackaged and configured OSCAR server. They offer a turnkey OSCAR solution that is relatively inexpensive to acquire and easy to operate. Medium and large sites and advanced users would invest in their hardware, this typically include a main server, a backup server that duplicates the main server and provides a failsafe option, LAN router and access points, Internet access device such as a DSL modem, in addition to the PCs, laptops and printers that are used by the medical staff.

An emerging hardware configuration is to host OSCAR in "the cloud" by leveraging virtualization technology. This is done by installing OSCAR in a virtual server that is hosted at a big data center managed by a professional service provider. The platform as a service model (PaaS) relives the client from the cost and effort of buying and operating the hardware. Images of OSCAR server are available for popular virtualization platforms such as VMWare. OSCAR has been demonstrated to work in the Amazon EC2 cloud service. Privacy issues are more relevant in the cloud solution because data is hosted at a third party side and is accessible on the Internet.

OSCAR is built around a client-server model where different clients access a central server over local network or the Internet. OSCAR is written in server-side Java. It features a web interfaces written in HTML, JavaScript, Java Servlets and Java Server Pages Scripts. The default backend database is MySQL. The default installation of OSCAR runs on the top of Apache Web server with Java Runtime Environment enabled via Tomcat application server. Ubunutu Linux is the preferred operating system although OSCAR runs on other distributions of Linux, Windows and MacOS. OSCAR is modular software composed of various modules that implement different functions.

4.2 Functionality

OSCAR provides a wide spectrum of clinical, administrative and support functions that are comparable to offerings of other EMRs. Clinical functions include electronic chart, prescriptions, labs, consultations, disease registry, measurements, preventions, allergies, and clinical forms. Administrative functions consist of billing, patient management (patient registration and demographics), appointment and schedule management, and electronic document repository. Support functions include system functions such as user management and administrative settings. In addition, they include OSCAR report, a reporting tool that generates various types of reports, OSCAR messenger an email-like communication module that allows users to send and receive messages inside OSCAR, and OSCAR tickler a patient-

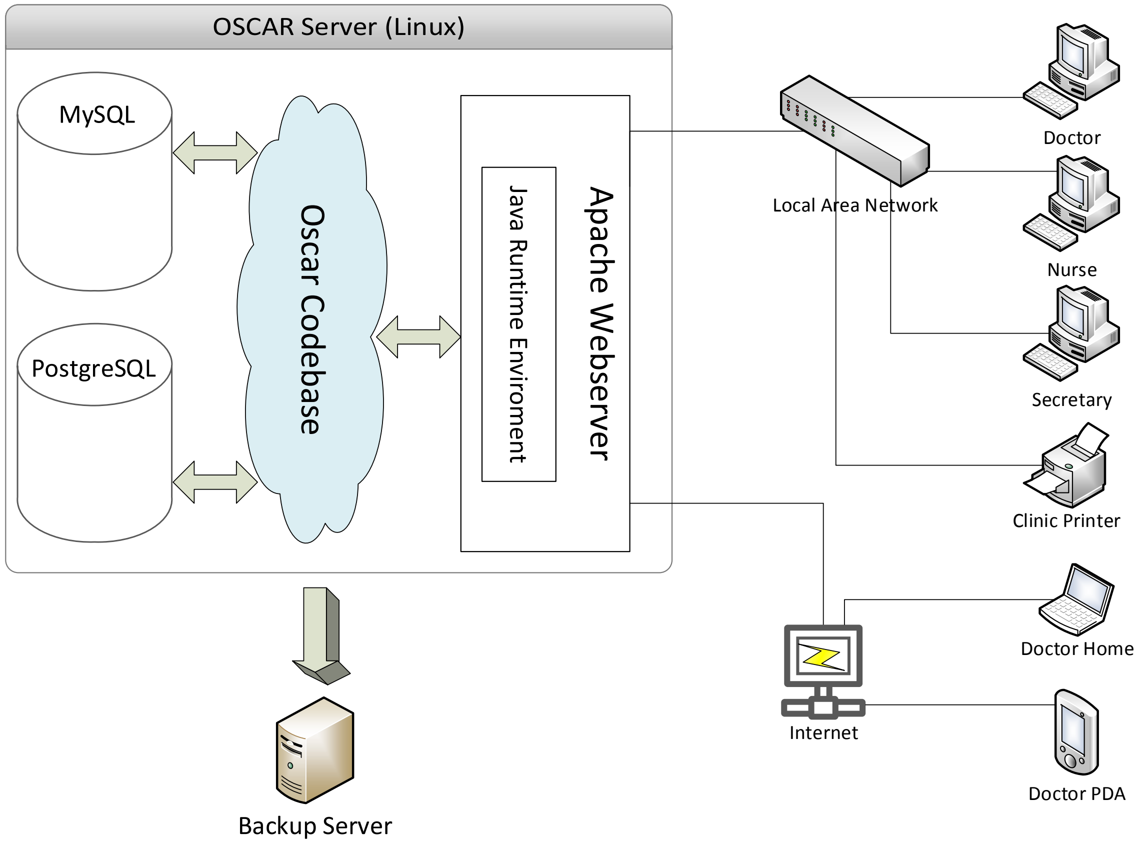


Figure 6: Typical OSCAR Installation

centric reminder and alarm tool. Finally, OSCAR allows the rapid development of plug-in medical templates and forms. Figure 7 demonstrates the various modules of OSCAR and the functions offered in them.

4.2.1 Clinical Functions

OSCAR provides an extensive collection of clinical functions. Although OSCAR's main focus is to provide functions relevant to ambulatory care and family medicine, it provides many other capabilities that are relevant to other specialties. The patient e-chart is the command and control center of clinical function (Figure 8). Clinical note for visits are recorded in chronological order in addition to other relevant medical information such as preventions, prescriptions, forms, labs, measurements, consultations, allergies and medical history. All medical notes are digitally signed before saved allowing for tracking the different practitioners that oversaw the patient.

OSCAR's prescription capabilities include integration with the Canadian Drugref database

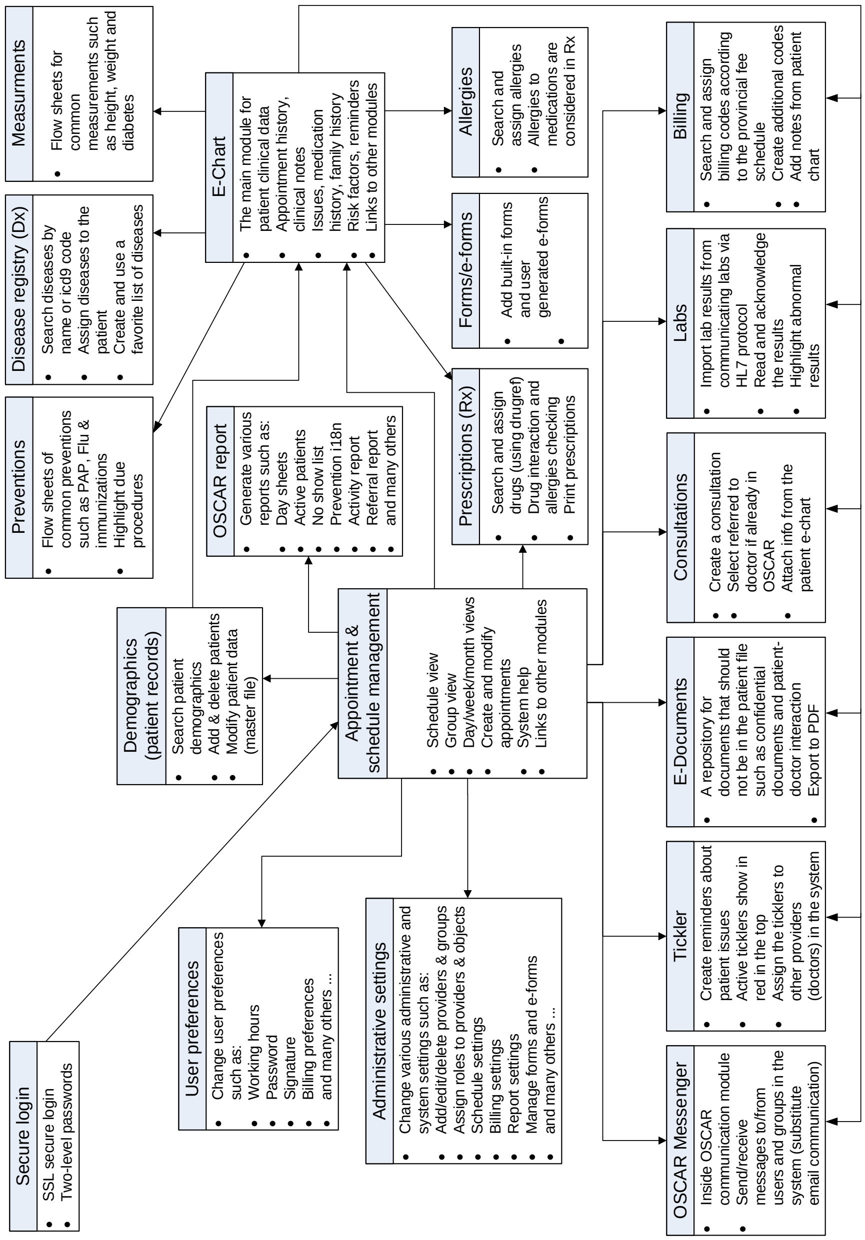


Figure 7: OSCAR modules and functions, arrows indicate logical flow

allowing for updating the list of drugs automatically. In addition OSCAR implements several decision support options such as drug-drug interaction, rental dosage adjustment, and drug allergy alerts. Furthermore, OSCAR is integrated with MyDrugRef a social network of trusted physicians allowing them to post related information about the drugs and share their prescriptions and decisions with their colleagues. Finally, disease registry is included to provide support in analyzing population health (OSCARCanada, 2012d).

OSCAR has extensive chronic disease management (CDM) functions including preventive care and immunization ta flow sheets for hypertension, diabetes, asthma and others. OSCAR has a built-in antenatal care record and planner. In addition, OSCAR's CDM include forms, reports and audit tools that can be customized for patients' preferences. Finally, OSCAR's CDM can be optimized to maximize desirable outcomes by continuously monitoring and alerting for screening and best practices interventions (OSCARCanada, 2012c).

OSCAR can import lab results via HL7 connection from hospitals and private laboratories. OSCAR automatically directs incoming labs to the responsible physician and flags abnormal results. Imaging results can be imported in a similar fashion. OSCAR has an intuitive referral system that allows pulling relevant information from the patient chart and including it in the consultation. Finally, OSCAR built-in functions are easily customizable with templates and e-forms (OSCARCanada, 2012c).

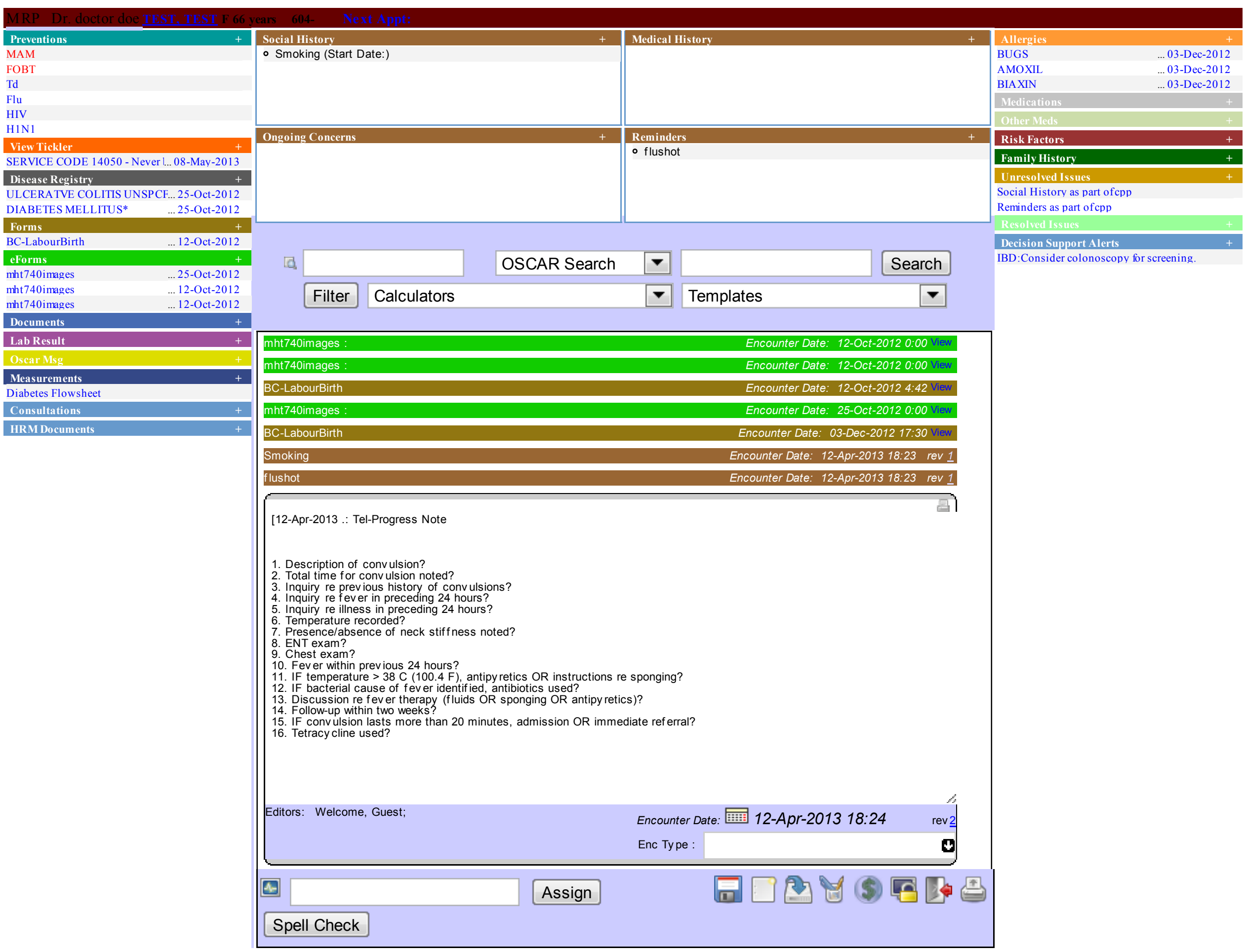


Figure 8: OSCAR electronic patient chart

4.2.2 Administrative Functions

OSCAR has a comprehensive demographic module allowing for registering and accessing patients' information. It also links to patient appointment history, billing and document. A search interface is provided to look up patients quickly. The appointment interface shows the daily appointments in the clinic and has a calendar that can be used to look up previous or allocate future appointments. Every appointment has a status. In addition the appointment links to the patient chart, demographic, prescriptions and billing.

The billing module in OSCAR conforms to the Canadian codes and standard and currently supports billing in the provinces of Ontario and British Columbia. The report interface comes handy when looking up for statistics. It can generate day sheets, chart lists, no show appointments, consultations and visits. It also generates reports on demographics, visits, laboratory requisitions and many others. In addition it features a Query by Example mechanism that makes it easier to generate reports from the database.

4.2.3 Support Functions

The administrative settings have various configuration options that include: adding, modifying and deleting users, groups and providers, appointments and schedule settings, billing options, report settings, e-form management and many other settings.

In addition to built-in functions, OSCAR allows the rapid development of medical forms to extend the built-in functions and forms. An e-form is simply an HTML form with additional tags that allow it to communicate with the database pulling and posting information there. Once an e-form is uploaded, the e-form module parses it and creates a table in the DB with corresponding fields. This translation is seamless, and the only requirement to write e-form is basic knowledge of HTML in addition to learning the specific OSCAR tags. This feature helps users customize the system without much technical knowledge. In addition, the "e-form generator", a community-developed module, provides an easy WYSISYG tool to develop e-forms based on scanned medical forms. Finally, users can share their e-forms in an online repository making it very easy to pick from a wide selection of user-created e-forms.

OSCAR has a secure login module that supports authorization using user names and passwords, authorization using user roles, and privacy using an encrypted communication channel between the OSCAR server and the client. Every user in OSCAR has unique user name and two passwords and is assigned a role that defines his privileges in the system. The list of roles includes: admin, doctor, secretary, clerk, nurse, vaccine provider and others.

The list can be tweaked from the Administrative preference menu.

Oscar has an internal messaging system that allows users to exchange message in OSCAR. The exchanged messages can draw patient data from the database. This eliminates the need for external data exchange channels such as email. Indeed the messaging interface resembles a web email interface. In addition new messages show up in red on the main screen which draws attention to them. Ticklers are internal reminders that are assigned to a user and a patient. They remind the user to perform a specific task on the patient such appointment booking, procedures and bills. They can be set up as personal reminders or reminders to other staff. Ticklers eliminate the need to store to-do lists. They are internal and can be share among staff.

4.3 Development Model and Ecosystem

OSCAR is based on community driven development model where multiple contributors and institution contribute to the project development. The contributors include the main development team at McMaster University, large institutions such as McGill University, user groups and independent clinics and doctors, and independent service providers who are interested in making business by providing support to OSCAR clients (OSCARCanada, 2012d). The core development team is located at McMaster University that supports and coordinates the development of OSCAR. The core team is responsible of maintaining the code repository and overseeing other members' contributions (OSCARCanada, 2012d).

4.3.1 Community Involvement

OSCAR has a vibrant community of medical users and professional developers. Most users are members of OCUS (OSCAR Canada Users Society) that is a non-for-profit organization that represent the interests of the user community in the OSCAR project (OCUS, 2012a). The community and maintains an active mailing list with an average of seven messages per day. The community holds regular national and regional meetings where members can meet face to face and exchange their ideas and expertise. The community also contributes to the OSCAR project. These contributions range from simple additions such as templates and medical forms to sophisticated enhancements and plug-ins.

4.3.2 OSCAR Service Providers (OSPs)

OSCAR Service Providers (OSP) are independent computer and software companies who are interested in making business by providing installation, training and support services to OSCAR users. OSPs are also encouraged to contribute fixes and improvements to the OSCAR project (OSCARCanada, 2012d). Because OSCAR is open-source, OSPs do not have the right to sell licenses to use the software. However, they can sell related services such as installation, training and support. They can also sell software extensions built around OSCAR. Currently there are over fifteen OSPs in Canada and their number is growing rapidly (OSCARCanada, 2012d).

4.4 Does Oscar Realize the Benefits and Respond to the Challenges of Open-Source?

4.4.1 Cost of OSCAR

High cost of EMRs is perhaps the biggest challenge to their wide adoption in small and medium size clinics. Based on the values of academic research and knowledge sharing, OSCAR is not designed to profit the developers. However, value generated processes are required to enlist interested third party developers and service providers. The availability of service providers is a must to gain a critical mass of healthcare users to adopt and continually use OSCAR (OSCARManual, 2012a). Although the OSCAR itself comes at no cost, there is a cost associated with getting support as well as acquiring the hardware infrastructure and maintaining it. Therefore, the cost depends on the type of installation and the size of the site. Typical costs for a small clinic as provided by Indivica (Indivica, 2011), one OSACR support provider based in Ontario include: $1,000 for server hardware and $2,700 for annual support. Other tasks are billed per hour. Here is cost estimation for a typical installation:

1. Servers: Depending on the clinic size server architecture is suggested. The cost is $1000 for a server suitable for a small clinic (1-5 practitioners), $1500 for a server suitable for a medium size clinic (6-15 practitioners), $3500 for a large clinic (up to 50 practitioners) and $7500 for a very large site (up to 100 practitioners).

2. Annual support: The annual support includes telephone and email support, one software update per year and 24/7 emergency coverage. The cost is $2,700/year for the first physician per clinic and $1,500 for each additional physician per clinic.

3. Training: Recommended staff training is 2 x 3 hour sessions to cover basic functionality and 2 x 3 hour sessions to cover basic billing. Recommended practitioner training consists of 2 x 3 hour sessions to cover basic usage. Training is billed by hour at $125.

4. Other tasks: Data conversion from other EMRs, installing OSCAR into an existing architecture, and other tasks are billed by hour at $125.

4.4.2 Support and Maintenance

There are several OSCAR Service Providers (OSPs) that specialize in providing OSCAR installation, maintenance and support in Canada. In addition, many IT consultants provide such services for small clinics. In addition, the user online community is an excellent starting point for looking up local companies and consultants; in addition, it provides support and help for a wide range of requests and questions. OSCAR's active user community, the availability of independent companies supporting it, and its affiliation with large teaching institutions such as McMaster University and McGill University are all factors contributing to the increasing visibility, recognition and adoption of OSCAR in Canada.

4.4.3 Community Contribution

The community of OSCAR users contributes actively to the project. We outline here two examples of users' contributions:

1. The e-form generator (OCUS, 2012b) is a WYSIWYG module (What You See Is What You Get) that allows OSCAR users to quickly build new medical e-forms by scanning paper forms and overlaying data placeholders such as textboxes and checkboxes over the scanned form (Figure 9). The generator then generates HTML code that can be added to OSCAR as an e-form. The generator cuts down the development time and reduces the difficulty of writing new forms considerably. Furthermore, many users share the e-forms they develop. As a result a wide variety of medical forms is freely available for OSCAR.

2. The OSCAR manual (OSCARManual, 2012c) utilizes wiki technology allowing users to contribute by creating and modifying the content of the manual. As a result the OSCAR manual is a collaborative product of both users and developers available for the public under the Creative Commons License. The spirit of collaboration and sharing in the OSCAR project extends beyond the source code to support and documentation. Those two contributions become two essential features of OSCAR. The OSCAR manual is the place to get information about OSCAR online. It is a very extensive manual with new material added and updated continuously. The e-form generator has become an integral part of OSCAR and comes included with default installation of OSCAR. Physicians can share their e-forms online and import other people e-forms to their

e-form library. The e-form generator turns out to be a powerful addition to OSCAR that is now one of its important built-in features (OSCARManual, 2012b).

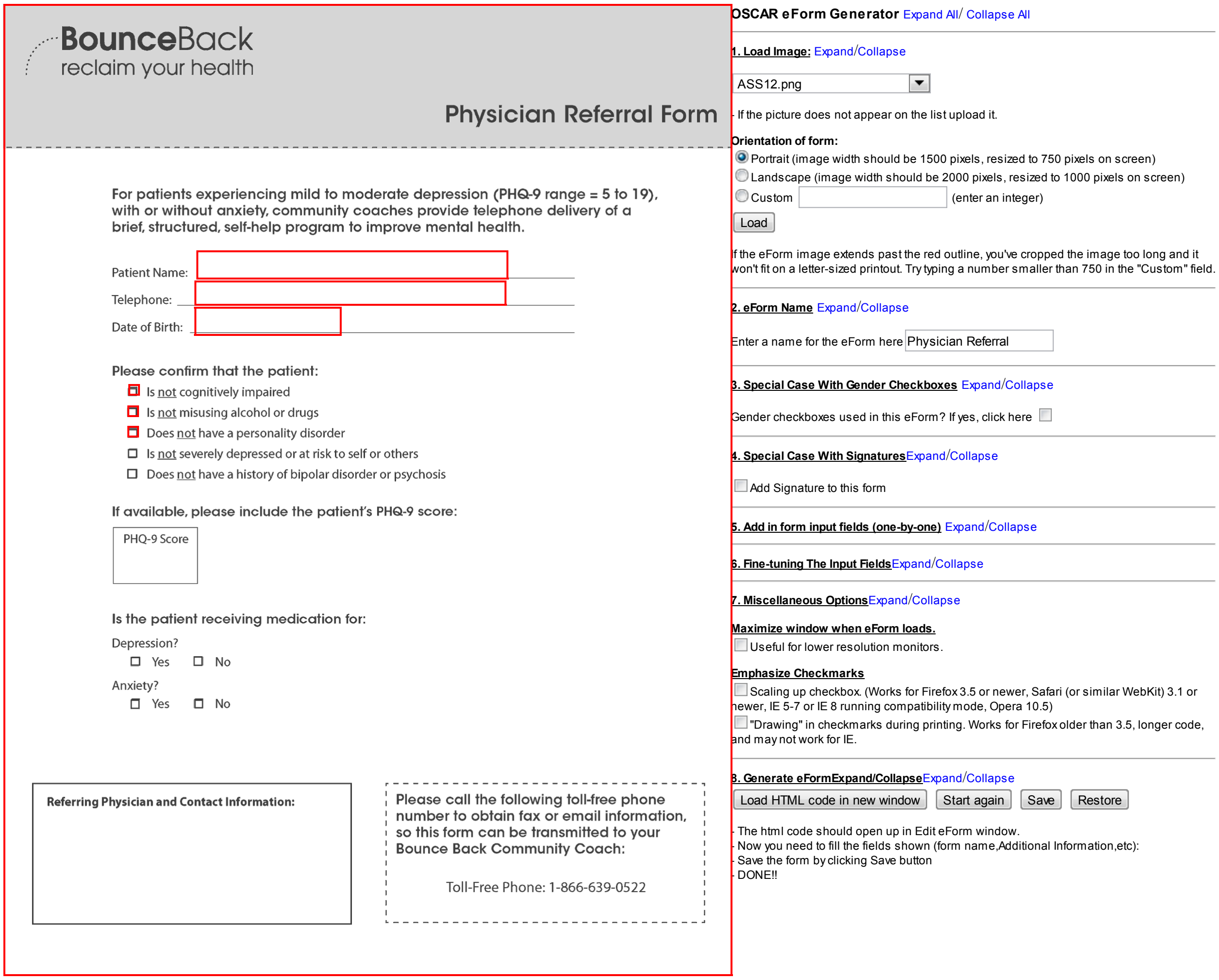


Figure 9: Interactive electronic form creation using the e-form generator tool

4.4.4 Open Standards & Ports

OSCAR complies with different standards and certifications. OSCAR meets the requirements of open source defined by the Open Source Initiative. The process of development complies with ISO 9000 standards of software development. Furthermore, OSCAR has passed provincial EMR certification in Ontario and Quebec. OSCAR has passed all rounds of Conformance testing in Ontario, including the most recent and rigorous CMS Specification v3.0 from OntarioMD (OntarioMD, 2012a,b). OSCAR can import and export clinical data and labs via HL7. OSCAR McMaster has registered for OntarioMD Specification 4.0 Validation Testing, and has achieved ISO 13485:2003 Certification (OSCARCanada, 2012b).

Oscar has an Excelleris interface for lab download and is in the process of completing its interface with British Columbia's Interior Health Authority's Physician Office Integration program so that Regional reports can be directly downloaded. However, in British Columbia there are still tensions between the government's Physician Information Technology Office (PITO) and OSCAR users over patient privacy issues because of PITO's mandate of uploading patients' information to a centralized government database. As a result OSCAR is not certified by PITO in BC however the BC OSCAR user group is independently managing and driving the introduction of OSCAR in the province (OSCARCanada, 2012c).

As a Canadian EMR, OSCAR comes with English and French localization out of the box. OSCAR multilingual support is flexible and allows the same installation to run in different languages depending on the browser language settings. This flexibility allows the Doctor to run OSCAR in English and the secretary to run it in French if they decide to do so. The localization is centered in a central properties file that contains labels and button texts, this allows for easy translation to other languages. Indeed OSCAR has been localized to Spanish and Polish and has considerable communities in South America and Poland.

4.4.5 Availability of Coded Data

OSCAR includes a variety of coded data databases including diseases, drugs, and billing. OSCAR supports the International Statistical Classification of Diseases and Related Health Problems (ICD) with version nine currently in place. OSCAR also supports adding new diseases to the disease registry. OSCAR includes two billing databases and modules for Ontario and British Columbia. The administrative settings allow uploading billing codes and modifying the existing ones. Finally, OSCAR integrates with drugref[6] a drug database containing the Health Canada drug database and a drug-drug interaction reference based on a collaborative open content database. The database contains around 25,000 products which companies have notified Health Canada as being marketed (OSCARCanada, 2012c).

OSCAR includes two billing databases and modules for Ontario and British Columbia. The administrative settings allow uploading billing codes and modifying the existing ones. However, this feature does not change the billing procedure and therefore cannot be used to customize the billing for other provinces. More effort is required to create billing modules that comply with the provincial guidelines in other provinces than Ontario and British Columbia.

[6] http://drugref.org

4.4.6 Security and Privacy

OSCAR takes security and privacy of medical information very seriously. There are three layers of security in order to insure that patients' confidential medical data is safe. The first layer is hardware security that makes sure that even if the server is accessed by intruders the data stored on it cannot be read. Hard disk and database encryption are recommended. In addition, all automatic backups that are done by OSCAR are encrypted so that the data will be useless if accessed accidently by intruders. The second layer of security is connection security that takes care of securing communication between the server and the clients accessing the server over local network or via the Internet. OSCAR utilizes Transport Layer Security (TLS) to encrypt the communication between the server and the clients. In addition to encryption, TLS ensures communication integrity and authentication. TLS protects against eavesdropping breaches and is the widely used approach for secure point-to-point communication over the Internet (Dierks, 2008). The third layer off security is application-level security. OSCAR access is secured via two-password login process. One password is set by the user and the second is set by the system administrator. The passwords themselves are stored in the database in an encrypted form which means that even the system administrator cannot look at them. Failing to enter the correct password for three consecutive times will lock the system for a while for that user.

It is worth noting that all of the previous mechanisms need active participation from the users in order to work. OSCAR encourages physicians to take an active role in promoting best practices for securing their patients' data such as changing default passwords, selecting strong passwords, updating passwords frequently, and not sharing passwords. In addition, setting screen locks, updating antivirus and security software is strongly recommended. Finally, it is advised to store the server in a physically secure place such as a special server cabinet or a locked room in the clinic with an alarm system (Dawes, 2010).

4.5 Summary

The OSCAR EMR case study demonstrates the promise of open-source software in primary care. The main advantage of OSCAR over proprietary EMRs is the reduced cost of acquisition and maintenance. In terms of complexity, OSCAR has a considerable learning curve. However, this training stage is facilitated by the many resources available online and also by the existence of an enthusiastic users' community. Moreover, a huge collection of downloadable e-forms and templates is available online making customization a very easy process. Finally, OSCAR favors open standards and interfaces including support for HL7, ICD-9 & 10, Canadian drug database, and billing codes. More effort is required to integrate OSCAR with legacy

HIT systems of hospitals and pharmacies.

The OSCAR project grew from a small program to organize patients' information written by a techie doctor to a well-rounded electronic medical record system gaining substantial market share in Canada. This growth could not have been achieved without the support and contributions of an active community of users and developers. Alternative models of technology innovation, software development and service providing are needed to address the issues challenging the wide adoption of health IT. Open-source has been identified as an alternative production model for creating goods and services that combines aspects of private investment and common action production models (Von Hippel and Von Krogh, 2003).

Next, we examine the determinants of members' contribution in the OSCAR community.

5 Toward a Local Perspective on Online Collaboration

Social capital is one predominant perspective to study social collaboration. It rests on the premise that the position of actors within a social network is closely related to their performance variation (Borgatti and Foster, 2003). Performance includes various social outcomes such as knowledge distribution, wages, job placement, promotion opportunities, creativity, innovation, political success, social support, and productivity (Aral and Van Alstyne, 2011). For example, the strength of relationships among people is an important determinant of social outcomes. Strong ties and frequent interactions among team members lead them to build mutual understanding of their expertise and achieve better performance (Wegner et al., 1985; Ren and Argote, 2011). On the other hand, weak ties among individuals are shown to be important for reaching opportunities beyond local communities that are not accessible via strong ties (Granovetter, 1973). Combining the two findings, a great influence can be gained by bridging a gap between and establishing strong ties with weakly tied individuals who have complementary resources or information (Burt, 1992).

In this paper, we focus on how the embeddedness of members in an online community of open-source software relates to their knowledge contribution to the community. We depart from previous research by bringing a different perspective to examine position within a network: we focus on the *local position* of members rather than the traditionally considered *global position*. This differentiation has significant theoretical and methodological implications. On the theoretical front, a local view of position implies a more confined and local organization of work in online communities than previously thought. From a methodological perspective, evaluating the local structure of large networks involves different algorithms that have only recently become feasible with the increase of processing power.

5.1 Social Capital in Social Networks

A social network is an apparatus used to model and study social phenomena by focusing on social actors and the configuration of their relationships. Social actors which can be persons, teams or organizations are represented with nodes (Borgatti et al., 2009). Relationships among actors are represented with ties connecting nodes. These ties are also referred to as edges or links. This representation allows for asking various kinds of theoretical questions depending on whether the focus is on the nature of ties, the characteristics of social outcomes and the temporal aspect of the network (Borgatti and Foster, 2003; Borgatti and Halgin, 2011a). Moreover, a social network representation allows for a quantitative analysis of the studied phenomena by leveraging methods and tools developed in multiple disciplines including social science, physics and computer science. Social network analysis (SNA) is the term used to refer to the set of methods and tools used to analyze and study social networks (Wasserman, 1994; Scott and Carrington, 2011).

One predominant perspective to study social network is structural social capital. This perspective focuses on the benefits that social actors acquire by occupying unique positions in their social networks (Cook and Emerson, 1978). Structural characteristics of actors are determinant of social outcomes at the individual level and also at the aggregate network level (Sparrowe et al., 2001). Structural cohesion, similarity and centrality are three sources of power in social networks (Friedkin, 1993). Under this perspective, social actors are seen as rational active agents who realize the importance of ties and exploit their position in the network in order to maximize their outcomes (Borgatti and Foster, 2003, p. 1004). Moreover, it is assumed that network structure is a good proxy for network processes under certain assumptions. The behavior of social actors is reflected in their positions with the resulting social network (Burt, 2010, ch. 2).

Early research on social capital focused on the importance of frequent interactions to create strong ties. In teams, strong ties lead to build shared understanding and joint expertise resulting in better performance (Wegner et al., 1985; Wegner, 1987). Research also examined the nuances of ties in social networks. The pioneering study of Granovetter (1973) examined the effect of weak ties on social achievement. Because strong ties among social actors increase the chance of overlap in their connections, strong ties are less likely to be sources of novel information. Weak ties, on the other hand, allow social actors to gain novel information and resources that are not available to their strong-tied connections. As a result, weak ties are more important in determining social capital than strong ties. Ronald Burt's extensive work on social capital focused on the idea of structural holes. Social actors who bridge between clusters of strongly tied social actors are occupant of structural holes in a social network.

A structural hole allows its occupant to obtain novel and non-redundant information and resources from disjointed clusters in the network resulting in a great influence and brokerage power (Burt, 1992).

5.2 Social Capital in Online Communities

Research has examined the structural characteristics of new forms of organization including online communities in order to understand what makes them succeed in certain domains where traditional organizations are failing (Tapscott and Williams, 2008). By the virtue of their design, online communities have unique characteristics that differentiate them from traditional organizations. First, online communities are built on open membership and voluntary participation and as a result lack central hierarchical authority in contrast to other organizations (Dahlander and Frederiksen, 2012). Moreover, leadership is emergent and temporal leading to changing leadership and sustainability regimes over time (Butler, 2001; O'Mahony and Ferraro, 2007; Ridings and Wasko, 2010). Fluidity is a theoretical construct used to refer to the dynamicity and perplexity of online communities (Faraj et al., 2011).

Studying social capital in online communities can be seen as an extension of studying social capital in *offline* social networks. Social network representation and analysis are widely used to study online communities (Huffaker, 2010; Faraj and Johnson, 2011). However, recent research indicates that this extension is a simplification as *online* communication changes completely the nature and dynamics of social interaction (Faraj et al., 2013; Kane et al., 2014; Levina and Arriaga, 2015). As a result, it is important to focus on the unique characteristics of online communities when studying social capital.

A landmark feature of online communities is the unbalanced composition of contribution. Few members generate most useful contributions while most other members are lurkers. Referred to as, *long tail* or *power law* distribution (Johnson et al., 2014), this feature is also found in other online nonsocial settings (Oestreicher-Singer and Sundararajan, 2012). Centrality is an important determinant of social capital in online communities. Members who occupy central positions in the social networks representing their communities entertain stronger ties with other members and contribute more knowledge to the community (Wasko and Faraj, 2005). A small core of devoted members account for the majority of contribution in online communities (Dahlander and Frederiksen, 2012).

The rule of the few is not unchallenged. Research indicates the importance of casual members in online communities. Peripheral members play an important role by bringing novel ideas that central members are not aware of (Jeppesen and Lakhani, 2010). Members who span boundaries between various communities can transfer, translate, and transform experiences and knowledge from one community to another (Dahlander and Frederiksen, 2012). Central members on the other hand tend to be homophilous by the virtue of their strong ties and continuous interaction (McPherson et al., 2001). Peripheral and boundary spanning members bring diversity to their communities which is key in enabling complex organisms to operate and grow (Page, 2001).

There is still a debate on the constituents of social capital in online communities. What are structural characteristics of members within the community that determine their knowledge contribution? Is it having a central position within the core of the community, or having a peripheral position that allows to bring to bear more diverse resources? Or are the two involved in a co-creation process as Dahlander and Frederiksen (2012) suggest? We seek to open the black box and examine the constituents of structural social capital in online communities.

5.3 The Two Constituents of Social Capital

Early research on social capital focused on the ability of social actors to maintain frequent and sustained interactions with immediate peers resulting in forming strong ties. Strong ties lead to social cohesion and results in maximizing cooperations and minimizing conflict in organizations (Nelson, 1989). Frequent and strong interactions are antecedents for power, influence, and knowledge creation (Achrol, 1997; McFadyen et al., 2009). Later, the importance of less frequent casual interactions also known as weak ties has been revealed. Weak ties allow accessing novel resources and information unavailable through strong ties (Granovetter, 1973). Moreover, a strong tie is not necessary for obtaining help from others (Constant et al., 1996).

Expanding the perspective of social capital beyond the quality of ties entertained with others, Burt's work on structural holes pioneered the importance of examining the position of a social actor within the social network as a predictor for her social capital (Burt, 1992). A structural hole is a position characterized by having connections to multiple, otherwise disconnected, groups and individuals. Because information, opinion, and practice are more homogenous within than between groups, social actors who span structure holes have an advantage in early exposure to diverse information, a political advantage as hubs in the information flow, and a brokerage advantage as bridges between disconnected peers. As a result, occupying structural holes leads to superior evaluations and compensation as performance metrics (Burt, 2010, p. 22).

We argue that there are two sources of structural social capital: local and global. Local social capital is derived from immediate or close connections. Early research on social capital focused more on immediate connections as a source of social capital. At that time data availability was limited to small networks (Cook et al., 1983). Most studies focused on *ego networks* where a limited number of actors are selected and then their neighbors or *alters* are identified. Relationships among ego and alters form the social network (Everett and Borgatti, 2005). For example, the notion of a weak versus a strong tie is defined over an immediate tie between ego and his colleagues (Granovetter, 1973). Later research extended this notion to cover not only immediate ties but ties beyond. With the advent of online platforms both large social networks and complete data representation become available for researchers to tap into. As a result, it became feasible to examine ego's position within globally the whole network (Borgatti and Everett, 2006).

In Burt's original work there are two types of structural holes: direct and indirect. A direct structural hole is a position where a social actor has few contacts who belong to disconnected cohesive groups. An indirect structural hole occurs when a social actor has few connected contacts who in their turn have access to structural holes. An indirect structural hole results from connecting with peers with networks rich in brokerage opportunities, which are the structural holes to which the actor has indirect access. Social capital results for accessing a combination of direct and indirect structural holes (Figure 10).

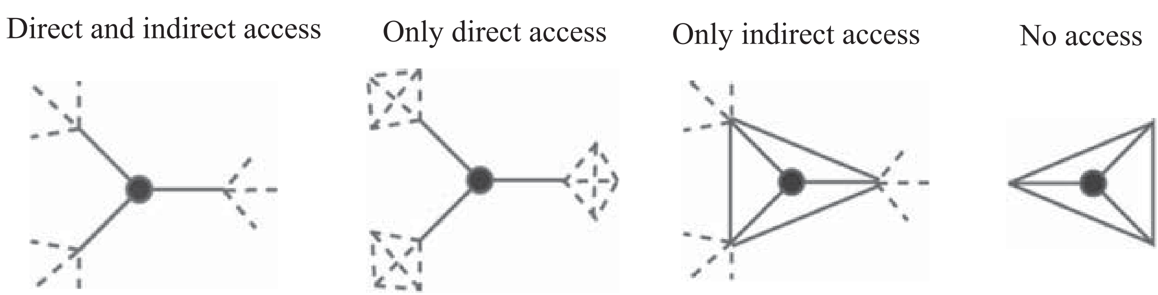


Figure 10: Direct and indirect access to structural holes (adapted from Burt (2010, p. 144))

The tradeoff between having immediate contacts with frequent and deep level of interaction and having distant contacts with less frequent communication applies to a larger extend in online communities. Online platforms make it easier to establish ties and cheaper to communicate with others. Aral and Van Alstyne (2011) offer a more nuanced view of information flow in email networks. The benefits of direct versus indirect contacts depend on the environment. In high dimensional information environments, brokers are disadvantaged as the low bandwidth of their weak ties is insufficient to communicate the large topic space. While structural holes offer better access to novel information, the rate of accessing information through structural holes is slow given the weak nature of bridging tie with distant contacts. On the other hand, immediate contacts offer less novel information but at higher bandwidth. In addition to social networks, global network position shapes outcomes in digital and economic networks (Singh et al., 2011a). Such outcomes include the flow of information and the evolution of the network (Sundararajan et al., 2013, p. 897).

Because of their large-scale nature, open access and transparency, research on online communities employ a global perspective apparatus to examine structural social capital. One way is to describe a member's structural capital is to examine her centrality within the community. Multiple conceptualizations of centrality exist. A general definition of centrality builds on that member's involvement in the walk structure of a network (Borgatti and Everett, 2006). Members who are central within the whole community have more social capital and contribute more knowledge and provide better answers to other members' requests (Wasko and Faraj, 2005). Another way to examine structural capital of a member is to assess her brokerage power. Members who span boundaries between different sub-communities broker knowledge transfer and generate many useful contributions (Dahlander and Frederiksen, 2012). A third way to study structural capital is to split the community into hierarchical layers and examine the member position within this *core-periphery structure* (Borgatti and Everett, 1999; Boyd et al., 2006). Core members are more persistent and longer tenured in their communities. They enjoy more prestige and respect from others (Cattani et al., 2014). Peripheral members are more dynamic and bring fresh ideas and solutions to problems that core members struggle solving (Jeppesen and Lakhani, 2010). Finally, members occupying an intermediate position between the core and the periphery are in a better position to achieve creative results (Cattani and Ferriani, 2008).

In contrast to early egocentric conceptualizations of structure in social networks where local neighborhoods among key actors are considered, concepts such as centrality, boundary spanning, and core-periphery are very global conceptualization of structure. They all assume a community that is connected and examine the relationships of each member in respect to everyone else in the community. Partly driven by the availability of complete network data and the computational power to process it, the resulting theoretical model implies a meritocracy hierarchy where positions depend on involvement but also limit future action. In open-source communities: "this division of roles and the corresponding decision-making power is consistent with practitioner accounts regarding the emergent, social basis of participation. In this context an individual's status in the community is not dictated by bureaucracy, but is a result of a proven (or unproven) level of skill, commitment and peer recognition" (Berdou, 2011, p. 12).

5.4 Toward a Local View of Structural Social Capital

Many people are involved in realizing work in crowdsourcing, large-scale collaboration, and mass-collaboration online settings (Estelles-Arolas and Gonzalez-Ladron-de Guevara, 2012). Empirical evidence from network science and the fact that everyone is eventually connected to everyone else in social networks (Watts and Strogatz, 1998; Newman et al., 2002) have contributed to promoting a global view of work, collaboration and social capital in online communities. The core periphery model has been used to study structural social capital in open source projects, Wikipedia, blogs, virtual organizations and online communities (Warmbrodt et al., 2008; Collier and Kraut, 2012). This model corresponds of a global way of organizing work. We do not argue that online communities do not exhibit a small world structure in aggregate. Rather, we argue that this structure emerges from local collaboration rather than induce global collaboration patterns. Each member has a local position and a global position within the community. However, based on arguments from organization science, online communities research and recent empirical from studying large scale online collaboration, we propose that the local position is what determines structural social capital in online communities.

The wisdom of the crowd premise rests on the assumption that under certain circumstances, the collective is smarter than the smartest member within (Surowiecki et al., 2007). Recent research demonstrates that not all types of coordination can benefit from mass-collaboration (Kittur and Kraut, 2008). Various arguments from organization science support this empirical finding. First, there is a cognitive limit on information processing at both individual and organization levels (March and Simon, 1958). Although collectively the crowd may reach a better decision, individuals and organizations rarely optimize for the best but rather satisfy their decision making needs by picking better alternatives. Second, the potential of conflict among individuals when making joint decisions calls for splitting organizational decision making into smaller units or groups (Tosi, 2008, ch. 7). Third, this division of labor leads various units to develop common language and shared knowledge. As a result, external communication with other units differs from communication within units (Cohen and Levinthal, 1990, p. 133). Finally, in knowledge intensive work settings where work is highly specialized and contextualized, coordination is required to manage knowledge because expertise is distributed among team members (Faraj and Xiao, 2006). Specialized knowledge is difficult to communicate externally. A higher level of coordination is required to integrate specialized knowledge (Grant, 1996b).

Although online communities bring together a large number of members, collaboration processes usually involve limited numbers of actors especially in knowledge creation settings. For example, in online communities focused on technical subjects, dyadic and triadic exchange patterns such as reciprocity and preferential attachment are the norm (Faraj and Johnson, 2011). While some online platforms eventually manage to incorporate the masses in their processes (Levina and Arriaga, 2015), many other platforms such as Wikipedia end up implementing decision rules similar to those in bureaucratic organizations (Butler et al., 2008). Recent empirical research provides evidence against global structural capital in online social networks. A central position within social media may be superfluous and will not necessarily translate to social influence (Aral, 2013). Furthermore, It has been suggested that contrary to the traditional argument that OSS development is the result of mass collaboration (Raymond, 1999), the majority of contributions comes from *lonely* developers who are working in *caves* and seldom interact with the rest of the community (Krishnamurthy, 2002). This has been supported by empirical evidence from OSS repositories such as SourceForge in which that larger the project is, the smaller the number of its administrators are (Krishnamurthy, 2002). Large-scale online collaboration is characterized with flatter hierarchies and more decentralized communication patterns with an uneven degree of participation (Crowston and Howison, 2005).

It is interesting to see that like early research in social networks (Cook et al., 1983), early research in online communities focused on immediate connections as determinants of social capital (Constant et al., 1996). This may have been driven by the scarcity of complete network data. We seek to bring the local perspective back when examining knowledge contribution in online communities. We argue that structural social capital in online communities is better described with local embeddedness in the community than global position. We propose that *the local position of a member in an online community is much important in determining her contribution to the community than her global position*. For example, Figure 11 depicts a hypothetical online community. We argue that we can learn more about the knowledge contribution of the starred member by examining her relationships within her neighborhood (the smaller circle) than by examining her relationships with every other member in the community. *What matters for this member to contribute to the community are her relationships with neighboring members rather than her global position within community*. Later, we elaborate more on the measurement of global and local properties of structure.

5.4.1 Global and Local Structural Capital in Online Communities

The structural social capital of one member in an online community stems from her structural position within the community. A global evaluation of position with respect to all other members in the community results in *global structural capital*. However, this evaluation could be done on a local scale, for example in the OSCAR community a member's position

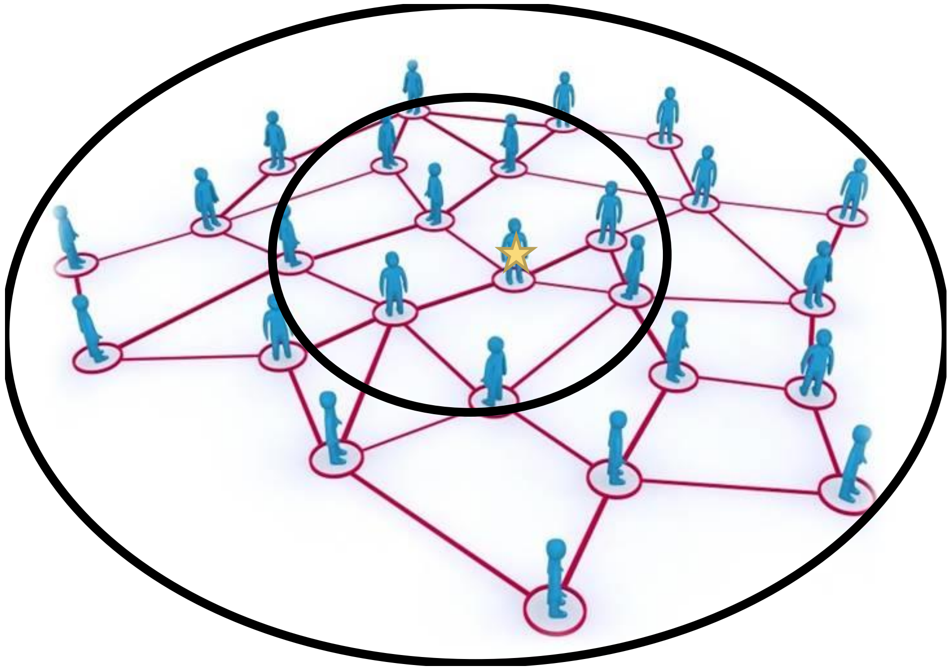


Figure 11: Local (within the small circle) and global position (within the big circle) of the starred member

could be evaluated within her provincial sub-community. An important case here is the Francophone and Anglophone sub-communities within the OSCAR community. This position reflects of the member's *local structural capital*.

This differentiation between local and global positions can be linked to different qualities of capital in the community. For example, because globally central members have long-tenure and expertise, it is more likely that their contributions are highly specialized. On the other hand, contributions coming from other members who are not globally central but perhaps locally central could be more novel and specialized. The nature of communication among members differs depending on whether it circulates in local sub-communities or the global community. Local communication may be more social as it happens among peers, while global communication may need more technical expertise. Local communication is frequent but less novel and diverse compared to global communication.

Although the local and global positions are correlated, as evident from the correlation analysis, they are not completely dependent. We can think of them as two dimensions to examine position in online communities. At one extreme, a member can be a peripheral member globally, but a core member locally. For example, in a regional Francophone OSCAR sub-community within the global OSCAR community. Another possibility is for a global core member not to have a local central position. Although this is a rare case because core members tend to be frequent communicators, this case could happen in expert-novice communities where few experts provide help to novice members but do not communicate among themselves. All in all, the local perspective adds a new dimension to examine structural social capital, and this new dimension could help solving the incommensurability of findings in core-periphery research.

The distinction between local and global structural capital has implication on how structural social capital is measured in social networks. Network structure can be evaluated both at the individual node-level and also at the aggregate network-level (Freeman, 1979). In this paper we focus at the individual-level evaluation of structure. The traditional approach to describe network structure is with measurements that take into account all nodes and links in the network. Those macroscopic measurements are called *global features*. Recent research suggests that global measures are insufficient for precisely describing network structure (Milo et al., 2002; Pržulj et al., 2004). Even if the distribution of links among nodes is similar across two networks, the local arrangement of the links may differ leading to a different local structure and potentially a different function (Janjić and Pržulj, 2012). Those local arrangements are called *local features*.

In this research we use egocentric network centrality in order to measure local structure capital. Unlike most SNA methods that focus on the relationships among all actors in the network, egocentric network analysis focuses on the relationship between an ego and her connections or alters in the network. As such, an egocentric network defines a locale centered on an ego actor. The size of the locale is chosen depending on the context of the research. A commonly used size is 1.5 which defines an egocentric network contains an ego, her alters and all relationships among them. We do not have a present locale size; one goal of this research is to examine various locales and consider how structural social capital varies by changing the size of the locale.

5.5 Knowledge Contribution in Online Communities

Knowledge has been characterized as one of the most important resources in an organization (Barney, 1991; Grant, 1996a; Conner and Prahalad, 1996). Knowledge creates capabilities (Kogut and Zander, 1992), and helps sustain the competitive advantage of the firm (Clemons and Row, 1991; Mata et al., 1995). The successful acquisition, coordination and integration of knowledge lead to positive organizational outcomes (Jarvenpaa and Leidner, 1997; Faraj and Sproull, 2000; Tiwana and Mclean, 2003; Sabherwal and Becerra-Fernandez, 2005).

Knowledge contribution processes in online communities are different than those in bureaucratic organizations (Tapscott and Williams, 2008). For example, the traditional method of problem solving relies on the availability of local expertise in in-house research and development teams. A radical departure from this method is to enlist many peripheral solvers to compete on providing a solution (Lakhani, 2006; Spencer, 2011). Another example is open-source software developed by communities of spatially, temporally, and organizationally distributed developers (Cummings et al., 2009; Feller et al., 2005). Traditional software development requires a careful plan typically done by software engineers and field experts for eliciting the requirements, designing the software, implementing, testing and releasing it (Sommerville, 2007). Open-source development on the other hand is an emergent process that starts with coders writing code without necessarily a laid-out plan. Afterwards, the code gets peer reviewed, tested and improved by other developers and users in the community (Feller and Fitzgerald, 2002). Finally, Wikipedia's approach for creating an encyclopedia relying on voluntary contribution is a third example (Butler et al., 2008; Hill, 2012).

Knowledge contribution to online communities is a central concept that is researched thoroughly in the literature. After all, such communities exist and are important because they tap the expertise and experience of their members toward creating new knowledge and realizing a common goal (Wasko and Faraj, 2000). Online communities produce various tangible artifacts such as software code in open-source communities and articles in wikis. Because knowledge creation is perhaps the most important outcome of online communities, the question of why members contribute knowledge has been a central question in online communities research. The altruistic behavior online has often described as extension to other forms of prosocial behavior such as helping bystanders (Sproull et al., 2005). However, other non-altruistic factors are documented including professional motivations, becoming visible, gaining self-esteem and connecting to potential members (Butler et al., 2013; Wasko and Faraj, 2005). Regardless of why members contribute, their contribution is the modus operandi in online communities.

5.5.1 Structural Social Capital and Knowledge Contribution

Knowledge has been viewed as a duality by many researchers (Hildreth et al., 2002). These conceptualizations include formal and informal knowledge (Conklin, 1996), individual and collective knowledge (Rulke et al., 1998), and tacit and explicit knowledge (Nonaka and Takeuchi, 1995). The duality perspective acknowledges that knowledge has a second component that is less tangible and more implicit. For example, informal knowledge is difficult to represent in traditional sources of knowledge such as books and articles. Collective knowledge spans many sources and is not represented in one place. Tacit knowledge is context specific and difficult to communicate to others.

Under the duality of knowledge perspective, knowledge evolves overtime (Nonaka and Takeuchi, 1995; Von Krogh et al., 2000). This evolution follows repetitive phases at different levels. Knowledge can be explicit or tacit. Explicit knowledge is easy to formalize and transfer whereas tacit knowledge is personal, context specific, hard to formalize and difficult to communicate. Knowledge creation is the transformation of tacit knowledge into explicit knowledge. Online communities of innovation are a place in which such knowledge transformation happens. Through communication and interaction afforded by the online community platform, the tacit knowledge of individual members is transformed into an explicit knowledge in different forms such as articles in wikis and code in open-source software communities.

Many studies have looked at the contingencies including structural ones leading to knowledge contribution in online communities. A strong relationship between two members is not a prerequisite for them to interact and collaborate online (Constant et al., 1996). On the other hand, strong ties are beneficial for the output of the community because they lead to a higher level of knowledge creation (McFadyen et al., 2009). Technology features can facilitate online interaction by providing a reliable platform that facilitates members' knowledge seeking and providing behavior in the community (Phang et al., 2009; Ma and Agarwal, 2007). The location of members within the community is a determinant to various types of contribution. Boundary spanning between communities and a core position within the community characterize innovators (Dahlander and Frederiksen, 2012). Peripheral members also play an important role in providing solutions to problems that core members cannot solve by bringing expertise from different fields than the core field of the problem and by coming from a different social stratum than core members (Jeppesen and Lakhani, 2010).

We focus on individual knowledge contribution in open-source online communities. In this intermediate phase, several artifacts are produced by community members toward creating the final knowledge artifact. For example, in writing a wiki article, members perform revisions by adding, deleting and modifying some of the content of the article. In OSS, developers commit small chunks of code to modify the existing source code files. Previous research has typically studied the final outcome of collaboration (Ransbotham et al., 2012). In this study there are two reasons for examining an intermediate form: First, we focus on the individual-level knowledge creation and the structural factors contingent on it. Final collaboration outcomes are by definition collective efforts and do not suit this task. Second, because the structure of online communities is continuously changing, we need to capture the temporal dimension when examining the relationship between structure and knowledge creation. Individual contributions are situated in shorter time spans than collective artifacts that are the result of longer periods of collaboration.

Individual contributions to online communities can be examined with two dimensions: quantity and quality. The distribution of contributions is extremely uneven and follows a power-law or a long-tail distribution (Johnson et al., 2014). The bulk of contribution comes from few members. The question of who contributes more is not settled yet. On one hand, the core of the community which is composed of dedicated long-tenured members is the source of much information and expertise in communities of practice (Wasko and Faraj, 2005). On the other hand, in many crowdsourcing settings, contributions come from casual members who are peripheral to the community (Jeppesen and Lakhani, 2010). Furthermore, contribution can be anonymous without expectations of reciprocity (Moon and Sproull, 2008; Faraj and Johnson, 2011). We think that this gap in findings can be explained by incorporating a local perspective of position into the equation. Following findings of previous research on online communities we propose that:

Proposition 5.1 *global centrality is positively related with knowledge contribution in the community, and*

Proposition 5.2 *global boundary spanning is positively related with knowledge contribution in the community.*

The division of labor among members in online communities is the prime source of differentiation in the quality of contribution. In open-source communities in particular, it is well known that members specialize in different tasks. For example, while casual users may contribute toward finding issues and problems (also known as *bugs*), developers specialize in coding modules because it is personally rewarding. They apply their domain knowledge to modules and features in the emerging software architecture at lowcost. However, some developers prefer generalization or low specialization by contributing broadly to many modules (Von Krogh et al., 2003, p. 1230). The *onion model* is often used to describe the division of labor in open-source communities: "In this model, the core group of skilled developers consists of people with the greatest authority and decision-making power with regard to how the project evolves. This group includes the project founders; the maintainers, that is, the developers in charge of the smooth operation of certain parts of the code base; the most engaged contributors; and the developers, who assume critical coordinating roles, such as that of managing the release process for each new version of the program. The next layer of the onion includes those programmers who contribute relatively small changes or patches to the code base. Their contributions are usually subject to review by the maintainers before acceptance. The next layer is populated by involved users, who provide feedback on how the program works by reporting faults (bugs in the language of software development) and by suggesting improvements either by participating in the project forums or by employing more specialized tools, such as bug databases. The outermost layer of the onion includes the constituency of the program's users and individuals (often described as lurkers) who observe the community's online discussions and have an interest in development, but who do not contribute." (Berdou, 2011, p. 12).

In such layered and hierarchical organizational structure, the relationships among peers within the same level are important for realizing work. We propose that:

Proposition 5.3 *local centrality is positively related with knowledge contribution in the community, and*

Proposition 5.4 *local boundary spanning is positively related with knowledge contribution in the community.*

What remains unknown and is a goal this research explores is whether contribution is better determined by global position or local position. We argue that local properties matter more relevant than global properties in determining knowledge contribution to online communities. Recent research has shown that collaboration in open-source software is more decentralized and localized that previously thought (Crowston and Howison, 2005; Krishnamurthy, 2002). Members who do not necessarily occupy a global position but entertain a central local position benefit the community by bringing novel knowledge that central members are unaware of. As a result, what matters for an individual to contribute to the community is her relationships with neighboring members rather than the absolute position in the whole community. We propose that:

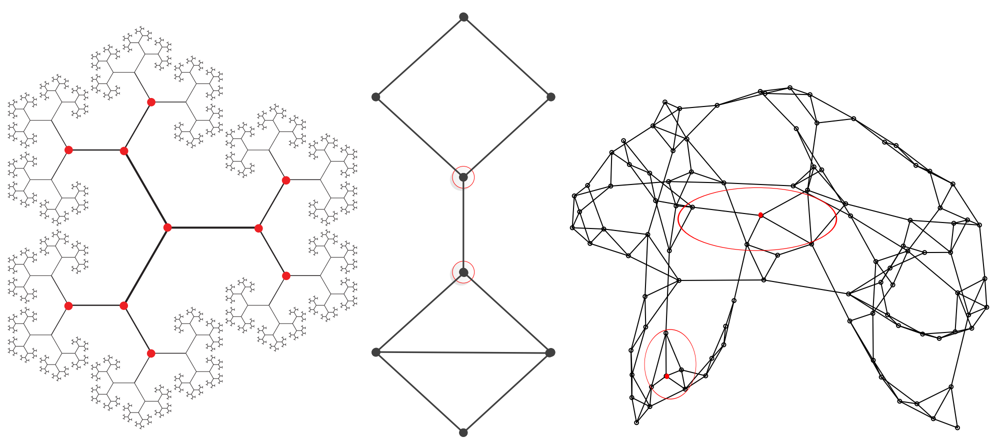
Proposition 5.5 *knowledge contribution is more positively related with local centrality that with global centrality in the community, and*

Proposition 5.6 *knowledge contribution is more positively related with local boundary spanning that with global boundary spanning in the community.*

5.6 Measuring Structural Capital in Social networks

The traditional global view to study structure evaluates the structural position of a node with regard to all other nodes in the network. In contrast, a local view requires only information about the immediate neighbors of a node in order to evaluate its structure. The structure of a node is defined by its relationships with nodes within its neighborhood rather than within the whole network. For example, in a large organization such as a university, to evaluate the global position of a professor in the faculty of management, we need to consider her relationships with everyone else in the university. A local view requires evaluating his relationships with immediate peers such as professors in the faculty and her research team. Beyond metaphors, we propose to operationalize the difference in structure by incorporating the local properties perspective.

Lets consider the example in Figure 12. In the first network all nodes (except the leaves) have exactly the same local properties. Each node has three neighbors, each in its own connects to three other nodes. However, nodes in network A have different global properties. There is a clear core periphery structure and nodes differ in their centralities as well. In network B, the two highlighted nodes have similar global properties but different local properties. They both span boundaries between the other nodes and are central in the network. However, the lower node is part of a denser set of nodes. A more real example is found in network C where the highlighted nodes have similar local features but different global positions.



(a) Same local, different global adapted from Reed (2013) **(b)** Same global, different local **(c)** Similar local but different global (real network)

Figure 12: Examples of local and global features

Another example is presented in Figure 13. In this example, actor Z is globally central and occupies a structural hole in the network bridging between two disconnected groups. However, locally Z is not very well connected. On the other hand, actors X and Y are locally central in their corresponding groups. However, they are both less central in the whole networks than Z. As we see from the two examples, both global and local positions are evaluated at a node-level. Each node has a global and a local position. In the next two sections, we operationalize global and local structural position in social networks.

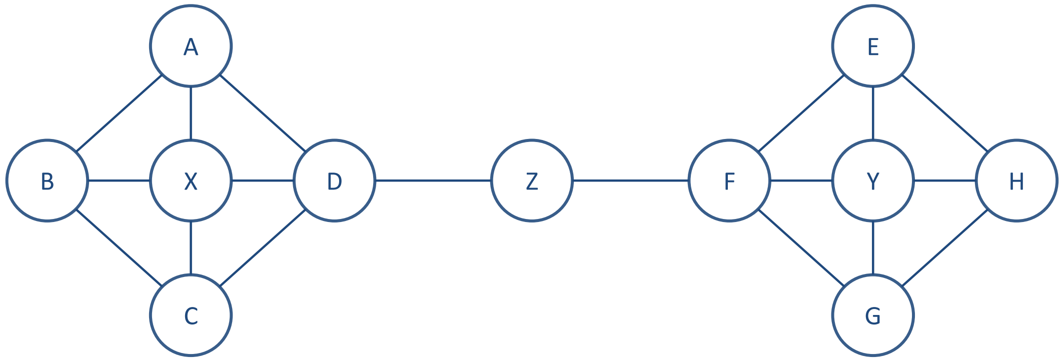


Figure 13: Example of global centrality (Z) and local centrality (X & Y)

5.6.1 Global Structural Capital

In online settings, given the lack of face to face interaction, the characteristics of members are formed primarily based on how online contributions are made, how new ties are formed, and how those ties influence others' impressions (Donath, 2007). The social network perspective has been applied both to the larger question of the structural role of members as well as leadership in online communities. The emergent consensus is that influencing members and leaders score highly in various centrality measures and also play a boundary spanning role in order to acquire information or resources (Balkundi and Kilduff, 2006).

Under this perspective, the influence of a member represented by a node in the network is determined by the set of relationships this member has with others. It is worth mentioning though that this computed influence is not social influence in the strict sense because it does not imply social behavior (Aral and Walker, 2011). A well-known example of an influence measure is Google's PageRank that is used to rank pages in Google's search results. Along with other undisclosed parameters, the rank of a page is computed primarily by looking at the incoming link to the page from other pages (Brin and Page, 1998).

Participation in an online community is typically extremely uneven and follows a power law distribution, with a small number of participants engaging in most interactions (Faraj and Johnson, 2011). Previous literature has debated the process of knowledge creation in online communities and the actors involved in that process. Some research emphasized role of core members in generating most useful contributions (Dahlander and Frederiksen, 2012), while others claimed that peripheral members are more useful in bringing fresh ideas and solutions (Jeppesen and Lakhani, 2010). In this research, we consider three structural determinants of members' influence: network centrality, boundary spanning and core/periphery structure.

Global features of network structure are metrics that take into account in order to estimate all the nodes and edges in the network. We consider three global features: centrality, boundary spanning and core-periphery structure. Figure 14 demonstrates three global properties: A) centrality in which central nodes in the network are colored with a more red color, B) betweeness in which nodes that arbitrate paths in the network are highlighted in a more red color, and C) coreness in which nodes that are part of cohesive components in the network are highlighted in red.

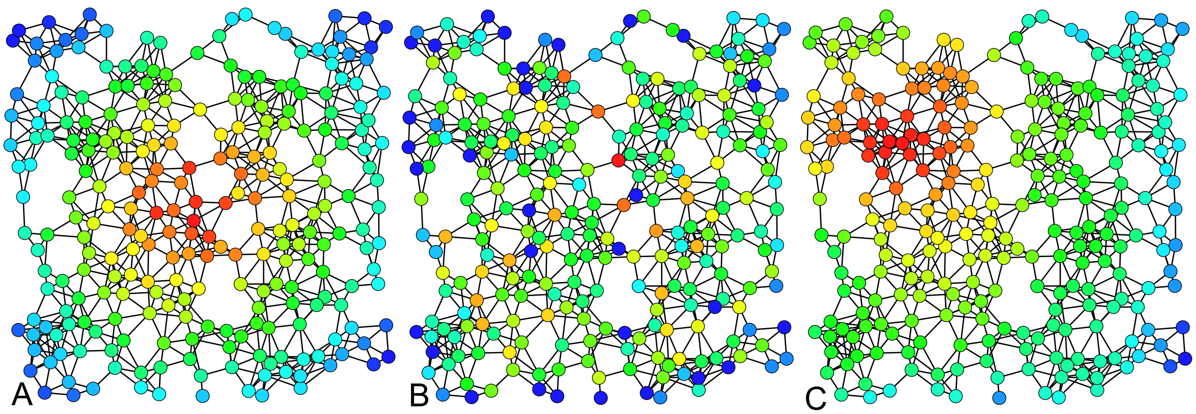


Figure 14: Examples of global features of nodes in a network A) centrality, B) boundary spanning and C) coreness (adapted from Rocchini (2012))

Centrality

In online communities there is support for influential members and leaders being associated with having a central position in the community although results diverge on the type of centrality. Huffaker (2010) analyzed 16 Google Groups discussion forums to identify the characteristics associated with the leadership impacts of triggering communication responses and influencing language usage. He found that expansiveness (out-degree centrality) was associated with leadership behaviors but brokering (betweenness centrality) was not. Looking at virtual collaboration supported by Second Life and in text-based chat rooms, Sutanto et al. (2011) found that both degree and betweenness centrality were associated with emergent leadership, but closeness centrality was not.

Multiple centrality measures exist in the literature (Brandes and Pich, 2007). We adopt the closeness centrality metric that measures how long it takes to sequentially disseminate a message to all other nodes in the network. A central node is close to many nodes in the network and hence it takes a short amount of time to diffuse a message from this node through the network. We operationalize the centrality of a node by computing the average distance from it to all other nodes in the network. We then take its reciprocal to measure the centrality of the node:

$$C(n) = \frac{1}{\sum_{m \in nodes} distance(n,m)/|nodes|}. \tag{5.1}$$

Boundary Spanning

As the community grows and more members join, heterogeneity of membership increases. Boundaries between subgroups in the community emerge as members blend around certain shared interests. Boundary spanning members are of particular importance because they bridge the boundaries between these subgroups by engaging with members within them and also link the core, the periphery and the outside world as well. Boundary spanning has been associated with leadership in multiple domains including knowledge-intensive work (Levina and Vaast, 2005) and open innovation communities (Fleming and Waguespack, 2007). Boundary spanning can be examined on a global scale where spanners bridge between communities and on a local scale where spanners bridge between sub-communities and topics within the community.

To measure boundary spanning, we consider the betweenness of the node based on its control over the paths between other pairs of nodes (Brandes, 2001). Let $\sigma(s,t)$ denotes the shortest path between nodes s and t in the network and $\sigma(s,t|n)$ denotes a shortest path between s and t that contains node n. The boundary spanning of n is the fraction of all pair shortest paths in the network containing it over all pair shortest paths in the network:

$$B(n) = \sum_{s \neq n \neq t} \frac{\sigma(s,t|n)}{\sigma(s,t)}. \tag{5.2}$$

Core/Periphery

Closely related to centrality, the concept of core/periphery provides a complementary understanding of the structure of a network (Borgatti and Everett, 1999). Compared to continuous centrality, core/periphery suggests that there are distinct sub-groups of members with jointly occupied, structurally equivalent positions. Core/periphery structures have been identified in smoking cessation (Cobb et al., 2010) and video-bloggers online communities (Warmbrodt et al., 2008). Membership in the core is associated with leadership in open source software developer communication networks (Crowston and Howison, 2005) and in Wikipedia (Collier and Kraut, 2012). Peripheral members are important for sparkling innovation by cultivating the community with external and fresh ideas (Jeppesen and Lakhani, 2010).

We operationalize the coreness of a node (i.e. being a core vs. peripheral) using the k-core decomposition. The decomposition is base on assigning a number to each node in the network depending on its core-periphery position. The k-core number (K) was used to divide the network into layers of cores. The cores are sub-networks with k connectivity. For example, it is possible to fragment the 1-core by deleting one edge, while at least two edges need to be deleted in order to fragment the 2-core (Figure 15). The k-core number is used to measure cohesively in social network with higher core as more densely connected parts of the network (Seidman, 1983).

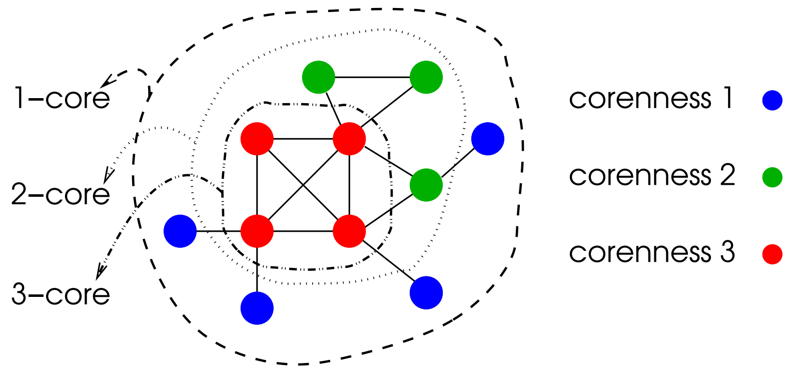


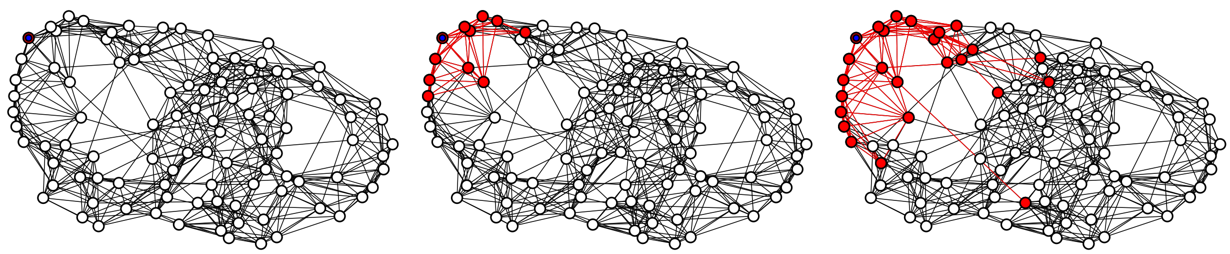
Figure 15: k-core decomposition for a small graph. Each closed line contains the set of vertices belonging to a given k-core, while colors on the vertices distinguish different k-shells (taken from Alvarez-Hamelin et al. (2006))

5.6.2 Local Structural Capital

Network science research has recently started to uncover the importance of local structural features when examining the function of various physical networks. Those *local* patterns give insight into the *microscopic* topology of networks rather than the *macroscopic* picture given by global features (Janjić and Pržulj, 2012). Among those examined patterns are network motifs which are patterns of interconnections among set of nodes that occur in the studied network at numbers significantly higher than expected in random networks (Milo et al., 2002). Different motifs distinguish different types of networks (that share similar global features) such as ecological food webs, genetics networks, and the World Wide Web. At the node-level, topological motifs extend the concept of network motifs by considering the position of the nodes relative to the motifs (Berg and Lässig, 2004).

In this research, we define the concept of a local neighborhood by expanding the notion of ego-centric networks (Everett and Borgatti, 2005; Vehovar et al., 2008). The n^{th} local neighborhood of a node m is a sub-network that includes m and its first, second, ..., n^{th} degree connections and their relationships. Computationally, this is realized using Breadth-First Search (BFS) starting from m with a depth of n (Sedgewick, 2001, ch. 18). Once a local neighborhood is determined, the node embeddedness is evaluated using the same algorithms used to evaluate its global position. This permits to compare global and local positions on the same grounds. The only difference is the scope of comparison which is the whole network in the case of global position and a sub-network in the case of local position.

Figure 16 demonstrates an example in which we want to evaluate the position of the blue node in the network. A global perspective utilizes the whole network [a], a local perspective of degree one uses first degree connections of the blue node (highlighted in red in [b]), while a local evaluation of degree two considers the position of the blue note within the sub-network of its first and second degree connections (highlighted in red in [c]).



(a) Global position **(b)** First degree local **(c)** Second degree local

Figure 16: All network is used to evaluate the position of the blue node in (a) while only highlighted sub-networks are used in (b) and (c)

5.6.3 Example of Global and Local Features

We demonstrate here the calculation of global and local features of structure at the node level for node g in the example network in Figure 17. We compute g's global position as well as its first and second-degree local position. In this example, for the sake of simplicity, the distance between each two connected nodes is one. The methods still work when the links are weighted differently which is the case for networks representing online communities.

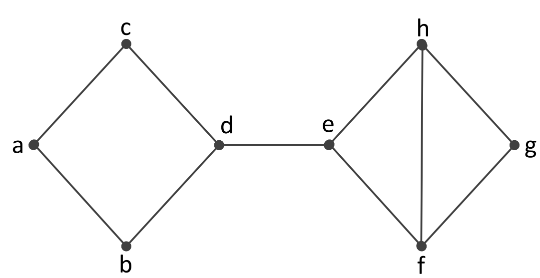


Figure 17: Example network

To compute the centrality of node g we compute its average distance to all other nodes and take its reciprocal.

$$C(g) = \frac{1}{(1+1+2+3+4+4+5)/7} = 0.35.$$

The boundary spanning of node g is zero as no shortest path between any two nodes in the network contains g. Finally, the core number of node g is two because two nodes are needed to disconnect the node from the network.

A first degree local neighborhood of g includes nodes h and f and edges $g-h$, $g-h$, and $h-f$. In this sub-network the local centrality of g is:

$$c_1(g) = \frac{1}{(1+1)/2} = 1.$$

g is very central in this locale but it is not a boundary spanner, however, as it does not arbitrate the shortest path between h and g, its local spanning is zero.

g's second degree neighborhood includes nodes h, f and e and edges $g-h$, $g-h$, $h-f$, $h-e$, and $f-e$. In this neighborhood the centrality of g is:

$$c_2(g) = \frac{1}{(1+1+2)/3} = 0.75.$$

Again as g does not mediate shortest paths between h, f, and e, its local spanning is zero.

5.7 Research Setting

The OSCAR EMR project has a vibrant community of users. Interaction among members is done mostly online in addition to one national meeting and several regional meetings every year. The mailing list is an excellent place for novice users to get assistance and help. Moreover, members in the community often contribute to the OSCAR project by sharing additions such as templates, medical forms and plug-ins. We have collected data from the online mail list of the OSCAR user community. The community maintains five mail lists hosted on SourceForge[7]. We collected seven years of data from the inception of the community in January 2006 through January 2013. The collected data consists of over 55,000 messages written by approximately 1,000 unique members in 35,000 threads. Figure 18 shows an excerpt from one communication thread [8].

Affiliation networks are natural representations of threaded discussion communities (Hansen et al., 2010, Ch. 9). In affiliation networks, members participate to common activities or

[7] `http://sourceforge.net/mail/?group_id=66701`
[8] taken from `http://sf.net/p/oscarmcmaster/mailman/message/22980338`

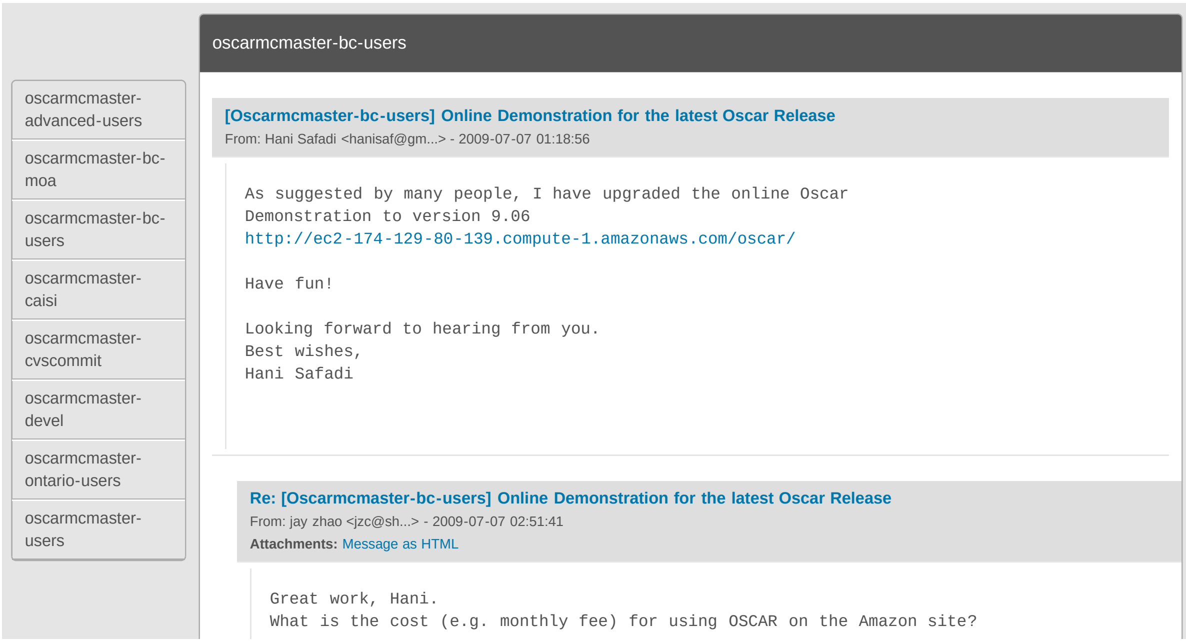


Figure 18: An excerpt from a thread in OSCAR maillist

affiliations. There are two types of nodes with links connecting two nodes from different types. In threaded discussion networks, the two types of nodes are members and threads. A link between a member and a thread is established when the member posts a message to the thread. While it is possible to analyze an affiliation network directly (Faust, 1997; Borgatti and Halgin, 2011b), it is typically transformed to a unimodal network because most algorithms assume a homogenous set of nodes (Figure 19). There are a couple of methods to implement this transformation. All of them establish a link between two members if they are affiliated with same threads. However, transformation methods differ in how they assign weights to links based on the number of shared threads (Borgatti and Halgin, 2011b).

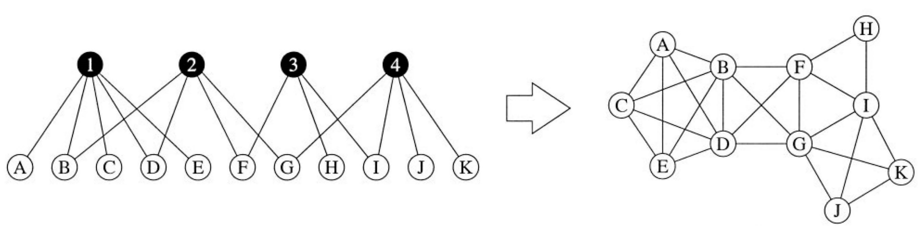


Figure 19: Affiliation network transformation (taken from Newman et al. (2002))

We adopt Newman (2001b)'s method of transforming affiliation networks. Designed to transform academic co-authorship networks, this methods fits well the context of this research as we can consider each thread in the mailing list to be collaboratively authored by the members posting in it. In this model, two members participating in many same threads have stronger links than other two members who participated in few threads. To compute the weight of a link between two members (w_{ij}) we add up the number of shared threads (δ_i^k denotes that member i posted in thread k) but also account for the number of members posting in one thread (n_k). If many members post in one thread we discount the importance of participation in this thread:

$$w_{ij} = \sum_k \frac{\delta_i^k \delta_j^k}{n_k - 1}. \tag{5.3}$$

Figure 20 demonstrate and example where members A and B participated in three threads. However, thread 1 has two other members posing in it, while thread 3 has another member posing in it. The weight of the link between A and B is the summation of their normalized participation in the three threads.

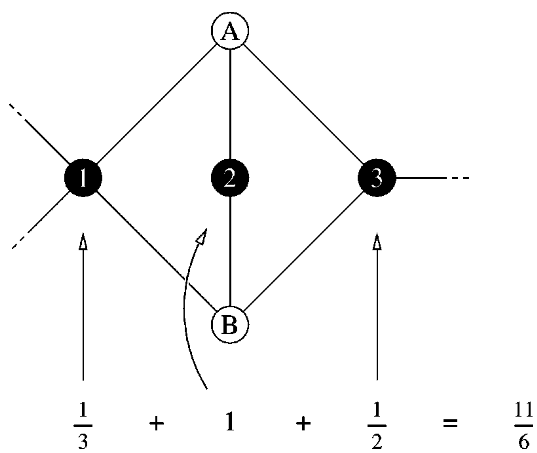


Figure 20: Assigning weights in affiliation network transformation (taken from Newman (2001b))

Metrics of the transformed network are presented in Table 2. Figure 21 visualizes the transformed network. For the sake of clarity the figure only plots members who have more than two messages and edges with weight more than one. The node size is proportional to the number of messages a member contributed and the thickness of an edge is proportional to its weight. As immediately seen in the figure, few members contribute a lot the the community. Furthermore, they entertain strong relationships with each other.

5.7.1 Measuring Knowledge at the Individual Level

The OSCAR community online forums are primarily a place for exchanging knowledge and expertise among members. Much of knowledge sharing is achieved by writing messages and engaging in a conversation with other members. To measure individual knowledge

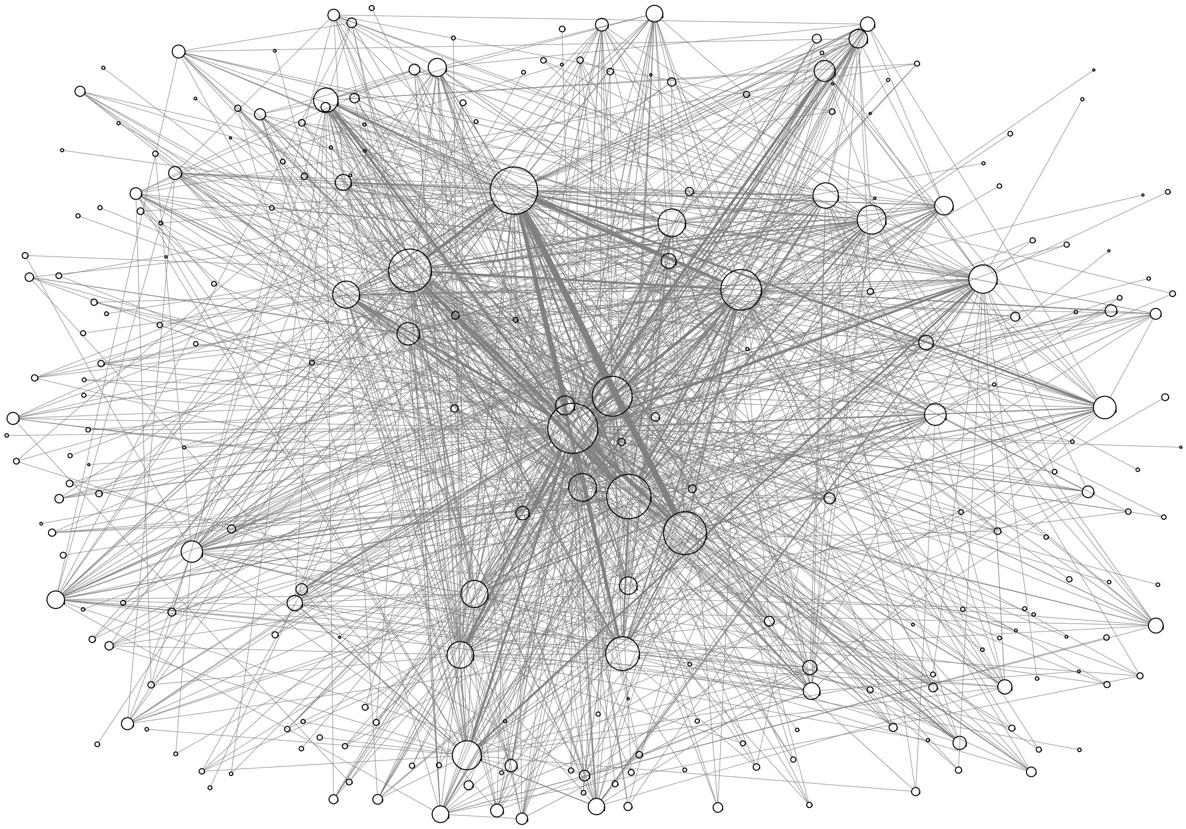


Figure 21: OSCAR network

contribution we distinguish between the quantity and the quality of contribution. We operationalize the quantity of contribution of a member by counting the number of messages posted by this member to the mailing list. Occasionally however, some members contribute codified knowledge by creating attachments to their messages. Codified knowledge contribution in attachments ranges from simple additions such as templates and medical forms to sophisticated enhancements and plug-ins. We measure a member's codified knowledge contribution using the number of attachments in his/her posted messages. Our data crawler takes care of superfluous attachments such as attached business cards, calendar entries, and previous messages included as attachments. Finally, to measure the degree of specialization in a member's postings we count the number of unique words, also known as dictionary size or richness of vocabulary.

5.7.2 Research Model and Analysis

To examine how members' contribution to the OSCAR community is related to their structural properties, we regress members' knowledge contribution measured by the three variables (quantity, codification and richness) on their global properties (centrality and boundary

Metric	Value
Network type	Undirected
Number of nodes	894
Number of edges	6105
Number of connected components	296
Nodes in the giant component	582
Diameter	4
Average geodesic distance	2.22
Density	0.015
Minimum degree	0
Maximum degree	403
Average degree	13.673
Median degree	3.000
Minimum clustering coefficient	0.000
Maximum clustering coefficient	1.000
Average clustering coefficient	0.506
Median clustering coefficient	0.643

Table 2: Metrics of the transformed unimodal network representing the OSCAR mailing list

spanning) and local properties (local centrality and local boundary spanning). Table 3 summarizes the research variables. There are three knowledge contribution variables: number of messages, number of attachments, and richness of postings. Independent variables are centrality and spanning variables. There are two global variables one for centrality (cfull) and one for boundary spanning (sfull). In considering local position we evaluate the position of each member in locales of increasing size from one to 24. This results in 24 variables for local centrality and 24 variables for local boundary spanning.

First we correlate local and global centrality and boundary spanning with the three knowledge contribution variables. This analysis serves to outline whether local and global position is related to knowledge contribution. Furthermore, we can examine the relationship between the size of the locale used to define local centrality and local boundary spanning with knowledge contribution. For a comparative analysis, we perform two kinds of regression analyses: exclusive and incremental. Furthermore, we repeat each analysis for each dependent variable. In exclusive regression analysis, we regress each dependent variable of each independent centrality and spanning variables for each locale separately. This analysis allows for an absolute comparison between embeddedness in each locale ($1-24$) and global embeddedness as well.

$$
\begin{aligned}
messages &= centrality_i + spanning_i + \epsilon \\
attachments &= centrality_i + spanning_i + \epsilon \\
richness &= centrality_i + spanning_i + \epsilon \\
i &\in \{1, 2, \ldots, 24, full\}.
\end{aligned}
$$

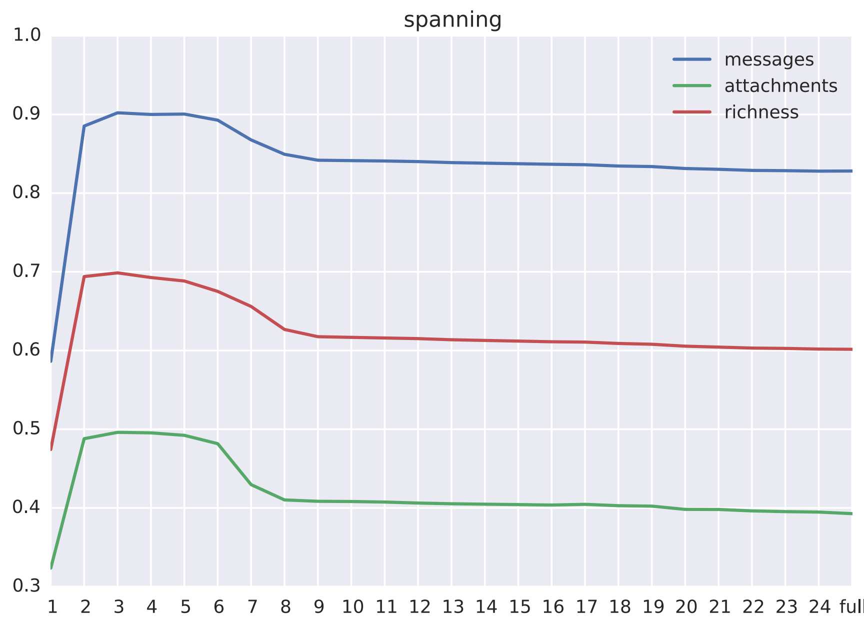
In inclusive analysis, we create nested models where model n includes locales from $1-n$. This analysis allows for examining the marginal benefit of increasing the size of local neighborhood.

$$
\begin{aligned}
messages &= \sum_{j=1}^{i}(centrality_j + spanning_j) + \epsilon \\
attachments &= \sum_{j=1}^{i}(centrality_j + spanning_j) + \epsilon \\
richness &= \sum_{j=1}^{i}(centrality_j + spanning_j) + \epsilon \\
i &\in \{1, 2, \ldots, 24, full\}.
\end{aligned}
$$

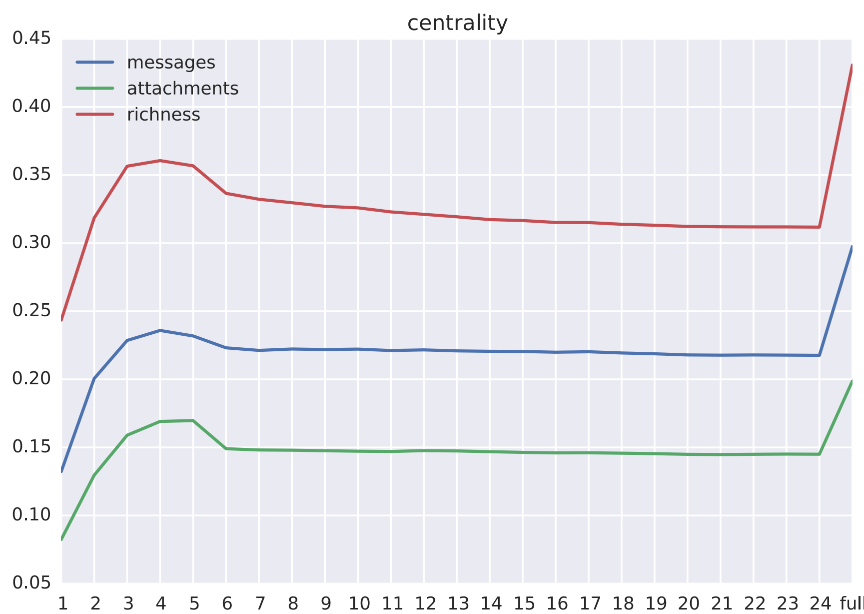
5.8 Results

Correlation analysis (Figure 22) between the knowledge contribution dependent variables and local independent variables for locales 1-14 and global independent variables shows that correlation plateaus at 2, 3 and 4. In fact correlation drops down when considering locals beyond 4 and only increases when considering the global position in the case of centrality. Results of regression analysis (Figure 23) confirm those found in the correlation analysis. R^2 plateaus at local 3 in exclusive regressions. Incremental regressions illustrates that the marginal benefit from adding more locales diminishes after 3 as well. Finally, one argument that could undermine the results is that most of the community is reachable in four degree of separation. This is not the case in the OSCAR community in which only 20% of nodes and 40% of edges are reach on average at degree four (Figure 24).

To summarize, results indicate that global centrality and global boundary spanning are positively correlated with knowledge contribution. The results hold for the three operationalizations of knowledge contribution that measure the quantity and the quality of contribution. In addition, local centrality and local boundary spanning are both positively correlated with knowledge contribution. Propositions 1, 2, 3, and 4 are supported. In



(a) Spanning



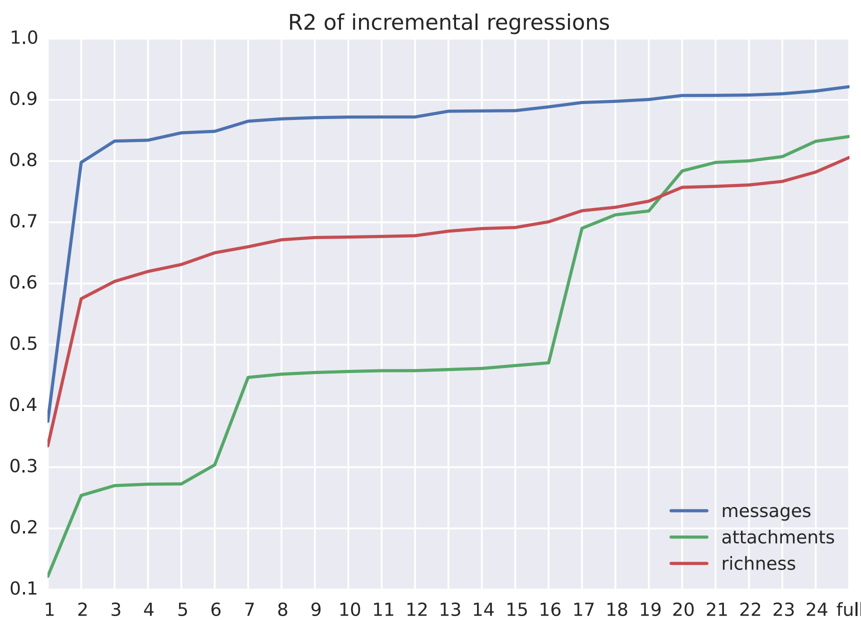
(b) Centrality

Figure 22: Correlation between independent variables and dependent variables

comparing global and local position, we find that the correlation between global centrality and knowledge contribution is more than the correlation between local centrality and knowledge contribution. This disconfirms proposition 5. However, the correlation between local boundary spanning and knowledge contribution is more than the correlation between global boundary spanning and knowledge contribution, supporting proposition 6 as a result.

Table 4 summarizes the findings and research propositions as well.

(a) Exclusive regression analysis



(b) Incremental regression analysis

Figure 23: Regression analysis

5.9 Discussion

We focus on online communities of innovation where knowledge contribution is a key outcome. There is an ongoing debate about the structure of social collaboration in online

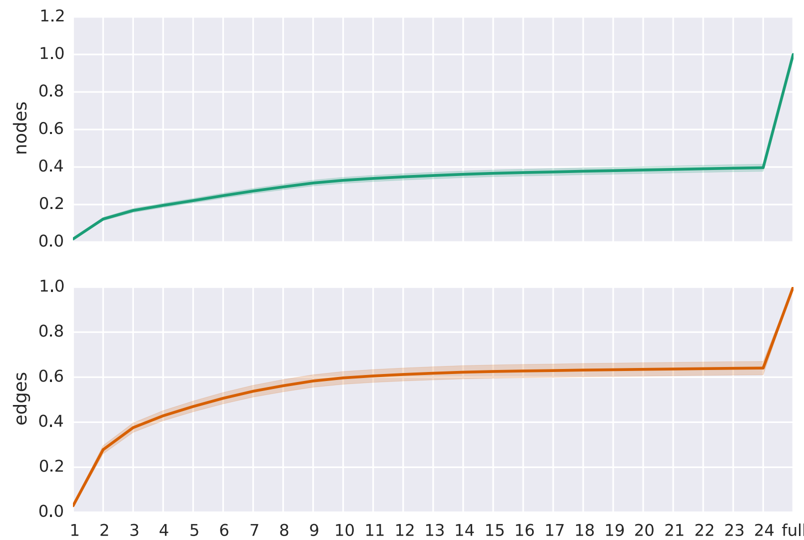


Figure 24: Percentage of network reached by increasing locals with 95% confidence interval

communities and social networks. Members' contribution to online communities depends on their position within the community. Bridging between different groups in the community in addition to having a central position within the community characterize contributors in online communities (Dahlander and Frederiksen, 2012). Interestingly, research has also found that important contributions come often from peripheral members in the community. Peripheral members play an important role in providing solutions to problems that core members could not solve by bringing expertise from different fields than the core field of the problem and by coming from a different social stratum than core members (Jeppesen and Lakhani, 2010). Central members have more social capital and contribute more knowledge and provide better answers to others' requests (Wasko and Faraj, 2005).

What determines structural social capital in online communities is an important question to ask from a theoretical standpoint and a practitioner perspective. Theoretically, it has become important to examine networked behavior and determine causality in online networks (Sundararajan et al., 2013). Structural capital is both a cause and consequence of networked behavior. In practice, determining structural capital helps induce desirable behavior in social networks. For example, in the context of marketing, in order to convince a subset of members to adopt a new product or innovation, and to trigger a large cascade of further adoptions, a certain set of members should be targeted (Kempe et al., 2003).

Global position summarizes a great deal of network structure and are relatively easy to compute. However, global features suffer from many shortcomings. First, Global features require complete network data in order to compute. Data about real-life large networks including social networks and biological networks is seldom complete. Global features are shown to perform poorly under incomplete network data settings resulting in biased and misleading results (Przulj, 2007). Second, in large networks that exhibit structure locality it may be irrelevant to consider the entire network. Social networks exhibit a high degree of locality. For example, in a big organization such as a university, the relationships among professors in the faculty of music are irrelevant when studying scientific collaboration in physical science. Finally, recent research shows that global features are insufficient to precisely describe structure. Two networks with similar global features may exhibit different local structures (Pržulj et al., 2004).

Our empirical findings show that most of information relevant to contribution in the OSCAR online community is embedded in locales of size three. Indeed going beyond three adds noise and reduces the correlation with all contribution measurements. These results clearly indicate that when knowledge contribution in the OSCAR online community is closely related to a member's local position rather than her global position in the community. Another interesting observation is that spanning correlates higher with all contribution measurement than centrality. This suggests that spanning and brokering plays a more important role in the OSCAR community. Finally, in the case of centrality, global centrality beats local centrality but only when all network is considered. The OSCAR network is not fully connected and composed of several disconnected components (Table 2). This suggests that global measurements may not be suitable when complete network representation is not available.

5.10 Conclusions

A network representation of social collaboration is often employed to study the structure of social collaboration in online communities. A social network is a collection of nodes that represent social actors and edges (or links) linking the nodes in order to represent interactions among the actors. This representation serves to abstract the complexity of the social phenomenon and allows for a mathematical treatment of the research question as the properties of the network including its structure are computed from its mathematical representation.

This study seeks to deepen our understanding of the structural properties of online communities and the role these properties play in shaping the production of the communities. We complement previous research by proposing a comprehensive apparatus to examine the structure of the OSCAR network on both the macro and the micro scales. We propose that the local features of structure in online communities of innovation are more important in influencing knowledge contribution than global features. This goes in tandem with recent advances in understanding large-scale collaboration (Crowston and Howison, 2005;

Krishnamurthy, 2002). Furthermore, novel research in network science has already outlined the importance of locality in determining the function of biological networks (Milenković and Przulj, 2008; Przulj, 2007).

In this study we extend previous research on the structure of online communities by bringing a different perspective to examine and study structure. First we study structure at the individual member-level rather than the aggregate network-level. Second, we focus on the *local properties* of structure rather than the traditionally considered *global properties*. While assessing global properties of structure requires the consideration of all members and their relationships in the community, the assessment of local properties requires information about a limited number of members and their relationships. This differentiation has significant theoretical implications. A local view of structure implies a more confined and local organization of work. Previous research has already suggested that open-source development is more disjointed and individual that previously thought (Krishnamurthy, 2002). As Coase (1937) reminds us, the transaction cost of coordination may exceed its benefits. This is why organizations exist in free market economies. This is perhaps true for online communities where despite the fact that technology democratizes communication, it is still more efficient to communicate and collaborate on a local scale.

This study is not without limitations. First, the concept of social capital is not without its critics. Social capital is a flawed concept as it overlooks the complexity and diversity of network ties that is required to understand the impact of personal connections (Warde and Tampubolon, 2002). Second, a structural perspective overlooks personal characteristics of members that affect their participation patterns in the community. For example, research has shown that members with longer tenure tend to participate more than other members (Wasko and Faraj, 2005). Professional affiliation also plays an important role in determining participation online. In the particular context of health IT, one could expect that professional users who are mostly doctors will have different participation behavior than tech-savvy members such as developers and service providers. In addition to affiliation and tenure, geographic location, expertise and gender are other characteristics that may be relevant in studying participating patterns that lead to knowledge creating in online communities.

Finally, other factors than structural position within the community contribute to social outcomes. Empirical studies indicate that online leaders tend to be long-term members of the group, entertain more ties with different others, and post frequently (O'Mahony and Ferraro, 2007). Such members are not necessarily more frequent posters than others if the community is based on knowledge sharing. In a study of a legal community, Wasko and Faraj (2005) found that experts while being more central were suspicious of the validity of content provided by non-expert others and engaged in exchanges with little expectation of reciprocity. Those limitations will be addressed in the next two studies where these factors will be taken into consideration.

Proposition	Supported	Comments
P1: global centrality is positively related with knowledge contribution in the community	Yes	global centrality is correlated with knowledge contribution
P2: global boundary spanning is positively related with knowledge contribution in the community	Yes	Global boundary spanning is correlated with knowledge contribution
P3: local centrality is positively related with knowledge contribution in the community	Yes	Local centrality is correlated with knowledge contribution
P4: local boundary spanning is positively related with knowledge contribution in the community	Yes	Local boundary spanning is correlated with knowledge contribution
P5: knowledge contribution is more positively related with local centrality that with global centrality in the community	No	Global centrality is more correlated with knowledge contribution than local centrality
P6: knowledge contribution is more positively related with local boundary spanning that with global boundary spanning in the community	Yes	Local boundary spanning is more correlated with knowledge contribution than global boundary spanning

Table 4: Summary of findings and research propositions

variable	type	mean	std	min	25%	50%	75%	max
messages	dependent	29.206	148.920	1	1	1.000	7.000	2700.000
attachments	dependent	3.164	24.997	0	0	0.000	0.000	475.000
richness	dependent	238.762	742.249	0	13	31.000	122.500	8929.000
c1	independent	0.579	0.598	0	0	0.492	1.000	3.432
c2	independent	0.357	0.354	0	0	0.309	0.570	2.206
c3	independent	0.316	0.299	0	0	0.276	0.514	1.442
c4	independent	0.298	0.279	0	0	0.269	0.474	1.275
c5	independent	0.284	0.264	0	0	0.265	0.453	1.199
c6	independent	0.271	0.249	0	0	0.263	0.429	1.000
c7	independent	0.261	0.239	0	0	0.260	0.411	1.000
c8	independent	0.255	0.233	0	0	0.257	0.403	1.000
c9	independent	0.250	0.229	0	0	0.256	0.400	1.000
c10	independent	0.247	0.225	0	0	0.254	0.394	1.000
c11	independent	0.244	0.223	0	0	0.254	0.389	1.000
c12	independent	0.242	0.221	0	0	0.253	0.385	1.000
c13	independent	0.241	0.220	0	0	0.253	0.384	1.000
c14	independent	0.240	0.219	0	0	0.253	0.383	1.000
c15	independent	0.239	0.218	0	0	0.253	0.382	1.000
c16	independent	0.239	0.218	0	0	0.253	0.382	1.000
c17	independent	0.238	0.217	0	0	0.252	0.381	1.000
c18	independent	0.238	0.217	0	0	0.252	0.380	1.000
c19	independent	0.238	0.217	0	0	0.252	0.380	1.000
c20	independent	0.238	0.217	0	0	0.251	0.379	1.000
c21	independent	0.237	0.216	0	0	0.251	0.379	1.000
c22	independent	0.237	0.216	0	0	0.251	0.378	1.000
c23	independent	0.237	0.216	0	0	0.251	0.378	1.000
c24	independent	0.237	0.216	0	0	0.251	0.378	1.000
cfull	independent	0.135	0.116	0	0	0.135	0.232	0.334
s1	independent	0.011	0.063	0	0	0.000	0.000	0.873
s2	independent	0.005	0.039	0	0	0.000	0.000	0.849
s3	independent	0.004	0.037	0	0	0.000	0.000	0.850
s4	independent	0.004	0.037	0	0	0.000	0.000	0.849
s5	independent	0.004	0.035	0	0	0.000	0.000	0.841
s6	independent	0.003	0.035	0	0	0.000	0.000	0.842
s7	independent	0.003	0.033	0	0	0.000	0.000	0.843
s8	independent	0.003	0.032	0	0	0.000	0.000	0.842
s9	independent	0.003	0.031	0	0	0.000	0.000	0.842
s10	independent	0.003	0.031	0	0	0.000	0.000	0.841
s11	independent	0.003	0.031	0	0	0.000	0.000	0.837
s12	independent	0.003	0.031	0	0	0.000	0.000	0.836
s13	independent	0.003	0.031	0	0	0.000	0.000	0.837
s14	independent	0.003	0.031	0	0	0.000	0.000	0.836
s15	independent	0.003	0.031	0	0	0.000	0.000	0.835
s16	independent	0.003	0.031	0	0	0.000	0.000	0.835
s17	independent	0.003	0.031	0	0	0.000	0.000	0.836
s18	independent	0.002	0.031	0	0	0.000	0.000	0.835
s19	independent	0.002	0.031	0	0	0.000	0.000	0.834
s20	independent	0.002	0.030	0	0	0.000	0.000	0.830
s21	independent	0.002	0.030	0	0	0.000	0.000	0.828
s22	independent	0.002	0.030	0	0	0.000	0.000	0.829
s23	independent	0.002	0.030	0	0	0.000	0.000	0.825
s24	independent	0.002	0.030	0	0	0.000	0.000	0.825
sfull	independent	0.001	0.012	0	0	0.000	0.000	0.332

Table 3: Descriptives of dependent (knowledge contribution) variables and independent (centrality and spanning) variables, $n = 894$

6 Code Integration in Online Communities

In the first study we have looked at knowledge creation in online communities at the individual (member) level. Taking a structural perspective we argued that local structural embeddedness influences members' knowledge creation. In this study, we complement the first study by examining how the community puts together individual contributions. In the context of open-source software, this corresponds into integrating code written by members for fixing bugs and implementing new functions. To an external observer, what is important is the codified knowledge that the community inscribes in textual documents and digital artifacts. However, from an internal perspective, the dynamics of interaction and individual learning are more important. Therefore, it is interesting to examine how internal dynamics and individual interactions are transformed into digital artifacts.

6.1 Code Integration in Online Communities

Writing program code is the main activity of members in open-source online communities. The goal of such communities is to evolve and share technological artifacts in the form of computer software and programs. Computer code is a form of explicit knowledge that could be transformed and shared with both members of the community and the users of the software. It is also a collective effort of the programmers and developers who are collaborating towards producing it. Because of the fundamental differences between traditional organizations and online communities such as the lack for formal hierarchical structure and the emergent nature of leadership (O'Mahony and Ferraro, 2007; Faraj et al., 2011), it becomes important to examine how online communities transform their members' individual contributions into a collective good.

Knowledge has been viewed as a duality by many researchers (Hildreth et al., 2002). These conceptualizations include formal and informal knowledge (Conklin, 1996), individual and collective knowledge (Rulke et al., 1998), and tacit and explicit knowledge (Nonaka and Takeuchi, 1995). The duality perspective acknowledges that knowledge has a second component that is less tangible and more implicit. For example, informal knowledge is difficult to represent in traditional sources of knowledge such as books and articles. Collective knowledge spans many sources and is not represented in one place. Tacit knowledge is context specific and difficult to communicate to others.

Under the duality of knowledge perspective, knowledge evolves overtime (Nonaka and Takeuchi, 1995; Von Krogh et al., 2000). This evolution follows repetitive phases at different levels. Knowledge can be explicit or tacit. Explicit knowledge is easy to formalize and transfer whereas tacit knowledge is personal, context specific, hard to formalize and difficult to communicate. Knowledge creation is the transformation of tacit knowledge into explicit knowledge. Online communities are a place in which such knowledge transformation happens. Through communication and interaction afforded by the online community platform, the tacit knowledge of individual members is transformed into an explicit knowledge in different forms such as articles in wikis and code in open-source software communities.

Knowledge creation is more central in online communities than in other organizations because the primary focus of these communities is to produce digital artifacts and sharable public goods rather than physical goods and services. Members in online communities have different incentives to participate and contribute to their online communities. Some members contribute out of primarily out of interest in the community (Wasko and Faraj, 2005), some because of pure altruistic behavior (Sproull et al., 2005) while others may have self-interested motivations (Butler et al., 2013; Wasko and Faraj, 2005). The altruistic behavior online has often described as extension to other forms of prosocial behaviors such as helping bystanders (Sproull et al., 2005). However, other non-altruistic factors are documented including professional motivations, becoming visible, gaining self-esteem and connecting to potential members (Butler et al., 2013; Wasko and Faraj, 2005). Regardless of why members contribute, their contribution is the pillar of online communities.

The different incentives of members to participate lead to different dynamics of participation. They also imply that online collaboration may not be as utopic and harmonious as previously thought (Faraj et al., 2011). In many open-source projects, the competing ideologies, discourses and pressure groups shape the diffusion process (Barrett et al., 2013). In SourceForge, a prominent code repository and development platform for open-source projects, many developers transition between projects. For any particular project, a large percentage of developers fail to become minimally active. However, a small percentage stick and overtime forms a core development group for the project (David and Rullani, 2008).

Because membership in online communities of innovation is diverse and open-ended, change in membership, contributions, community structure and processes is inevitable. Online communities can operate despite the heterogeneity and diversity of their members. Indeed, these two characteristics are identified as the motor of operation in complex systems (Page, 2001). However, it becomes essential to online communities to manage members' contributions. Overtime, bureaucratic managerial structures and autocratic leadership emerge in many growing online communities such as Wikipedia and Debian Linux (O'Mahony and Ferraro, 2007; Butler et al., 2008). Although status and power in most online communities are obtained by the virtue of participating and contributing (Levina and Arriaga, 2015), these emergent organizational structures determine how to orchestrate other members' contributions.

Members' contribution to online communities is contingent on many internal and external factors. These factors include the quantity and quality of other members' contribution, the degree of moderation in the community, and also the size of the community (Ren and Kraut, 2014). Communities achieving a certain *critical mass* provide their members resources leading them to contribute even if every single interaction does not produce value for them personally (Butler et al., 2015). Finally, the technology platform on which an online community is implemented plays an essential role in shaping interaction and contribution of its members (Faraj et al., 2013; Levina and Arriaga, 2015).

Lastly, specific to the context of this study, coordinating contributions is very relevant in online communities of healthcare IT. First, traditionally healthcare institutions are bureaucratic organizations. Healthcare communities of innovation bridge old and new organization structures (Lim et al., 2010). Second, health IT communities of innovation bring together members who have different professional backgrounds, training experience and incentives to participate. For example, doctors are highly paid and trained in the medical profession yet they are the primary users of health IT. Programmers and developers have a different background. Many contribute for the joy of hacking and loving technology. Other may have material incentives to participate such as getting paid for offering services to other members. We seek to examine the emergent administration that coordinate code integration in community driven development as well as the factors on which coordination is contingent.

6.2 Innovation in Open-Source Communities

Organizations change continuously over time. An organization is like a living organism that passes through different phases in its life with different structure, activities, and processes (Cameron et al., 1993, p. 45). Change rather than stability is advocated as a way of organizational life (Orlikowski, 1996). Different theories and perspectives that explain how and why change occurs have been proposed (de Ven and Poole, 2005; Demers, 2007). The same is true for communities, Faraj et al. (2011) propose that online communities are fluid: they morph and change over time yet retain a basic coherence and shape that defines them. Community members evolve different status hierarchies overtime (O'Mahony and Ferraro, 2007; Butler et al., 2013).

The classic literature in organizational learning argues that the process of innovation in organizations is composed of two consequent steps: exploration and exploitation (March, 1991). To succeed, an organization should balance the time and resources spent on exploration and exploitation. Organizations that favor exploration over exploitation will suffer the costs of experimentation without gaining its benefits and end up with underdeveloped ideas. On the other hand, organizations that favor exploitation will fail to maintain competitive advantage (Gupta et al., 2006). *Organizational ambidexterity* refers to the ability of an organization to engage in incremental and discontinuous innovation at the same time (Tushman and O'Reilly III, 1996; Junni et al., 2013). Research on innovation at the team-level also suggests that both divergence and then convergence among team members is needed to achieve closure and is a precursor for innovation (Sutton and Hargadon, 1996). This perspective has been extended for new settings including virtual and large scale collaboration (Dennis and Valacich, 1999; Pinsonneault et al., 1999).

Not only in traditional organizations but also in online communities, venturing in exploring new ideas and possibilities is a precursor of innovation. However, the capability to establish a common ground and converge is also essential for innovating. A stable environment and well-defined organizational actor roles are needed to converge. Innovation requires novelty. However, novelty is challenging because it may not necessarily lead to (Carlile and Lakhani, 2011). This view has been challenged by research that considered the divergence of interests, passions and time as the mode of operation in online communities (Faraj et al., 2011). Under this perspective, online communities are nexus of tensions and conflicts rather than agreement and harmony. Divergence rather than convergence is a main characteristic of online communities.

Research has adopted the classical view to study the process of innovation in new organizational forms such as open-source software development and online communities of innovation. For example, David and Rullani (2008) examined developers' movement across projects in open-source repository SourceForge. They found that developers explore several niche projects before deciding on a project to spend most their efforts on. As such developers couple the two subprocesses of exploration and exploitation at the individual level to optimize their productivity. O'Mahony and Bechky (2008) note that organizations of divergent interests can collaborate by creating a boundary organization and focusing on areas of mutual interests. Boundary organizations help both sides respond to managing challenges of governance, membership, ownership, and control over production.

Open-source online communities elicit contributions from a large body of members. This open-ended nature of participation results in a large number of contributions that compete for being integrated into the software. One perspective that suits this context is evolutionary change in which innovation is the results of the sub-processes of variation, selection and retention (Demers, 2007, ch. 7).

6.2.1 Variation

In traditional organizations, the process of variation includes changing current routines and competencies. Variation can be intentional resulting from conscious responses to both internal difficulties and the external environment or blind occurring independently of conscious planning (Aldrich, 1999, ch. 2). Variation in team processes involves eliciting ideas from all team members toward solving a problem, also known as brainstorming (Figure 25). In the process, seed questions related to problems at hand are presented. Then participants are probed to throw ideas related to solving the problem and answering the questions. (Sutton and Hargadon, 1996). In the context of open-source community, variation is natural and comes from the breadth of new ideas and contributions by all members.

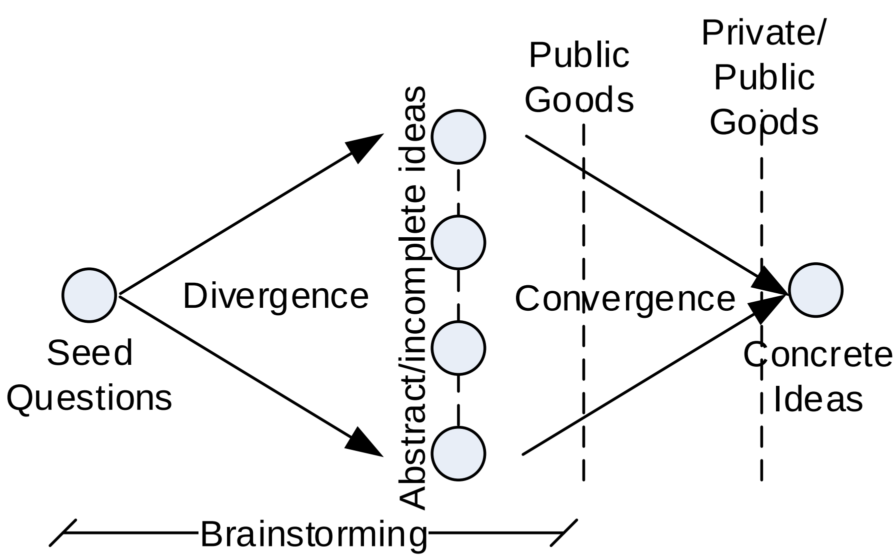


Figure 25: The idea generation process

6.2.2 Selection

In the process of selection, some of the variations got eliminated while others are selected. In organizations, the forces of selection could be internal to the organization or external set by the market forces and competitive pressure (Aldrich, 1999, ch. 2). Selection is more challenging than variation. For example in brainstorming, sessions are not documented and many participants may feel they wasted their time and scarified their precious ideas without gaining any privilege or reward after the session ends. In online communities, selection involves deciding what suggested changes and novel features are to be integrated into the software.

6.2.3 Retention

In organizations, during the retention phase, selected variations are preserved, duplicated, and reproduced (Aldrich, 1999, ch. 2). Selection after variation and divergence is needed for innovation in organizations. Retention is what makes innovation persists (Lim et al., 2010). In open-source communities, the digital distribution of the code as a non-rival, non-excluding, and free good well as the accompanied artifacts such as documentation ensures retention (Rayna, 2008). Retention is simply maintained by the growing user base of the software.

6.2.4 Variation and Selection in Online Communities

Research on online communities has focused on the convergence of collaboration toward producing goods and services. Online communities have been looked at as harmonious collections of individuals who collaborate because of various utilitarian and communitarian motives (Wasko and Faraj, 2005). Recently, researchers started questioning this assumption of shared interests, convictions and goals and called for looking at divergence, conflict, and tension in online communities as the underlying motor of production and collaboration (Faraj et al., 2011). Some researchers challenged the efficiencies of new forms of organization and shed light on the potential disadvantages of new collaborative platforms (Kane and Fichman, 2009). The debate is also echoed in recent organization science research that called for a reconsideration of theoretical and methodological issues for studying organizational networks and argued for a perspective that emphasizes the drivers of network change as well as the role of time in this process (Ahuja et al., 2012).

Online communities are built on open membership and voluntary participation and as a result provide an opportunity for all members to participate and contribute. At the surface, they lack central hierarchical authority (Dahlander and Frederiksen, 2012). Online communities are designed to "attract a critical mass of members who are willing to engage even if every single interaction does not produce value for them personally" (Butler et al., 2015, p. 2). Leadership in online communities is emergent and bottom-up (Butler, 2001; O'Mahony and Ferraro, 2007). This incentivizes members to participate and contribute in order assume positions of power and prestige (Levina and Arriaga, 2015).

Proposition 6.1 *All members of the community participate equally in proposing new features.*

Different roles are involved in the process of code integration in online communities. In more traditional information exchanges there are mainly two parties involved: knowledge seekers and knowledge providers (Constant et al., 1996). However, this simple classification evolves into an elaborate hierarchy overtime (Butler et al., 2008). Emerging leadership and bureaucratic hierarchies are important to filter and integrate individual contributions. The process of knowledge creation has a temporal dimension. On one hand, common ground is often needed to facilitate the sharing of expertise and resources (Borys and Jemison, 1989). On the other hand, diversity and heterogeneity of goals and interests may be generative for online knowledge sharing (Faraj et al., 2011; Page, 2001). Novelty of ideas is important for stimulating innovation. At the same time, novelty is risky because of the inherent ambiguity within it (Carlile and Lakhani, 2011). Most likely, convergence and new knowledge can be created via intense dialogic interactions in the community (Tsoukas, 2009). The trade-off between experimenting with new ideas and exploiting existing ones is a key process in organizational learning (March, 1991).

In addition to resolving conflicts, there is a need to filter individual contributions. Employing a resource-based perspective, Butler (2001) studied how sustainable social structures form in online communities (listserv). Despite the fact that members contribute time, energy, and other resources, such resource comes with an associative cost. Resources need to be processed by members in order to create benefits. Social structure in online communities forms as a balance between the opposing impacts of membership size and communication activity. These limitations of online collaboration are confirmed by other research in different contexts. For example, in peer-to-peer file sharing, the negative externalities associated with community growth in addition to free riding limits the size of the network (Asvanund et al., 2004). In virtual investment-relating communities, there is a diminishing marginal cost of additional information posted by members. The cost incurred by a user increases with the total number of postings in the community and is convex with regard to the total number of postings (Gu et al., 2007).

If divergence is the main theme in online communities and convergence is difficult to achieve, then how does the community manage to integrate knowledge? The availability of large number of members who are interested in pursing certain novel ideas and topics is important for creating variations. At the same time, core members in the community who assumed roles of administration and leadership are often responsible to curate contributions and generated content by others (Levina and Arriaga, 2015). Team members who occupy core positions in their organizations have arbitration power and as a result a higher chance of being selected (Cattani and Ferriani, 2008). At the same time, a core position limits the flexibility and novelty of ideas that its occupants have. An intermediate position between the core and the periphery is advantageous for creative contributions (Cattani et al., 2014). In online communities, members who occupy core positions in their communities as well as span boundaries across communities combine both arbitrage and brokerage power (Dahlander and Frederiksen, 2012). Those members have a higher chance to be nominated as leaders by other members (Johnson et al., 2015).

Leaders in online communities tend to be core members who provide help to other members (Faraj et al., 2015). However, this argument extend to other core members who happen to be experts in their field and provide help and answer other members' requests in the community (Wasko and Faraj, 2000). In open-source communities, where arbitration is required to accept code commits and change requests, experts are in a better position to have their word heard not only because of their expertise but also because of their reputation in the community. We propose:

Proposition 6.2 *Features proposed by members closer to the core have a higher chance of being selected.*

Proposition 6.3 *Features proposed by members with higher boundary spanning have a higher chance of being selected.*

One main assumption made in organizational learning literature is that divergence is temporary and convergence is possible afterwards. This is a reasonable assumption in a tightly controlled environment with an established hierarchy and a clear chain of command. In an online community, however, we argue that divergence is always increasing and convergence may not be easily achieved. First, membership in online communities is open and voluntary. As a result members will follow their interests and leave the community if they cannot pursue them. In addition, as more members join the networks the quantity and scope of interests in the community increase. In this respect, online communities are much like an open-system (Luhmann et al., 1995) in which entropy is always increasing (Von Bertalanffy, 1950). With time, variations and ideas retain supporters and legitimacy and increase their chance of being selected.

Proposition 6.4 *Features that have been suggested for longer duration have a higher chance of being selected.*

Instead of converging existing members on a certain idea, online communities attract members who already share interests and allow them to organize and push forward their interests. Because online communities have the capability of attracting a large number of members, it is not necessary to convert existing members, instead if enough members are interested in pursing a certain topic they can get together and collaborate on it. On the other hand, attracting members to participate and contribute requires a substantial effort and the availability of novel ideas that spark members' interests. Knowledge creation in online communities results from the concentration of many members around certain topics of interests. Those members who are already convinced by their cause carry it forward. The availability of critical mass is one of the forces behind knowledge creation in online communities (Ren and Kraut, 2014; Butler et al., 2015).

Proposition 6.5 *Features that are specialized have a higher chance of being selected.*

6.3 Research Setting

We examine the process of integration in the OSCAR community. The community promotes itself as platform built on trust, transparency and respect to empower and promote its members to collaborate towards developing and sustaining the OSCAR EMR. Indeed, the community uses the metaphor of a heart to depict itself (Figure 26). In this heard ideas, innovations from anyone are welcome. Those ideas follow a review process in which a technical committee and a product management committee approve new requests. Approved requests are contributed to the code based and are integrated into OSCAR after code review and testing.

Contributing a new feature to OSCAR is performed via SourceForge feature request platform (Figure 27). In addition, to comply with ISO 13485:2003 standard, the contributor should fill and submit a feature request form indicating the details of the proposed contribution (Figure 28). The request then follows the review process indicated above, informally via the OSCAR heard or formally in Figure 29. We list below the description of the process as obtained via private communication with the community manager:

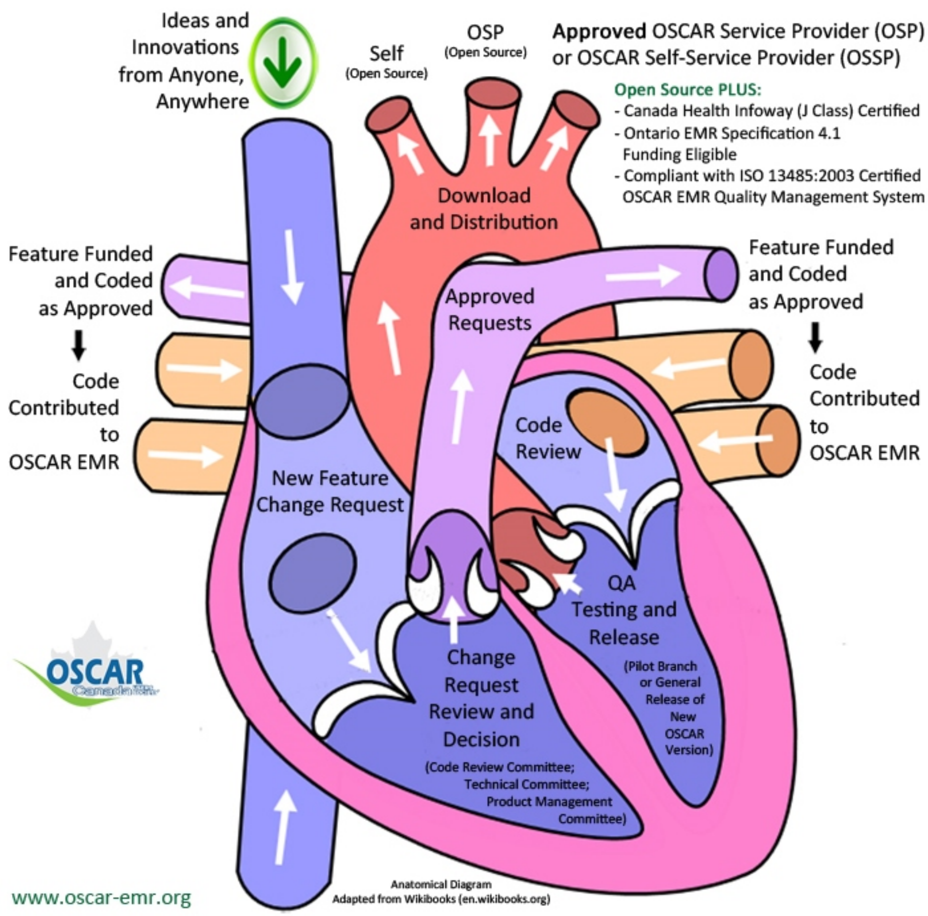


Figure 26: Idea selection and retention in the OSCAR community

The Change Notice Request form can be found in the Feature Requests section of the Tracker on the SourceForge website (`http://sourceforge.net/tracker/?group_id=66701&atid=515435`). This form includes a description of the new feature, where it interacts with OSCAR, any risks to integrating the new feature, any risks from not integrating the new feature, testing requirements and training requirements. Figure 3 provides a sample of the Change Notice Request Form.

In order to submit a change notice request, Approved OSPs must register on SourceForge (`http://sourceforge.net/projects/oscarmcmaster/`). Once registered, access will be available to submit new change requests at the following site: `http://sourceforge.net/tracker/?group_id=66701&atid=515435`.

Once the Change Notice Request form has been accepted by the OSCAR Change Request Committee, the contributor can begin pushing commits to GIT/Gerrit. In order to commit to GIT/Gerrit, Approved OSPs must sign up for Gerrit at `https://source.oscartools.org`:

#	Summary	Milestone	Status	Owner	Created	Updated	Pric
915	Billing RAMQ features for QC Certification	None	open		2013-07-18	2013-07-18	5
864	New Pain Measurements	Main Trunk	closed	John Wilson	2013-05-08	2013-05-08	5
839	Health Tracker	Clinical Features	closed	John Wilson	2013-03-25	2013-11-25	5
771	Wailist: email trigger for new online application	CAISI A - CAISI Collaborative	closed	zhoulm	2012-11-05	2012-12-20	7
759	PMM default login screen	None	open	annie	2012-10-12	2012-10-12	6
754	Anonymous Notes for Phone Encounters	CAISI D - Waitlist Management	open	tiger	2012-09-20	2012-12-28	7
714	Add comment as tooltip on druglist	None	closed	John Wilson	2012-05-30	2013-05-13	5
635	Allow change of invoice type	spec4	open	Randy Jonasz	2012-01-25	2012-01-25	5
634	Allow to bill for Pall Care Case Management similar to INR	spec4	open	Randy Jonasz	2012-01-25	2012-01-25	5
633	Prompt to bill once-a-year codes	spec4	open	Randy Jonasz	2012-01-25	2012-01-25	5
632	Improve OHIP simulation report	spec4	open	Randy Jonasz	2012-01-25	2012-01-25	5
631	Improvements to preventions	spec4	open	Randy Jonasz	2012-01-25	2012-01-25	5
630	Add UI for modifying stop sign rules	spec4	open	Randy Jonasz	2012-01-25	2012-01-25	5
629	Improve supervising of resident billing	spec4	open	Randy Jonasz	2012-01-25	2012-01-25	5
628	Create Document Viewer for eChart	None	open	Randy Jonasz	2012-01-20	2012-01-20	5

Figure 27: Feature request page `http://sf.net/p/oscarmcmaster/feature-requests/`

`8080`. Once signed up for Gerrit, they must make themselves known on the OSCAR Developers Listserv (`https://lists.sourceforge.net/lists/listinfo/oscarmcmaster-devel`) and request to be added to the OSCAR group. Once added to the group, they will be able to commit code, referencing the title or ID# of the Change Notice request as listed on the SourceForge Tracker.

Code that has been submitted will be reviewed and approved before being incorporated into OSCAR. Any other commits to GIT/Gerrit will be rejected if the code reviewer cannot ascertain the intended purpose for the commit.

If an Approved OSP has a client who would like to be the Beta Test site for a particular

OSCARMcMaster

OSCAR Change Request Form

Date:

Change Description (incl. Benefits)

OSCAR Interaction points

Integration risks

Risk if not incorporated

Minimum testing requirements/process

Training requirements to support changes

This is the Document Control ID # and revision status of the document.

DEV-F01-Change Notice Request Form.doc Rev. 1.1Rev. 1.1 – 8/10/2011 Page 1 of 1

Figure 28: Change Notice Request Form - DEV-F01

Change Request, the Approved OSP is required to complete the Beta Test Approval Form (Appendix C) and upon completion of the Beta test phase complete a Beta Test Completion Form (Appendix D). Both forms must be submitted to the Quality Management Coordinator at OSCAR EMR once completed.

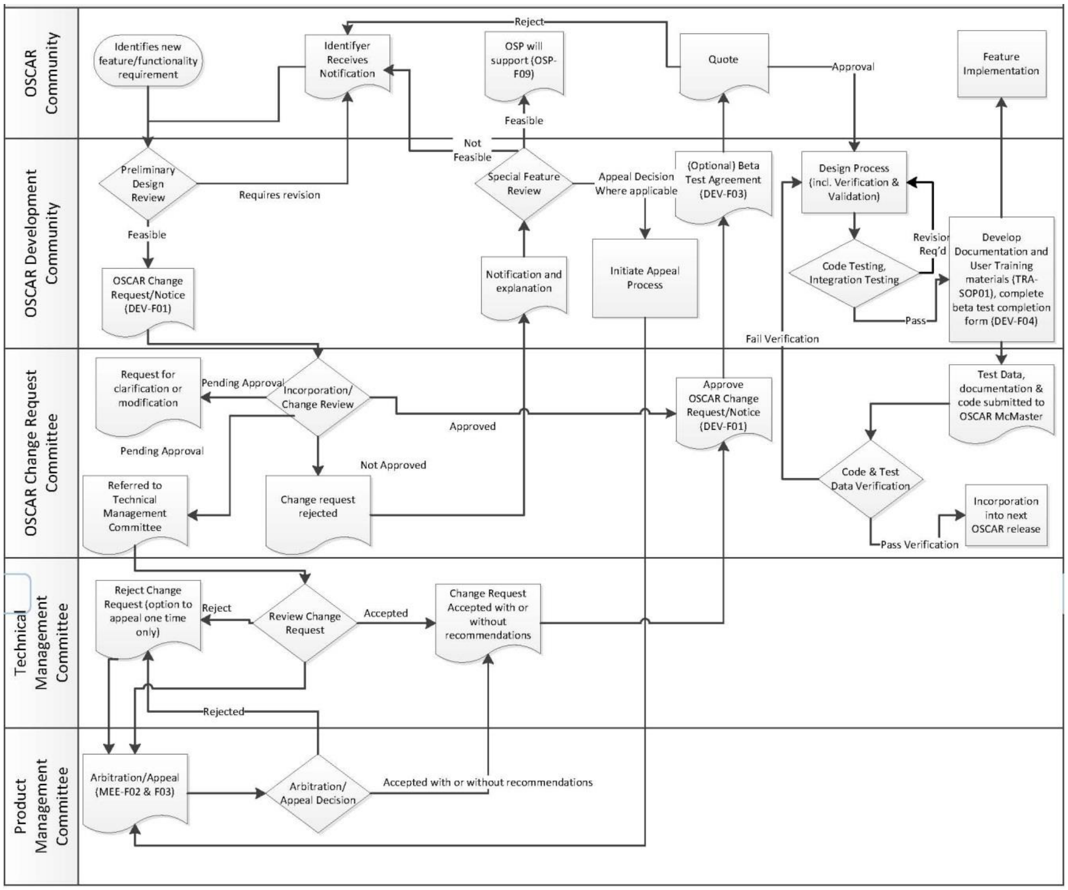


Figure 29: OSCAR Feature Design Process - DEV-SOP02

6.4 Measurements

We examine the contributions submitted to the feature requests' platform on SourceForge (Figure 30). Each feature is proposed by a member identified by her SourceForge username. During its lifetime, a feature may have different statuses:

- Open Resolution-None: the request hasn't been looked at and is still considered new.
- Pending Resolution-None: the reviewer has read the request and is requesting a clarification.
- Pending Resolution-accepted: a green light from the technical committee to move ahead with development.
- Pending Resolution-remind: the request has been scheduled for review at the next Technical Committee meeting.

- Pending Resolution-postponed: the technical Committee has reviewed the request and is requesting more information about the change.
- Closed Resolution-fixed: the code has been committed to the repository and has been approved to be included in the next release.
- Deleted Resolution-rejected: the request has been rejected.

We consider those requests that were fixed to be fulfilled. This is the binary dependent variable in this research that we try to explain based on the contributing member characteristics and the feature request characteristics as well (Table 5 summarizes the research variables).

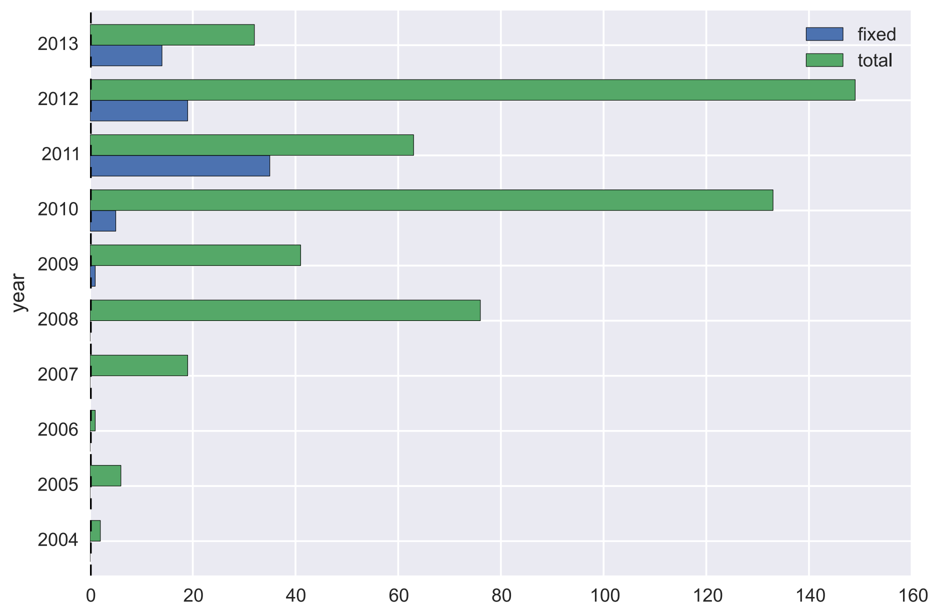


Figure 30: Feature requested and fulfilled

Contributing member characteristics are measured by considering the activity of the member on the discussion mailing list of the community. It is worth noting though that many requests are submitted by anonymous members. In addition, there is no explicit match between the member name on the feature request, which is the username in SourceForge, and the identification on the mailing list, which is a masked email address used by the member to send messages. We matched member names across the two platforms (feature requests and mailing list) manually, some members could not be match and regarded them as anonymous members. Next, we constructed a network representing communication in the mailing list using the same methods of the previous study. We consider the following characteristics of the contributing member:

	type	description
fixed	dependent	A binary variable indicating whether the feature has been fulfilled and integrated into the code or not
age	independent / feature	The number of days since the submssion of the request until 4 October 2014 (which is the day we collected the data)
richness	independent / feature	The number of unique words expressed in describing the request
specialization	independent / feature	The degree of specialization of the request measured using the topic model of all requests
mcentrality	independent / member	The closeness centrality of the requesting member in the network representing the maillist communication of the community
mspanning	independent / member	The boundary spanning of the requesting member (measured with betweeness centrality) in the network representing the maillist communication of the community
mcoreness	independent / member	The core number of the member

Table 5: Research variables

- Centrality: The closeness centrality of the requesting member in the network representing the maillist communication of the community.

- Boundary spanning: The boundary spanning of the requesting member (measured with betweeness centrality) in the network representing the maillist communication of the community.

- Coreness: The core number of the member.

The exact implementation of these measurements is described in the previous study. Finally, we considered an anonymous contributor to be very peripheral members whose centrality, spanning and coreness are zero[9].

Next, we consider the characteristics of the request itself. First, we look at the age of the request defined as the number of days since the submission of the request until 4 October 2014 (which is when we collected the data from the repository). We also examine the body of text used to describe the request. Because communication among members is

[9]Similar results were obtained when we excluded anonymous members from the analysis

primarily written and because modern technology affords the permanent archival and quick access of communication, the dynamics of code integration are reflected in the discourse of communication within the community. We consider two linguistic characteristics related to a feature request:

First, we examine the richness of the request defined as the number of unique words used to describe it. Second, we consider the degree of specialization of the request. This is a trickier measure than the first one because it depends on the quality of words used to describe the request. A document from a collection of documents is shown. The content of the document is known. From this information, topics which are probability distributions over words are computed. For that purpose we employ a Latent Dirichlet allocation (LDA) topic model (Blei et al., 2003; Blei and Lafferty, 2007). In this model, the descriptions of requests are a corpus of documents each composed of a set of words. Each word is associated probabilistically with a set of topics. Figure 48 demonstrates the idea of a topic model. This model definition is cyclical because its components: documents, words and topics, are all interdependent. We only observe words' association with documents. To estimate the model parameters, Gibbs sampling, a method that estimates the distribution of authors over topics and words over topic is used (Blei et al., 2003). Once the model parameters are estimated, the resulting model can be used for both generative and discriminative purposes.

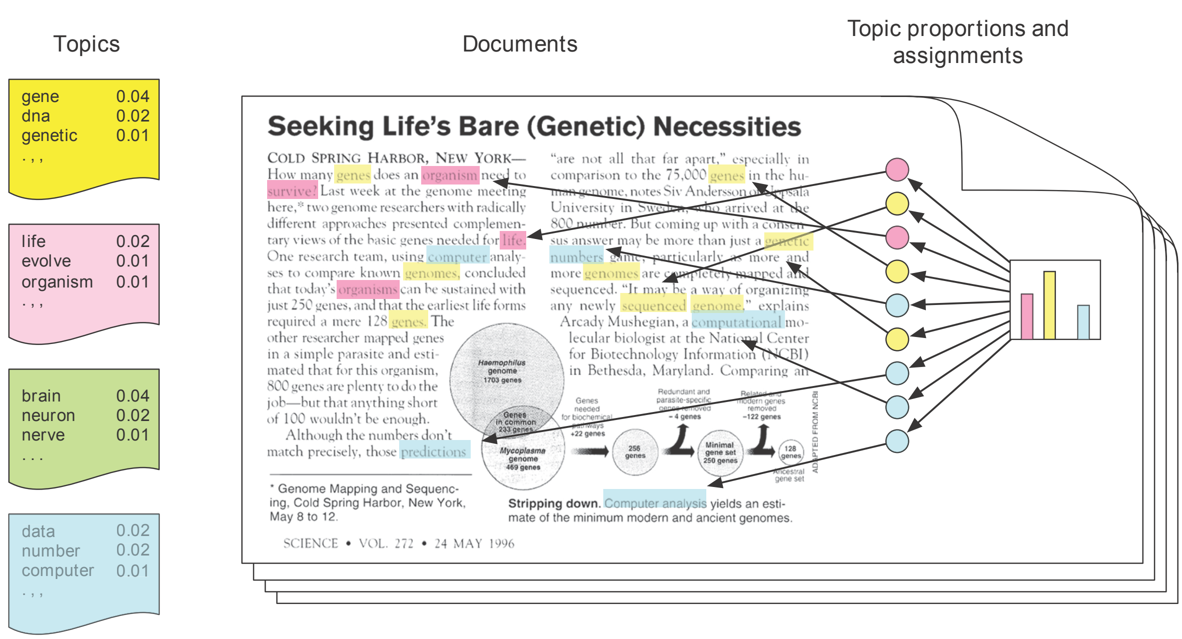


Figure 31: An example of a topic model (adapted from from Blei (2012))

We used the GenSim package to estimate the LDA model (Rehurek and Sojka, 2010) and the Termite Package to visualize the model (Chuang et al., 2012) as shown in Figure 32. As the figure demonstrates, words have various associations with topics. Some words have high attachments with few topics. We consider this a sign of specialization as those words are indicative of few topics. For example, the word “process” is a very general word while the word “catalina” is very specialized. The same is true for documents representing requests. A request is probabilistically assigned with each topic depending on the words it contains. The specialization of a request is the mean of its probability distribution over topics:

$$specialization(request) = \frac{\sum_{T \in topics} P(request|T)}{|topics|}.$$

6.5 Results

The distribution of independent variables are shown in Figure 33. The descriptive statistics of all variables are listed in Table 6. There are 74 only requests that are fulfilled among 522 (Figure 34). Looking at the distribution of coreness of members (Figure 35), half of feature requests come from core members (Members at cores 8 and 7 contribute 51.5% of requests). This 50-50 split is atypical in online communities where a power-law distribution of contributions result roughly in a 80-20 split.

	count	mean	std	min	25%	50%	75%	max
fixed	522	0.1418	0.3491	False	0	0	0	True
age	522	1462	617.6	610	929	1456	1857	3779
richness	522	27.8	27.83	1	9	23.5	38	334
specialization	522	0.3221	0.1986	0.03333	0.1965	0.2471	0.4067	0.9923
mcentrality	522	0.2605	0.08617	0	0.2308	0.2714	0.3055	0.485
mspanning	522	0.01403	0.04063	0	0	0.001827	0.01073	0.3238
mcoreness	522	5.462	2.612	0	3	6	8	8

Table 6: Descriptive stats

Next we examine, the correlations among the dependent variable (`fixed`) and independent variables. Because of the binary nature of the dependent variable, we report both person correlation coefficients (Figure 36) and Spearman correlation coefficients (Figure 37). In both scales, `fixed` is negatively correlated with `age` and positively correlated with `mcoreness` at approximately 30%. All other dependent variables are not significantly correlated with `fixed`.

Next we perform a logistic regression analysis (Long and Freese, 2006). We perform two regression analyses incorporating variables related to the request characteristics and requesting member characteristics individually and a model including both. We exclude

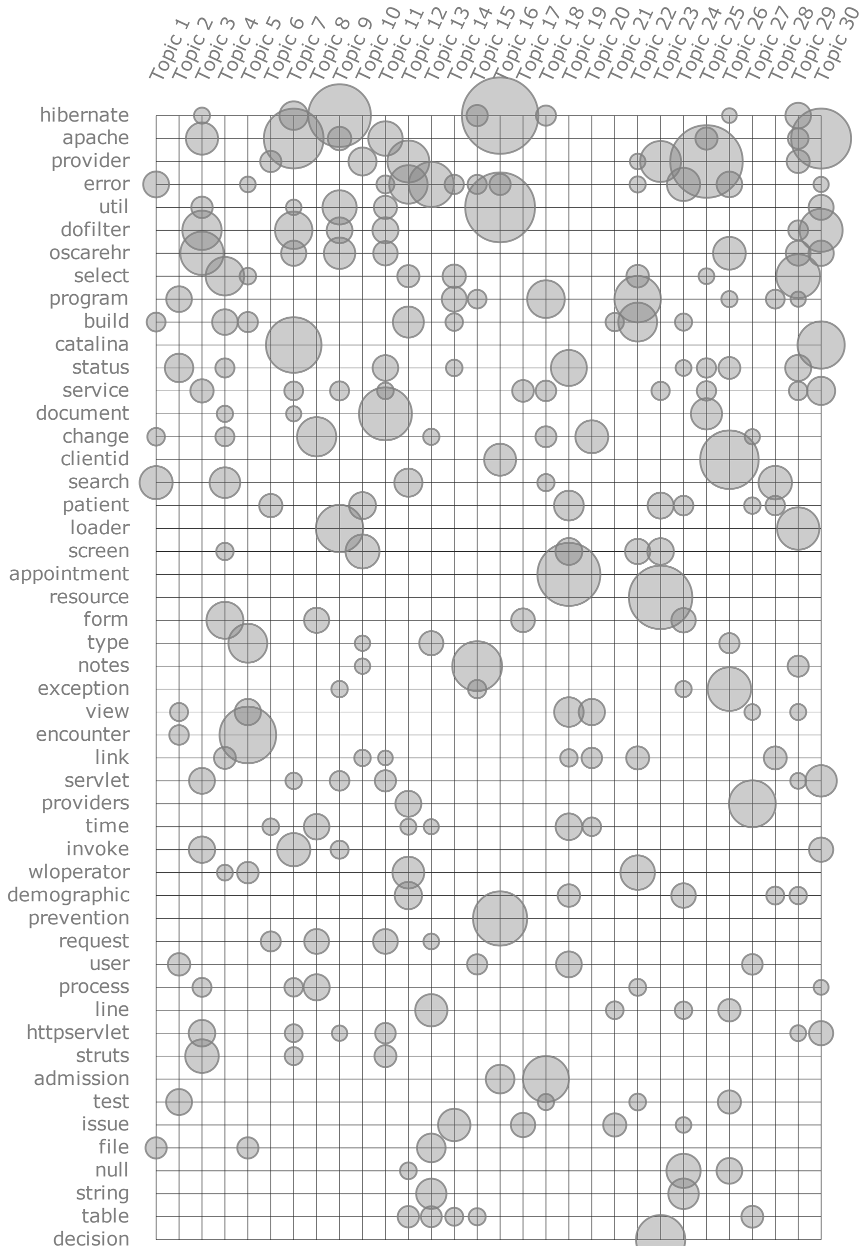


Figure 32: topic model

mcentrality from the analysis because of its collinearity with mspanning. Results are reported in Table 7. Results confirm those found in the correlation analysis. Examining the full model, mcoreness is the only variable that is significant statistically and economically. age is statistically significant but its coefficient is close to zero.

Except for the position of the member in the core of the community, no independent variable was significantly correlated with code integration. This includes both request characteristics such as specialization and novelty and requestor characteristics as well. In addition, core

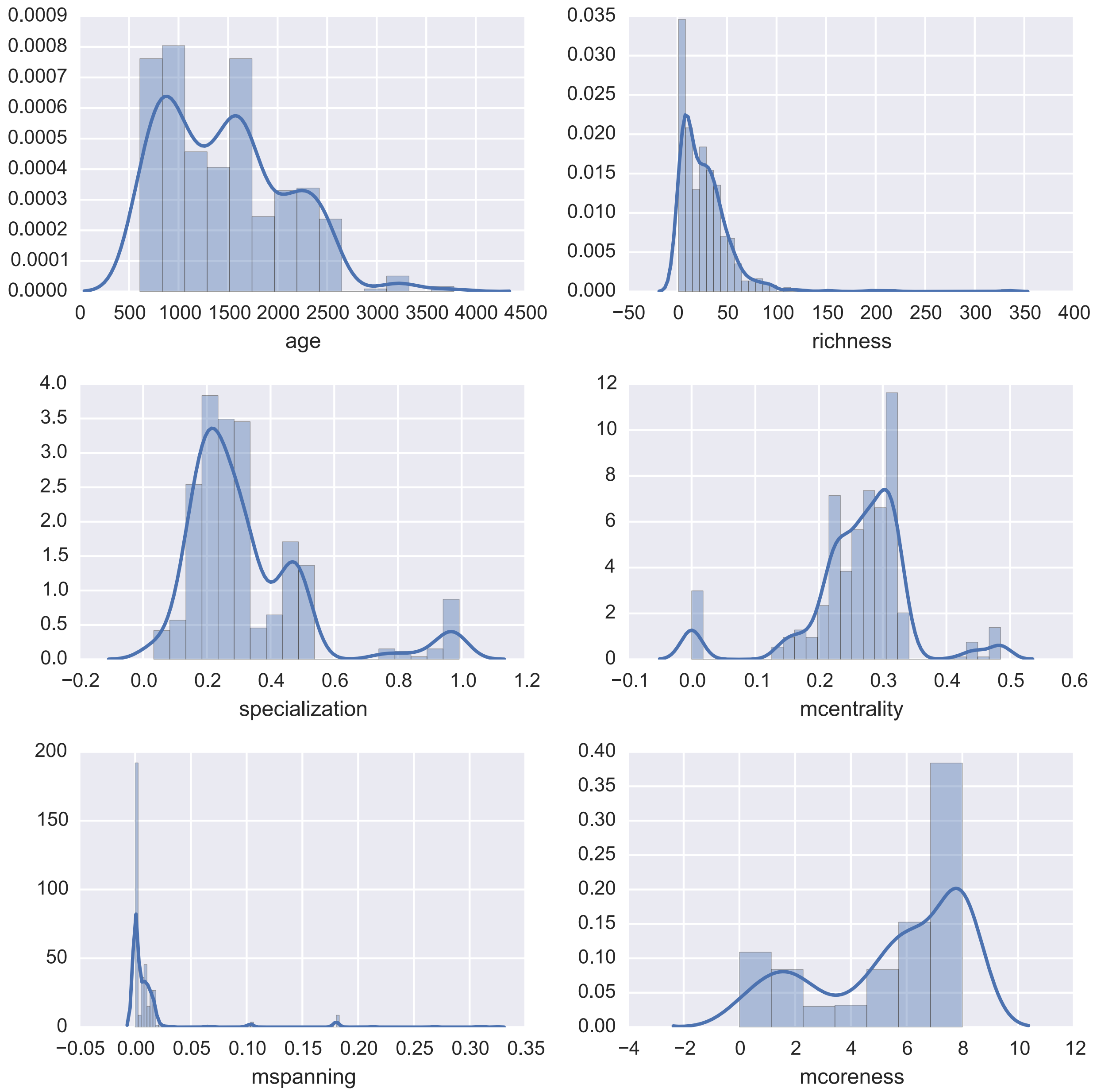


Figure 33: Distribution of variables

members submit more than half of new features. The results, indicate the primacy of core members in submitting and overseeing code integration requests in the community. Results and proposition summary is presented in Table 8.

6.6 Discussion

Online communities excel in providing an opportunity for members to contribute and participate. The diverse contributions are channeled into a digital good via a process of

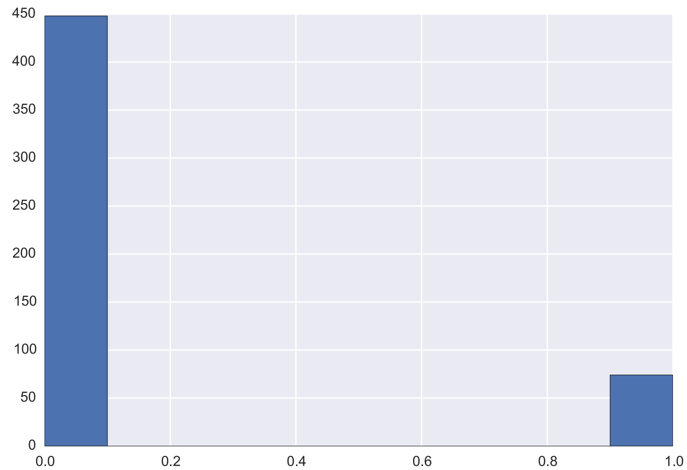


Figure 34: Distribution of the dependent variable (request is fulfilled)

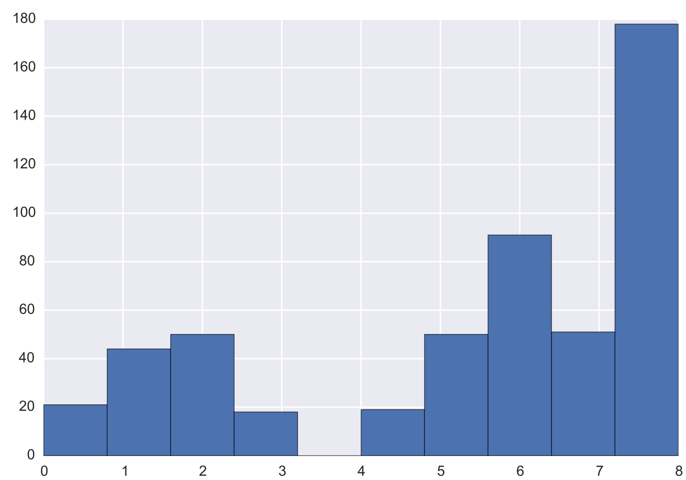


Figure 35: Distribution of members' coreness

code integration in which few ideas are selected and retained. As we found in the previous study, the OSCAR community encompasses a large and heterogeneous body of members collaborating on a local scale. The community has set a platform to formalize and collect members' contributions and channel them toward the next release of the software via a review process (Figure 30). The community views itself as a democracy where every member can contribute and also participate in selecting others' contributions.

Results hint to a different picture. Not only do half of contributions (i.e. variations) come from core members of the community, but also their contributions have twice the chance to be selected and retained. Code integration is tightly controlled by the core members in the community. One plausible explanation is that code integration has inherent difficulties that are salient for offline and online collaboration. One issue is evaluation apprehension which

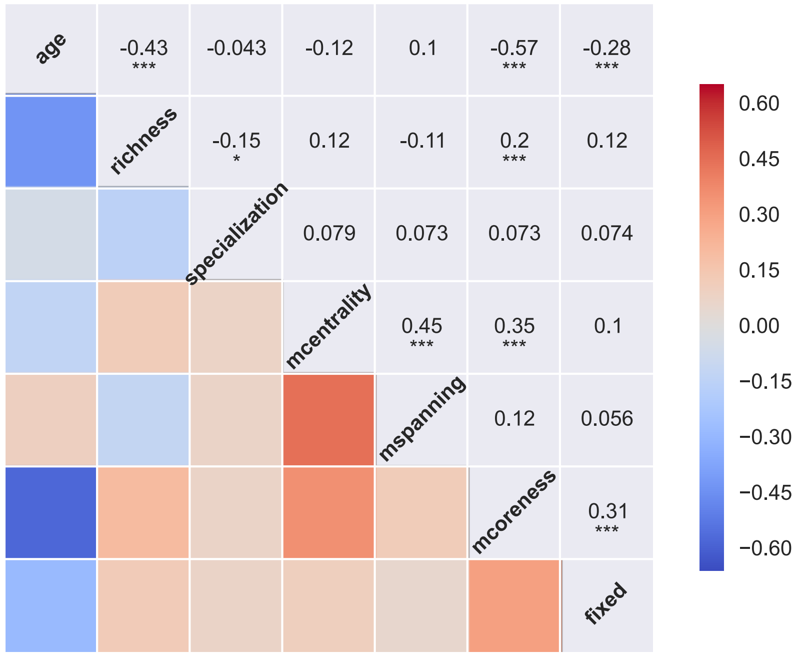


Figure 36: Pearson correlation matrix

is the fear of negative evaluations from other members (Diehl and Stroebe, 1987). Another is free riding which suggests that members perform less productively when they work in a group than when they work alone (Albanese and Van Fleet, 1985; Diehl and Stroebe, 1987).

Whether members are free riding on core members in the community is debatable. Research on open collaboration emphasizes the role of the periphery in contributing (Jeppesen and Lakhani, 2010). Another explanation is that while many members contribute variations and new ideas, they do not explicitly channel their contributions towards being selected and integrated. The fact that it is possible to customize the OSCAR software via pluggable forms and templates without changing the code may also explain why peripheral members shy from channeling their contributions into the platform. A third explanation, which we explore in the next study, is that members participate in the community for personal goals that are not necessarily related to coding and producing software.

6.7 Conclusion

New organizational forms such as communities of innovation are flat rather than hierarchical (Lim et al., 2010), more fluid and emergent (Faraj et al., 2011). Membership is diverse and distributed temporally and physically (Cummings et al., 2009), leadership structure is emergent and implicit (O'Mahony and Ferraro, 2007; Johnson et al., 2015), and collaboration

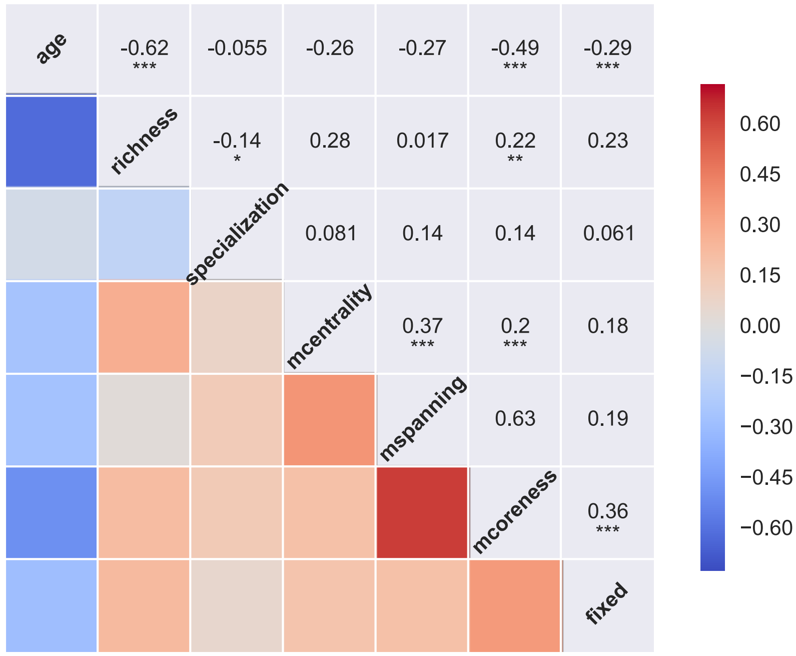


Figure 37: Spearman correlation matrix

processes are not stable and change over time (Crowston and Howison, 2005). Contrary to previous research, we focus on code integration within one community. We look for the determinants of code integration at the project-level rather than in a large collection of projects (David and Rullani, 2008). How do members come together and combine their ideas and experiences in online communities? To answer this question we examine the two processes of variation and selection in the OSCAR community.

We examine several determinants of code integration in the OSCAR community. These determinants include members' position in the community and the characteristics of their contributions. Interestingly, what matters for getting a contribution selected is the position of its requesting member in the core-periphery structure of the community. Moreover, half of variations are contributed by core members. While this finding confirms pervious studies that emphasized the role of the few in online communities (Wasko and Faraj, 2005), it is challenging the proposed fair and democratic nature of open-source development (Von Krogh et al., 2003).

This study is not without limitations. First, while the OSCAR project is hosted in SourceForge, not all communication and activity happen in the online sphere. The community hosts yearly national and regional meetings to discuss and promote the development of OSCAR. The offline component is absent in this study which focuses on development and communication happening online.

	request	member	full
intercept	-2.1921***		-2.8249***
	(0.1756)		(0.2802)
age	-0.0018***		-0.0015***
	(0.0003)		(0.0004)
richness	0.0002		-0.0008
	(0.0052)		(0.0048)
specialization	0.7556		0.0710
	(0.6127)		(0.6726)
mcoreness		0.1684***	0.5546***
		(0.0355)	(0.1195)
mspanning		0.7122	2.9022
		(2.2195)	(3.1902)
Log-Likelihood	-187.398	-349.607	-168.596
BIC	399.828	711.730	374.739
AIC	382.797	703.215	349.193
N	522	522	522
all variables are mean-centered			
Standard errors in parentheses			
* p<.1, ** p<.05, ***p<.01			

Table 7: Logistic regression results

Second, technological characteristics that facilitate interaction online by providing a reliable and easy to use platform are a precursor to members' information seeking and providing behavior in the online community (Phang et al., 2009). Specific feature such as virtual co-presence support, persistent labeling, self-presentation, and deep profiling are enablers of online knowledge contribution (Ma and Agarwal, 2007). Modern online collaboration is more nuanced than technologies typically studied in communication studies such as telephony and email (Carlson and Zmud, 1999). Technology is getting better at a fast pace and is affording more collaborative synchronous and asynchronous capabilities (Dennis et al., 2008) that allow members to mix and remix their contributions (Hill and Monroy-Hernández, 2012). For example, open-source software development tools such as Github allows programmers to easily collaborate on a project, fork it to create another project, and mix it with other projects in the platform (Dabbish et al., 2012). We examine the role of the platform plays in shaping members' behavior in the next study.

Proposition	Supported	Comments
P1: All members of the community participate equally in proposing new features	No	Half of new features are proposed by core members
P2: Features proposed by members closer to the core have a higher chance of being selected.	Yes	A core position increases the chance of a feature to be fulfilled
P3: Features proposed by members with higher boundary spanning have a higher chance of being selected	No	Effect is not statistically significant
P4: Features that have been suggested for longer duration have a higher chance of being selected	No	With time the chance of a feature to be fulfilled diminishes
P5: Features that are specialized have a higher chance of being selected	No	Effect is not significant

Table 8: Research Propositions

7 The Power of Words: New Members' Engagement in Online Communities

A fundamental question in online communities is to understand how members establish and maintain mutual relationships that grow the community overtime. This is an important question to ask because although online communities appear as anarchic collections of individuals largely devoid of a formal authority, their members develop strong group norms and leadership structure (O'Mahony and Ferraro, 2007; Johnson et al., 2015), successfully generate common goods, and outperform traditional organizations in many domains (Tapscott and Williams, 2008). Online communities bring people who share similar interests and motivations to collaborate toward achieving a common goal that is not institutionally mandated (Sproull and Arriaga, 2007; Coakes and Smith, 2007). Collaboration among members is facilitated by IT platforms that afford various communication capabilities (Dennis et al., 2008). Recent research suggests that different IT affordances have great consequences on the outcomes of collaboration in online communities such as knowledge creation (Faraj et al., 2013).

This paper focuses on understanding how new members in online communities establish and maintain relationships with existing members. We examine members' engagement in an open-source community over seven years. We differentiate between short-term engagement defined as interactions and ties established by new members in the period following joining the community and long-term engagement defined as ties and interactions entertained by the new members in the long run. Furthermore, we study short-term and long-term engagement among members in online communities built on technology platforms that are based primarily on the exchange of written communication. Examples of such platforms include SourceForge, Wikipedia, and the majority of discussion forums and online bulletin board. Recently, with the advent of Web 2.0, social media and mobile applications, there has been a shift to incorporate richer features into online collaboration platforms (e.g. GitHub &

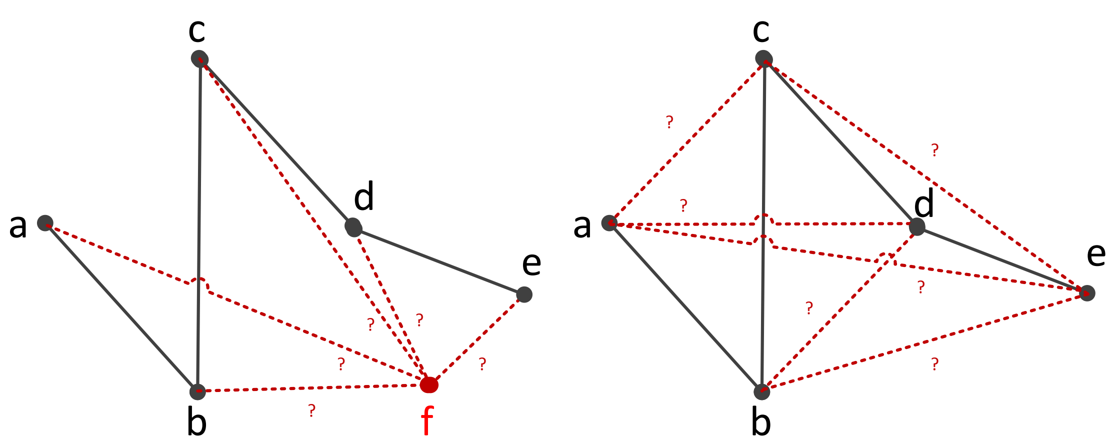
Knol). This trend is built on the premise that such features will be advantageous over pure text communication. However, this premise is not grounded in research as we do not know much about on how written words shape the dynamics of online communities including tie formation and tie persistence (Levina and Arriaga, 2015).

It is interesting to note that most research on community evolution focuses on how a community emerges or dies (Backstrom et al., 2006; Kairam et al., 2012; Hallerbach et al., 2013; Danescu-Niculescu-Mizil et al., 2013). There is a gap in examining the processes that sustain and nurture interaction and collaboration (Dahlander and McFarland, 2013). We propose that new members in an online community utilize different strategies for short-term and long-term engagement with existing members. Those strategies are partially shaped by the technological platform on which the community is implemented, but also by the new members needs and skills when joining the community. We examine three classes of strategies: structural, demographical, and behavioral and discuss how they apply within different phases of tenure in the community. Furthermore, we examine how these strategies shape the evolution of the community overtime.

7.1 Tie Formation and Sustained Engagement in Online Communities

Extant literature has so far focused on identifying the reasons that drive members to participate in online communities. A wide variety of factors have been identified: motivational factors based on self-interest, access to information, advice seeking, reputation building, expertise signaling, altruism, empathy, reciprocity, bonding with others, and commitment to the community goals (Constant et al., 1996; Lakhani and Wolf, 2005; Wasko and Faraj, 2005; Kudaravalli and Faraj, 2008). The nature of participation on the other hand is affected by a variety of structural and social factors such as message volume, content, and contributors' characteristics (Butler et al., 2013; Ridings and Wasko, 2010). Online communities are dynamic and fluid (Faraj et al., 2011), evolve practices that support their sustainability and growth (Kudaravalli and Faraj, 2008; Ridings and Wasko, 2010), and have governance and leadership structures that are emergent and highly situated (O'Mahony and Ferraro, 2007). The fluidity of online communities calls for a more focused consideration of how members establish their relationships and interact online. A perspective distinguishing short-term from long-term interactions suits well the dynamic context of online communities. Furthermore, this perspective has been employed to study relationships in traditional organizations.

A recent study by Dahlander and McFarland (2013) examined academic collaboration of new faculty in Stanford University. Their findings suggest that hired professors utilize different strategies for forming new ties and maintaining existing collaboration ties in the longer term: “ties form when unfamiliar people identify desirable and matching traits in potential partners. By contrast, ties persist when familiar people reflect on the quality of their relation-ship and shared experiences” (Dahlander and McFarland, 2013, p. 69). Because online communities are characterized with open and dynamic membership, we differentiate between two interrelated questions: (1) how do new members, when joining an online community, establish ties with existing members in the community, and (2) overtime, how do those members evolve long-term ties with others. Figure 38 illustrates the difference between the two questions. We refer to the first question as short-term engagement and to the second one as long-term engagement. We argue different mechanisms explain the two types of tie formations in online communities.



(a) Short-term engagement **(b)** Long-term engagement

Figure 38: Long term and short-term engagement (ties at time t & $t+1$ are colored black and red respectively)

7.1.1 Short Term Engagement in Online Communities

The question of how a new entrant picks others to link to has been extensively studied in social networks (Snijders, 2001; Burt, 2002; Kaiser and Hilgetag, 2004). In contrast to long-term engagement among existing members who are assumed to know each other, initial tie formation brings a new member who is not known to the community. In addition, depending on the context, this new member may partially observe the characteristics and relationships of existing members. There are not mutual expectation of direct or indirect reciprocity. The new member’s decision to establish a tie with another existing member is based on this partial observation of members and their interaction in the community.

"Newcomers were attracted to those modules where entry points were present and visible and contribution barriers were low" (Von Krogh et al., 2003, p. 1235).

Joining a community and establishing relationships with others may stem from utilitarian needs such as the desire to acquire resources (Butler, 2001), entertain professional relationships with experts (Kudaravalli and Faraj, 2008), and obtain brokerage power (Burt, 1992). However, communitarian needs such as the desire to belong to a group or to help others are very common at this stage (Constant et al., 1994, 1996). In most online settings, short-term engagement occur under a veil of ignorance where members' identities, activities and position are not fully disclosed. Indeed, the capability of members to hide, show and manipulate their online identities is a prime feature that creates unique dynamics in online settings (Schultze, 2012). Furthermore, in many online communities, experts provide advice and help to novice members without a prerequisite relationship (Faraj et al., 2015).

7.1.2 Long Term Engagement in Online Communities

In online communities, establishing a relationship typically involves cooperation or providing help within the context of the community such as answering a question in forums (Constant et al., 1996) or resolving a bug in open-source communities (Von Krogh et al., 2003). Long term tie formation rests on the assumption the existence of a form of reciprocation. Research in social exchange theory uncovered various mechanisms to explain how social actors establish relationships (Willer et al., 2002). In direct reciprocity, a first member may help a second member if the second member has helped the first member in the past or is expected to help in the future (Nowak, 2006). On the other hand, in indirect reciprocity (informally referred to as karma), a first member helping a second member increases the likelihood of the second member helping a third member and also the likelihood of the first member to be helped by someone else than the second member (Nowak and Roch, 2007). The two mechanisms are special cases of generalized exchange, also commonly known as gift economy, where members are willing to provide help without immediate expectations but only long-term expectations of getting helped in the future (Mashima and Takahashi, 2008).

Direct reciprocity and indirect reciprocity mechanisms are employed by members in online communities focused on technical topics (Faraj and Johnson, 2011). Surprisingly, these communities tend to steer away from preferential attachment, a mechanism that explains the formation of power laws in physical and social phenomena including online (Capocci et al., 2006). Whereas the tie formation process occurs among unfamiliar strangers seeking points of similarity, the tie persistence process happens among people who are familiar with each other to sustain and extend their association. In this process, members will reflect on their past engagement experience as well as their common background and interests (Dahlander and McFarland, 2013, p. 97). Next, we present several mechanisms of online engagement and reflect on how they are associated with short-term and long-term engagement among members in online communities.

7.2 Engagement Mechanisms in Online Communities

Research examining network evolution has typically focused on the structural factors that shape the evolution of networks. "Experiments on large coauthorship networks suggest that information about future interactions can be extracted from network topology alone, and that fairly subtle measures for detecting node proximity can outperform more direct measures." (Liben-Nowell and Kleinberg, 2007, p. 1019). Or as Kossinets and Watts (2006, p. 88) put it "Network evolution is dominated by a combination of effects arising from network topology itself and the organizational structure in which the network is embedded." This approach has predictive validity in the context of big data where network topology summarizes a great deal about actors' behavior. However, it does not answer the question of how members establish ties nor help understanding behavior at the individual level. Recently, researchers started to consider various contextual factors affecting the evolution of online communities (Butler et al., 2015; Faraj et al., 2015). In this paper, we take three classes of mechanisms of members' engagement into consideration: structural mechanisms, demographical mechanisms, and behavioral mechanisms.

7.2.1 Structural mechanisms

Structural mechanisms explain the formation of new ties in social networks based on existing ties and more generally the structure of the social network. The current structural configuration is a predictor of future configurations (Kaiser and Hilgetag, 2004; Capocci et al., 2006; Backstrom et al., 2006). The most well known growth mechanism is *preferential attachment* (Barabási and Albert, 1999) in which new nodes (representing members in online communities) link to other existing nodes with probabilities proportional to the popularity of those existing nodes. In online communities, preferential attachment predicts that new members prefer to cooperate with other popular members (Barabási et al., 2002). Preferential attachment leads to a structure shown to exist in real-world networks in a wide variety of domains such as Internet routers' networks and academic authorship networks (Barabási and Frangos, 2002). The popularity of a node is measured by counting the number of links it has in the network, also refereed to as node *degree*.

A power-law distribution is a specific case of a long-tail distribution in which very few nodes have a high share of links while the rest of the nodes have limited number of connections. Many real-world networks including online communities are characterized with a long-tail degree distribution, a considerable clustering, and a small average distance between pairs of nodes (Klemm and Eguiluz, 2002). Because such characteristics cannot be explained with preferential attachment alone (Johnson et al., 2014), other structural mechanisms are suggested. Local search is a mechanism in which a new node preferentially meets existing nodes in the network and then links to some of their neighbors (Jackson and Rogers, 2007). Triad formation is a mechanism where when a new member establishes a tie with an existing member, she is likely to establish another tie with an acquaintance of that member founding a triad as a result. Other hybrid approaches include hyperbolic similarity (Papadopoulos et al., 2012) and recursive search (Vázquez, 2003).

7.2.2 Demographical mechanisms

The second breed of mechanisms explains members' engagement based on demographics similarity. "Birds of a feather flock together" suggests that social actors group around same salient characteristics and backgrounds (McPherson et al., 2001). Homophily explains the emergence of clusters of similar nodes, a phenomenon found in many social and physical networks (Newman, 2001a). Such characteristics may include geographic location, language, profession, tenure, gender, and age (Kossinets and Watts, 2006). Depending on the platform and also the tenure of a member, such characteristics may be partially observable by the new member. Homophily suggests that new members will favor other existing members who share similar demographics.

While the majority of structural mechanisms produce networks with structure similar to that of many real-world networks, they still do not generalize well to explain the growth of online social networks (Jin et al., 2001; Jackson and Rogers, 2007). Part of the problem is that social actors have agency and are much more nuanced and dynamic than physical entities (Borgatti and Foster, 2003). Homophilic mechanisms address this issue by incorporating social actors' characteristics into the equation. In real-life settings a combination of structural and homophilic mechanisms come together into play (Schaefer et al., 2010; Johnson et al., 2014).

7.2.3 Behavioral mechanisms

While in demographical mechanisms, new members observe other members' traits, in behavioral mechanisms, new members observe other members' behavior in the community and decide to engage with those they find their actions interesting and compatible with their own. Observing such actions is highly contingent on the technology platform. Platform designers can make various choices that affect how users interact online (Levina and Arriaga, 2015). One choice, for example, is to allow user to partially or fully disclose their identities or remain anonymous (Levina and Vaast, 2005). In online communities built on asynchronous communication such as the one in this study, textual content is the major communication medium and produced artifact. In spite of that, research on online communities has been content agnostic (Hansen, 1999; Sundararajan et al., 2013). A novel mechanism we propose in this paper is based on the fact that behavioral cues are reflected the content of communication in online communities.

Research in communication studies has long identified communication processes as constitutive of property of organizing that can shape the activities of collectives (Taylor and Van Every, 1999). The communication-as-constitutive perspective is instrumental in describing how people get organized and how organizations come to be reenacted and reproduced through analysis of organizational communication (Cooren et al., 2011). For example, the choice of vocabulary is related with social and organizational categorization (Loewenstein, 2014). The different vocabularies used to refer to men and women in media assibilates people to pre-existing gender discrimination (Fairclough, 1995).

Written words and language use are more important in online communities than in other social settings because members in online communities interact almost exclusively through written communication. Recent research has confirmed the important role of language in online communication. For example, Yoo and Alavi (2004) report that leaders of virtual teams write longer and more frequent emails than other executive team members. Danescu-Niculescu-Mizil et al. (2013) show that members in online communities adapt their language and written communication to fit community norms overtime. Huffaker (2010) identifies that online leaders have more linguistic diversity and are more likely to use affective and assertive language. Zhu et al. (2012) find that different types of language are associated with different leadership styles in Wikipedia.

Recent research sheds light on the importance of content in shaping online communities' dynamics (Ferguson et al., 2012). Distinction and status production are affected by the flow of content in online platforms (Levina and Arriaga, 2015). Topic consistency overtime is a factor that increases the likelihood of new members to stay in the community (Butler et al., 2015). In this paper, we argue that the content of written communication in online communities is a key factor determining short-term and long-term tie formation. New members observe other members' communication and establish new ties based on similarity of interests expressed via written text. As shared interests are implied by what members write and post in the community, engagement is also determined by what members write in the community.

7.2.4 Affordances of Technology Platform

Tie formation mechanisms in online communities cannot be examined independently from the IT platform on which the community is built. What the platform affords members to do shape their access to information about others and constrain their behavior online (Majchrzak and Markus, 2015). Structural, demographical and behavioral engagement mechanisms depend on whether the platform makes visible and accessible the community structure, members' demographics and behaviors. This varies hugely depending on the platform (Levina and Arriaga, 2015). Older platforms afforded less visibility, synchronicity and editability. Newer online collaboration platforms incorporate social features resulting in richer affordances for users such as personal identification, visibility, searchability, association and socialization (Treem and Leonardi, 2012).

The inclusion of novel features is driven by the advancement in communication technology and the availability of high interactivity devices such as smartphones and tablets. However, it is assumed that complementing traditional online collaboration platforms with richer features is advantageous and will improve the outcomes. This assumption is challenged by recent research that started to shed light on the implications of the inclusion of social media features in online communities and knowledge collaboration platforms. The choice to include and exclude a set of features can be detrimental for online communities (Hallerbach et al., 2013). Some features, when activated, will have positive effects on knowledge sharing, but others will have unintended adverse consequences (Faraj et al., 2013).

The following comparison serves to illustrate the difference. SourceForge is among the first web platforms that offered a free platform for hosting, developing, and managing open-source projects. It is built mainly on older technologies favoring textual communication among members identified by their usernames (Figure 39). On the other hand, GitHub, a recent popular open-source development platform incorporate social media features allowing its users to have personal profile pages, read customized news feeds, follow other users and send personal messages (Figure 40).

In this study we focus on a community developing an open-source project built using SourceForge. Communication among members is asynchronous and implemented via *GNU mailman* an open-source electronic list manager [10]. Mailman offers a web interface for

[10] `http://www.list.org/`

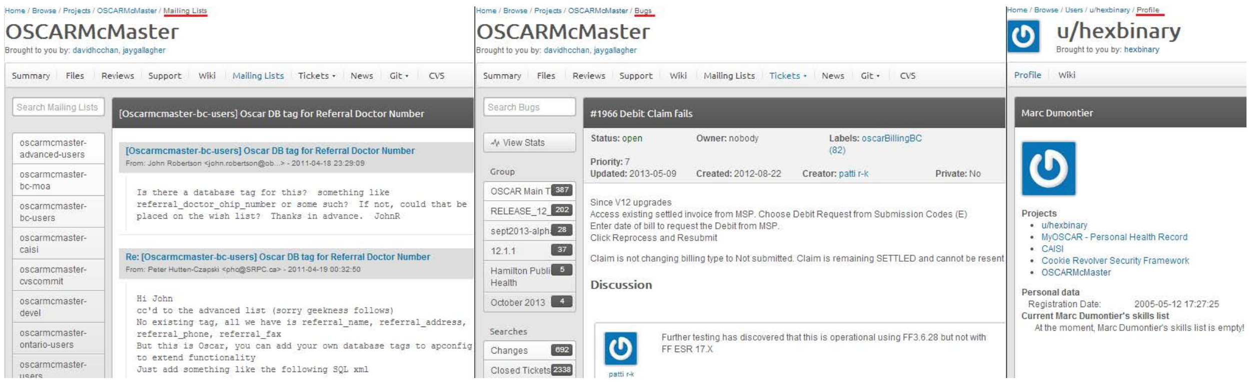


Figure 39: Discussions, bug tracking and user profiles in SourceForge

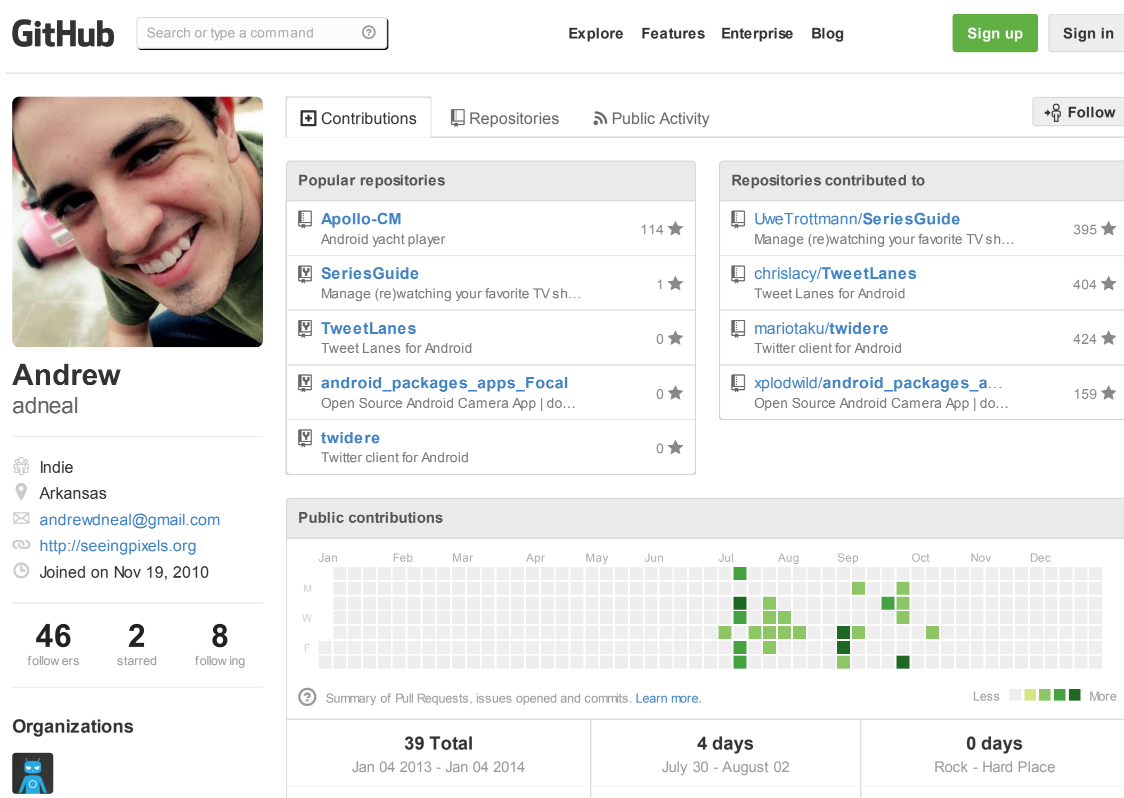


Figure 40: GitHub offers an open-source development platform with social media inspired features

all aspects of mailing list functions including subscription, settings and browsing. At the same time, mailman relays messages via email (Viega et al., 1998). Communication among members in mailman is text-based where members are identified by their email addresses. Furthermore, email addresses are partially masked in the web interface in order to avoid spam bots and email harvesting (Figure 39).

The communication as a constituent of organizing (CCO) frameworks considers that linguistic artifacts are not merely representational but also constituent of social and organizational realities (Alvesson and Kärreman, 2000; Demers et al., 2003). This perspective is even more relevant in knowledge work settings that constitute over 80% of jobs in developed countries (Fairclough, 2004). We extend CCO to examine how communication in online communities shapes the activities of members. Because communication among members in online communities is primarily written, and because members identities are partially disclosed (e.g. avatars and user names), how members get organized in online communities can largely be considered through written exchanged communication in the community.

Research has started recently to examine the role of language in shaping the dynamics of online communities. For example, the content of communication moderates the effect of structural position on contribution (Aral and Van Alstyne, 2011). Communication is closely related to behavior in Wikipedia (Zhu et al., 2012). Leaders in Wikipedia and other online communities use language that is substantially different than language used by other members (Johnson et al., 2015; Huffaker, 2010; Yoo and Alavi, 2004). Building on these findings, we posit that language and communication are of particular importance in shaping engagement and tie formation in open-source online communities.

As a new member joins the community she starts receiving emails from the mailing list. Because all communication is relayed to all members by default, the new member will receive postings from other existing members covering various topics in the community. At first, it is difficult to know who is who in the community. The platform does not afford rich member profiles. The only way to get to know others at first is to read through their communication. The new member will have certain interests and information needs. This will push her to interact with others who seem to share same interests. We propose that:

Proposition 7.1 *Behavioral mechanisms are used for short-term engagement.*

With time, the new member will accumulate more information on other members in the community. This is possible as some members may disclose information about themselves in their communication. For example, members may identify themselves and where they work. It is also possible to learn more about others by looking at their email addresses and signatures. Names also may be revealing of some personal characteristics. In addition, while short-term engagement may be dictated by a need to resolve problems and issues with the software, long-term engagement is less problem-oriented and more social. Long tenured members have an opportunity to learn more about each others and socialize accordingly. We propose that:

Proposition 7.2 *Demographical mechanisms are used for long-term engagement.*

Finally, in open-source communities, it is typically easy to identify the community leaders, administrators and active members (Johnson et al., 2015). They are active participants and information providers in the community. In the short-term, a new member will observe a large number of messages originating from active members. They may even be among the first to answer a request by a newbie. In the long-term, leaders, administrators and active members are known to the rest of the community. We propose that:

Proposition 7.3 *Structural mechanisms are used for both short-term and long-term engagement.*

7.3 Methods

We study the growth of the OSCAR online community over seven years from its inception on SourceForge in January 2006 through January 2013. Community membership is heterogenous and enlists multiple individuals and institutions including the main development team at McMaster University, large institutions such as McGill University, independent clinics and doctors, and independent developers, programmers and service providers. Members communicate predominantly online using the discussion mailing lists in SourceForge[11]. Over the study period there were over 25,000 messages exchanged written by approximately 1,000 unique members in 7,000 threads in seven mailing lists (Table 9).

Mail-list	**Members**	**Threads**	**Messages**
Advanced users	91	680	2,823
BC Medical Office Assistants	99	321	715
BC Users	604	3,497	14,944
CAISI Users	25	40	69
Developers	699	2,293	8,366
Ontario Users	9	8	11
Other Users	55	86	18
Total	-	**6,925**	**26,946**

Table 9: Data collection from OSCAR online mail-lists from January 2006 to January 2013

7.3.1 Modeling Members' Engagement

Affiliation networks are natural representations of threaded discussion communities (Hansen et al., 2010, Ch. 9). In affiliation networks, members participate to common activities or

[11] `http://sourceforge.net/projects/oscarmcmaster/`

affiliations. There are two types of nodes with links connecting two nodes from different types. In threaded discussion networks, the two types of nodes are members and threads. A link between a member and a thread is established when the member posts a message to the thread. Despite being an intuitive representation, the affiliation network is typically transformed to a network because most algorithms assume a homogenous set of nodes. Fortunately, this transformation is straightforward (Newman et al., 2002). We establish a link between two members if they both post to the same thread. Moreover, this link or edge is weighted by the number of times these two members post to same threads.

There are two caveats on the transformation process. First, as in the first study (Section 5.7), we weigh edges using Newman (2001b)'s formula of transforming collaboration networks. Second and most relevant to this study, because all messages are timestamped we can track the evolution of new member engagement overtime. To compute the short-term engagement of a new member in the community, we consider this member's participation during the first three months after joining the community (Figure 41). In other words, we consider the threads that he posted in during the first three months of his tenure. Doing this for every member results in a network representing members' short-term engagement. Because there is no hardline that differentiates short-term and long-term engagement, using the same method, we can compute long-term engagement (six to twelve months).

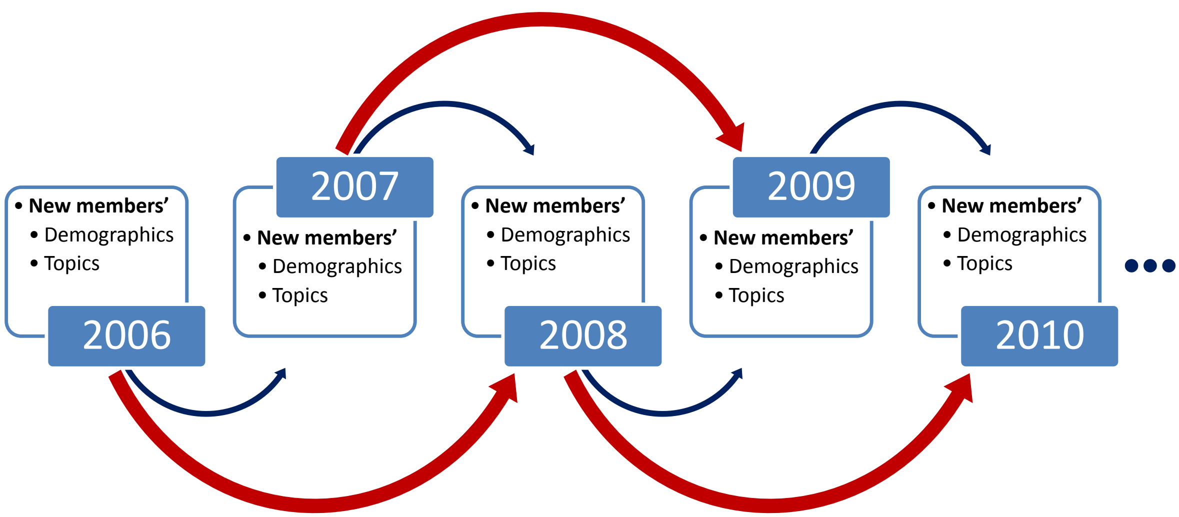
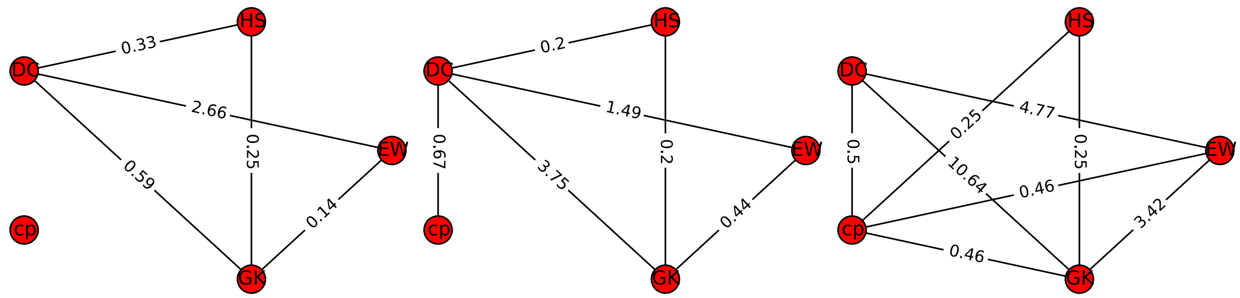


Figure 41: Model of short-term (thin arrow) and long-term engagement (thick arrow)

Figure 42 shows a miniature example focusing on five members in the OSCAR community. We move now to examining mechanisms that could explain the short-term and long-term engagement in the community. Next we perform two experiments. We examine the effect of various mechanisms at the individual and community level. First, we look at what explains engagement among new and existing members among structural, behavioral, and demographical mechanisms. Second, we consider how those mechanisms explain the aggregate structure of the community.



(a) First three months **(b)** Three to six months **(c)** Six to twelve months

Figure 42: Example of short-term and long-term engagement

7.3.2 Determinants of Short-Term and Long-Term Engagement

In order to explain members' engagement in the community at different stages (short-term and long-term), we examine each tie formed by a new member joining the community at each stage during her tenure (Figure 42). The strength of each tie reflects the degree of engagement between the incoming member and the existing member depending on the number of common threads they participate in. Next, we explain the strength of each tie by examining structural, behavioral, and demographical determinants.

Structural determinants depend on the position of the existing member in the community. This stems from a preferential attachment argument where it is more likely for the new member to engage with existing members who occupy prestigious positions in the community. We consider the centrality of the existing members measured with closeness centrality, the brokerage power or spanning of the existing member measured with betweenness centrality (Brandes, 2001), and also the prestige of the existing member measured with PageRank (Brin and Page, 1998). The measures are operationalized with closeness centrality, betweenness centrality and eigenvector centrality respectively (Borgatti and Everett, 2006). Because these three measures are correlated we use closeness centrality as the structural independent variable.

Demographical determinants are personal characteristics of members in the community. In contrast to structural determinants where we only consider the structural characteristics of the existing members, we consider how similar are the personal characteristics of the new member to those of existing members. This is a homophily argument where "birds of a feather flock together." As a community of professional members, the professional affiliation of members in the community is the main personal characteristic that drives demographical engagement (Burt, 2010, p. 350). We examine four groups in the community: doctors, administrators, developers and service providers. Membership in the same group is regarded as a demographic similarity independent variable.

Because members do not self-identify their profession in this community, we had to code profession manually. This is a very involved process that required examining members' messages in the community looking for titles, affiliations and other cues of profession. For this purpose we developed a database module to scroll over all postings of each member (Figure 43). As a second step, for those members whose profession could not be uncovered from their messages, we performed Internet searches on Google and LinkedIn in order to uncover their professions[12]. Results are in Table 10. The 802 members whose professions were not coded include many lurkers who posted one or two messages, used free email accounts, and did not leave much information in their signatures.

profession	**count**
Uncoded	802
Administrators	34
Developers	52
Doctors	67
OSPs (Oscar Service Providers)	43

Table 10: Members' professions

Finally, to evaluate behavioral mechanisms, we consider how similar are the postings of the new members to those of the existing member during the period under consideration. We use various measurement to evaluate textural content, including a basic cosine similarity between the two word vectors representing two posts (Huang, 2008), and a topic similarity measurement based on a topical modeling of postings in the community (Blei and Lafferty, 2007). We elaborate more on the details of these measurements in a dedicated section.

The analysis is at the dyadic level we regress the weight of ties representing engagement on structural, demographical and behavioral determinants. Only ties among members whose professions are coded are considered.

$$weight = \underbrace{centrality + spanning}_{\text{Structural}} + \underbrace{same\_profession}_{\text{Demographical}} + \underbrace{cosine\_sim + topic\_sim}_{\text{Behavioral}}.$$

Table 11 summarizes the research variables.

[12]It helped that as a member of the community I already had many members as LinkedIn contacts, because LinkedIn allows searches up to third degree connections

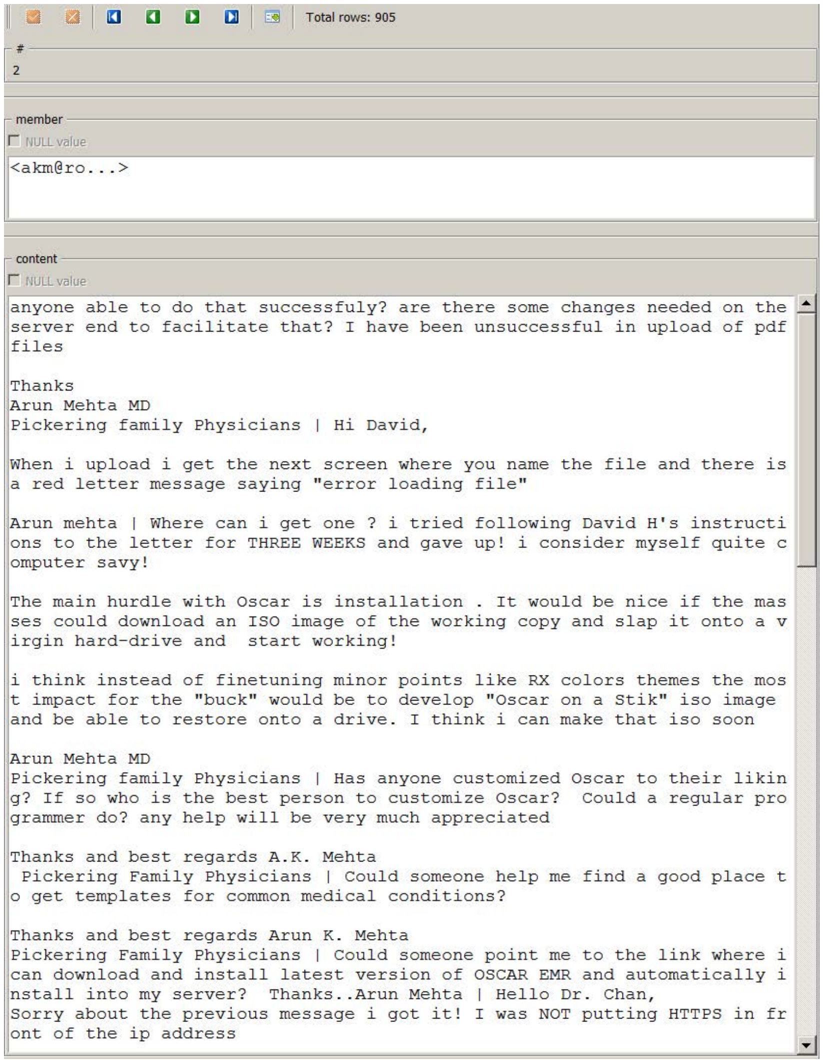


Figure 43: Coding members' professions

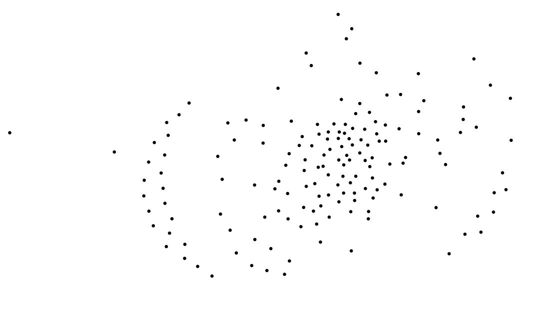
7.3.3 How do Various Mechanisms Shape the Evolution of the Community?

Second, we consider how various engagement mechanisms shape the growth of the community overtime. In order to answer the question, we take a simulation approach and try to recreate the structure of the OSCAR community with information about the messages exchanged in the community (Figure 44). We discard information about messages' embeddedness in threads and any information related to the receivers messages. In other words we only know the member who wrote the message, when it was written (timestamp), and its content (Table 19). Discarding message embeddedness in threads and receivers of messages effectively eliminates the structure of the community.

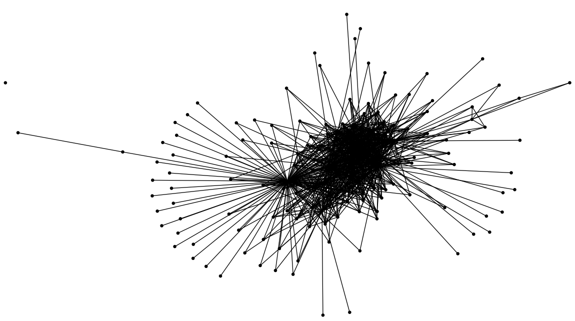
We recreate the structure using variants of the three mechanisms: structural, demographical

	type	strategy	description
weight	dependent	-	The weight of the tie between the new member and the existing member
cosine_sim	independent	behavioral	the cosine similarity between the posting of the new member and the postings of the existing member
topic_sim	independent	behavioral	the topic similarity between the posting of the new member and the postings of the existing member
centrality	independent	structural	the centrality of the existing member in the community
spanning	independent	structural	the spanning of the existing member in the community
same_profession	independent	demographical	whether the new member and the existing member have the same profession

Table 11: Research variables



(a) Nodes without structure



(b) Recreating structure

Figure 44: Recreating structure of the OSCAR network using messages' content and timestamps

and behavioral engagement. Furthermore, we test a hybrid mechanism combining two engagement processes. We compare the resulting networks with the original OSCAR network to see how faithfully each mechanism recreated the real structure. In addition, we compare the networks to a randomly generated network that serves as a baseline. Networks are represented using the NetworkX package in Python (Hagberg et al., 2008). The following mechanisms are considered:

1. Structural mechanisms:

   - Popularity attachment (PA): The new node preferentially attached to another node following Barabási and Albert (1999)'s model.

- Popularity with triad formation (TIE): a new node is attracted to other popular nodes. However, when attaching to another node, it also probabilistically attaches to one of that node's neighbors forming a triad (Holme and Kim, 2002).

2. Demographical mechanisms:

   - Homophily based on personal characteristics (HOM): a new node probabilistically links to an another node according to the similarities in demographics between the two members represented by the two nodes.

3. Behavioral mechanisms:

   - Similarity attachment (SA): a new node probabilistically links to a similar node according to the content similarity measurement.

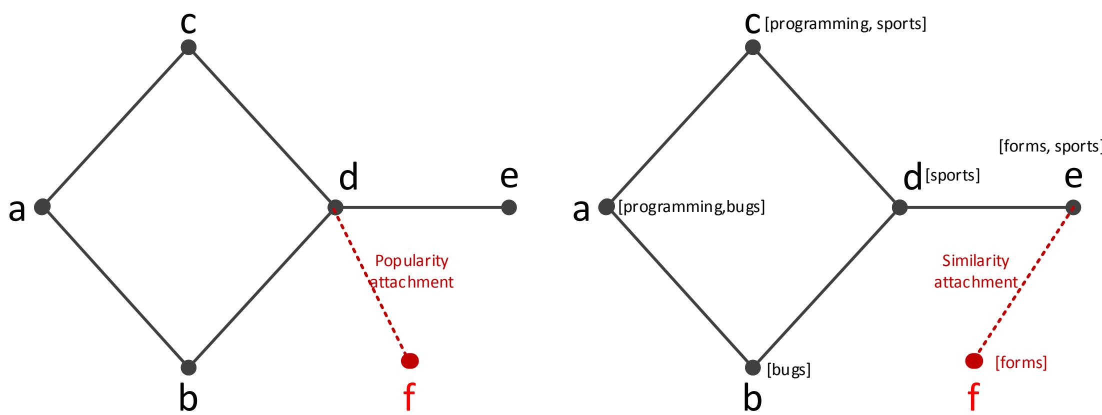
4. Other mechanisms:

   - Random walk (RAN): the new node links with a randomly selected node creating a random graph (Erdos and Rényi, 1960).
   - Similarity×popularity attachment (SPA): a hybrid model combining PA and SA using Papadopoulos et al. (2012)'s algorithm in which an entering node decides what other nodes to connect to based on minimizing the hyperbolic distance (Krioukov et al., 2010). We incorporate nodes popularity and content similarity into the distance function.

We grow the four networks incrementally using the timestamped messages. When a new member, represented as a node, joins the network he/she evaluates the possibility to link to other existing members based on their popularity and/or the similarity of their posted messages to the new member's first message (Figure 45). Table 12 summarizes the simulation mechanisms.

mechanism	**type**	**model**
RAN	baseline	(Erdos and Rényi, 1960)
PA	structural	(Barabási and Albert, 1999)
TIE	structural	(Holme and Kim, 2002)
HOM	demographical	personal attributes similarity
SA	behavioral	content similarity
SPA	hybrid	(Papadopoulos et al., 2012)

Table 12: Comparing the degree and clustering distributions



(a) Popularity attachment **(b)** Similarity attachment

Figure 45: Mechanisms of establishing new ties

Comparing Simulated Networks

One intuitive idea to compare the structure of two networks is to examine whether the same nodes and edges in one network exist in the second one. However, performing such a comparison is computationally infeasible as it requires solving the graph isomorphism problem (Köbler et al., 1994; Arvind et al., 2012). In order to make structural comparison feasible, the nodes and edges characteristics are aggregated into a set of features that are used as a basis for comparison. In this paper, we compare the simulated networks to the OSCAR network using traditional features such as the *degree distribution* and the *clustering distribution* (Albert and Barabási, 2002). In addition, we incorporate the *graphlet distribution*, a novel way to examine the microscopic structure of networks (Pržulj et al., 2004).

The degree distribution is the first property looked at when comparing two networks. In addition, many real life networks exhibit degree distributions very close to the mathematical bounds. The *degree* of a node is simply the number of edges it has. The degree distribution is a probability distribution of degrees over all the nodes in the network. Clustering refers to the tendency of the nodes in the network to group together forming triads, a property of social networks where people bind together forming close knit communities (Holland and Leinhardt, 1971).The *clustering coefficient* of a node is number of actual links among its neighbors over the number of all possible links among them (Watts and Strogatz, 1998). The clustering distribution is the distribution of clustering coefficients over the degrees of nodes in the network.

Recent research examined local patterns in order to describe the *microscopic* structure of networks (Milo et al., 2002; Pržulj et al., 2004). Among those examined patterns are small connected non-isomorphic induced subgraphs of large networks called *graphlets* (Figure 46). The graphlet distribution in a network is represented by the relative frequency of 2, 3, 4, and 5-node graphlets in the network. The graphlet distribution is shown to better describe the structure of biological networks than do degree and clustering distributions (Janjić and Pržulj, 2012). Moreover, the average logarithmic distance between the graphlet distributions of two networks serve to measure their similarly (Milenković and Przulj, 2008). Figure 47 plots the degree distribution, the clustering distribution and the graphlet distribution of the OSCAR network.

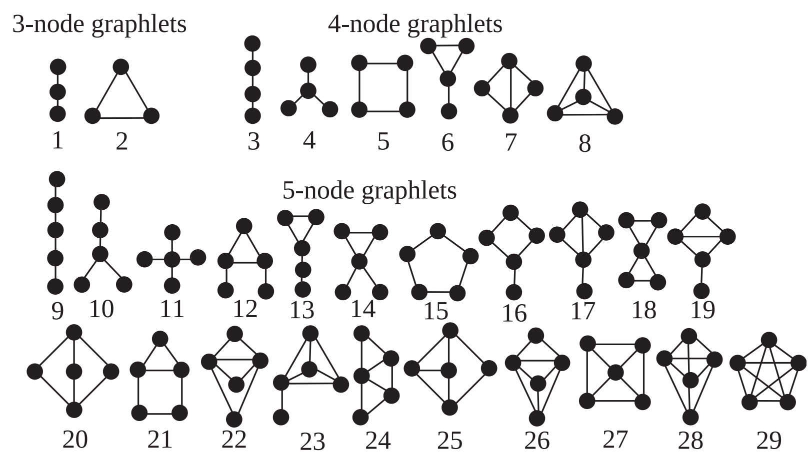
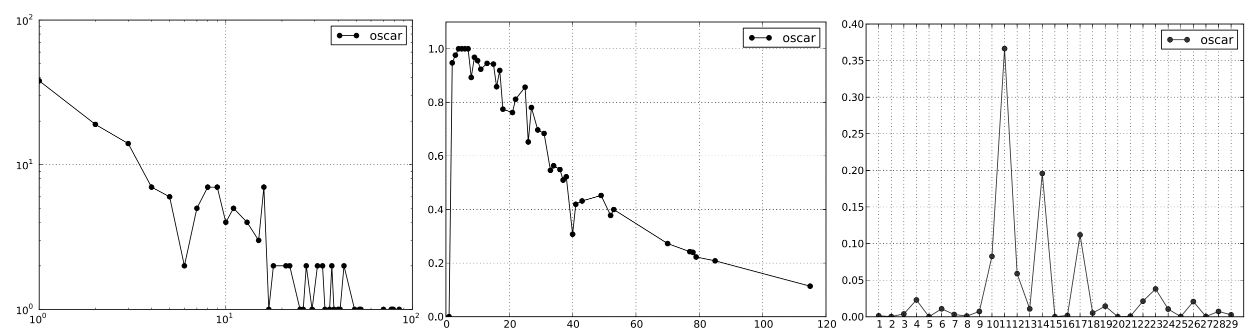


Figure 46: The 29 graphlets of 3, 4, and 5 nodes, taken from Pržulj et al. (2004, p. 3509)



(a) Degree distribution **(b)** Clustering distribution **(c)** Graphlet frequency

Figure 47: Metrics used to compare networks: degree, clustering and graphlet distributions (of the OSCAR network)

Once networks are simulated using the three growth mechanisms and the degree, clustering, and graphlet distributions are calculated for each, we compare each network type to the OSCAR network by comparing the three distributions to those of the OSCAR network respectively. In order to compare two distributions, we first compare their statistical moments, and then perform a two-sample Kolmogorov–Smirnov (K-S) test (Smirnov, 1948). The null hypothesis in the test is assumed to be that the two samples come from the same distribution. The test also measures the distance between the two samples (Young, 1977). The distance function serves to sort the simulated networks depending on their similarity to the OSCAR network.

7.3.4 Measuring Content Similarity

In this paper, we model the communication in the community using an author-topic model (Rosen-Zvi et al., 2010). Author topic modeling considers a collection of documents written collaboratively by a set of authors. In the context of threaded communication in online communities, documents are the discussion threads and authors are community members. Author topic modeling has been used to models the discourse in scientific research including the groups of interest and changing trends (Griffiths and Steyvers, 2004; Rosen-Zvi et al., 2010). The model has been typically used to explore the emergence of topics as well the changing relationships between authors and topics. In this paper, we use the model to quantify and measure diversity and novelty as well as explore the dynamics of discourse in the community.

The model assumes that several authors collaborate to write a document in the same way academics collaborate in writing a paper (represented by a binary matrix A). Under this assumption, we do not know exactly what word each author wrote in the document. Instead, each author is associated with a multinomial distribution over topics (represented with a probabilistic matrix Θ) and each topic is associated with a multinomial distribution over words (represented with a probabilistic matrix Φ). Represented by a binary matrix P, the document is a bag of words that can be seen as a distribution over topics which is in its turn a mixture of the distributions associated with the authors:

$$P = \Phi \times \Theta \times A. \tag{7.1}$$

This model definition is cyclical because its components, authors, documents, words and topics, are all interdependent. We only observe documents' authorship and their content. To estimate the model parameters, Gibbs sampling, a method that estimates the distribution of authors over topics and words over topic is used (Blei et al., 2003). Once the model parameters are estimated, the resulting model can be used for both generative and discriminative purposes.

Figure 48 demonstrates the idea of an author-topic model. A document from a collection of documents is shown. The authors of the document as well as its content are known. From this information, topics which are probability distributions over words and authors are computed (the two columns toward the left).

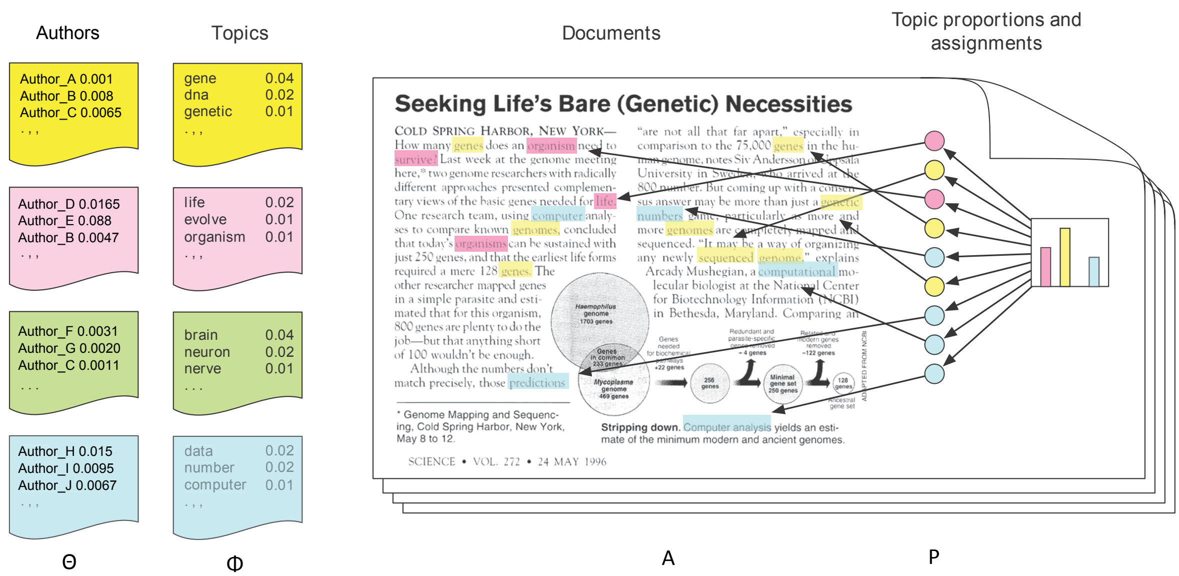


Figure 48: An example of an author topic model (adapted from from Blei (2012))

For the purpose of analyzing communication in online communities, we consider each thread to be a document with members posting messages into the thread as its authors. A more direct representation is considering each message a document with one author. The former representation, however, assumes that all members posting into a thread are sharing the same interest in the topic. For example if a member starts a thread with a question then another member responds with just a link to solve the issue, we still consider that this second member is associated with the words describing the problem posted by the first one. We have used the Matlab Topic Modeling Toolbox (Griffiths and Steyvers, 2004) to train an author-topic model with 50 topics using the OSCAR community (6,925 threads, 980 members, 32,331 unique words). Table 20 shows ten topics with top 7 words and authors associated with each topic. We can see that the first topic is about electronic forms while the second is about organizing meetings. In addition, different members are associated with each topic.

A qualitative examination of word and author probabilities over topics is useful to explore the landscape of the community (Rosen-Zvi et al., 2010). In this paper we use the model to quantify the (semantic) similarity between an incoming member and other existing members in the community. We consider the messages posted by the incoming member as a representation of his/her interests. We examine what topics are mostly related to the words of the message and who has interests in these topics. Formally, we multiply the word vector, the word over topics matrix, and topics over author matrix. Because the result is a probability vector, we use probability to represent similarity or semantic distance between the new member and the existing members:

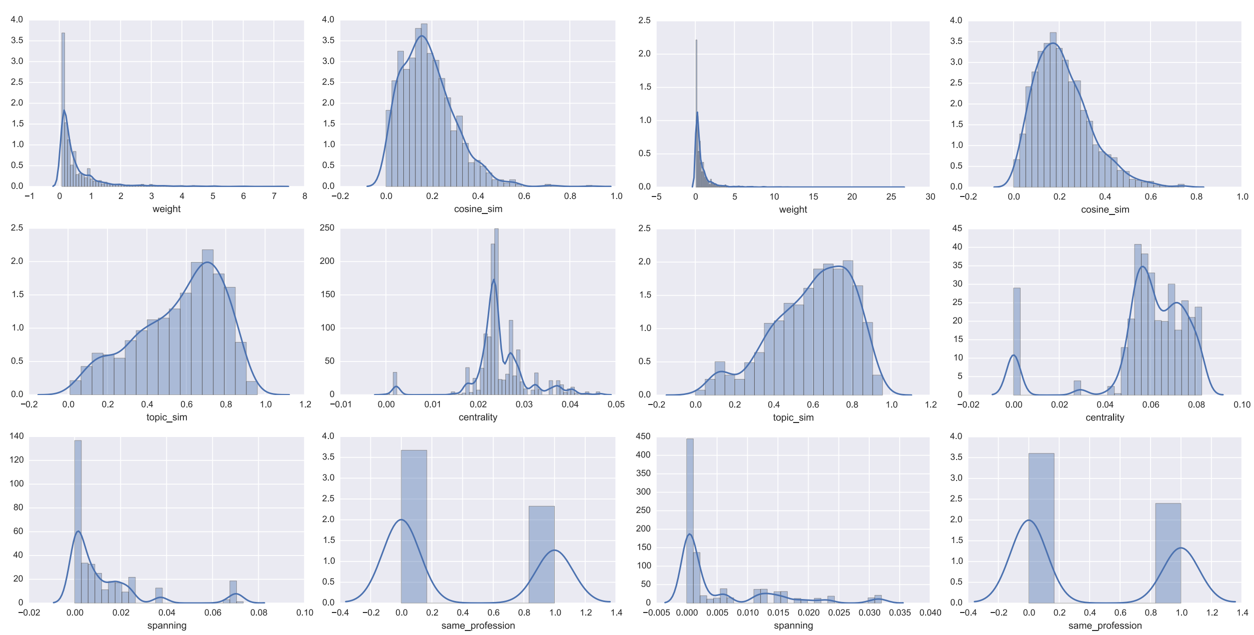
$$\theta_{st} = \pi(1 - \underset{t}{\operatorname{argmax}}\{w_s \times \Phi \times \Theta\}). \tag{7.2}$$

7.4 Results

We present here the results of the two experiments examining the various strategies used my members for short-term and long-term engagement, and also their effect on shaping the structure of the community overtime.

7.4.1 Determinants of Short-Term and Long-Term Engagement in the Community

The descriptive statistics of research variables for short-term and long-term engagement are illustrated in Table 13. The distributions of the variables are plotted in Figure 49. The first analysis we perform is a correlation analysis (Figure 50). During short-term engagement, tie weight is positively correlated with cosine similarity, topic similarity and boundary spanning. During long-term engagement, tie weight is positively correlated with all independent variables. Furthermore, the correlation with cosine similarity, topic similarity and boundary spanning is higher in long-term engagement than correlation in short-term engagement.



(a) Short-term engagement **(b)** Long-term engagement

Figure 49: Variable distributions

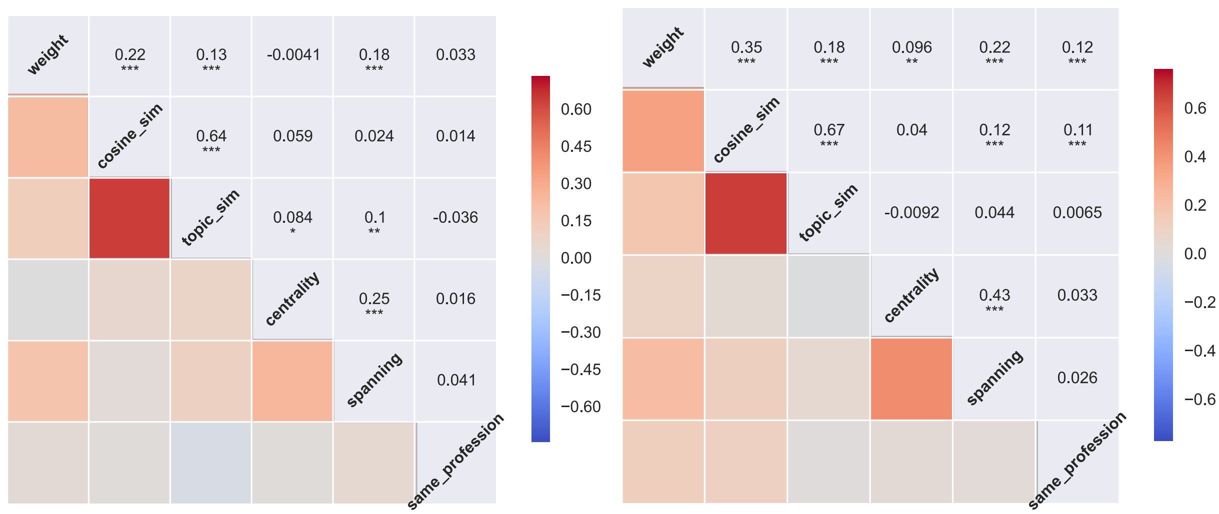
short-term	count	mean	std	min	25%	50%	75%	max
weight	1428	0.582	0.793	0.083	0.143	0.297	0.667	7.186
cosine_sim	1428	0.182	0.115	0.000	0.098	0.169	0.246	0.897
topic_sim	1428	0.566	0.218	0.010	0.413	0.615	0.736	0.961
centrality	1428	0.025	0.006	0.000	0.023	0.024	0.027	0.047
spanning	1428	0.013	0.018	0.000	0.001	0.006	0.018	0.073
same_profession	1428	0.388	0.487	0.000	0.000	0.000	1.000	1.000
long-term	count	mean	std	min	25%	50%	75%	max
weight	1581	0.863	1.441	0.071	0.167	0.417	1.000	26.256
cosine_sim	1581	0.214	0.121	0.000	0.126	0.197	0.285	0.747
topic_sim	1581	0.597	0.203	0.000	0.462	0.630	0.758	0.957
centrality	1581	0.058	0.020	0.000	0.054	0.061	0.071	0.082
spanning	1581	0.005	0.008	0.000	0.000	0.001	0.007	0.032
same_profession	1581	0.400	0.490	0.000	0.000	0.000	1.000	1.000

Table 13: Descriptive statistics of short-term and long-term variables

Next, we perform OLS regression analysis for short-term and long-term engagement (Table 14). Adjusted R^2 is around 40% indicating a good fit of the model. Not all variables are statistically significant. Furthermore, the coefficient of topic similarity in long-term regression is negative despite the positive correlation with the dependent variable. Those inconsistencies stem from the multicollinearity between cosine and topic similarities and centrality and boundary spanning as well. To resolve this issue we perform another regression analysis with a reduced set of variables where we only include one variable for each strategy (structural, behavioral and demographical). Results are reported in Table 15. While adjusted R^2 has dropped to lower thirties, the coefficients of independent variables are consistent with the values of correlation as reported in Figure 50.

The results indicate that new members utilize various strategies to engage with existing member in the short and long-terms. What is important here is the magnitude of difference between the coefficient in short-term and long-term engagement. While the coefficient of the same profession is almost zero in short-term, it jumps to 0.3 in long-term. This indicates that a new member may not engage with those having her same profession in the short-term but rather in the long-term. The coefficient of topic similarity rises from 0.6 to 0.9 indicating an increasing tendency to engage with those interested in similar topics in the community. The coefficient of centrality, on the other hand, halves from 7 to 3.5 which hints to a less dependence of engaging with central members in the community.

We proposed that behavioral mechanisms will be used in short-term engagement, however, the regression analysis demonstrates that behavioral mechanisms are used for both short-



(a) Short-term **(b)** Long-term

Figure 50: Correlation matrix

	short-term	long-term
cosine_sim	1.6970***	4.4668***
	(0.2298)	(0.3707)
topic_sim	0.0452	-0.5940***
	(0.1121)	(0.1794)
centrality	3.9779*	0.0031
	(2.0788)	(1.3376)
spanning	8.1619***	31.4249***
	(1.1648)	(4.4431)
same_profession	0.0682*	0.2210***
	(0.0410)	(0.0673)
R2-adj	0.396	0.384
F	187.880	198.291
BIC	3291.472	5363.668
N	1428	1581
Standard errors in parentheses		
* p<.1, ** p<.05, ***p<.01		

Table 14: OLS regression analysis results

term and long-term engagement. The correlation with long-term engagement is three times more than the correlation with short-term engagement. The second proposition is supported, demographical attachment is significantly more correlated with long-term than short-term engagement. Finally, the third proposition is not supported, structural attachment (centrality) correlates only with short-term engagement. Table 18 summarizes the results and the research propositions.

	short-term	long-term
topic_sim	0.6319***	0.9294***
	(0.0845)	(0.1258)
centrality	7.0802***	3.5304***
	(2.0778)	(1.2896)
same_profession	0.0875**	0.3102***
	(0.0423)	(0.0712)
R2-adj	0.355	0.299
F	262.651	226.064
AIC	3382.798	5565.997
N	1428	1581
Standard errors in parentheses		
* p<.1, ** p<.05, ***p<.01		

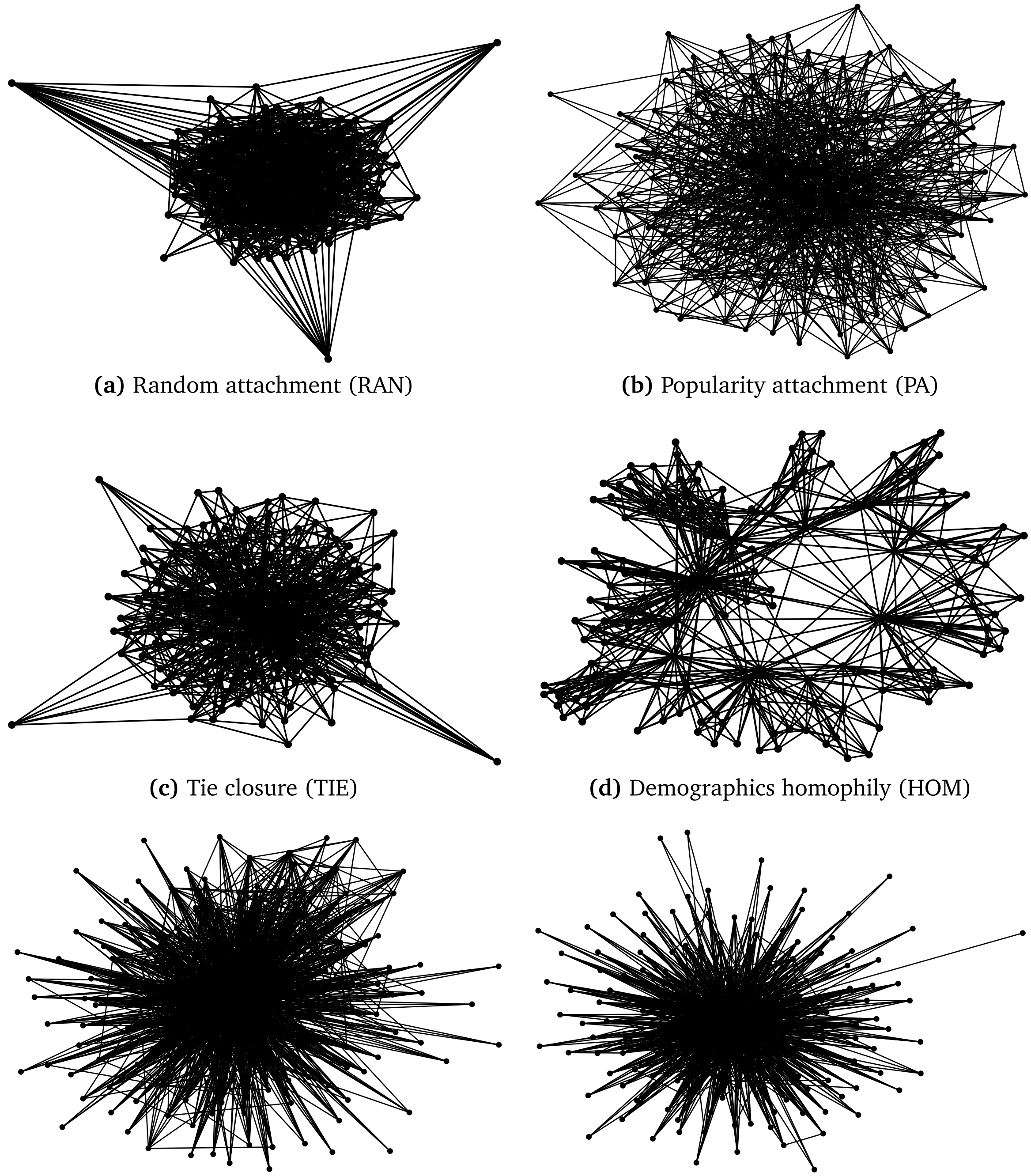
Table 15: OLS regression analysis results for the reduced model

7.4.2 How do Various Mechanisms Shape the Evolution of the Community Overtime?

We compare the four resulting networks (Figure 51) with the OSCAR network (Figure 44). A qualitative examination of the network plots shows SA and SPA networks are most similar to OSCAR. In particular, they model well the core periphery structure that is visible in the OSCAR network: there is a densely connected core with many peripheral nodes attached to its surface (the spikes in the plots). Least similar to OSCAR are PA, RAN and HOM networks. RAN and PA are very regular while HOM is more clustered than OSCAR. Needless to say, a more rigorous numerical comparison is needed. We compare the networks using the degree, clustering, and graphlet distributions.

Figure 52 plots the degree and clustering distributions of the resulting networks. Table 16 summarizes their four statistical moments (mean, variance, skewness, and kurtosis) and reports the two-sample Kolmogorov–Smirnov test (Young, 1977) to compare those distributions against those of the OSCAR network. Looking at the degree distribution, the SPA and SA are closest to OSCAR followed by TIE, HOM, PA and RAN. In articular a K-S coefficient of 0.38 represents the probability that SPA degree distribution and OSCAR degree distribution comes from different probability distributions. Clustering results are slightly different. HOM is the most closer to OSCAR followed by SPA and SA. TIE, PA and RAN are the least similar.

Figure 53 plots the graphlet distributions of the resulting networks. To quantify graphlet distribution similarity, let $N_i(G)$ denote the number of graphlets of type i in network G ($i \in \{1, ..., 29\}$). Let $T(G)$ denote the total number of graphlets in a network G: $T(G) =$



(a) Random attachment (RAN)

(b) Popularity attachment (PA)

(c) Tie closure (TIE)

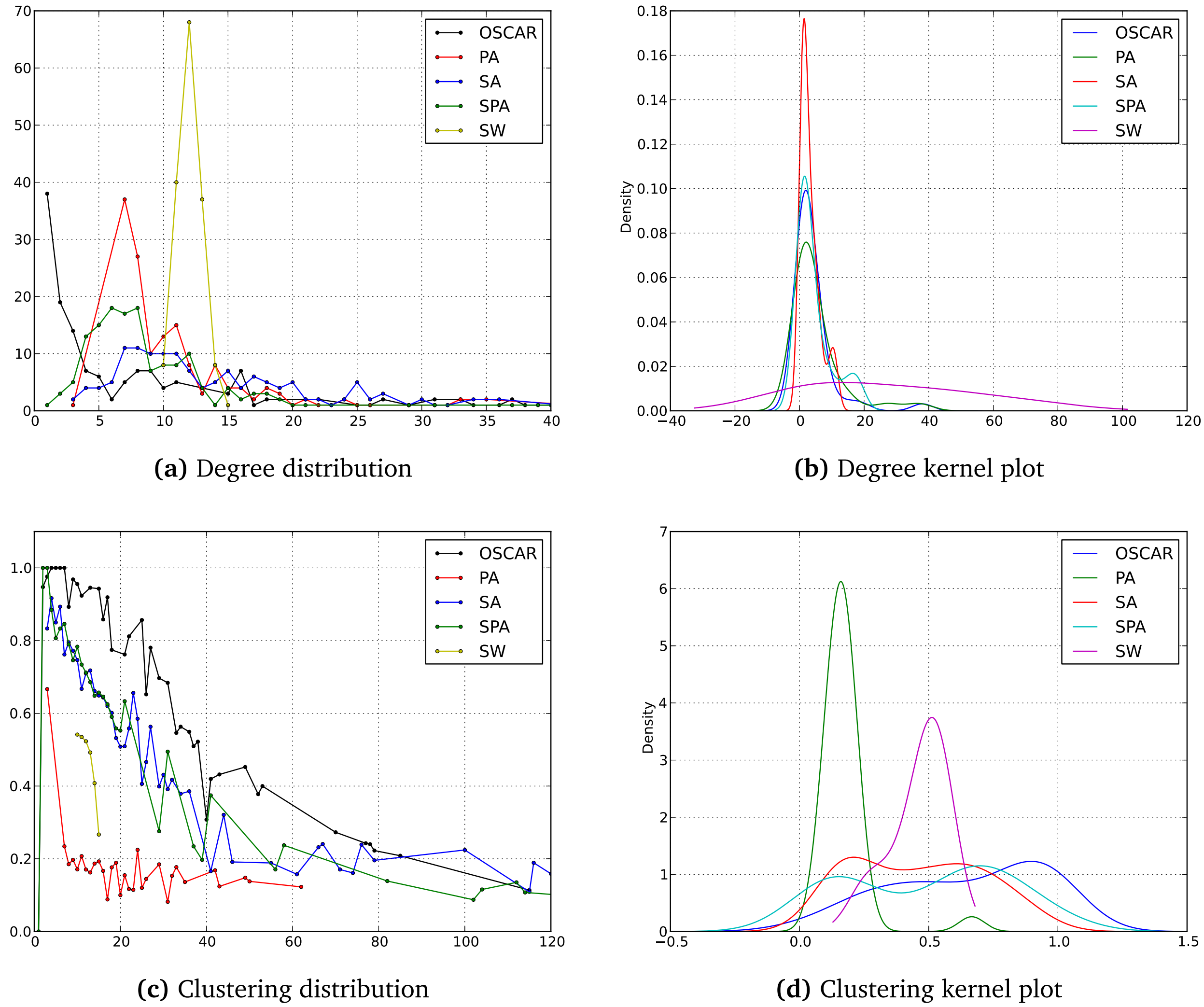
(d) Demographics homophily (HOM)

(e) Similarity attachment (SA)

(f) Similarity & Popularity attachment (SPA)

Figure 51: The OSCAR network and the other simulated networks

$\sum_{i=1}^{29} N_i(G)$. Then, we compute the relative frequency of each graphlet type $F_i(G) = N_i(G)/T(G)$. Finally, to measure the distance between the local features of two networks G and H, we use an l_1 distance function between the frequency of graphlets in the two



(a) Degree distribution **(b)** Degree kernel plot

(c) Clustering distribution **(d)** Clustering kernel plot

Figure 52: Degree and clustering distributions

networks:

$$D(G,H) = \sum_{i=1}^{29} |F_i(G) - F_i(H)|. \tag{7.3}$$

Table 17 lists the frequency of graphlets with similarity measures. The SA network is the most similar to the OSCAR network followed by SPA, HOM and TIE.

7.5 Discussion

Results of the first experiment indicate that new members in the OSCAR community utilize various strategies to engage with existing members in the community. In the short-term, new members rely on structural strategies and behavioral strategies as indicated in the

	Degree			Clustering			Graphlet
	Average	Variance	K-S	Average	Variance	K-S	distance
OSCAR	12.68	345.08	-	0.64	0.17	-	-
Random walk	13.46	10.81	0.54	0.07	0.00	0.75	1.58
Preferential attachment	13.40	95.81	0.52	0.19	0.01	0.75	1.15
Triad formation	11.43	107.86	0.50	0.31	0.01	0.62	0.74
Homophily	11.87	84.63	0.51	0.76	0.05	**0.28**	0.69
Similarity attachment	22.58	717.18	**0.45**	0.61	0.05	0.44	**0.36**
SPA	15.38	651.17	**0.38**	0.73	0.05	**0.34**	**0.68**

Table 16: Comparing the degree and clustering distributions

correlation matrix (Figure 50) and regression results (Table 15). This implies that new members engage with centrally positioned members in the community and to members who exchange messages with similar content to their messages. However, in the long-term, those new members gravitate toward using demographical strategies. In addition, in the short-term all strategies are used at the same time. With time, members get more information about others in the community which allows them to engage on more personal (based on demographics and behaviors) and less structural (based on centrality) grounds.

The results of the second study demonstrate that content similarity attachment approximates better the degree distribution and graphlet distribution than other mechanisms and is second behind homophily in explaining clustering in the resulting networks. In addition, a combined similarity and popularity attachment performs better than individual mechanisms alone. This confirms what recent research reports about the need to combine mechanisms in order to explain network evolution (Johnson et al., 2014). However, the main finding is that the content of communication exchanged among members is the main driver of new ties. Members of online communities establish mutual relationships to promote common interests and discuss topics and issues of same concerns. Moreover, a network representation of online communities and other social collaboration settings often strips the phenomenon from its richness and reduces it to links between entities.

The OSCAR project is hosted on SourceForge, a mainstream open-source development platform that offers a wide spectrum of collaborative development capabilities including a code repository, bug tracking, project wiki, and mailing lists (Figure 39). In contrast to more novel open-source development platforms such as GitHub and Google Code that incorporate social media capabilities affording for richer and more personalized communication among members, SourceForge relies on older technologies that are based upon textual communication. For example, forums in SourceForge are implemented with email lists. As a result, members are known with a masked version of their email addresses. There is no way to

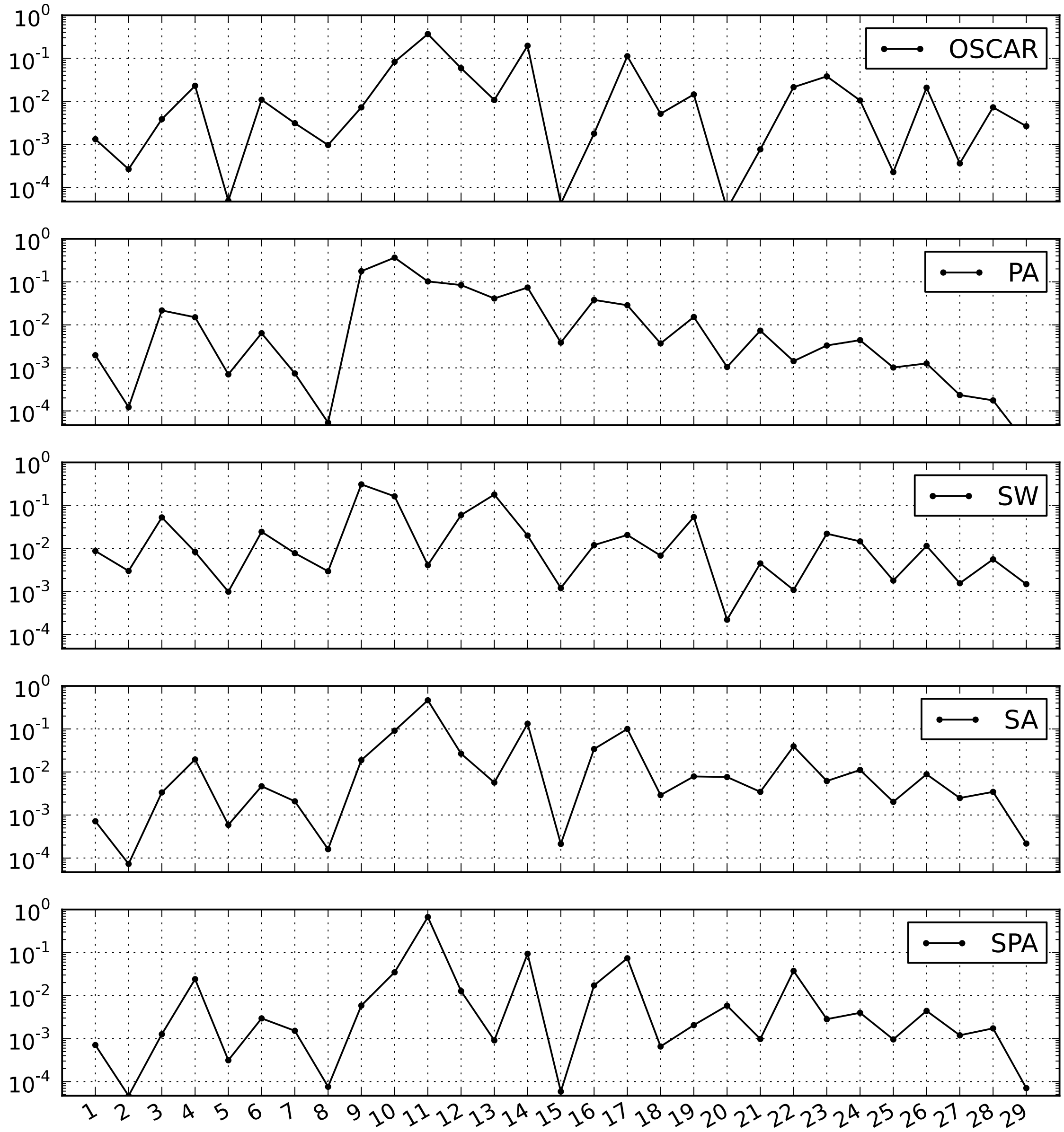


Figure 53: Graphlet distributions

obtain contextual information about a member within the forums and no way to send a personal message. Those affordances of SourceForge are perhaps what influences behavioral engagement strategies relying on textual similarities.

7.6 Conclusion

Online collaboration including in online communities is a novel form of organizing enabled by information and communication technology. In this study, we answer the question of how new members in online communities establish their ties with existing members in the

	OSCAR	RAN	PA	TIE	HOM	SA	SPA
1	0.001318	0.003426	0.001968	0.001782	0.002135	0.000715	0.000706
2	0.000265	0.000091	0.000122	0.000126	0.000489	0.000073	0.000047
3	0.003853	0.038282	0.021631	0.014586	0.014154	0.003344	0.001263
4	0.022962	0.012740	0.015023	0.020448	0.019214	0.019560	0.024043
5	0.000049	0.000873	0.000706	0.000308	0.000021	0.000589	0.000313
6	0.010868	0.003058	0.006400	0.006988	0.012804	0.004659	0.002946
7	0.003097	0.000119	0.000742	0.000772	0.002105	0.002082	0.001515
8	0.000963	0.000001	0.000053	0.000050	0.000625	0.000160	0.000076
9	0.007230	0.390751	0.177000	0.087562	0.044226	0.018861	0.005850
10	0.082520	0.388802	0.365249	0.309792	0.286196	0.091246	0.034698
11	0.366340	0.032184	0.102381	0.233750	0.173804	0.463909	0.668664
12	0.058881	0.031359	0.084154	0.070803	0.066453	0.026631	0.012694
13	0.010740	0.031158	0.041094	0.032003	0.073415	0.005684	0.000908
14	0.195840	0.015520	0.073727	0.129518	0.172182	0.132466	0.093493
15	0.000042	0.007085	0.003875	0.001267	0.000016	0.000213	0.000059
16	0.001779	0.035333	0.038026	0.018288	0.001011	0.034188	0.017193
17	0.111829	0.002411	0.028630	0.038258	0.049095	0.100001	0.073525
18	0.005134	0.000626	0.003705	0.005747	0.016975	0.002902	0.000654
19	0.014455	0.002413	0.015252	0.011542	0.024873	0.007867	0.002054
20	0.000031	0.000545	0.001049	0.000345	0.000000	0.007629	0.005810
21	0.000764	0.002845	0.007339	0.003614	0.000160	0.003455	0.000978
22	0.021299	0.000038	0.001434	0.002411	0.004834	0.039450	0.037379
23	0.038020	0.000042	0.003318	0.003276	0.017892	0.006164	0.002824
24	0.010488	0.000188	0.004411	0.004720	0.008475	0.011131	0.003964
25	0.000226	0.000099	0.001021	0.000435	0.000022	0.002020	0.000956
26	0.020759	0.000005	0.001269	0.001297	0.006747	0.008848	0.004402
27	0.000360	0.000005	0.000234	0.000161	0.000068	0.002482	0.001191
28	0.007245	0.000000	0.000176	0.000140	0.001561	0.003447	0.001725
29	0.002642	0.000000	0.000017	0.000008	0.000442	0.000217	0.000070
l_1 dist	0.000000	1.581548	1.152946	0.747934	0.692921	0.368092	0.685468

Table 17: Comparing the graphlet distributions

short-term and maintain these ties in the long-term. Several mechanisms are considered: structural, demographical and behavioral. A new member using a structural mechanism has tendency with popular members occupying prestigious positions in the community. Preferential attachment to popular members is a structural mechanism that yields a network where few members have the majority of links. In demographical mechanisms, new members will gravitate toward interacting with existing members who share same characteristics. Homophily, one demographical mechanism, reflects the common wisdom that "birds of a feather flock together." New members establish social ties with other members whom they find similar creating as a result clusters within the community.

In contrast to demographical mechanisms that are based on members' personal attributes,

Proposition	Supported	Comments
P1: Behavioral mechanisms are used for short-term engagement	Partial support	Behavioral mechanisms are used for both short-term and long-term engagement. The correlation with long-term engagement is three times more than the correlation with short-term engagement
P2: Demographical mechanisms are used for long-term engagement	Yes	demographical attachment is significantly more correlated with long-term than short-term engagement
P3: Structural mechanisms are used for both short-term and long-term engagement	No	structural attachment correlates only with short-term engagement

Table 18: Research Propositions

behavioral mechanisms imply that new members watch other members' behavior in the community and engage with those whose behaviors are compatible. Members' actions and behaviors are expressed almost exclusively in the virtual space via written communication. Despite this fact, research has often overlooked the role of written communication in shaping the dynamics in online communities including how new members establish ties to existing members. As communication in online communities is conveyed via written text, a new member will examine what other existing members have written before communicating with them. New members establish ties with other existing members based on similar interests and ideas expressed in the content of written communication.

Our results indicate that new members utilize various strategies simultaneously to engage with others in both the short and long-terms. However, new members gravitate toward structural and behavioral mechanisms in the short run. In the long-term, those new members reduce their dependency on structural mechanisms and start engaging with others based on their demographics. On an aggregate level, engagement based on behavioral mechanisms better explains the degree distribution and graphlet distribution induced by new ties and is second only to homophily in explaining the clustering distribution. Furthermore, combining several mechanisms yields better fit indicating the simultaneous existence of various processes simultaneously.

The contribution of this research is both theoretical and methodological. First, it expands our understanding of how new members engage with others in online communities in different time frames during their tenure. This question has been overlooked in previous research in favor for examining general network exchange patterns. Online communities are fluid and dynamic (Faraj et al., 2011) with an open membership, and an emergent leadership structure

(O'Mahony and Ferraro, 2007). Members evolve their interests and develop new ones as a result of conversation and dialogic interaction with other members in the community. Such contextual dynamics are not taken into consideration by tie formation mechanisms that assume previous configuration to be a predictor of future ties. Second, a social network representation of online collaboration models all communication between two members as a relationship without taking into account the content of communication. For example a message from a member to another is modeled as a relationship among the two regardless whether this message conveyed collaboration or tension. The important of text and content exchanged via online relationships is underemphasized as a driver of relationship formation. The proposed mechanism in this paper fills this gap and puts content back into the equation. This goes in tandem with recent studies in organization change that stresses the importance of text, conversation and discursive performances in inducing organizational categorization and change (Loewenstein, 2014; Demers, 2007).

The importance of written communication in online communities has implications for designers of online technology platforms. Newer online collaboration platforms incorporate social features resulting in richer affordances for users such as personal identification, visibility, searchability, association and socialization (Treem and Leonardi, 2012). For example, GitHub, a recent popular open-source development platform incorporate social media features allowing its users to have personal profile pages, read customized news feeds, follow other users and send personal messages (Figure 40). Future work includes testing tie formation mechanisms in a richer and a comparative setting. For example, it is interesting to compare the dynamics of tie formation and network evolution between communities hosted in different platforms (e.g. SourceForge & GitHub). This will give insights into the role technology affordances play in shaping the growth and production of open-source online communities.

Methodologically, the paper highlights the importance of considering various determinants when evaluating the structure of a social network. Recent research still focuses on degree distribution as a determinant of network similarity (Johnson et al., 2014). The incorporation of degree, clustering and graphlet distributions provide a combined macroscopic and microscopic perspective in analyzing the structure of networks. Finally, the paper contributes methodologically by leveraging advances in computational linguistics to automatically analyze the content of the messages. Recent advances in computational linguistics and natural language processing has allowed the automatic processing and treatment of speech and text, two data sources that have been exclusively analyzed with qualitative techniques. This is great news for online communities research because of the large size of communication text exchanged online among members.

This research is not without limitations. First the study is confined the context of the OSCAR community which is very professional and not representative of other open-source communities. Notwithstanding this limitation, similar communities with a unified interest, such as open-source communities would display this parallel structure, but divergent communities that run on comments, such as communities on social media platforms, may not see the same effect because the content of their comments is not sufficiently descriptive enough of mutual interests.

timestamp & member	message
2009-12-19 3:10 Bharat K	...i am solo but i belong to FGH , the FGH GROUP has decided to go with health screen, but i would like to use oscar. so will ontario md consider our case? ...
2009-12-19 3:24 David D	...Perhaps. I believe they are looking at all the restrictions, but first starting with the case of the solos and duos ...
2009-12-19 4:43 Santiago D	...yo estoy usando OSCAR, es buenisimo. Instale varios todo ambulatorio, pero creo que tranquilamente sirve para internación. Estoy en Mar del Plata Cualquier cosa escribime ...
2009-12-19 5:41 John F	...Que sorte das características você está procurando? Pesaroso, meu português é muito pobre. Eu terei que escrever em inglês ...
2009-12-19 7:46 Tracy M	...Local electronic patient record systems are often more effective than larger scale projects, according to a new study by the academic leading the independent evaluation of the Summary Care Record ...
2009-12-19 13:27 David C	...The CAISI component of OSCAR has shelter bed management and case management but I am not sure if it is enough to run a hospital. There is also an Infirmary in Toronto using OSCAR. There is a hospital here in Hamilton thinking about moving all their out-patient clinics to OSCAR ...
2009-12-19 16:09 Peter HC	...To ensure that you are capable of restoring OSCAR you will need the following 1) Your Mysql database - this is the most valuable and changing part of your data, and is your "chart" for practical purposes 2) Your Oscar Documents which contains scanned documents, eForm templates, downloaded laboratory files, downloaded files from the ministry ...
2009-12-20 2:46 Jel C	...ubuntu - firefox I have never really had the eform check box issue that some seem to have had with firefox on some different operating system. but, I have just updated my Ubuntu to the Karmic Koala (9.10) which has firefox 3.5.5 and something great has happened ...
2009-12-20 3:56 John Y	...The problem seems to be primarily Windoze-FF3 related. Mac OS-X and FF3 get along just fine. I believe most using WinXP have downgraded to FF2 as a work-around. Not sure if Vistas or Win7 with FF3 is any better. Has anyone tried? ...
2009-12-20 4:09 Bharat K	...what are the instructions to install oscar on Ubuntu 9.10. Are they different from David Daley 's install guide? Is it possible to get an explanation of the various enteries in the Oscar properties file? is their any documentation some where about the oscar properties file? ...
2009-12-20 4:47 Jel C	...Don't do that! 9.10 is one of the ubuntu 6 monthly releases and to move up from there one has to upgrade it every 6 months. I was talking about what I use on my desktop/laptop. Even the desktops in our office, I use the LTS for, so they are still on 8.04 ...
2009-12-20 5:32 Peter HC	...To start I should reiterate that the default OS to install OSCAR on is Ubuntu Hardy Heron 8.04 LTS 32 bit server. Don't deviate from this unless you have an idea of what you are doing and why you are deviating. Unlike some other operating system the Ubuntu 32 bit server can handle 64 bit processors and memory above 4 G ...

Table 19: Sample timestamped messages used to recreate structure

Word	W. Prob.	Author	A. Prob.	Word	W. Prob.	Author	A. Prob.
	Topic Prob. = 0.03396				**Topic Prob. = 0.02888**		
form	0.05169	John Y	0.41963	user	0.02642	Gunther K	0.10203
eform	0.04223	John R	0.16744	information	0.01827	Colleen K	0.09462
johny	0.03447	David P	0.09782	people	0.01820	David C	0.08858
html	0.02492	Peter HC	0.03619	source	0.01329	Jel C	0.06896
eforms	0.01870	Gunther K	0.03072	attend	0.01280	Earl W	0.04738
print	0.01668	Jel C	0.02352	meeting	0.01224	John F	0.04495
johnyap	0.01643	Darius OS	0.01885	systems	0.01183	David P	0.03208
	Topic Prob. = 0.01884				**Topic Prob. = 0.02075**		
patients	0.02577	Jel C	0.20468	provider	0.04535	Gabriel J	0.08715
hospital	0.01930	David P	0.13422	search	0.03039	Peter HC	0.08553
care	0.01867	Tracy M	0.12141	report	0.02730	Kris VS	0.06429
heart	0.01774	Peter E	0.11880	admin	0.02688	Patti RK	0.04659
disease	0.01514	Martin D	0.11328	original	0.02657	Earl W	0.04608
treatment	0.01277	David C	0.03414	demographic	0.02594	David P	0.03950
free	0.01237	John Y	0.03241	enter	0.02411	Ian P	0.03343
	Topic Prob. = 0.01438				**Topic Prob. = 0.01663**		
search	0.04406	Peter HC	0.17050	string	0.05106	Ted L	0.30871
phone	0.02421	Gunther K	0.07576	lang	0.04934	Marc M	0.08183
writing	0.02015	David C	0.07094	javac	0.03283	Ronnie C	0.05647
sprint	0.01602	Joe R	0.04306	code	0.02934	Dan B	0.05476
code	0.01571	Roy M	0.04189	build	0.02737	Jay G	0.05432
service	0.01548	Earl W	0.03416	source	0.02395	David D	0.04776
html	0.01333	Jel C	0.03065	classes	0.02165	Peter HC	0.03104
	Topic Prob. = 0.02475				**Topic Prob. = 0.01072**		
care	0.02159	Jel C	0.15384	caisi	0.07549	Annie Z	0.27048
source	0.01539	David C	0.07764	insert	0.03670	Tony W	0.10984
canada	0.01522	Gunther K	0.07620	values	0.03367	Jay G	0.08566
folks	0.01504	John F	0.05402	table	0.02478	Randy J	0.05776
electronic	0.01326	Peter HC	0.03443	description	0.02269	Joel L	0.04307
project	0.01182	Tracy M	0.03337	client	0.02164	Patrice D	0.03524
hospital	0.01055	Joe R	0.02896	column	0.02123	David D	0.02810
	Topic Prob. = 0.02555				**Topic Prob. = 0.01192**		
left	0.06647	Shelter L	0.29145	date	0.05191	Mark R	0.10403
style	0.06149	Herb C	0.26122	text	0.05023	Patti RK	0.07886
input	0.04427	John R	0.11316	charset	0.04705	Clare H	0.07402
type	0.04393	TCFP O	0.05352	plain	0.04668	MaryMOAS	0.04559
position	0.04283	Gunther K	0.05282	messageid	0.04603	Santiago D	0.04471
absolute	0.04119	Henry L	0.03623	contenttype	0.04528	Henry L	0.03186
fontsize	0.04068	Carole W	0.02473	attachment	0.02558	my fly	0.02456

Table 20: Ten topics with associated word and author probabilities

8 Synthesis and Contributions

Online communities of innovation are a novel form of organizing enabled by information and communication technology. These platforms involve a large number of distant members toward achieving a common goal. Collaboration among members happens almost exclusively in the virtual space via written communication. The goal of this research centers on a deeper understanding of how communities of innovation are successful at involving users, integrating various inputs, and organizing interactions toward creating common goods. Contrary to previous studies that prioritize programmer led open-source efforts, this work focuses on actual end-users as they use, modify, and participate in the process of evolving the software. In particular we focus on an open-innovation healthcare community that is developing OSCAR an open-source electronic medical record.

The OSCAR community is primarily Canadian, has been in existence for a decade, and has developed an EMR that is rapidly diffusing (currently used by over 1,500 Canadian doctors to follow over a million patients). OSCAR EMR is freely available and is gaining market share against commercial products typically costing $25,000 per year per user. Given the complexity of such software, the mission-critical nature of patient records, and the fact that the vast majority of users are not computer savvy, this success is unusual and significant. Contrary to other open-source projects, OSCAR is embedded in a community of doctors rather than a community of programmers. Such a project deviates from the stereotypical open-source project where geek developers are mostly in charge of developing and building the software.

This thesis focused on the dynamics of interactions that enable a loosely managed community to produce a digital artifact that is competing with commercial solutions. In addition to reviewing literature on online communities, open-source and healthcare IT, we presented the case study of OSCAR focusing on its history and also its differentiating factors as an open-source health IT. We then pursued three empirical studies to understand the processes of knowledge creation and integration in the OSCAR community.

In the first study, we asked the question of knowledge creation from a pure structural perspective at the individual level. What are the determinants of knowledge creation in online communities of innovation? Previous research has established that network position is a determinant of social capital. However we proposed that social collaboration in online communities exhibit a unique locality that is not adequately described by existing structural perspectives that rely on global position. Results indicate that collaboration among members of the community is very local and decentralized allowing members to produce diverse contributions. The contribution of a member is dictated by her centrality and boundary spanning in her locale rather than in the community on a whole.

In the second study, we examined the process of knowledge integration in the OSCAR community. How does the community put together individual members' contributions into the final digital artifact? Over a decade the OSCAR project remained unified producing a major software release every two years. We examined the role that core and peripheral members play in controlling the process of integrating new features into OSCAR. Although contribution at the individual level is very local and decentralized, an emerging committee of developers and users coordinates code integration and makes sure the software remains unified and consistent over time. In this process, a core position within the community increases the chance of a member requesting a new feature to have her request fulfilled. Interestingly, neither the characteristics of the request nor its age affect its chance of being accepted. When it comes to integration, the OSCAR committee is highly centralized.

In the third study we focused on members' short-term and long-term engagement in the community and the mechanisms by which new members establish and sustain relationships with others. How does the community grow overtime? We showed the vital role of shared interests in driving the evolution of relationships and growth of online communities. In addition to structural and behavioral mechanisms such as preferential attachment and homophily, in the context of collaboration in online communities where ideas, debates and conversations are conveyed in written words, new relationship between two members can be predicted by examining similarity of ideas, themes and topics expressed in the written communication of the two members.

This thesis fills theoretical and methodological gaps in the literature of healthcare IT and open innovation communities in healthcare. Healthcare IT is a highly multidisciplinary field. Each stream of research brings a certain focus and contributions to understanding the role of technology in healthcare (Safadi and Faraj, 2011). Despite this diversity of research sources, there is still a lack of research on community driven development of healthcare IT. Research on open-source software has been driven by classical examples such as GNU/Linux and

Apache which research in health IT focused on proprietary health IT systems. This thesis fills this gap by studying community-driven development of open-source health IT. Previous literature has acknowledged that online communities are trusty sources for generating reliable knowledge but the question on the internal dynamics that lead to knowledge creation remains a black box. Opening this black box not only allows for answering the research question on how knowledge is created but also allows to compare knowledge creation process between traditional organizations and new forms of collaboration.

8.1 Theoretical Contributions

Contrary to stable organizations, online communities are characterized as fluid because they morph and change their boundaries, norms, interactions, and foci yet retain their basic shape and characteristics (Faraj et al., 2011). New members join and bring fresh ideas that established actors cannot conceive (Jeppesen and Lakhani, 2010). Online communities evolve practices that allow for their sustainability, governance and growth (Kudaravalli and Faraj, 2008; Ridings and Wasko, 2010; O'Mahony and Ferraro, 2007). Homophily, which is the tendency of connected people to share similar interests has been considered the main organizational principle in social networks and online communities (McPherson et al., 2001). Later diversity rather than similarity has been identified as a key feature in organizational dynamics (Ferguson et al., 2012). This diversity in online communities discourse leads to divergence of goals, processes and solutions (Daniel et al., 2013). The alteration between convergence and divergence allow communities of innovation to transform contestation into collaboration (O'Mahony and Bechky, 2008; Butler and Wang, 2012).

The continuous evolution of online communities calls for focusing on the dynamics of their growth. One intriguing characteristic of innovation communities is their ability to operate despite the heterogeneity and diversity of their members. On one hand, common ground is often needed to facilitate the sharing of expertise and resources (Borys and Jemison, 1989). On the other hand, diversity and heterogeneity of goals and interests may be generative for online knowledge sharing (Faraj et al., 2011; Page, 2001). New knowledge is created via intense dialogic interactions in the community (Tsoukas, 2009; Zhang et al., 2013). In addition, Knowledge integration in online communities is achieved via a different process than that of traditional organizations. Convergence is necessary for achieving closure when solving a problem or developing a new innovation. However, achieving convergence in online communities is difficult for multiple reasons. First, membership is open and voluntary. Members are free to participate in whatever they find interesting or leave. Second, leadership in online communities is emergent and assumes different roles and responsibilities than traditional leadership. As a result, divergence is always increasing in online communities.

The findings of the first study add more nuances to the core-periphery debate in online communities. In the OSCAR community, both global centrality and local centrality lead to higher levels of contribution. A central position in the core of the community is important, however, peripheral members who are locally central within their egocentric networks bring contribution to the community as well. Members in the OSCAR community have a choice in investing their time within their local sub communities or the global communities. Local communication may be more social as it happens among peers, while global communication may need more technical expertise. Local communication is frequent but less novel and diverse compared to global communication.

Although the local and global positions are correlated, as evident from the correlation analysis, they are not completely dependent. We can think of them as two dimensions to examine position in online communities. At one extreme, a member can be a peripheral member globally, but a core member locally. For example, in a regional Francophone OSCAR sub-community within the global OSCAR community. Another possibility is for a global core member not to have a local central position. Although this is a rare case because core members tend to be frequent communicators, this case could happen in expert-novice communities where few experts provide help to novice members but do not communicate among themselves. All in all, the local perspective adds a new dimension to examine structural social capital, and this new dimension could help solving the incommensurability of findings in core-periphery research.

Contributions in the OSCAR community are local and decentralized. However, the integration of contributions is centralized and overseen by a committee of core members. On the surface this can be seen as limiting the effectiveness of the community and the incentives of peripheral members to contribute. However, OSCAR allows the development of plugins, eforms and templates that do not need to be integrated in order to be used. This gives all members a high degree of freedom in pursuing their own interests without needing a consent from to committee overseeing the code of OSCAR. Indeed, the developed plugins are shared online and could be downloaded and plugged into OSCAR [13]. OSCAR is seen by many members as a platform that is robust and affords easy customizations. While most contributions are not channeled into the core code of OSCAR, they are not wasted as they are diverted into electronic forms, templates and other plugins. This confirms with recent findings on the role of peripheral developers in open-source projects (Setia et al., 2012). While peripheral developers may not be developing core code, they are contributing in other ways and also

[13] http://oscarcanada.org/oscar-users/emr-resource/eform/eform-examples

learning via the process (Singh et al., 2011b; Mehra and Mookerjee, 2012).

One recent perspective to study organizational change is to focus on the discourse and language use in speech and written communication among organizational actors. This approach construes organizations as texts, conversations and discursive performances where change is considered a linguistic process (Demers, 2007, Ch. 8). For example, Demers et al. (2003) used a discursive perspective to analyze official merger announcements. They found that top managers involve wedding narratives in merger-acquisitions announcements in order to increase the legitimacy of these decisions. The analysis of language posits that the linguistic artifacts such as text and speech are not merely representational but rather constituent of social and organizational reality (Alvesson and Kärreman, 2000). As such, the dynamics of discourse dictate the process of change in the organization (Young and Fitzgerald, 2006, Ch. 5).

Language plays a more significant role in shaping modern knowledge-driven societies (Fairclough, 2004). A discursive approach is even more relevant for studying change and dynamics of online communities that are built primarily around written communication among members. Face-to-face contact, handshakes and other social cues are very rare in online communities whose members span spatial and temporal boundaries (Cummings et al., 2009). Moreover, from a research perspective online communities present a golden opportunity because their history is documented and waiting to be explored. This thesis demonstrates the importance of discourse in shaping the growth of the OSCAR community. A better understanding of how online communities operate, evolve and eventually produce common goods and knowledge is achieved by emphasizing the temporal dimension and the role of language in shaping the dynamics of interaction among members.

8.2 Methodological Contributions

With the advent of big data new methods and tools are needed to make use of big data sets in social science research. The three studies contribute methodologically by presenting novel approaches to treat network and textual data. The first and third studies present a comprehensive apparatus to examine social network structure globally and locally at both the individual level and the aggregate level. While most research on social networks has ignored the role of content, the second and third studies put content back into the picture. In fact in the context of online communities, all communication is documented and archived online. This is both a challenge and an opportunity. In the second study, inspired from computational linguistics, we propose an apparatus to examine and measure the specialization of feature requests in the community. In the third study, we show how structure and content can be examined simultaneously when studying the growth of the community.

At the individual level, we show local structural centrality and boundary spanning is more determinant of knowledge contribution than global centrality and boundary spanning. Those local patterns give insight into the roles member play in online communities than do global features. This calls for a reconsideration of structural measurements in online communities and network research. At the aggregate level, social networks exhibit unique structural characteristics that are atypical in other settings (Jackson and Rogers, 2007). Therefore, it is important to consider various measurements when evaluating the structure of social networks. Unfortunately, even recently most research considers only degree distribution as the main fingerprint of a network (Johnson et al., 2014). Looking at different fingerprints of structure such as the clustering distribution may yield different results and conclusions. More importantly, considering local characteristics such as the frequency of graphlets extends comparison based on degree and clustering based on the relationship of a node with its immediate connections (Milenković and Przulj, 2008).

Relationships in online communities are formed almost exclusively through written communication. Therefore, the fundamental building block of online community is the use of language. Little research has addressed how the comparative use of language shapes community dynamics. Recent advances in computational linguistics have allowed the automatic processing and treatment of speech and text, two data sources that have been exclusively analyzed with qualitative techniques. Perhaps the best example that demonstrates the power of the machine in treating human language is the recent victory of IBM Watson as the world champion of Jeopardy (Ferrucci et al., 2010). This is great news for online communities research because of the large size of communication text exchanged online among members. This thesis demonstrates the prime importance of content in defining the behavior of members in online communities and shaping their growth.

8.3 Practical Contributions

In the context of healthcare IT, the thesis enriches our understanding of open-source EMRs and will shed light on how open-source is comparable to proprietary systems in healthcare. Previous research has emphasized the importance of developers' in developing and supporting open-source software. By examining the dynamics of the OSCAR community, we broaden the previous perspective and include the role of members who are neither developers nor programmers but medical practitioners who use the software in their profession in testing, fixing, and supporting the software.

The OSCAR EMR demonstrates the promise of OSS in primary care. The main advantage of OSCAR over proprietary EMRs is the reduced cost of acquisition and maintenance. In terms of complexity, OSCAR has a considerable learning curve. However, this training stage is facilitated by the many resources available online and also by the existence of a community of enthusiastic users. Moreover, a huge collection of downloadable e-forms and templates is available online making customization a very easy process. Finally, OSCAR favors open standards and interfaces including support for HL7, ICD-9 & 10, Canadian drug database, and billing codes. More effort is required to integrate OSCAR with legacy HIT systems of hospitals and pharmacies.

As open-source software, OSCAR offers an accessible and open platform for various stakeholders. While primarily developed and maintained by a team of developers at McMaster University, OSCAR enlists contributions from its users' community. The community represents the interests of users in the project. Communication between users and developers online via the mailing list and during user group meetings ensures that OSCAR stays aligned to medical users' needs. In addition, the user community allows users to share their knowledge and expertise creating a space for obtaining support. Moreover, the user community contributes non-trivial extensions to OSCAR such as e-forms, templates, the e-form generator and the OSCAR manual. In the studied clinic, customizability of OSCAR was essential for its acceptance. Customized forms matched clinic paper forms and customized templates helped to match work practices. Customized templates for consultations and referrals improved the external communication with other clinics and specialists leading to better consultation and referral practices. Finally, allowing qualified commercial businesses to provide support and training services ensures that OSCAR reaches beyond tech savvy users to all medical practitioners accustomed to the traditional model of software acquisition.

Compared to other OSS, health OSS needs many more customizations to fit the requirement at the implementation site. While software packages such as browsers, operating systems and even programming language and development tools work for the majority of users with its default settings, this is rarely the case of health software that need to comply with local, regional and provincial requirements. Customization is both an individual and collective effort that stems from the need of particular members and support from the community. However, not only do members actively participate in shaping the evolution of the EMR, but also they do differ in opinions and preferences. The lack of a central authority limits the constraints on members' modification and customization. As a result, the EMR could be customized in multiple ways to achieve the same task. While this may be puzzling for novice members and users, it is a powerful affordance for existing members who have more freedom in fitting the system to their needs. We illustrate this with the following case about developing a billing module in OSCAR.

In a growing and expanding open-source EMR, a need for implementing billing functions emerged as more clinics and doctors who have adopted the EMR complained about the lack of billing in the EMR. Moreover, this function is essential to compete with proprietary EMRs that already offer it. There were multiple discussions in the online forums of how best to implement billing. On one hand, there was a group of developers who thought it is best to implement billing a new core module. A new module will allow more flexibility in implementing billing functions. In addition, developers already have the needed experience to write the require code. On the other hand, there was a group of advanced users who are mostly doctors but had gained much experience writing medical forms by using the forms module. This group thought that it is better to implement billing as a fillable form. Doing so will not require changing neither the database nor the installation base. The billing form can be plugged in online.

After much debate convergence was not achieved. There was a clear conflict of interests between the two groups. Developers have the incentives to write code and keep the EMR tightly integrated by not allowing two many plugins along with the unstructured data that come with them. Advanced users on the other hand, cannot write code. They know how to design forms using the forms module. They are not very interested in data integrity or the wholeness of the project and they do not want to be left at the mercy of developers. Overtime, however, developers went ahead and implemented the billing core module. Advanced users could not realize their implementation because of lack of time and because the group was not big enough to carry on such an implementation. As a result, billing emerged as a core module in the EMR.

One year later, however, advanced users resumed their activity in writing a pluggable billing form. Because billing depends a lot on the jurisdiction in which the clinic operates (provincial, private, public, for profit, etc...). The function implemented in the core billing module did not satisfy the need of all users. Nor were developers able to implement all needed functions and satisfy the needs of users in all jurisdictions. The best they could have done was to write two billing module: one for Ontario and one for British Columbia. Fillable forms on the other hand are much more flexible and can be customized by the end users. Advanced users were able to write a sophisticated billing form and make it available for the rest of the community. With time different customizations for that form were contributed by users in different jurisdictions and as a result many users ended up subsisting the rigid billing core module with a lighter but more flexible billing form that suits their needs.

The final outcome of the billing story is that there are two different implementations of billing in the open-source EMR with duplication functions. It is up to the user to choose the one that suits his/her needs or to remain confused between the two.

Second, the findings of this paper have important implications for both members and designers of online communities. As Aral et al. (2013) remind us, understanding the design of social media platforms is important because it impacts interaction, enable and constrain social and economic phenomena (p. 6). The OSCAR community is hosted in SourceForge, a platform that limits member identification and affords threaded textual communication. The strategies employed by new members reflect both their personal interaction preferences and also the platform affordances. Designers of online communities' platforms should take into consideration features and capabilities in order to achieve the desirable outcomes of the community (Bateman et al., 2011; Oestreicher-Singer and Zalmanson, 2013; Levina and Arriaga, 2015).

The inclusion of novel features is driven by the advancement in communication technology and the availability of high interactivity devices such as smartphones and tablets. However, it is assumed that complementing traditional online collaboration platforms with richer features is advantageous and will improve the outcomes. This assumption is challenged by recent research that started to shed light on the implications of the inclusion of social media features in online communities and knowledge collaboration platforms. The choice to include and exclude a set of features can be detrimental for online communities (Hallerbach et al., 2013). Some features, when activated, will have positive effects on knowledge sharing, but others will have unintended adverse consequences (Faraj et al., 2013).

From members' perspective, members who want to achieve a position of authority may engage in various strategies to attract newcomers. A member wishing to employ a structural mechanism may want to be a frequent poster and also communicate with many other existing members. This will give him/her more visibility and increase the chance of receiving new ties from new members. On the hand, a member who wants to exploit homophily may want to expose more personal information online in return for future ties with similar peers. Our results indicate that more important than the two previous mechanisms is the ability of this member to post meaningful messages that express topical interests. Such messages will allow future members to evaluate their shared interests and communicate with the existing members.

References

Achrol, R. S., 1997. Changes in the theory of interorganizational relations in marketing: toward a network paradigm. *Journal of the academy of marketing science*, 25(1):56–71. (cited in p. 54)

Agarwal, R. and Angst, C. M., 2006. Technology-Enabled Transformations in U.S. Health Care. *Human-computer interaction and management information systems: applications*, 6: 357. (cited in p. 17)

Ahuja, G., Soda, G., and Zaheer, A., 2012. The Genesis and Dynamics of Organizational Networks. *Organization science*, 23(2):434–448. (cited in p. 2, 4, 34, 90)

Alavi, M. and Leidner, D. E., 2001. Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, pages 107–136. (cited in p. 12)

Albanese, R. and Van Fleet, D. D., 1985. Rational behavior in groups: The free-riding tendency. *The Academy of Management Review*, 10(2):244–255. (cited in p. 105)

Albert, R. and Barabási, A.-L., January 2002. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74(1):47–97. doi: 10.1103/RevModPhys.74.47. URL `http://link.aps.org/doi/10.1103/RevModPhys.74.47`. (cited in p. 126)

Aldrich, H., 1999. *Organizations evolving*. Sage. ISBN 9780803989191. (cited in p. 89, 90)

Alvarez-Hamelin, J. I., Dall Asta, L., Barrat, A., and Vespignani, A., 2006. Large scale networks fingerprinting and visualization using the k-core decomposition. *Advances in neural information processing systems*, 18:41. (cited in p. xiii, 69)

Alvesson, M. and Kärreman, D., 2000. Taking the Linguistic Turn in Organizational Research Challenges, Responses, Consequences. *The journal of applied behavioral science*, 36(2): 136–158. (cited in p. 11, 118, 149)

Anderson, J. G., 2007. Social, ethical and legal barriers to e-health. *International journal of medical informatics*, 76(5-6):480–483. (cited in p. 20)

Angst, C. and Agarwal, R., 2006. Overcoming personal barriers to adoption when technology enables information to be available to others. (cited in p. 20)

Aral, S. and Walker, D., 2011. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science*, 57(9):1623–1639. (cited in p. 66)

Aral, S., 2013. What Would Ashton Do–and Does It Matter?. *Harvard Business Review*, 91 (5):25–27. ISSN 00178012. (cited in p. 58)

Aral, S. and Van Alstyne, M., 2011. The Diversity-Bandwidth Trade-off. *American Journal of Sociology*, 117(1):90–171. (cited in p. 51, 55, 118)

Aral, S., Dellarocas, C., and Godes, D., 2013. Introduction to the special issue-social media and business transformation: A framework for research. *Information Systems Research*, 24 (1):3–13. (cited in p. 153)

Ardichvili, A., Page, V., and Wentling, T., 2003. Motivation and barriers to participation in virtual knowledge-sharing communities of practice. *Journal of knowledge management*, 7 (1):64–77. (cited in p. 9, 27)

Argote, L. and Miron-Spektor, E., 2011. Organizational learning: From experience to knowledge. *Organization Science*, 22(5):1123–1137. (cited in p. 1)

Arrow, K. J., 1963. Uncertainty and the welfare economics of medical care. *The American economic review*, 53(5):941–973. (cited in p. 4, 13, 18)

Arvind, V., Köbler, J., Kuhnert, S., and Vasudev, Y., 2012. Approximate Graph Isomorphism. *Mathematical Foundations of Computer Science 2012*, pages 100–111. (cited in p. 126)

Asvanund, A., Clay, K., Krishnan, R., and Smith, M. D., 2004. An empirical analysis of network externalities in peer-to-peer music-sharing networks. *Information systems research*, 15(2): 155–174. (cited in p. 12, 91)

Backstrom, L., Huttenlocher, D., and Kleinberg, J., 2006. Group Formation in Large Social Networks : Membership , Growth , and Evolution. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 41–54. ISBN 1595933395. (cited in p. 110, 113)

Balkundi, P. and Kilduff, M., 2006. The ties that lead: A social network approach to leadership. *The Leadership Quarterly*, 17(4):419–439. URL `http://www.sciencedirect.com/science/article/pii/S1048984306000038`. (cited in p. 66)

Barabási, A.-L. and Albert, R., 1999. Emergence of scaling in random networks. *Science*, 286(5439):509–512. (cited in p. 113, 124, 125)

Barabási, A.-L. and Frangos, J., 2002. *Linked: The New Science Of Networks Science Of Networks*. Basic Books. (cited in p. 113)

Barabási, A.-L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., and Vicsek, T., 2002. Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311(3):590–614. (cited in p. 113)

Barney, J., 1991. Firm resources and sustained competitive advantage. *Journal of management*, 17(1):99–120. (cited in p. 60)

Baron, R. J., Fabens, E. L., Schiffman, M., and Wolf, E., 2005. Electronic health records: just around the corner? Or over the cliff? *Annals of internal medicine*, 143(3):222–226. (cited in p. 19, 24)

Barrett, M., Heracleous, L., and Walsham, G., 2013. A RHETORICAL APPROACH TO IT DIFFUSION: RECONCEPTUALIZING THE IDEOLOGY-FRAMING RELATIONSHIP IN COMPUTERIZATION MOVEMENTS. *MIS Quarterly*, 37(1). (cited in p. 86)

Bateman, P. J., Gray, P. H., and Butler, B. S., 2011. Research Note-The Impact of Community Commitment on Participation in Online Communities. *Information Systems Research*, 22 (4):841–854. (cited in p. 153)

Benkler, Y., 2006. *The wealth of networks: How social production transforms markets and freedom*. Yale University Press. (cited in p. 1)

Berdou, E., 2011. *Organization in Open Source Communities: At the Crossroads of the Gift and Market Economies*, volume 15. Taylor & Francis US. (cited in p. 8, 15, 56, 64)

Berg, J. and Lässig, M., 2004. Local graph alignment and motif search in biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(41): 14689–14694. (cited in p. 69)

Berg, M. and Bowker, G., 1997. The multiple bodies of the medical record. *The Sociological Quarterly*, 38(3):513–537. (cited in p. 4, 23)

Blei, D. M., 2012. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84. (cited in p. xiv, 100, 129)

Blei, D. M. and Lafferty, J. D., June 2007. A correlated topic model of Science. *The Annals of Applied Statistics*, 1(1):17–35. ISSN 1932-6157. doi: 10.1214/07-AOAS114. URL `http://projecteuclid.org/euclid.aoas/1183143727`. (cited in p. 100, 122)

Blei, D. M., Ng, A. Y., and Jordan, M. I., 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022. (cited in p. 100, 128)

Blumenthal, D. and Glaser, J. P., 2007. Information technology comes to medicine. *New England Journal of Medicine*, 356(24):2527–2534. (cited in p. 17, 19)

Borgatti, S. P. and Everett, M. G., 1999. Models of core/periphery structures. *Social Networks*, 21(4):375–395. URL `http://www.sciencedirect.com/science/article/pii/S0378873399000192`. (cited in p. 56, 68)

Borgatti, S. P. and Everett, M. G., October 2006. A Graph-theoretic perspective on centrality. *Social Networks*, 28(4):466–484. ISSN 03788733. doi: 10.1016/j.socnet.2005.11.005. URL `http://linkinghub.elsevier.com/retrieve/pii/S0378873305000833`. (cited in p. 55, 56, 121)

Borgatti, S. P. and Foster, P. C., 2003. The network paradigm in organizational research: A review and typology. *Journal of management*, 29(6):991–1013. URL `http://dx.doi.org/10.1016/S0149-2063(03)00087-4`. (cited in p. 51, 52, 114)

Borgatti, S. P. and Halgin, D. S., October 2011a. On Network Theory. *Organization Science*, 22(5):1168–1181. ISSN 1047-7039. doi: 10.1287/orsc.1100.0641. URL `http://pubsonline.informs.org/doi/abs/10.1287/orsc.1100.0641`. (cited in p. 52)

Borgatti, S. P. and Halgin, D. S., 2011b. Analyzing affiliation networks. *The Sage handbook of social network analysis*, pages 417–433. (cited in p. 72)

Borgatti, S. P., Mehra, A., Brass, D. J., and Labianca, G., 2009. Network analysis in the social sciences. *science*, 323(5916):892–895. (cited in p. 52)

Borys, B. and Jemison, D. B., 1989. Hybrid arrangements as strategic alliances: Theoretical issues in organizational combinations. *Academy of Management Review*, pages 234–249. (cited in p. 91, 147)

Boyd, D. M. and Ellison, N. B., 2007. Social network sites: Definition, history, and scholarship. *Journal of Computer Mediated Communication*, 13(1):210–230. (cited in p. 7)

Boyd, J. P., Fitzgerald, W. J., and Beck, R. J., May 2006. Computing core/periphery structures and permutation tests for social relations data. *Social Networks*, 28(2):165–178. ISSN 03788733. doi: 10.1016/j.socnet.2005.06.003. URL http://www.sciencedirect.com/science/article/pii/S0378873305000456. (cited in p. 56)

Brandes, U., 2001. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2):163–177. (cited in p. 68, 121)

Brandes, U. and Pich, C., 2007. Centrality estimation in large networks. *International Journal of Bifurcation and Chaos*, 17(07):2303–2318. URL http://www.worldscientific.com/doi/abs/10.1142/S0218127407018403. (cited in p. 67)

Brin, S. and Page, L., 1998. The anatomy of a large-scale hypertextual Web search engine. *Computer networks and ISDN systems*, 30(1):107–117. (cited in p. 66, 121)

Brown, J. S. and Duguid, P., 2001. Knowledge and organization: A social-practice perspective. *Organization science*, pages 198–213. (cited in p. 27, 34)

Burt, R., 1992. *Structural holes: The social structure of competition*. Harvard University Press. (cited in p. 51, 53, 54, 112)

Burt, R. S., October 2002. Bridge decay. *Social Networks*, 24(4):333–363. ISSN 03788733. doi: 10.1016/S0378-8733(02)00017-5. URL http://linkinghub.elsevier.com/retrieve/pii/S0378873302000175. (cited in p. 111)

Burt, R. S., 2010. *Neighbor networks: Competitive advantage local and personal*. Oxford University Press. (cited in p. xiii, 52, 54, 55, 122)

Butler, B. S., 2001. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information systems research*, 12(4):346–362. (cited in p. 11, 53, 91, 112)

Butler, B., Joyce, E., and Pike, J., 2008. Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 1101–1110, 1357227. ACM. (cited in p. 10, 58, 61, 87, 91)

Butler, B., Kiesler, S., and Kraut, R., 2013. Community Effort in Online Groups: Who Does the. *Leadership at a Distance: Research in Technologically-Supported Work*, page 171. (cited in p. 8, 25, 61, 86, 88, 110)

Butler, B. S. and Wang, X., 2012. The Cross-Purposes of Cross-Posting: Boundary Reshaping Behavior in Online Discussion Communities. *Information Systems Research*, 23(3-part-2): 993–1010. doi: 10.1287/isre.1110.0378. URL `http://pubsonline.informs.org/doi/abs/10.1287/isre.1110.0378`. (cited in p. 147)

Butler, B. S., Bateman, P. J., and Gray, P. H., 2015. An Attraction-Selection-Attrition Theory of Online Community Size and Resilience. *Mis Quarterly*, X(X):1–30. (cited in p. 87, 91, 93, 113, 115)

Byrne, C. M., Mercincavage, L. M., Pan, E. C., Vincent, A. G., Johnston, D. S., and Middleton, B., 2010. The value from investments in health information technology at the US Department of Veterans Affairs. *Health Affairs*, 29(4):629–638. (cited in p. 23)

Cameron, K. S., Sutton, R. I., and Whetten, D. A., 1993. *Readings in organizational decline: Frameworks, research, and prescriptions*. Ballinger Publishing Company. ISBN 0887302238. (cited in p. 87)

Capocci, A., Servedio, V. D. P., Colaiori, F., Buriol, L. S., Donato, D., Leonardi, S., and Caldarelli, G., September 2006. Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia. *Phys. Rev. E*, 74(3):36116. doi: 10.1103/PhysRevE.74.036116. URL `http://link.aps.org/doi/10.1103/PhysRevE.74.036116`. (cited in p. 112, 113)

Carlile, P. and Lakhani, K., 2011. Innovation and the Challenge of Novelty: The Novelty-Confirmation-Transformation Cycle in Software and Science. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, (11-096). (cited in p. 88, 91)

Carlson, J. R. and Zmud, R. W., 1999. Channel expansion theory and the experiential nature of media richness perceptions. *Academy of management journal*, pages 153–170. (cited in p. 14, 107)

Castonguay, C., Marcotte, J., and Venne, M., 2008. Task Force on the Funding of the Health System. URL `http://www.groupes.finances.gouv.qc.ca/financementsante/en/groupe/membres.asp`. (cited in p. 19)

Cattani, G., Ferriani, S., and Allison, P. D., February 2014. Insiders, Outsiders, and the Struggle for Consecration in Cultural Fields: A Core-Periphery Perspective. *American Sociological Review*. ISSN 0003-1224. doi: 10.1177/0003122414520960. URL `http://asr.sagepub.com/cgi/doi/10.1177/0003122414520960`. (cited in p. 56, 92)

Cattani, G. and Ferriani, S., December 2008. A Core/Periphery Perspective on Individual Creative Performance: Social Networks and Cinematic Achievements in the Hollywood

Film Industry. *Organization Science*, 19(6):824–844. ISSN 1047-7039. doi: 10.1287/orsc.1070.0350. URL `http://pubsonline.informs.org/doi/abs/10.1287/orsc.1070.0350`. (cited in p. 56, 92)

Chuang, J., Manning, C. D., and Heer, J., 2012. Termite: Visualization Techniques for Assessing Textual Topic Models. In *Advanced Visual Interfaces*. URL `http://vis.stanford.edu/papers/termite`. (cited in p. 100)

Clemons, E. K. and Row, M. C., 1991. Sustaining IT advantage: the role of structural differences. *Mis Quarterly*, pages 275–292. (cited in p. 60)

Coakes, E. and Smith, P., 2007. Developing communities of innovation by identifying innovation champions. *Learning Organization, The*, 14(1):74–85. (cited in p. 1, 109)

Coase, R. H., 1937. The nature of the firm. *economica*, 4(16):386–405. (cited in p. 81)

Cobb, N. K., Graham, A. L., and Abrams, D. B., 2010. Social network structure of a large online community for smoking cessation. *American Journal of Public Health*, 100(7): 1282–1289. (cited in p. 68)

Cohen, W. M. and Levinthal, D. A., 1990. Absorptive capacity: a new perspective on learning and innovation. *Administrative science quarterly*, 35(1). (cited in p. 57)

Collier, B. and Kraut, R., 2012. Leading the Collective: Social Capital and the Development of Leaders in Core-Periphery Organizations. In *Collective Intelligence*. (cited in p. 57, 68)

Conklin, J., 1996. Designing organizational memory: preserving intellectual assets in a knowledge economy. *Group Decision Support Systems*, 1:362. (cited in p. 61, 85)

Conner, K. R. and Prahalad, C. K., 1996. A resource-based theory of the firm: Knowledge versus opportunism. *Organization science*, 7(5):477–501. (cited in p. 60)

Constant, D., Kiesler, S., and Sproull, L., 1994. What's mine is ours, or is it? A study of attitudes about information sharing. *Information systems research*, 5(4):400–421. (cited in p. 112)

Constant, D., Sproull, L., and Kiesler, S., April 1996. The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice. *Organization Science*, 7(2):119–135. ISSN 1047-7039. doi: 10.1287/orsc.7.2.119. URL `http://pubsonline.informs.org/doi/abs/10.1287/orsc.7.2.119`. (cited in p. 8, 16, 54, 58, 62, 91, 110, 112)

Cook, K. S. and Emerson, R. M., 1978. Power, equity and commitment in exchange networks. *American Sociological Review*, 43(5):721–739. (cited in p. 52)

Cook, K. S., Emerson, R. M., Gillmore, M. R., and Yamagishi, T., 1983. The distribution of power in exchange networks: Theory and experimental results. *American journal of sociology*, pages 275–305. (cited in p. 55, 58)

Cooren, F., Kuhn, T., Cornelissen, J. P., and Clark, T., 2011. Communication, organizing and organization: An overview and introduction to the special issue. *Organization Studies*, 32 (9):1149–1170. (cited in p. 11, 115)

Cross, R. and Sproull, L., 2004. More Than an Answer: Information Relationships for Actionable Knowledge. *Organization science*, 15(4):446–462. URL `http://www.jstor.org/stable/30034748`. (cited in p. 9)

Crowston, K. and Howison, J., 2005. The social structure of free and open source software development. *First Monday*, 10(2-7). (cited in p. 58, 64, 68, 80, 106)

Cummings, J., A., E., and C.K., P., 2009. Crossing Spatial and Temporal Boundaries in Globally Distributed Projects: A Relational Model of Coordination Delay. *Information Systems Research*, 20(3):420–439. (cited in p. 61, 105, 149)

Cutler, D. M., 2005. *Your money or your life: strong medicine for America's health care system*. Oxford University Press, USA. ISBN 0195181328. (cited in p. 13, 17)

Dabbish, L., Stuart, C., Tsay, J., and Herbsleb, J., 2012. Social coding in github: transparency and collaboration in an open software repository. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 1277–1286. ACM. (cited in p. 107)

Dahlander, L. and Frederiksen, L., August 2012. The Core and Cosmopolitans: A Relational View of Innovation in User Communities. *Organization Science*, 23(4):988–1007. ISSN 1047-7039. doi: 10.1287/orsc.1110.0673. URL `http://pubsonline.informs.org/doi/abs/10.1287/orsc.1110.0673`. (cited in p. 9, 53, 54, 56, 62, 66, 79, 90, 92)

Dahlander, L. and McFarland, D. A., 2013. Ties That Last: Tie Formation and Persistence in Research Collaborations over Time. *Administrative Science Quarterly*, 58(1):69–110. doi: 10.1177/0001839212474272. URL `http://asq.sagepub.com/content/58/1/69.abstract`. (cited in p. 110, 111, 113)

Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., and Potts, C., 2013. No country for old members: User lifecycle and linguistic change in online communities. *Proceedings of the 22nd International World Wide Web Conference*. (cited in p. 34, 110, 115)

Daniel, S., Agarwal, R., and Stewart, K. J., 2013. The effects of diversity in global, distributed collectives: A study of open source project success. *Information Systems Research*, 24(2): 312–333. (cited in p. 147)

David, P. A. and Rullani, F., 2008. Dynamics of innovation in an open source collaboration environment: lurking, laboring, and launching FLOSS projects on SourceForge. *Industrial and Corporate Change*, 17(4):647–710. (cited in p. 86, 88, 106)

Davidson, E. J. and Chismar, W. G., 2007. The interaction of institutionally triggered and technology-triggered social structure change: An investigation of computerized physician order entry. *MIS quarterly*, 31(4):739–758. (cited in p. 19)

Dawes, M., 2010. What is OSCAR Security. URL `http://www.martindawes.ca/Oscar/WhatisOscar&Security.pdf`. (cited in p. 48)

de Ven, A. H., 2005. Running in packs to develop knowledge-intensive technologies. *Mis Quarterly*, 29(2):365–377. (cited in p. 11)

de Ven, A. H. and Poole, M. S., 2005. Alternative approaches for studying organizational change. *Organization Studies*, 26(9):1377–1404. (cited in p. 88)

Demers, C., 2007. *Organizational change theories: A synthesis*. Sage Publications, Inc. ISBN 0761929320. (cited in p. 10, 11, 88, 89, 140, 149)

Demers, C., Giroux, N., and Chreim, S., 2003. Merger and acquisition announcements as corporate wedding narratives. *Journal of Organizational Change Management*, 16(2): 223–242. (cited in p. 118, 149)

Dennis, A. R. and Valacich, J. S., 1999. Research Note. Electronic Brainstorming: Illusions and Patterns of Productivity. *Information Systems Research*, 10(4):375. (cited in p. 88)

Dennis, A. R., Fuller, R. M., and Valacich, J. S., 2008. Media, tasks, and communication processes: A theory of media synchronicity. *Mis Quarterly*, 32(3):575–600. (cited in p. 14, 107, 109)

DesRoches, C. M., Campbell, E. G., Rao, S. R., Donelan, K., Ferris, T. G., Jha, A., Kaushal, R., Levy, D. E., Rosenbaum, S., and Shields, A. E., 2008. Electronic health records in ambulatory care: a national survey of physicians. *New England Journal of Medicine*, 359 (1):50–60. (cited in p. 2, 4, 17, 18)

Diehl, M. and Stroebe, W., 1987. Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of personality and social psychology*, 53(3):497–509. (cited in p. 105)

Dierks, T., 2008. The transport layer security (TLS) protocol version 1.2. URL `http://wiki.tools.ietf.org/html/rfc5246`. (cited in p. 48)

Donath, J., October 2007. Signals in Social Supernets. *Journal of Computer-Mediated Communication*, 13(1):231–251. ISSN 10836101. doi: 10.1111/j.1083-6101.2007.00394.x. URL `http://doi.wiley.com/10.1111/j.1083-6101.2007.00394.x`. (cited in p. 66)

Erdos, P. and Rényi, A., 1960. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl*, 5:17–61. (cited in p. 125)

Estelles-Arolas, E. and Gonzalez-Ladron-de Guevara, F., March 2012. Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2):189–200. ISSN 0165-5515. doi: 10.1177/0165551512437638. URL `http://jis.sagepub.com/cgi/doi/10.1177/0165551512437638`. (cited in p. 57)

Everett, M. and Borgatti, S. P., January 2005. Ego network betweenness. *Social Networks*, 27 (1):31–38. ISSN 03788733. doi: 10.1016/j.socnet.2004.11.007. URL `http://linkinghub.elsevier.com/retrieve/pii/S037887330400067X`. (cited in p. 55, 69)

Fairclough, N., 1995. *Media discourse*. Arnold Publishers. (cited in p. 11, 115)

Fairclough, N., 2004. Critical discourse analysis in researching language in the new capitalism: Overdetermination, transdisciplinarity and textual analysis. *Systemic functional linguistics and critical discourse analysis*, pages 103–122. (cited in p. 118, 149)

Faraj, S. and Johnson, S. L., 2011. Network exchange patterns in online communities. *Organization science*, 22(6):1464. URL `http://dx.doi.org/10.1287/orsc.1100.0600`. (cited in p. 9, 53, 58, 63, 66, 112)

Faraj, S. and Xiao, Y., 2006. Coordination in fast-response organizations. *Management Science*, 52(8):1155–1169. (cited in p. 57)

Faraj, S., Jarvenpaa, S. L., and Majchrzak, A., 2011. Knowledge collaboration in online communities. *Organization science*, 22(5):1224–1239. (cited in p. 4, 5, 21, 25, 53, 85, 86, 88, 90, 91, 105, 110, 139, 147)

Faraj, S. and Sproull, L., 2000. Coordinating expertise in software development teams. *Management science*, 46(12):1554–1568. (cited in p. 60)

Faraj, S., Kane, G. C., and Azad, B., 2013. The Contradictory Influence of Social Media Affordances on Online Communal Knowledge Sharing. *Journal of Computer-Mediated Communication*, 19(1):38–55. ISSN 1083-6101. doi: 10.1111/jcc4.12030. URL `http://dx.doi.org/10.1111/jcc4.12030`. (cited in p. 53, 87, 109, 116, 153)

Faraj, S., Kudaravalli, S., and Wasko, M., 2015. Leading Knowledge Collaboration in Online Communities. *MIS Quarterly*, X(X):1–20. (cited in p. 92, 112, 113)

Faus, S. A. and Sujansky, W., 2008. Open-Source EHR Systems for Ambulatory Care: A Market Assessment. Technical report, California HealthCare Foundation. URL `http://www.chcf.org/publications/2008/01/opensource-ehr-systems-for-ambulatory-care-a-market-assessment`. (cited in p. 13, 18)

Faust, K., 1997. Centrality in affiliation networks. *Social networks*, 19(2):157–191. (cited in p. 72)

Feller, J. and Fitzgerald, B., 2002. *Understanding open source software development*. Addison-Wesley Professional. ISBN 0201734966. (cited in p. 16, 61)

Feller, J., Fitzgerald, B., Hissam, S. A., and Lakhani, K. R., 2005. *Perspectives on free and open source software*. MIT Press Cambridge, MA. ISBN 0262062461. (cited in p. 4, 15, 61)

Feller, J., Finnegan, P., Fitzgerald, B., and Hayes, J., 2008. From peer production to productization: A study of socially enabled business exchanges in open source service networks. *Information systems research*, 19(4):475–493. (cited in p. 9)

Fenton, N. E. and Pfleeger, S. L., 1998. *Software metrics: a rigorous and practical approach*. PWS Publishing Co. (cited in p. 31)

Ferguson, J., Huysman, M., and Soekijad, M., 2012. You take the high road and I'll take the low road: unfolding parallel discourse in epistemic communities. *Proceedings of the 72nd Academy of Management Annual Meeting*. (cited in p. 115, 147)

Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A. A., Lally, A., Murdock, J. W., Nyberg, E., Prager, J., and Others, 2010. Building Watson: An overview of the DeepQA project. *AI magazine*, 31(3):59–79. (cited in p. 150)

Fleming, L. and Waguespack, D. M., 2007. Brokerage, boundary spanning, and leadership in open innovation communities. *Organization Science*, 18(2):165–180. (cited in p. 68)

Flowers, S. and Henwood, F., 2010. *Perspectives on User Innovation*, volume 16. Imperial College Pr. ISBN 1848166990. (cited in p. 7)

Foster, D. M., 2010. Open Source Adoption in Worldwide Governments. URL `http://software.intel.com/en-us/blogs/2010/01/19/open-source-adoption-in-worldwide-governments/`. (cited in p. 14, 17)

Freeman, L. C., 1979. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239. (cited in p. 60)

Friedkin, N. E., 1993. Structural bases of interpersonal influence in groups: A longitudinal case study. *American Sociological Review*, 58:861. (cited in p. 52)

Glynn, E., Fitzgerald, B., and Exton, C., 2005. Commercial adoption of open source software: an empirical study. In *2005 International Symposium on Empirical Software Engineering, 2005.*, pages 217–226. IEEE. ISBN 0-7803-9507-7. doi: 10.1109/ISESE.2005.1541831. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1541831. (cited in p. 17)

Granovetter, M. S., 1973. The strength of weak ties. *American journal of sociology*, pages 1360–1380. (cited in p. 51, 52, 54, 55)

Grant, R. M., 1996a. Toward a knowledge-based theory of the firm. *Strategic management journal*, 17:109–122. (cited in p. 27, 34, 60)

Grant, R. M., 1996b. Prospering in dynamically-competitive environments: organizational capability as knowledge integration. *Organization science*, 7(4):375–387. (cited in p. 57)

Griffiths, T. L. and Steyvers, M., 2004. Finding scientific topics. *Proceedings of the National academy of Sciences of the United States of America*, 101(Suppl 1):5228–5235. (cited in p. 128, 129)

Gu, B., Konana, P., Rajagopalan, B., and Chen, H. W. M., 2007. Competition among virtual communities and user valuation: The case of investing-related communities. *Information systems research*, 18(1):68–85. (cited in p. 12, 13, 91)

Gupta, A. K., Smith, K. G., and Shalley, C. E., August 2006. the Interplay Between Exploration and Exploitation. *Academy of Management Journal*, 49(4):693–706. ISSN 0001-4273. doi: 10.5465/AMJ.2006.22083026. URL http://amj.aom.org/cgi/doi/10.5465/AMJ.2006.22083026. (cited in p. 88)

Hagberg, A., Swart, P., and S Chult, D., 2008. Exploring network structure, dynamics, and function using NetworkX. Technical report, Los Alamos National Laboratory (LANL). (cited in p. 124)

Hallerbach, I. P., Barrett, M., and Faraj, S., 2013. Emergence in Nascent Online Communities: An Affordance Perspective. In *International Conference of Information Systems (ICIS)*. (cited in p. 110, 116, 153)

Hannan, M. T. and Freeman, J., 1984. Structural inertia and organizational change. *American sociological review*, 49(2):149–164. (cited in p. 10)

Hansen, D., Shneiderman, B., and Smith, M. A., 2010. *Analyzing social media networks with NodeXL*. Morgan Kaufmann. (cited in p. 71, 119)

Hansen, M. T., 1999. The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative science quarterly*, 44(1):82–111. (cited in p. 34, 115)

Hartwick, J. and Barki, H., 2001. Communication as a dimension of user participation. *Professional Communication, IEEE Transactions on*, 44(1):21–36. (cited in p. 17)

Hayek, F. A., 1945. The use of knowledge in society. *The American Economic Review*, 35(4): 519–530. (cited in p. 14, 27)

Hildreth, P. M., Kimble, C., and Others, 2002. The duality of knowledge. *Information Research*, 8(1):1–8. (cited in p. 61, 85)

Hill, B. M. Almost Wikipedia: What eight early online collaborative encyclopedia projects reveal about the mechanisms of collective action. 2012. URL `http://cyber.law.harvard.edu/events/luncheon/2011/10/makohill`. (cited in p. 61)

Hill, B. M. and Monroy-Hernández, A., 2012. The remixing dilemma: The trade-off between generativity and originality. *American Behavioral Scientist*. (cited in p. 107)

Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R., Scoville, R., and Taylor, R., 2005. Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Affairs*, 24(5):1103–1117. (cited in p. 17)

Hillman, J. M. and Given, R. S., 2005. Hospital implementation of computerized provider order entry systems: results from the 2003 leapfrog group quality and safety survey. *Journal of healthcare information management: JHIM*, 19(4):55. (cited in p. 17)

Hodge Jr, J. G., Gostin, L. O., and Jacobson, P. D., 1999. Legal issues concerning electronic health information. *JAMA: the journal of the American Medical Association*, 282(15): 1466–1471. (cited in p. 20)

Hof, R., June 2005. The power of us. *Business Week*, 3938:74–82. (cited in p. 13)

Holland, P. W. and Leinhardt, S., May 1971. Transitivity in Structural Models of Small Groups. *Small Group Research*, 2(2):107–124. ISSN 1046-4964. doi:

10.1177/104649647100200201. URL http://sgr.sagepub.com/cgi/doi/10.1177/104649647100200201. (cited in p. 126)

Holme, P. and Kim, B. J., 2002. Growing scale-free networks with tunable clustering. *Physical Review E*, 65(2):26107. (cited in p. 125)

Huang, A., 2008. Similarity measures for text document clustering. In *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand*, pages 49–56. (cited in p. 122)

Huffaker, D., 2010. Dimensions of Leadership and Social Influence in Online Communities. *Human Communication Research*, 36(4):593–617. URL http://dx.doi.org/10.1111/j.1468-2958.2010.01390.x. (cited in p. 53, 67, 115, 118)

Indivica, 2011. What You Need To Consider When You Are Getting Started. URL http://indivica.com/?page_id=118. (cited in p. 44)

Jackson, M. O. and Rogers, B. W., June 2007. Meeting Strangers and Friends of Friends: How Random Are Social Networks? *American Economic Review*, 97(3):890–915. ISSN 0002-8282. doi: 10.1257/aer.97.3.890. URL http://pubs.aeaweb.org/doi/abs/10.1257/aer.97.3.890. (cited in p. 114, 150)

Janjić, V. and Pržulj, N., 2012. Biological function through network topology: a survey of the human diseasome. *Briefings in Functional Genomics*. (cited in p. 60, 69, 127)

Jarvenpaa, S. L. and Leidner, D. E., 1997. An information company in Mexico: extending the resource-based view of the firm. In *Proceedings of the eighteenth international conference on Information systems*, pages 399–410. Association for Information Systems. (cited in p. 60)

Jeppesen, L. B. and Lakhani, K. R., October 2010. Marginality and Problem-Solving Effectiveness in Broadcast Search. *Organization Science*, 21(5):1016–1033. ISSN 1047-7039. doi: 10.1287/orsc.1090.0491. URL http://pubsonline.informs.org/doi/abs/10.1287/orsc.1090.0491. (cited in p. 5, 8, 9, 20, 27, 53, 56, 62, 63, 66, 68, 79, 105, 147)

Jha, A. K., DesRoches, C. M., Campbell, E. G., Donelan, K., Rao, S. R., Ferris, T. G., Shields, A., Rosenbaum, S., and Blumenthal, D., 2009. Use of electronic health records in US hospitals. *New England Journal of Medicine*, 360(16):1628–1638. (cited in p. 4, 18)

Jin, E. M., Girvan, M., and Newman, M. E. J., 2001. Structure of growing social networks. *Physical review E*, 64(4):46132. (cited in p. 114)

Johnson, S. L., 2010. Should I Stay or Should I Go? Continued Participation Intentions in Online Communities. In Toombs, L. A., editor, *Seventieth Annual Meeting of the Academy of Management*. (cited in p. 10, 27)

Johnson, S. L., Faraj, S., and Kudaravalli, S., 2014. Formation of Power Law Distributions in Online Communities. *MIS Quarterly*. (cited in p. 53, 63, 114, 136, 140, 150)

Johnson, S. L., Safadi, H., and Faraj, S., 2015. The Emergence of Online Community Leadership. *Information systems research*. (cited in p. 92, 105, 109, 118, 119)

Junni, P., Sarala, R. M., Taras, V., and Tarba, S. Y., July 2013. Organizational Ambidexterity and Performance: A Meta-Analysis. *Academy of Management Perspectives*, 27(4):299–312. ISSN 1558-9080. doi: 10.5465/amp.2012.0015. URL `http://amp.aom.org/cgi/doi/10.5465/amp.2012.0015`. (cited in p. 88)

Kairam, S., Wang, D., and Leskovec, J., 2012. The life and death of online groups: Predicting group growth and longevity. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, pages 673–682. ISBN 9781450307475. doi: http://dx.doi.org/10.1145/2124295.2124374. URL `http://dl.acm.org/citation.cfm?id=2124374`. (cited in p. 110)

Kaiser, M. and Hilgetag, C. C., 2004. Spatial growth of real-world networks. *Physical Review E*, 69(3):36103. (cited in p. 111, 113)

Kane, G. C. and Fichman, R. G., 2009. The shoemaker's children: Using wikis for information systems teaching, research, and publication. *Mis Quarterly*, 33(1):1–17. (cited in p. 90)

Kane, G. C., Alavi, M., Labianca, G. J., and Borgatti, S. P., 2014. WHAT'S DIFFERENT ABOUT SOCIAL MEDIA NETWORKS? A FRAMEWORK AND RESEARCH AGENDA. *MIS Quarterly*, 38(1). (cited in p. 53)

Kantor, G. S., Wilson, W. D., and Midgley, A., 2003. Open-source software and the primary care EMR. *Journal of the American Medical Informatics Association*, 10(6):616. (cited in p. 18, 21, 22)

Kaplan, B. and Harris-Salamone, K. D., 2009. Health IT success and failure: recommendations from literature and an AMIA workshop. *Journal of the American Medical Informatics Association*, 16(3):291–299. (cited in p. 20)

Kapur, T., Pieper, S., Whitaker, R., Aylward, S., Jakab, M., Schroeder, W., and Kikinis, R., 2012. The National Alliance for Medical Image Computing, a roadmap initiative to build a free and open source software infrastructure for translational research in medical image analysis. *Journal of the American Medical Informatics Association*, 19(2):176–180. (cited in p. 22)

Kempe, D., Kleinberg, J., and Tardos, E., 2003. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM. (cited in p. 79)

Kilduff, M. and Brass, D. J., 2010. Organizational social network research: Core ideas and key debates. *The Academy of Management Annals*, 4(1):317–357. (cited in p. 7)

Kittur, A. and Kraut, R. E., 2008. Harnessing the wisdom of crowds in wikipedia: quality through coordination. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 37–46. ACM. (cited in p. 57)

Klemm, K. and Eguiluz, V. M., 2002. Growing scale-free networks with small-world behavior. *Physical Review E*, 65(5):57102. (cited in p. 114)

Knoke, D. and Yang, S., 2008. *Social network analysis*. Sage Publications, Inc. ISBN 1412927498. (cited in p. 34)

Köbler, J., Schöning, U., and Torán, J., 1994. *The graph isomorphism problem: its structural complexity*. Birkhauser Verlag. (cited in p. 126)

Kogut, B. and Zander, U., 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization science*, 3(3):383–397. (cited in p. 60)

Kossinets, G. and Watts, D. J., January 2006. Empirical analysis of an evolving social network. *Science (New York, N.Y.)*, 311(5757):88–90. ISSN 1095-9203. doi: 10.1126/science.1116869. URL `http://www.ncbi.nlm.nih.gov/pubmed/16400149`. (cited in p. 113, 114)

Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A., and Boguñá, M., 2010. Hyperbolic geometry of complex networks. *Physical Review E*, 82(3):36106. (cited in p. 125)

Krishnamurthy, S., 2002. Cave or community?: An empirical examination of 100 mature open source projects. *First Monday*, 7(6). (cited in p. 58, 64, 81)

Kudaravalli, S. and Faraj, S., 2008. The structure of collaboration in electronic networks. *Journal of the Association for Information Systems*, 9(10/11):706–726. (cited in p. 5, 110, 112, 147)

Kurbel, K. E., 2008. *The making of information systems: software engineering and management in a globalized world*. Springer. ISBN 3540792600. (cited in p. 16)

Kush, R. D., Helton, E., Rockhold, F. W., and Hardison, C. D., 2008. Electronic health records, medical research, and the Tower of Babel. *New England Journal of Medicine*, 358(16): 1738–1740. (cited in p. 20)

Lakhani, K. and Wolf, R. 2005, *Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects*, chapter 1. MIT Press. (cited in p. 1, 2, 7, 14, 15, 27, 110)

Lakhani, K. R., 2006. Broadcast Search in Problem Solving: Attracting Solutions from the Periphery. In *Technology Management for the Global Future, 2006. PICMET 2006*, volume 6, pages 2450–2468. IEEE. (cited in p. 61)

Lapointe, L. and Rivard, S., 2005. A multilevel model of resistance to information technology implementation. *Mis Quarterly*, 29(3):461–491. (cited in p. 23)

Laurent, A. M. S., 2004. *Understanding open source and free software licensing*. O'Reilly Media. (cited in p. 15)

LeapfrogGroup, 2004. Leapfrog Group Hospital Quality and Safety Survey. URL `http://www.leapfroggroup.org/media/file/Leapfrog-Survey_Release-11-16-04.pdf`. (cited in p. 17)

Lebart, L., Salem, A., and Berry, L., 1998. *Exploring textual data*. Kluwer Dordrecht. (cited in p. 34)

Leonardi, P. M., 2007. Activating the informational capabilities of information technology for organizational change. *Organization Science*, 18(5):813–831. (cited in p. 16)

Lerner, J. and Schankerman, M., 2010. *The comingled code: Open source and economic development*. MIT press. (cited in p. 7, 15, 27)

Lerner, J. and Tirole, J., 2002. Some simple economics of open source. *The journal of industrial economics*, 50(2):197–234. (cited in p. 15, 21)

Levina, N. and Arriaga, M., 2015. Distinction and Status Production on User-Generated Content Platforms: Using Bourdieu's Theory of Cultural Production to Understand Social Dynamics in Online Fields. *Information Systems Research*, page Forthcoming. (cited in p. 8, 12, 53, 58, 87, 91, 92, 110, 115, 116, 153)

Levina, N. and Vaast, E., 2005. The emergence of boundary spanning competence in practice: implications for implementation and use of information systems. *MIS Quarterly*, 29(2): 335–363. (cited in p. 68, 115)

Liben-Nowell, D. and Kleinberg, J., 2007. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7):1019–1031. ISSN 1532-2890. doi: 10.1002/asi.20591. URL http://dx.doi.org/10.1002/asi.20591. (cited in p. 113)

Lim, M., Griffiths, G., and Sambrook, S., 2010. Organizational structure for the twenty-first century. In *annual meeting of The Institute for Operations Research and The Management Sciences, Austin*. (cited in p. 87, 90, 105)

Lincke, R., Lundberg, J., and Löwe, W., 2008. Comparing software metrics tools. In *Proceedings of the 2008 international symposium on Software testing and analysis*, pages 131–142. ACM. (cited in p. 31)

Liu, C. T., Yang, P. T., Yeh, Y. T., and Wang, B. L., 2006. The impacts of smart cards on hospital information systems–an investigation of the first phase of the national health insurance smart card project in Taiwan. *International journal of medical informatics*, 75(2):173–181. (cited in p. 19)

Loewenstein, J., 2014. Take my word for it: How professional vocabularies foster organizing. *Journal of Professions and Organization*. doi: 10.1093/jpo/jot004. URL http://jpo.oxfordjournals.org/content/early/2014/01/06/jpo.jot004.abstract. (cited in p. 115, 140)

Long, J. S. and Freese, J., 2006. *Regression models for categorical dependent variables using Stata*. Stata press. (cited in p. 101)

Luhmann, N., Bednarz, J., Baecker, D., and Knodt, E. M., 1995. *Social systems*, volume 1. Stanford University Press Stanford. (cited in p. 92)

Ma, M. and Agarwal, R., 2007. Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Information systems research*, 18(1):42. (cited in p. 9, 62, 107)

Majchrzak, A. and Markus, M. L., 2015. Technology affordances and constraints in management information systems (MIS). *Encyclopedia of Management Theory*. URL http://ssrn.com/abstract=2192196. (cited in p. 116)

March, J. G., 1991. Exploration and exploitation in organizational learning. *Organization science*, 2(1):71–87. (cited in p. 88, 91)

March, J. G. and Simon, H. A., 1958. Organizations. (cited in p. 57)

Mashima, R. and Takahashi, N. 2008, The emergence of generalized exchange by indirect reciprocity. In *New Issues and Paradigms in Research on Social Dilemmas*, pages 159–176. Springer. (cited in p. 112)

Mata, F. J., Fuerst, W. L., and Barney, J. B., 1995. Information technology and sustained competitive advantage: A resource-based analysis. *MIS quarterly*, pages 487–505. (cited in p. 60)

McFadyen, M., Semadeni, M., and Cannella, A. A., 2009. Value of strong ties to disconnected others: Examining knowledge creation in biomedicine. *Organization science*, 20(3): 552–564. (cited in p. 8, 54, 62)

McPherson, M., Smith-Lovin, L., and Cook, J. M., 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444. (cited in p. 54, 114, 147)

Mehra, A. and Mookerjee, V., 2012. Human capital development for programmers using open source software. *MIS quarterly*, 36(1):107–122. (cited in p. 149)

Mehra, A., Dewan, R., and Freimer, M., 2011. Firms as Incubators of Open-Source Software. *Information Systems Research*, 22(1):22–38. (cited in p. 15)

Milenković, T. and Przulj, N., January 2008. Uncovering biological network function via graphlet degree signatures. *Cancer informatics*, 6:257–73. ISSN 1176-9351. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2623288&tool=pmcentrez&rendertype=abstract`. (cited in p. 81, 127, 150)

Miller, D. and Friesen, P. H., 1982. Structural change and performance: Quantum versus piecemeal-incremental approaches. *The Academy of Management Journal*, 25(4):867–892. (cited in p. 10)

Miller, J. H. and Page, S. E., 2007. *Complex adaptive systems: An introduction to computational models of social life*. Princeton Univ Pr. ISBN 0691127026. (cited in p. 19)

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U., 2002. Network motifs: simple building blocks of complex networks. *Science Signalling*, 298(5594):824. (cited in p. 60, 69, 126)

Mintzberg, H., 2002. Managing care and cure-up and down, in and out. *Health Services Management Research*, 15(3):193–206. (cited in p. 19, 23)

Mockus, A., Fielding, R. T., and Herbsleb, J., 2000. A case study of open source software development: the Apache server. In *Software Engineering, 2000. Proceedings of the 2000 International Conference on*, pages 263–272. IEEE. (cited in p. 15)

Monge, P. R. and Contractor, N. S., 2003. *Theories of Communication Networks*. Oxford University Press. (cited in p. 34)

Moon, J. Y. and Sproull, L. S., 2008. The role of feedback in managing the Internet-based volunteer work force. *Information systems research*, 19(4):494–515. (cited in p. 10, 63)

Natter, M. D., Quan, J., Ortiz, D. M., Bousvaros, A., Ilowite, N. T., Inman, C. J., Marsolo, K., McMurry, A. J., Sandborg, C. I., Schanberg, L. E., and Others, 2013. An i2b2-based, generalizable, open source, self-scaling chronic disease registry. *Journal of the American Medical Informatics Association*, 20(1):172–179. (cited in p. 22)

Nelson, R. E., 1989. The strength of strong ties: Social networks and intergroup conflict in organizations. *Academy of Management Journal*, 32(2):377–401. (cited in p. 54)

Newman, M. E. J., 2001a. Clustering and preferential attachment in growing networks. *Physical Review E*, 64(2):25102. (cited in p. 114)

Newman, M. E. J., 2001b. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical review E*, 64(1):16132. (cited in p. xiii, 72, 73, 120)

Newman, M. E. J., Watts, D. J., and Strogatz, S. H., 2002. Random graph models of social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 1):2566–2572. (cited in p. xiii, 57, 72, 120)

Nonaka, I. and Takeuchi, H., 1995. *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford University Press, USA. (cited in p. 61, 62, 86)

Nowak, M. A., 2006. Five rules for the evolution of cooperation. *science*, 314(5805): 1560–1563. (cited in p. 112)

Nowak, M. A. and Roch, S., 2007. Upstream reciprocity and the evolution of gratitude. *Proceedings of the Royal Society B: Biological Sciences*, 274(1610):605–610. (cited in p. 112)

OCUS, 2012a. OSCAR Canada User Society - About OSCAR. URL `http://www.oscarcanada.org/about-oscar/brief-overview/`. (cited in p. 43)

OCUS, 2012b. OSCAR Canada User Society - eForm Generator. URL `http://www.oscarcanada.org/oscar-users/emr-resource/eform/eform-building-resources-1/eform-generator-v4.0`. (cited in p. 45)

OECD, O. f. E. C.-O. 2011, *Society at a Glance 2011: OECD Social Indicators*, chapter 7. Publications de l'OCDE. ISBN 9264098526. (cited in p. 2)

Oestreicher-Singer, G. and Sundararajan, A., 2012. Recommendation Networks and the Long Tail of Electronic Commerce. *MIS Quarterly*, 36(1). (cited in p. 53)

Oestreicher-Singer, G. and Zalmanson, L., 2013. Content or community? A digital business strategy for content providers in the social age. *MIS Quarterly*, 37(2):591–616. (cited in p. 153)

O'Leary, M. B. and Cummings, J. N., 2007. The spatial, temporal, and configurational characteristics of geographic dispersion in teams. *Mis Quarterly*, 31(3):433–452. (cited in p. 14, 25)

Oliver, P. E. and Marwell, G., 2001. Whatever happened to critical mass theory? A retrospective and assessment. *Sociological Theory*, 19(3):292–311. (cited in p. 12)

Olivera, F., Goodman, P. S., and Tan, S. S., 2008. Contribution behaviors in distributed environments. *Management Information Systems Quarterly*, 32(1):23. (cited in p. 9)

Olson, M., 1965. *The logic of collective action: public goods and the theory of groups*, volume 124. Harvard University Press. ISBN 0674537513. (cited in p. 12)

O'Mahony, S. and Ferraro, F., 2007. The emergence of governance in an open source community. *The Academy of Management Journal*, 50(5):1079–1106. (cited in p. 2, 5, 8, 10, 25, 53, 81, 85, 87, 88, 91, 105, 109, 110, 140, 147)

O'Mahony, S. and Bechky, B. A., 2008. Boundary organizations: Enabling collaboration among unexpected allies. *Administrative Science Quarterly*, 53(3):422–459. (cited in p. 5, 89, 147)

OntarioMD, 2012a. EMR Offerings. URL `https://www.ontariomd.ca/portal/server.pt/community/emr_offerings/`. (cited in p. 37, 46)

OntarioMD, 2012b. Ontario EMR Specification. URL `https://www.ontariomd.ca/portal/server.pt/community/emr_certification/228/emr_specification_3_0/334`. (cited in p. 23, 46)

O'Reilly, T., DiBona, C., Cooper, D., and Stone, M., 2005. The open source paradigm shift. *Open sources*, 2:253–271. (cited in p. 14)

Orlikowski, W. J., 1996. Improvising organizational transformation over time: A situated change perspective. *Information systems research*, 7:63–92. (cited in p. 10, 88)

OSCAR, 2012. Oscar community web site. URL `http://www.oscarcanada.org/`. (cited in p. 5)

OSCARCanada, 2012a. About OSCAR. URL http://www.oscarcanada.org/about-oscar/brief-overview/index_html. (cited in p. 37)

OSCARCanada, 2012b. Brief Overview: Everything You Wanted to Know About OSCAR, But Were Afraid to Ask. URL http://www.oscarcanada.org/about-oscar/brief-overview. (cited in p. 46)

OSCARCanada, 2012c. Drugref. URL http://www.oscarcanada.org/oscar-users/drugref/index_html. (cited in p. 41, 47)

OSCARCanada, 2012d. OSCAR Support. URL http://www.oscarcanada.org/support. (cited in p. 40, 43, 44)

OSCARCanada, 2012e. OSCAR Users around the World. URL http://www.oscarcanada.org/usersgroups/oscar-users-around-the-world. (cited in p. 37)

OSCARManual, 2012a. About Open Source. URL http://www.oscarmanual.org/introduction/about-open-source. (cited in p. 44)

OSCARManual, 2012b. eForms. URL http://new.oscarmanual.org/oscar_emr_12/clinical-functions/e-forms. (cited in p. 46)

OSCARManual, 2012c. OSCAR EMR Users Manual. URL http://new.oscarmanual.org/introduction. (cited in p. 45)

Page, S. E., 2001. Self Organization and Coordination. *Computational Economics*, 18(1):25–48. (cited in p. 54, 87, 91, 147)

Papadopoulos, F., Kitsak, M., Serrano, M. A., Boguñá, M., and Krioukov, D., 2012. Popularity versus similarity in growing networks. *Nature*, 489(7417):537–540. (cited in p. 114, 125)

Phang, C. W., Kankanhalli, A., and Sabherwal, R., 2009. Usability and sociability in online communities: A comparative study of knowledge seeking and contribution. *Journal of the Association for Information Systems*, 10(10):721–747. (cited in p. 8, 62, 107)

Pinsonneault, A. and Caya, O., 2005. Virtual teams: What we know, what we don't know. *International Journal of e-Collaboration (IJeC)*, 1(3):1–16. (cited in p. 14)

Pinsonneault, A., Barki, H., Gallupe, R. B., and Hoppen, N., 1999. Research Note. The Illusion of Electronic Brainstorming Productivity: Theoretical and Empirical Issues. *Information Systems Research*, 10(4):378. (cited in p. 88)

Preece, J. and Shneiderman, B., 2009. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, 1(1):13–32. (cited in p. 25)

Pržulj, N., Corneil, D. G., and Jurisica, I., 2004. Modeling interactome: scale-free or geometric? *Bioinformatics*, 20(18):3508–3515. (cited in p. xiv, 60, 80, 126, 127)

Przulj, N., January 2007. Biological network comparison using graphlet degree distribution. *Bioinformatics (Oxford, England)*, 23(2):e177–83. ISSN 1367-4811. doi: 10.1093/bioinformatics/btl301. URL `http://www.ncbi.nlm.nih.gov/pubmed/17237089`. (cited in p. 80, 81)

Ransbotham, S., Kane, G. C., and Lurie, N. H., 2012. Network characteristics and the value of collaborative user-generated content. *Marketing Science*, 31(3):387–405. (cited in p. 27, 62)

Raymond, E., 1999. The cathedral and the bazaar. *Knowledge, Technology & Policy*, 12(3): 23–49. (cited in p. 21, 27, 58)

Rayna, T., 2008. Understanding the challenges of the digital economy: The nature of digital goods. *Communications and Strategies*, 71(3):13–36. (cited in p. 90)

Reed, J. A., 2013. Visual Gallery. URL `http://mumble.net/~jar/visuals/fractal.html`. (cited in p. 65)

Rehurek, R. and Sojka, P., 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks*, pages 46–50, Valletta, Malta. University of Malta. ISBN 2-9517408-6-7. URL `http://www.fi.muni.cz/usr/sojka/presentations/lrec2010-poster-rehurek-sojka.pdf`. (cited in p. 100)

Ren, Y. and Argote, L., 2011. Transactive memory systems 1985–2010: An integrative framework of key dimensions, antecedents, and consequences. *The Academy of Management Annals*, 5(1):189–229. (cited in p. 51)

Ren, Y. and Kraut, R. E., April 2014. Agent-Based Modeling to Inform Online Community Design: Impact of Topical Breadth, Message Volume, and Discussion Moderation on Member Commitment and Contribution. *Human-Computer Interaction*, 29(4):351–389. ISSN 0737-0024. doi: 10.1080/07370024.2013.828565. URL `http://www.tandfonline.com/doi/abs/10.1080/07370024.2013.828565`. (cited in p. 87, 93)

Reynolds, C. J. and Wyatt, J. C., 2011. Open source, open standards, and health care information systems. *Journal of Medical Internet Research*, 13(1). (cited in p. 13, 18, 21)

Rice, T. H. and Unruh, L. 1998, *The economics of health reconsidered*, chapter 1. Health Administration Press Chicago. ISBN 1567930735. (cited in p. 13)

Ridings, C. and Wasko, M. M. L., 2010. Online discussion group sustainability: Investigating the interplay between structural dynamics and social dynamics over time. *Journal of the Association for Information Systems*, 11(2):1. (cited in p. 5, 10, 53, 110, 147)

Roberts, J. A., Hann, I. H., and Slaughter, S. A., 2006. Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the Apache projects. *Management science*, pages 984–999. (cited in p. 1, 15, 27)

Robertson, A., Cresswell, K., Takian, A., Petrakaki, D., Crowe, S., Cornford, T., Barber, N., Avery, A., Fernando, B., and Jacklin, A., 2010. Implementation and adoption of nationwide electronic health records in secondary care in England: qualitative analysis of interim results from a prospective national evaluation. *BMJ: British Medical Journal*, 341. (cited in p. 17, 19)

Rocchini, C., 2012. Examples of centrality. URL `http://en.wikipedia.org/wiki/File:Centrality.svg`. (cited in p. xiii, 67)

Romanow, D., Sunyoung, C., and Straub, D., 2012. Riding the Wave: Past Trends and Future Directions for Health IT Research. *MIS quarterly*, 36(3):iii–A18. (cited in p. 18)

Romanow, R. J., 2002. Building on Values [electronic Resource]: the Future of Health Care in Canada. URL `http://www.hc-sc.gc.ca/hcs-sss/hhr-rhs/strateg/romanow-eng.php`. (cited in p. 2)

Rosen-Zvi, M., Chemudugunta, C., Griffiths, T., Smyth, P., and Steyvers, M., 2010. Learning author-topic models from text corpora. *ACM Transactions on Information Systems (TOIS)*, 28(1):4. (cited in p. 128, 129)

Rothschild, M. and Stiglitz, J., 1976. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, pages 629–649. (cited in p. 13)

Rulke, D., Zaheer, S., and Anderson, M., 1998. Transactive knowledge and performance in the retail food industry. In *Paper delivered at the Stern School of Business conference on Managerial and Organizational Cognition, New York City. New York, NY: New York University, Leonard N. Stern School of Business*. (cited in p. 61, 86)

Sabherwal, R. and Becerra-Fernandez, I., 2005. Integrating specific knowledge: insights from the Kennedy Space Center. *Engineering Management, IEEE Transactions on*, 52(3): 301–315. (cited in p. 60)

Safadi, H. and Faraj, S., 2010. The Role of Workarounds during an Opensource Electronic Medical Record System Implementation. In *International Conference of Information Systems (ICIS)*. (cited in p. 19)

Safadi, H. and Faraj, S., 2011. IT in Healthcare: an Integrative Study of Organizational Change. In *AMCIS 2011 Proceedings*, page 21. AIS. URL `http://aisel.aisnet.org/amcis2011_submissions/21`. (cited in p. 4, 146)

Schaefer, D. R., Light, J. M., Fabes, R. A., Hanish, L. D., and Martin, C. L., 2010. Fundamental principles of network formation among preschool children. *Social Networks*, 32(1):61–71. (cited in p. 114)

Schultze, U., December 2012. Performing embodied identity in virtual worlds. *European Journal of Information Systems*, 23(1):84–95. ISSN 0960-085X. doi: 10.1057/ejis.2012.52. URL `http://www.palgrave-journals.com/doifinder/10.1057/ejis.2012.52`. (cited in p. 112)

Scott, J. and Carrington, P. J., 2011. *The SAGE handbook of social network analysis*. SAGE publications. (cited in p. 52)

Sedgewick, R., 2001. *Algorithms in C, Part 5: Graph Algorithms, Third Edition*. Addison-Wesley Professional, third edition. ISBN 9780768685329. (cited in p. 69)

Seidman, S. B., 1983. Network structure and minimum degree. *Social networks*, 5(3): 269–287. (cited in p. 69)

Setia, P., Rajagopalan, B., Sambamurthy, V., and Calantone, R., 2012. How peripheral developers contribute to open-source software development. *Information Systems Research*, 23(1):144–163. (cited in p. 148)

Shirky, C., 2010. *Cognitive surplus: Creativity and generosity in a connected age*. ePenguin. ISBN 1846142172. (cited in p. 12)

Singer, J. D. and Willett, J. B., 2003. *Applied longitudinal data analysis: modeling change and event occurrence: modeling change and event occurrence*. Oxford University Press, USA. (cited in p. 34)

Singh, P. V. and Phelps, C., 2013. Networks, Social Influence, and the Choice Among Competing Innovations: Insights from Open Source Software Licenses. *Information Systems Research*, 24(3):539–560. doi: 10.1287/isre.1120.0449. URL `http://pubsonline.informs.org/doi/abs/10.1287/isre.1120.0449`. (cited in p. 15)

Singh, P. V., Tan, Y., and Mookerjee, V. S., 2011a. Network Effects: The Influence of Structural Capital on Open Source Project Success. *MIS Quarterly*, 35(4):813–829. (cited in p. 56)

Singh, P. V., Tan, Y., and Youn, N., 2011b. A hidden Markov model of developer learning dynamics in open source software projects. *Information Systems Research*, 22(4):790–807. (cited in p. 149)

Smirnov, N., 1948. Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, 19(2):279–281. (cited in p. 127)

Snijders, T. A. B., 2001. The Statistical Evaluation of Social Network Dynamics. *Sociological Methodology*, 31(1):361–395. ISSN 1467-9531. doi: 10.1111/0081-1750.00099. URL `http://dx.doi.org/10.1111/0081-1750.00099`. (cited in p. 111)

Sommerville, I., 2007. *Software Engineering*. Addison Wesley. (cited in p. 15, 61)

South, S. J., 1995. Do you need to shop around? Age at marriage, spousal alternatives, and marital dissolution. *Journal of Family Issues*, 16(4):432–449. (cited in p. 34)

Sparrowe, R. T., Liden, R. C., Wayne, S. J., and Kraimer, M. L., 2001. Social networks and the performance of individuals and groups. *Academy of management journal*, 44(2):316–325. (cited in p. 52)

Spencer, R. W., 2011. Consistent Patterns in Large-Scale Collaboration. *Collaborative Computational Technologies for Biomedical Research*, pages 99–111. (cited in p. 61)

Sproull, L. and Arriaga, M., 2007. Online communities. *Handbook of Computer Networks*, pages 898–914. (cited in p. 1, 109)

Sproull, L., Conley, C., and Moon, J. Y., 2005. Prosocial behavior on the net. *The social net: The social psychology of the Internet*, pages 139–161. (cited in p. 7, 8, 61, 86)

Sujansky, W. V., 1998. The benefits and challenges of an electronic medical record: much more than a" word-processed" patient chart. *Western Journal of Medicine*, 169(3):176. (cited in p. 20)

Sundararajan, A., Provost, F., Oestreicher-Singer, G., and Aral, S., 2013. Research Commentary–Information in Digital, Economic, and Social Networks. *Information Systems Research*, 24(4):883–905. doi: 10.1287/isre.1120.0472. URL `http://pubsonline.informs.org/doi/abs/10.1287/isre.1120.0472`. (cited in p. 56, 79, 115)

Surowiecki, J., Silverman, M. P., and Others, 2007. The wisdom of crowds. *American Journal of Physics*, 75(2):190–192. (cited in p. 57)

Sutanto, J., Tan, C.-H., Battistini, B., and Phang, C. W., 2011. Emergent Leadership in Virtual Collaboration Settings: A Social Network Analysis Approach. *Long Range Planning*, 44 (5-6):421–439. (cited in p. 67)

Sutton, R. I. and Hargadon, A., 1996. Brainstorming groups in context: Effectiveness in a product design firm. *Administrative Science Quarterly*, 41(4). (cited in p. 88, 89)

Tapscott, D. and Williams, A. D., 2008. *Wikinomics: How Mass Collaboration Changes Everything*. Portfolio Trade. ISBN 1591841933. (cited in p. 1, 2, 20, 25, 53, 61, 109)

Taylor, J. R. and Van Every, E. J., 1999. *The emergent organization: Communication as its site and surface*. Routledge. (cited in p. 115)

Tiwana, A. and Mclean, E. R., 2003. Expertise integration and creativity in information systems development. *Journal of Management Information Systems*, 22(1):13–43. (cited in p. 60)

Tosi, H. L., 2008. *Theories of organization*. Sage. (cited in p. 57)

Treem, J. and Leonardi, P., 2012. Social media use in organizations: Exploring the affordances of visibility, editability, persistence, and association. *Communication yearbook*, 36:143–189. (cited in p. 116, 140)

Tsoukas, H., 2009. A dialogical approach to the creation of new knowledge in organizations. *Organization Science*, 20(6):941–957. (cited in p. 91, 147)

Tushman, M., Lakhani, K. R., and Lifshitz-Assaf, H., 2012. Open Innovation and Organizational Design. *Journal of Organization Design*, 1(1). doi: dx.doi.org/10.7146/jod.1.1.6336. (cited in p. 1, 2, 14, 20)

Tushman, M. L. and O'Reilly III, C. A., 1996. Ambidextrous organizations: managing evolutionary and revolutionary change. *California management review*, 38(4). (cited in p. 88)

Valimaki, M., 2003. Dual licensing in open source software industry. *Systemes d'Information et Management*, 8(1):63–75. (cited in p. 15)

Vázquez, A., 2003. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, 67(5):56104. (cited in p. 114)

Vehovar, V., Lozar Manfreda, K., Koren, G., and Hlebec, V., July 2008. Measuring ego-centered social networks on the web: Questionnaire design issues. *Social Networks*, 30(3):213–222. ISSN 03788733. doi: 10.1016/j.socnet.2008.03.002. URL `http://linkinghub.elsevier.com/retrieve/pii/S0378873308000142`. (cited in p. 69)

Venkatraman, N. and Lee, C.-H., 2004. Preferential Linkage and Network Evolution: A Conceptual Model and Empirical Test in the U.S. Video Game Sector. *The Academy of Management Journal*, 47(6):876–892. URL `http://www.jstor.org/stable/20159628`. (cited in p. 12)

Viega, J., Warsaw, B., and Manheimer, K., 1998. Mailman: The GNU Mailing List Manager. In *LISA*, pages 309–316. (cited in p. 117)

Von Bertalanffy, L., 1950. The theory of open systems in physics and biology. *Science*, 111 (2872):23–29. (cited in p. 92)

Von Hippel, E., 2005. *Democratizing innovation*. the MIT Press. ISBN 0262002744. (cited in p. 1, 3, 13, 20)

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. (cited in p. 1, 16, 49)

Von Krogh, G., 2012. How does social software change knowledge management? Toward a strategic research agenda. *The Journal of Strategic Information Systems*, 21(2):154–164. (cited in p. 1, 14)

Von Krogh, G., Spaeth, S., and Lakhani, K. R., 2003. Community, joining, and specialization in open source software innovation: a case study. *Research Policy*, 32(7):1217–1241. (cited in p. 1, 16, 17, 63, 106, 112)

Von Krogh, G., Ichijo, K., and Nonaka, I., 2000. *Enabling Knowledge Creation: How to Unlock the Mystery of Tacit Knowledge and Release the Power of Innovation: How to Unlock the Mystery of Tacit Knowledge and Release the Power of Innovation*. Oxford University Press, USA. (cited in p. 62, 86)

Walker, J., Pan, E., Johnston, D., Adler-Milstein, J., Bates, D. W., and Middleton, B., 2005. The value of health care information exchange and interoperability. *Health Affairs*, 24:5. (cited in p. 2, 17)

Wang, S. J., Middleton, B., Prosser, L. A., Bardon, C. G., Spurr, C. D., Carchidi, P. J., Kittler, A. F., Goldszer, R. C., Fairchild, D. G., and Sussman, A. J., 2003. A cost-benefit analysis of electronic medical records in primary care. *The American journal of medicine*, 114(5): 397–403. (cited in p. 19)

Warde, A. and Tampubolon, G., 2002. Social capital, networks and leisure consumption. *The Sociological Review*, 50(2):155–180. (cited in p. 81)

Warmbrodt, J., Sheng, H., and Hall, R., 2008. Social Network Analysis of Video Bloggers' Community. In *41st Annual Hawaii International Conference on System Sciences*, page 291. IEEE Computer Society. ISBN 0769530753. (cited in p. 57, 68)

Wasko, M. and Faraj, S., 2000. It is what one does?: why people participate and help othersin electronic communities of practice. *The Journal of Strategic Information Systems*, 9(2):155–173. (cited in p. 8, 61, 92)

Wasko, M. and Faraj, S., 2005. Why Should I Share: Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice. *Management Information Systems Quarterly*, 29(1):35–47. (cited in p. 7, 8, 21, 25, 53, 56, 61, 63, 79, 81, 86, 90, 106, 110)

Wasko, M., Teigland, R., and Faraj, S., 2009. The provision of online public goods: Examining social structure in an electronic network of practice. *Decision Support Systems*, 47(3): 254–265. (cited in p. 10, 12)

Wasserman, S., 1994. *Social network analysis: Methods and applications*, volume 8. Cambridge university press. (cited in p. 52)

Watts, D. J. and Strogatz, S. H., 1998. Collective dynamics of'small-world'networks. *Nature*, 393(6684):440–442. (cited in p. 57, 126)

Wegner, D. M. 1987, Transactive memory: A contemporary analysis of the group mind. In *Theories of group behavior*, pages 185–208. Springer. (cited in p. 52)

Wegner, D. M., Giuliano, T., and Hertel, P. T. 1985, Cognitive interdependence in close relationships. In *Compatible and incompatible relationships*, pages 253–276. Springer. (cited in p. 51, 52)

Wen, W., Forman, C., and Graham, S. J. H., 2013. Research Note-The Impact of Intellectual Property Rights Enforcement on Open Source Software Project Success. *Information Systems Research*, 24(4):1131–1146. (cited in p. 15)

West, J. and Lakhani, K. R., 2008. Getting clear about communities in open innovation. *Industry and Innovation*, 15(2):223–231. (cited in p. 2, 14)

Willer, D., Walker, H. A., Markovsky, B., Willer, R., Lovaglia, M., Thye, S., and Simpson, B., 2002. Network Exchange Theory. *New directions in contemporary sociological theory*, pages 109–144. (cited in p. 112)

Yoo, Y. and Alavi, M., 2004. Emergent leadership in virtual teams: what do emergent leaders do? *Information and Organization*, 14(1):27–58. (cited in p. 34, 115, 118)

Young, I. T., 1977. Proof without prejudice: use of the Kolmogorov-Smirnov test for the analysis of histograms from flow systems and other sources. *Journal of Histochemistry & Cytochemistry*, 25(7):935–941. (cited in p. 128, 133)

Young, L. and Fitzgerald, B., 2006. *The power of language: How discourse influences society*. Equinox Pub. (cited in p. 11, 149)

Zhang, C., Hahn, J., and De, P., 2013. Research NoteâĂŤContinued Participation in Online Innovation Communities: Does Community Response Matter Equally for Everyone? *Information Systems Research*, 24(4):1112–1130. doi: 10.1287/isre.2013.0485. URL `http://dx.doi.org/10.1287/isre.2013.0485`. (cited in p. 147)

Zhu, H., Kraut, R., and Kittur, A., 2012. Effectiveness of shared leadership in online communities. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 407–416. ACM. (cited in p. 115, 118)