DISSERTATION

MODELING SUSTAINABILITY OF PARTICIPATORY INFORMATION SYSTEMS FOR URBAN COMMUNITIES: A MIXED-METHOD APPROACH

by

Claudia López

Submitted to the Graduate Faculty of

the School of Information Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

SCHOOL OF INFORMATION SCIENCES

This dissertation was presented

by

Claudia López

It was defended on

November 24th, 2015

and approved by

Rosta Farzan, PhD, Assistant Professor

Yu-Ru Lin, PhD, Assistant Professor

Peter Brusilovsky, PhD, Professor

Brian Butler, PhD, Professor, University of Maryland

Dissertation Director: Rosta Farzan, PhD, Assistant Professor

DISSERTATION

MODELING SUSTAINABILITY OF PARTICIPATORY INFORMATION SYSTEMS FOR URBAN COMMUNITIES: A MIXED-METHOD APPROACH

Claudia López,

University of Pittsburgh, 2015

In the face of the growing challenge of low civic participation, various technology-based endeavors, such as hyper-local social media and open government initiatives, have emerged to facilitate citizens' involvement with their local communities. However, evidence-based guidelines about how to start and maintain viable information systems for urban communities are scarce and inconclusive. My dissertation aims to identify factors that affect the sustainability of these systems by conducting a mixed-method empirical investigation of the E-Democracy platform, a pioneer platforms of its kind. With this goal, I proposed a framework to model the influence of offline characteristics of the target urban communities, system design, and residents' online behaviors on the sustainability of participatory information systems for urban communities. Guided by this framework, I conducted analyses of urban communities' public data; longitudinal studies, content classification and social network analysis of archival data of the system; and a cross-sectional study of user surveys. The results indicate that (1) certain community characteristics, such as neighborhood instability, play a crucial role in the sustainability of these information systems, (2) both on-site and off-site communication among residents is key for the systems to have an impact on community involvement, and (3) particular design decisions can foster participation of underrepresented populations. My work bridges the gap between research on social computing and community informatics by providing a framework to analyze the sustainability of participatory information systems for urban communities. The findings have implications for information experts and technology developers seeking to study or design technologies for local communities.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Working to earn this degree was a fulfilling and rewarding experience. Many people have been influential in shaping up my life in ways that led to this goal. This is a small tribute to them.

I deeply thank my parents, Berta and Antonio, who have been a constant source of encouragement throughout my life and who strove to give me all possible opportunities to find my own way. I also want to thank my siblings, Roxanna, Jasna and Jorge, for advising me about everything I ever needed. Their love and support made me confident to pursue this goal.

I owe a gigantic amount of gratitude to Gonzalo, who shared with me this life journey and was always supportive of all the projects I had in mind. His company made this experience very joyful.

I am profoundly grateful of my academic mentors whom I admire greatly. I thank Hernan Astudillo for "catalyzing" my career as a researcher. I thank Yu-Ru Lin for teaching me how cool data mining could be. I thank Peter Brusilovsky for being an awesome academic "grandfather." I thank Brian Butler for tons of research and professional advice. Special thanks go to Rosta Farzan, my advisor and friend, who believed in my research ideas and helped me enormously to make them become a doctoral dissertation. Their guidance was fundamental to obtaining this degree.

I wish to express my gratitude to PAWS Lab members and associates. I was privileged to share my time in Pittsburgh with Sharon, Jennifer, Sherry, Jessica, Di, Yun, Roya, Denis, Julio, Xidao, Amir, and Michael. I thank you for our lunch and coffee/tea times, research discussions and talk rehearsals, parties and trips, yoga and a relay marathon, feminism and WIS, and so many other cool things. I also thank the school's staff for being so helpful with me, and my friends Daniel and Antonio for being always there. It was great to undergo this unique life experience with them.

I am hugely indebted to E-Democracy for partnering with us to make this research project possible. I also wish to acknowledge the financial support of Becas Chile, Fulbright, the School of Information Sciences, and Departmento de Informática, UTFSM.

1.0 INTRODUCTION

The Internet has greatly facilitated communication between people around the world, which has positively impacted on widely-valued social goals. Studies of well-known participatory information systems with global reach such as Usenet, Wikipedia, Facebook, Twitter, and Open Source Software have shown that Internet-mediated communication facilitates rapid information sharing [68] and coordination [79], supports collective action [55], and can eventually lead to the development of civic engagement [57] and social capital [15, 48].

At the same time, the idea that the Internet is changing our society in ways that endanger the existence of local community life has been long discussed [29, 123, 147]. As long-distance communication becomes easier and commute times become longer, researchers argue [123, 147], places are expected to lose their relevance in people's lives. Therefore, opportunities to maintain an active and engaged community life in one's neighborhood can be considerably reduced. While these predictions have been supported by some evidence [85, 120], they have also been challenged by other studies [76, 125, 147].

Nevertheless, the potential threats to local community life have inspired researchers and practitioners to pursue another agenda: the use of technology to prevent or mitigate the decline of local community life [135, 149]. Since the 1970s, local communities have used different technologies, ranging from electronic bulletin boards to social networking sites, in order to achieve this goal. Even the earliest local online information systems, before the emergence of social media, were designed as participatory information systems in which residents were both volunteer producers and consumers of local information [132]. This trend continues as newer information systems for local communities still rely heavily on user contributions.

1.1 PROBLEM STATEMENT

Prior research has shown that participatory information systems have the potential for helping local communities achieve social goals such as strong democracy, social capital, individual empowerment, sense of community and economic development [113]. Among them, social capital [36] has been more consistently investigated. Social capital refers to the value associated with the social relationships among people. It has been defined as the "*resources embedded in a social structure which are accessed and/or mobilized in purposive actions*" [92]. Its operationalization in terms of the number of social ties that a person has have been widely used in the assessments of online systems for local communities [135, 149]. Cross-sectional and longitudinal user surveys show that participation in community networks can be associated with increases in social capital [59, 61, 77]. Community networks can create opportunities to develop local social interaction [60]. Compared to non-users, users of a community network recognize more neighbors by name [59]. Conclusions drawn from an ethnographic study argue that the existence of a community network can also reduce barriers to collective action and community mobilization [59].

However, evidence across different studies is not consistent [113, 149]. While some projects have found evidence of social capital increase among users of local information systems, other projects have not [149]. As these findings come from case studies that focused on single local communities, it has been difficult to identify which factors might explain this difference across projects [42].

Although there is significant potential for these systems to have a positive social impact on local communities, creating sustainable participatory information systems for these communities has remained difficult in practice. Even though the urban setting provides a rich context for information-oriented technology endeavors [28], many local information systems that rely on user-generated content, such as discussion forums, social networking sites and digital public displays, struggle to remain viable over time [28]. Indeed, lack of enough participation and content has led several for-profit local community sites to close[1] [2] or become part of bigger multi-purpose systems.[3] On the other hand, research-funded projects to build systems for local communities have rarely reported

[1] http://www.nearbie.com/
[2] http://www.sharesomesugar.com/
[3] http://nabewise.com/

what has happened after the research was over. Therefore, it remains unknown if and how these systems have become viable "in the wild" afterwards [42].

From an information systems perspective, engaging enough contributors and maintaining a reliable stream of information in systems that rely on user-generated content is challenging [87]. Even online systems with worldwide reach struggle to become self-sustainable and many attempts have failed to do so [126, 86]. This is particularly difficult for local information systems because their potential audience is limited to people who live in or are visiting a specific place. Furthermore, creating enough relevant local content to attract this audience is not trivial. The frequency of new local information is generally low (few new items per day) and the community interaction that discusses this information online is often insufficient [28].

1.2 DISSERTATION WORK

To address this problem, my dissertation work proposes a conceptual framework to systematically analyze the sustainability of participatory information systems for urban communities. Furthermore, it reports on a mixed-method investigation that applies the proposed framework in order to investigate a sample of 35 online forums for neighborhoods and districts in the US that have remained active for several years "in the wild"

Grounded in related research and in collaboration with my co-authors in previous studies [100], we conceptualize sustainability of participatory information systems through three measurable aspects: *attraction* of new users, *retention* of current users, and *performance* of generating content and impact.

We draw from research on different fields to identify factors that might affect these sustainability measures. Research on participatory information systems with worldwide reach indicates that both individual [4, 13, 15, 73, 90, 129, 145, 151] and collective [21, 30, 137] online behavior affect different measures of the systems' sustainability. Studies on small groups [91] and volunteer associations [133] have provided evidence that different characteristics of their members and the relationships among them affect the performance and viability of these collectives. Literature on community informatics has suggested, but not yet empirically tested, that the sustainability of

participatory information systems for local communities might also be influenced by offline characteristics of the target local communities [77, 149]. Based on the findings from these fields, we propose a framework to analyze different aspects that can affect the sustainability of participatory information systems for urban communities.

Guided by this proposed framework, we address five major research questions to investigate if the sustainability of participatory information systems for urban communities is associated with the following aspects: (1) collective characteristics of the target urban communities; (2) collective features of the online activity, social network and shared content; (3) individuals' offline roles and demographics; (4) individuals' online activity, network positions and shared content; and, (5) system design decisions.

To empirically answer these research questions, we conducted a mixed-method investigation of the E-Democracy platform, which is a non-profit organization that provides online forums for urban communities "to support participation in public life, strengthen communities and build democracy."[4] Active since 1994, E-Democracy today hosts more than 40 online forums for neighborhoods and cities across three countries, including the US.[5]

In particular, we focused our investigation on 35 online forums that target neighborhoods and districts in the cities of Minneapolis and St. Paul in the state of Minnesota, US. We chose this area because it had the largest number of geographical areas using E-Democracy forums at the moment of data collection (September 2014). This sample allowed us to study the sustainability of online forums across communities that are comparable to each other, as they exist in similar geographical and cultural contexts. At the same time, these forums target geographical urban communities that vary with regard to their population size, racial diversity and population instability, according to the 2010 US Census. This aspect of the sample makes this dissertation's findings more generalizable.

Our data includes all post exchanges that had taken place in the 35 sampled neighborhood forums, along with all posts from the two city-wide forums that serve the cities of Minneapolis and St. Paul. The neighborhood forums varied in tenure from one to six years at the time of data collection. By then, the city-wide forums had been active for more than eight years. The archival data of the neighborhood forums has more than 75,000 posts that have been provided by more

[4] http://forums.e-democracy.org/about/ Last retrieved on November 7th, 2015.
[5] http://blog.e-democracy.org/posts/2647 Last retrieved on November 7th, 2015.

than 5,000 people. The data from city-wide forums comprise more than 92,000 messages that have been posted by slightly more than 2,000 users. We also have access to information about new members who joined any of these forums since 2010. Furthermore, we have data from the 2014 E-Democracy user survey, which was responded to by more than 1,300 platform users. These data were facilitated to us by E-Democracy.

Specifically, my dissertation work includes three major studies:

1. *Offline and online collective aspects and sustainability*: A longitudinal analysis of archival data of neighborhood forums that explores both online and offline collective characteristics of the forums and their association with collective measures of attraction, retention and performance. This study centers on the demographics of the forums' target urban communities and the global characteristics of a forum's online activity, social networks, and shared content. To be able to conduct this study, we supplement the archival analysis with social network analysis, manual and automatic classification of the content shared in the forums, and factor analysis of neighborhood demographic data.
2. *Offline and online individual aspects and sustainability*: A longitudinal analysis of archival data at the individual level and a cross-sectional study of user survey data. The archival analysis investigates how online behavior of an individual is associated with measures of sustainability. The survey analysis explores how the offline context (e.g., roles, demographics) of individuals relates to their expectations about the forums, their activities in the forums, and their perceptions of the forums' impact on their community involvement.
3. *Design decisions and sustainability*: An archival data analysis that compares the consequences of alternative design decisions of the E-Democracy forums on inclusion of women and people of color, two underrepresented populations on this civic platform. To undertake this study, public data is used to determine users' races and gender and the results are combined with the archival data of the forums.

Together, the results of my dissertation work provide evidence for the importance of considering the demographics of the target urban communities when assessing the sustainability of their participatory information systems. Various demographic factors are not only significantly influential for different aspects of sustainability, but accounting for demographics can also change the

effectiveness assessment of particular design decisions on sustainability. Specifically, residential instability of the neighborhoods populations turns out to be the most critical demographic factor on all measures of sustainability.

The results also indicate that online forums that have a mix of newcomers and old-timers among their contributors are more sustainable. At the individual level, this is explained by the patterns of online response and retention that these two kinds of users exhibit. On the one hand, newcomers are more likely to receive online responses (thus driving generation of more content), but they are less likely to continue contributing to the forums in the future. On the other hand, even though they are much less likely to receive online responses, old-timers are more likely to post again in the forums. Thus, a combination of newcomers and old-timers tends to create content and interaction at any given time and to have a user core that will continue creating content in the forums at subsequent times.

The findings of our survey analysis show that there is a great deal of off-site communication among the forums' users that are follow-ups to issues raised within the online forums. Off-site communication is also significantly and positively related to users' sense of satisfaction with the forums. This finding reveals that a critical part of users' interactions leaves no digital trace in the local information systems. This poses a challenge for the assessment of sustainability of information systems for urban communities that rely only on archival data of the users' online interactions.

The survey analysis also indicates that women are more likely than men to view the encouragement of community engagement as very important online forums for urban communities. The match between this goal and the goal of the E-Democracy neighborhood forums offers a feasible explanation as to why women make up the majority of the user base in the E-Democracy neighborhood forums. Beyond attracting more participation of women in the E-Democracy platform, the neighborhood forums are also associated with a broader participation of people of color, compared to the city-wide forums. This evidence is encouraging in regard to the impact of enacting design decisions that aim to mitigate gender and racial digital inequalities in the context of technologies for urban communities.

My dissertation work and its results have implications for information scientists and technology developers seeking to investigate and design information systems to increase community engagement of citizens within urban communities.

1.3 ORGANIZATION OF THIS PROPOSAL

This proposal is structured as follows. Chapter 2 introduces the problem that motivates my dissertation work. Chapter 3 summarizes related work. Chapter 4 introduces the proposed analysis framework and the research questions. Chapter 5 introduces a brief description of the history of our research platform: E-Democracy. Chapters 6, 7 and 8 describe the three studies that were conducted as part of my dissertation work. Chapter 6 reports on a longitudinal analysis of the the interplay of offline and online collective factors on the sustainability of participatory information systems for urban communities. This study combines archival data analysis with social network analysis, automatic content classification, and factor analysis of neighborhood demographic data. Chapter 7 uses archival and user-survey data to model the relationship between offline and online aspects of individuals and sustainability measures of their participation in the local information systems. Chapter 8 puts together archival data and public data in order to assess the impact of three particular design decisions on attracting under-represented populations to the E-Democracy platform. Chapter 9 discusses the results of the three studies in conjunction and their implications. Finally, Chapter 10 concludes this doctoral dissertation by summarizing its contributions.

2.0 MOTIVATION

This section introduces the topic of information systems for urban communities and states the problem that motivates my dissertation work.

2.1 PARTICIPATORY INFORMATION SYSTEMS FOR URBAN COMMUNITIES

Community networks were one of the earliest technological endeavors that aimed to serve local communities. Community networks were designed to support local communities' efforts to achieve social goals, such as encouraging community life [132]. Community networks were defined as "computer-based networks created by and for a local community" [25]. This definition emphasizes that while community networks were built in collaboration with public libraries or community informatics research projects, they originated from the grassroots efforts of the local communities.

However, the development and study of community networks in the context of urban communities declined considerably after the 90s, according to my review of the literature. Over time, community informatics shifted its focus to rural communities or developing countries [135], thus leaving the study of information systems for urban communities relatively understudied. At the same time, widespread household Internet availability reduced the relevance of public libraries as major providers of Internet services, which undermined the libraries' interest in supporting community networks for their cities.

Recently, a renewed interest in creating participatory information systems for urban communities has emerged. The popularity of social media, big data and open government has motivated a number of technology practitioners to get involved in grassroots or for-profit technological endeavors that target city residents. Many of these newer technological efforts have been categorized

under the umbrella term of "hyper-local", which aims to emphasize their focus on bounded geographical communities as opposed to other social media targeting worldwide audiences.

Aligned to the goals of community networks, a number of these hyper-local systems aim to facilitate community involvement as well. Nevertheless, newer goals have emerged. Some hyper-local systems attempt to increase political deliberation [40] and enhance awareness about local information (e.g. restaurants in Yelp[1], classifieds in Craigslist[2]).

My dissertation work focuses on investigating participatory information systems for urban communities that aim to influence community involvement. Therefore, the rest of this section is limited to the research within that scope.

2.1.1 Technology

In addition to the differences between the goals that community networks and hyper-local systems pursue, they also differ in their technologies.

Since their origins, community networks included a major software component that enabled residents to become both producers and consumers of local information (e.g., [132]). Over time, different technologies were used to implement these participatory media. In the 1970s, the project Community Memory made computer terminals available in different places within a city, so residents could share or find local data in an electronic bulletin board [37]. This attempt was later followed by discussion forums in the PEN Project [127] and E-Democracy [39]; question-and-answers sites in the Cleveland Free Net [10]; local email lists or listservs in Blacksburg [26] and Netville [59]; and websites that provided an integrated view of local information in Blacksburg [26] and other library-supported community networks [118]. Regardless of the specific technology, prior literature reports that community networks obtained most (if not all) of their content from the local residents [10, 37, 26, 61, 107, 118].

With the development of the Web, hyper-local systems embraced social computing technologies [23, 25] to implement participatory mechanisms for content creation. Nextdoor[3] and Neighbortree[4] provide social networking sites for neighborhoods where their residents can create user

[1] http://www.yelp.com/ Last retrieved on October 13th, 2015.
[2] https://pittsburgh.craigslist.org Last retrieved on October 13th, 2015.
[3] https://nextdoor.com/ Last retrieved on October 13th, 2015.
[4] http://www.neighbortree.com/ Last retrieved on October 13th, 2015.

profiles and network with other neighbors through a closed system. EveryBlock[5] offers an online space for community conversation with a strong focus on gathering and sharing local news. Neighborland[6] and Rebuild Your Community[7] enable users to collaborate online with other residents or local organizations in order to draw attention to problems in their communities and take action on these issues. Neighborhood Fruit,[8] NeighborGoods[9] and FavorTree[10] create marketplaces to foster offline interactions among neighbors. Whoo.ly [67] and CityBeat [150] integrate locally-relevant content from other social media sites such as Twitter in order to provide aggregated views of local information. Furthermore, many neighbors use independent email lists or blogs to create neighborhood-focused online spaces.

The evolution of technology has also broadened the number of geographical communities covered by the systems. Early community networks targeted a specific town or city, and sometimes particular neighborhoods. Few community networks were available in more than one city. On the contrary, hyper-local systems are more likely to be available in many urban communities within a country. Even though they target specific neighborhoods or cities, they can be used by local communities all around the nation in which they are available.

Figure 1 provides a timeline of the launches of the different participatory information systems for local communities that I have identified in my review of the literature.[11] [12] This figure shows which local information systems are still active (with darker font color). Three systems have remained active since the times of community networks, two of which have also adopted newer technologies over time. E-Neighbors[13] has evolved into a more modern social networking site. E-Democracy[14] has updated the user interface of their online discussion forums several times.

[5] http://www.everyblock.com/ Last retrieved on October 13th, 2015.
[6] https://neighborland.com/ Last retrieved on October 13th, 2015.
[7] http://rebuildyourcommunity.com/ Last retrieved on October 13th, 2015.
[8] http://neighborhoodfruit.com/ Last retrieved on October 13th, 2015.
[9] http://neighborgoods.net/ Last retrieved on October 13th, 2015.
[10] http://favortree.com/ Last retrieved on October 13th, 2015.
[11] Figure 1 aims to provide an overview of the history of local information systems, but it is not necessarily a comprehensive account of all of them.
[12] While Facebook, Foursquare, Yelp and Craigslist are used for information sharing within a local context, they do not aim to increase community involvement. Therefore, they are not included as part of the core of my literature review but they are included in Figure 1 as references on time.
[13] https://www.eneighbors.com/ Last retrieved on October 13th, 2015.
[14] http://forums.e-democracy.org/ Last retrieved on October 13th, 2015.

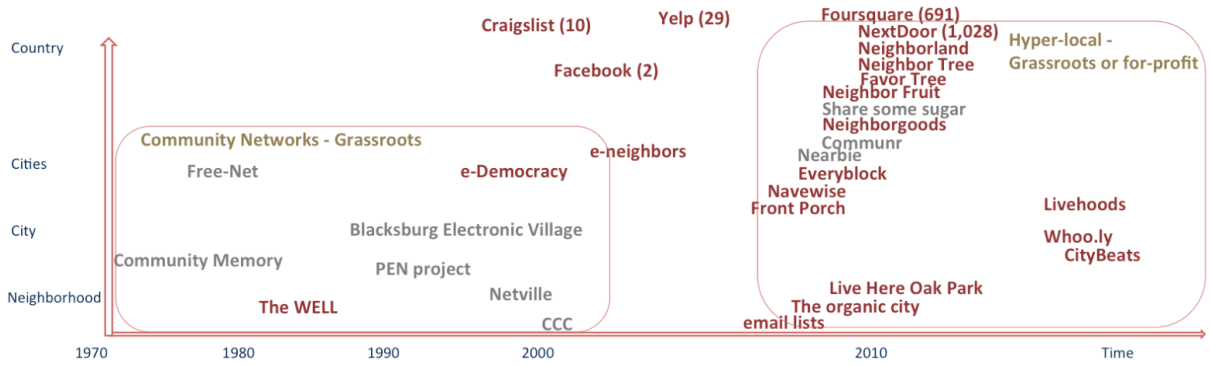


Figure 1: Timeline of participatory information systems for local communities

Overall, old and new information systems for local communities tend to rely heavily on residents to create local content. For this reason, I position my investigation of information systems for urban communities within the field of social computing.

2.1.2 Research

Research on community networks and hyper-local systems have focused on different aspects. Most research on community networks was oriented to assess the impact of local information systems on their users and target local communities (e.g. Netville [59], Blacksburg [77]). The focus on measuring impact made self-report methods popular. Surveys and interviews with users were common and, in some cases, ethnographic studies were conducted. Given that most community networks were available in a single city or town, this research was dominated by case studies that examine a particular information system that served a specific local community.

While still interested in studying impact [24], research on hyper-local systems has tended to focus on understanding the user experience (e.g., Whoo.ly [67], NextDoor [102], CityBeat [150]). Perceived usefulness of the information, satisfaction, and concerns about privacy are some of the topics that have been investigated. Self-reporting research methods are also pervasive in these projects, but they have sometimes been complemented with analysis of user online behavior.

Urban informatics is an emerging concept that refers to the research of people, technology and place in urban settings [51, 52, 53]. Urban areas and, particularly, its public places are the main focus of interest. Rather than focusing on communities as a whole, urban informatics focuses on individuals and their use of technology in public places. This strand of research has been categorized into the urban ideals they pursue [43]. Although six categories have been proposed [43], three of them seem the most comprehensive: 1) the *ubiquitous city* that should make the urban life easier to manage (e.g smart cities [34], navigation [139]) and rely on massive invisible data collection (e.g. sensors, public transportation cards); 2) the *urban flaneurs and situationists* in which the city is a stage for creative and cultural engagement (e.g. [2]); and 3) the *city as an operating system* where people consciously and purposefully exchange information through technology in order to increase their place involvement (e.g. e-participation [41], participatory urban planning [12], citizen science [116]). Similar to the study of hyper-local systems, urban informatics tends to focus on investigating the user experience. Figure 2 shows the trends of the goals, technology and research focus in the context of participatory information systems for local communities.

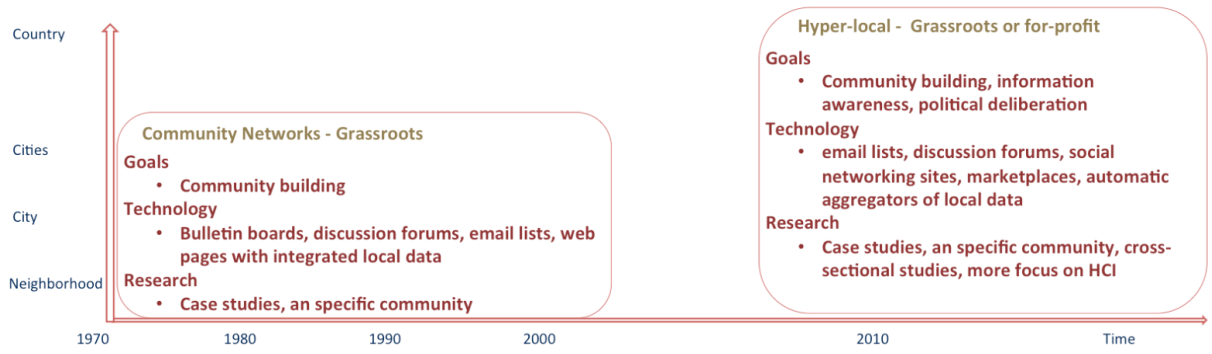


Figure 2: Trends of goals, technology, and research focus of PIS for urban communities

2.2 PROBLEM STATEMENT

In practice, similar to successful community networks such as those serving Netville [59] in Canada and Blacksburg [77] in the state of Virginia, newer hyper-local systems such as NextDoor have been able to engage a considerable number of residents in their served cities in the US. NextDoor is currently ranked the 298th site in the US after a consistently positive trend of user engagement, according to Alexa measures of web traffic data.[15] Nextdoor has also officially partnered with several US city governments to be used as a communication media between local governments and residents.

Nevertheless, many other information systems for urban communities struggle to remain sustainable. For example, EveryBlock offered a website for neighborhood discussion and hyper-local news for 19 US cities and had to suspend its operation in 2013 for economic reasons. Other applications for local communities, such as Communr[16] and Nearbie, have completely ceased operation. Furthermore, hyper-local information systems often report small rates of daily new content and little online discussion [25]. Insufficient new content and online interaction poses challenges to attract new users [84, 21] and maintain current ones [21]. Consequently, this may negatively affect the local systems' ability to remain sustainable over time.

Prior research provides scarce and scattered evidence about how participatory information systems for urban communities achieve sustainability. Literature reviews on community informatics have shown that not all projects have found evidence that community networks have met their goals in regard to impact (an aspect of sustainability), when impact has been measured as social capital increase [113, 135, 149]. However, it is not yet clear what defines success or failure with regard to this goal. The fact that research of community networks has been heavily dominated by case studies [42], and each case study is very dependent on the particular characteristics of the served local community [42], makes it difficult to compare findings across studies and discover patterns that can explain success or failure in this regard [149].

Two other studies provide additional evidence regarding the sustainability of local information systems. A study of the content shared in a community network concluded that discussion of poli-

[15] http://www.alexa.com/siteinfo/nextdoor.com Last retrieved on September 29th, 2015.
[16] http://www.communr.com/ Last retrieved on October 13th, 2015.

tics and local issues (related to danger and novelty) generated more participation in the discussion forums [107]. Another study that looks at a sample of local groups in e-Neighbors found that the length of discussion forums and the variety of discussion topics did not vary across neighborhoods with different poverty levels [54]. My dissertation work aims to contribute to this strand of research in order to model the factors that influence sustainability of participatory information systems for urban communities.

3.0 RELATED WORK

Information systems that rely on volunteers to generate content face challenges in order to remain active [126, 86]. Their existence relies on their ability to engage enough users to generate content, which will in turn attract other users [84]. The concept of sustainability has been introduced to capture this process. Users who are involved in participatory information systems become part of an online social structure that is able to generate outcomes such as online content. Sustainability refers to the ability of these social structures to "continue providing benefits for members over the long term" [21], thus ensuring survival of the social structure and the participatory information system over time.

A summary of the relevant literature that informs my dissertation work is presented in this chapter. First, the main research trends on the study of sustainability of participatory information systems with global reach are presented. These trends have often drawn from prior research on both work teams and volunteer associations [20]. Therefore, the literature in these fields is also reviewed. Then, a summary of the literature on participatory information systems in the context of urban communities is provided. The last section reviews research that explores how changes in the design of a participatory information system can influence its sustainability.

3.1 PARTICIPATORY INFORMATION SYSTEMS WITH GLOBAL REACH

Achieving better understanding about how participatory information systems function and remain active over long periods of time has been the goal of several research projects in the field of social computing. These studies have been mostly focused on systems with global reach such as Usenet, Wikipedia, Facebook and Twitter. The focus of these projects can be grouped into two major

categories: (1) factors that explain collective attraction and departure of users [21, 22, 30, 121, 137, 143, 152, 153], and (2) factors that influence an individual's decision to join or leave a participatory information system [4, 6, 13, 15, 73, 90, 129, 145, 151].

Regarding collective aspects, prior research provides evidence that several aggregated measures of users' characteristics and the dynamics of their online interactions are correlated to the survival of their online groups. These factors can be categorized as heterogeneity of users in terms of tenure or stability of participation [30, 121], membership overlap across online groups [143, 137, 152, 153], number of people who participate [21, 22, 137], and heterogeneity of the kinds of content that people share online [21, 144]. While heterogeneity of content has a negative impact on attraction and retention of users [30, 121], heterogeneity of user tenure has a mixed influence with curvilinear and negative effects in different studies [30, 121]. Membership overlap has been found to be positively related to the viability of online groups over time [143, 137, 152, 153], and the evidence about the effect of the number of users is mixed [21, 22, 137].

Among the factors that influence individual decisions, the probability of joining an online group was positively related to the number of social ties within current members of the online group and the level of connectivity among them (i.e., embeddedness in the social structure) [6]. Several other factors are associated with user retention. These factors can be classified as the level of online response or feedback received by a user [4, 13, 15, 73, 90, 129], the kinds of content shared in the system [145, 148], and characteristics of the user's online behavior [151]. Consistently, receiving online responses has been found to be a positive predictor of user retention [4, 13, 15, 73, 90, 129]. In health-related information systems, certain kinds of content (for example, social rather than informational) turned out to be positively related to longer participation in the systems [145]. In question-and-answers systems, users who tend to answer questions were more likely to remain active than users who tend to ask questions [151].

Both project categories have often used longitudinal archival data of the online systems in order to model user behavior either collectively or individually. This research method and the individual and collective factors that can affect different aspects of sustainability have been incorporated into my approach to investigate participatory information systems in the particular context of urban communities.

3.2 SMALL GROUPS AND TEAMS

Research on small groups [91] explores the dynamics of the interactions within small collectives of people (generally from 3 to 10 people). Although participatory information systems often target larger numbers of people, literature on small groups has been frequently used as a source of theories and evidence to guide the investigation on these systems. The study on work groups has been especially relevant to inform the way online groups function as a collective and generate different kinds of outcomes.

Although the concept of sustainability is not generally used in the research on teams, the literature on effectiveness of teams is particularly enlightening for the purposes of my research. Teams usually exist in contexts (e.g. organizations) where the of the group is very unlikely. Therefore, the focus of their research is to discover what makes them function more effectively. On the contrary, social structures that are created through the interactions among volunteer users of participatory systems can indeed become unsustainable (and disappear) under certain conditions (e.g., all contributors leave the system). The connection between these two areas of study is that the mechanisms that allow teams to work effectively can also help online social structures to work better, and in turn, ensure their sustainability, as argued in our prior work [100].

The study of team effectiveness, as summarized in [83], has categorized the measures of effectiveness into three aspects: performance, meeting the members needs, and viability (i.e, the members' will to remain in the team). These measures are affected both by inputs and processes. Inputs refer to the composition of the team (i.e., members, their characteristics, and their resources). Processes are the activities that the team members undertake to execute (or fail to execute) the team tasks. In a way, processes "mediate the translation of inputs to outcomes" [83].

Different factors have been studied as antecedents of team effectiveness, including the processes of information transfer among team members [106], collective cognitive processes in the teams [44], mechanisms of team building [81] (e.g. goal setting [82]), leadership styles and structures [109], and social ties among team members [8].

Given that local information systems have often been studied from the perspective of social capital, my dissertation work draws on the research on social ties, which are a common operationalization of social capital. A meta-analysis of studies on social ties within groups has shown

that density and centrality in the social networks affect two measures of team effectiveness: performance and viability [8]. The social network of small groups can represent both formal and informal communication ties [8]. Density of the networks of these two kinds of ties are moderately and positively related to team's performance. Teams with strong bonds among their members (i.e., group cohesiveness [66]) are generally associated with high performance [9]. Density in the two kinds of social networks is also positively and strongly associated with team viability [8]. Additionally, a match between network centrality of formal ties (i.e., leadership) and network centrality of informal ties is strongly and positively related to performance [8].

The study of social networks in teams has also introduced the idea that an individual's network position can generate benefits [16]. This perspective emphasizes the importance of brokering positions that can amend structural holes in a social network's connectivity [16, 17, 18]. Being a broker among somewhat disconnected networks is associated with individual benefits such as promotions and bonuses in teams [16]. Brokers also facilitate knowledge transfer between organizational units [140]. This benefit would be important for the study of the impact of membership overlap on the performance of online social structures.

Together, these findings complement the previously discussed studies of social computing by adding the perspective of team performance and the importance of social ties among the team members for performance and viability.

3.3 VOLUNTEER ASSOCIATIONS

Research on volunteer associations has been proposed as an alternative metaphor to guide the research on participatory information systems [20]. A volunteer association is a group of people "*relatively freely organized to pursue mutual and personal interests or to achieve common goals, usually non-profit in nature*" [133]. Although a participatory information system may lead to the creation of social structures that can match the definition of a volunteer association, this area of research has been less explored as a source of evidence for the study of these systems.

As argued in [20], online social structures are similar to volunteer associations in several dimensions such as size of membership, rates of growth and loss of membership and communi-

cation activity (e.g., sporadic distribution and episodic structure). Moreover, similar to volunteer associations, participatory systems for urban communities also target a population within a geographically-defined area.

Volunteer associations compete for members within a locale and vary their composition as a result of the dynamics of this competition [104]. Heterogeneity plays a significant role in these changes in composition. Atypical members and those who are also the target of other associations are more likely to leave [119]. In aggregate levels, town heterogeneity (especially race and education heterogeneity) has been found to be related to the number of memberships that their residents have. Prior research has also found that volunteer organizations are larger in larger cities [103]. Moreover, larger associations have more stable membership and are more central in networks of local organizations [103].

This research provides another perspective to my approach in which collective aspects of urban (geographically-defined) communities might also affect the sustainability of their associated participatory information systems.

3.4 BENEFITS FOR MEMBERS OVER THE LONG TERM

Along with maintaining an active stream of content, participatory information systems need to provide benefits for their members. In particular, systems for urban communities aim to positively affect the residents' community involvement. Social capital has been a dominant framework to study the impact of these systems. Social capital is a complex notion that is conceptualized and measured in diverse ways by different researchers. A recent definition describes it as "*resources embedded in a social structure which are accessed and/or mobilized in purposive actions*" [92].

Coleman [36] is often recognized as one of the first authors who attempted to model the concept of social capital. Coleman argues that closure in a social structure facilitates social capital in terms of the creation of obligations and expectations and the enactment of social norms [36]. However, researchers have found that Coleman's view does not account for other properties of a social structure that can also lead to social capital. For example, sparse social structures can facilitate access to new information [58]. Hence, other models of social capital have been proposed.

Lin's model [92] aims to further clarify the concept of social capital by its identifying antecedents and consequences. Lin argues that both collective assets (e.g., norms and trust) and individual aspects (e.g., level of education, income) determine an individual's network position and access to resources. This would explain why some people have more social capital than others. According to this model, individuals need to mobilize their social networks in order to take advantage of their social capital.

Adler proposed an alternative conceptual model [1], which considers the notion that social capital can generate benefits as well as risks. For example, closure may facilitate the enactment of social norms, but it places burdens on the abilities of its members to access new information. Therefore, Adler's model includes the concept of value that represents how certain forms of social capital are appreciated in a particular context. The value of social capital is also determined by other factors such as task and symbolic contingencies. The model also adds an explicit feedback mechanism. Value obtained through social capital can in turn affect the social structure. At the individual level, Adler's model suggests that the individuals' social relations also affect their opportunities, motivations and abilities to generate social capital.

Both models provide insights regarding the social capital concept. Lin clearly separates the collective and individual antecedents of social capital. Furthermore, this model also assigns a role to the individuals' agency. Individuals can decide whether to mobilize their social network or not. Attempts to mobilize their social network are attempts to capitalize the potential returns of this kind of capital. On the other hand, Adler highlights that the returns of social capital are not necessarily positive: there are also potential risks. The return value depends on the characteristics of the tasks at hand. In turn, the perceived value becomes feedback for the social structure as a whole. Moreover, Adler also includes other individual characteristics, such as motivations and abilities, that might affect the decision to capitalize the potential returns of social capital.

A combination of these models is useful to explain how research on participatory information systems has employed or investigated social capital. Prior research has widely explored:

1. How the use of participatory information systems affects the individuals' positions in their social network and the resources available to them [14, 48, 62, 59, 77, 78, 138, 141];
2. How individuals' characteristics such as demographics, motivations and abilities affect their use of participatory information systems and/or the benefits they get from it [14, 115, 138, 141].

However, other aspects of the potential connections between social capital and participatory information systems have been less explored. Only few projects have reported results about:

1. How the individuals' network position and the resources available to them influence their level of use of participatory systems [32, 146];
2. The relationship between the collective assets or characteristics of the target communities and their use of participatory information systems, which has been discussed in prior research [77, 149] but has not been yet systematically explored;
3. Whether people attempt and achieve the mobilization of resources from the social capital available through participatory information systems [89, 49].

Moreover, two underlying assumptions are common in the research at the intersection between social capital and participatory information systems for local communities. First, it is often assumed that the benefits of social capital are potentially the same for all residents who use the technology. Second, it is also commonly assumed that once the technology is made available to the local community, all residents will be equally inclined and able to engage with it. While the concept of the digital divide is important in the field of community informatics, most studies on community networks refrain from reporting their contribution to address digital inequalities within local communities.

Recent research has raised questions about the potential positive impact of technology on people who have little social capital to start with [45]. People from economically distressed communities have more difficulties obtaining the promised benefits of social capital, and see little potential impact of technology on realizing this promise. It is possible that technology can only positively influence social capital among those who already have high social capital [45].

On the other hand, both early and recent research on the digital divide has revealed that there have been systematic inequalities in access and use of the Internet across different populations in the US and worldwide [70, 112, 131]. While the access gap has narrowed in the US [117], inequalities in the different Internet uses and the production of online content across different platforms remain [63, 70, 131]. In particular, a study of civic technologies in the UK has found that women are less likely than men to participate in these platforms [56]. In the US, national surveys have revealed that social class and race are strongly associated with the use of technology

for civic goals. For example, white people are more likely than African-Americans and Hispanics to engage in civic communications, including political communications, both online and offline [136]. My dissertation work aims to complement these relatively understudied research topics in the context of participatory systems for urban communities.

3.5 SYSTEM DESIGN AND ITS INFLUENCE ON SUSTAINABILITY

Beyond the observational approaches to investigate the organically-developed dynamics of social interactions that are enabled by participatory information systems, a number of research projects have focused on assessing whether system design has an effect on user behavior, and therefore, on the sustainability of the systems. Research on social computing has identified several issues that challenge the survival of participatory information systems: (1) the start paradox in which there are few users who can create online content and there is little content to attract new users; and (2) the complexities of managing a collective of people, which includes developing commitment, encouraging contributions, reducing rate of user attrition, recruiting and socializing newcomers, developing leaders, regulating behavior, and managing coordination [84]. With the goal of assessing if the design of a system can address these problems, an evidence-based approach for the design of participatory information systems has been proposed and undertaken by several projects [84]. For example, research studies have compared the effectiveness of alternative design decisions to socialize newcomers [33, 50] and leverage contribution rates [11, 88]. This research method consists of implementing alternative principles through different system design decisions. The experimental designs test the hypotheses about the effectiveness of the proposed mechanism(s) compared to a control group. The online behaviors of users in different groups are then compared to make conclusions about the effectiveness of the alternative designs. A number of design decisions have been evaluated this way, such as:

- broadcasting a request for specific contributions [38],
- asking specific people to do specific tasks [11, 33],
- emphasizing the uniqueness of a user's contributions [11, 101],
- providing social information and feedback [31, 124],

- assigning people to competing groups [11, 46],
- setting personal or group goals [11, 46], and
- reducing the effort required to identify tasks that are likely to be done by a user. [38, 64]

From this strand of research, I borrow the perspective that system design might affect the dynamics of online social interactions and, therefore, alternative design decisions might have different effects on the measures of sustainability of participatory information systems.

Grounded in these areas of related work and their findings, the next chapter will explain a framework that proposes a conceptualization of the sustainability of participatory information systems for urban communities and the factors that can influence it.

4.0 ANALYSIS FRAMEWORK

This chapter introduces a framework to analyze the sustainability of participatory information systems for urban communities. The chapter formulates a conceptualization of sustainability. It also summarizes the potential antecedents of sustainability. The framework integrates perspectives from social computing, community informatics, and the study of teams and volunteer associations.

4.1 CONCEPTUALIZING SUSTAINABILITY

Grounded in related research, we propose an analysis framework (Figure 3) to investigate the sustainability of participatory information systems for urban communities. Research on participatory information systems with worldwide reach suggests that their sustainability depends on (1) the availability of resources, and (2) the ability to convert resources into tangible and intangible benefits for the members [21]. Research on small groups considers performance and viability as measures of team effectiveness [8]. In an earlier collaborative work [100], we have proposed that these two perspectives can be connected. In participatory information systems such as those targeting urban communities, the main resources are the users that can contribute content to the system. Team viability can be operationalized as the ability to attract new members and to retain current members, which in turn ensures resource availability in online communities. Performance can measure the team's ability to take advantage of the available resources to achieve the team's goals of gathering content and generating an impact on the users. Hence, by incorporating these two different theoretical perspectives, we characterize sustainability through the following aspects:

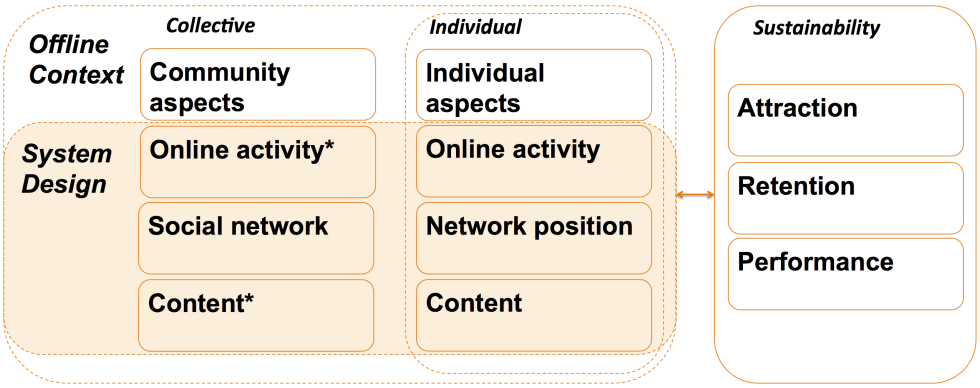


Figure 3: Framework to analyze sustainability of PIS for urban communities

1. *attraction*: the ability to attract new members, and

2. *retention*: the ability to retain existing members, and

3. *performance*: the ability to gather content and generate an impact.

Participatory information systems for urban communities exist in an underlying social context where the target users have already developed dynamics of coexistence and communication that might affect the adoption of another communication media. Furthermore, once some residents have decided to use a participatory information system, a series of online behaviors takes place. These behaviors are enabled by the system design and can in turn influence whether the system will continue to be used. Thus, both the offline context and the design of the system are expected to play a role on the sustainability of participatory information systems for urban communities.

Prior research (see chapter 3) provides evidence about the influence of several factors on different aspects of our conceptualization of sustainability. We categorize this evidence into collective and individual levels (see Tables 1 and 2, respectively). The next sections summarize prior findings according to the proposed analysis framework.

Table 1: Collective variables in the framework

Factors	Variables	Attraction	Retention	Performance
Community aspects	Population size	+ [103]	+ [103]	
	Heterogeneity	- [128]		
Online activity	Tenure heterogeneity		+/- [30]	
	Membership overlap	- [143] + [137]		+ [153]
Social network	Size	+ [21, 137] - [143]	- [72, 142]	- [134]
	Connectedness	- [6] + [110]	+ [8]	+ [9]
	Centralization	- [137]		
Content	Content	+ [21]	-[21, 144]	
	Online response			

4.2 OFFLINE CONTEXT

Given that this dissertation work focuses on the problem of maintaining sustainable participatory information systems for urban communities, a critical goal is to identify specific features that make urban communities different from (or similar to) other kinds of communities that participatory information systems may serve. Two major features of this particular context are explored in this dissertation: offline community and individual characteristics.

Community aspects: Similar to volunteer associations, the characteristics of the population in the urban community and the dynamics of competition among groups that target their residents as potential members might affect the sustainability of participatory information systems. Residents can decide to join zero, one or multiple online groups that focus on the urban community. Once a resident has joined, her continuous participation is not ensured. Atypical members and those who are also the target of other groups might be more likely to leave, as happens among volunteer associations [119, 104]. In aggregated levels, town heterogeneity (especially race and education

Table 2: Individual variables in the framework

Factors	Variables	Attraction	Retention	Performance
Individual aspects	Roles			[19]
	Demographics	[5, 136, 3]	[119, 104]	[70, 56]
Network position	Centrality	+ [6]	+ [146]	+ [146]
	Broker			+ [16, 17, 18]
Online activity	Tenure		+ [4]	+ [4]
	Overlap			
Content	Kind of content		[145, 151]	[89]
	Online response		+ [4, 13, 15, 73, 90, 129]	

heterogeneity) is negatively related to the number of memberships in volunteer associations that their residents have [128]. Besides, the size of the population can also have an impact. Larger cities tend to have larger volunteer organizations [103], and larger associations have more stable memberships and are more central in networks of local organizations [103]. Thus, it is plausible to expect that collective aspects of the target communities, such as heterogeneity and population size, can influence the adoption and sustained use of their participatory information systems.

User roles and characteristics: Compared to participatory information systems with global reach, there is a higher probability of an underlying social structure among potential users of systems for urban communities. The fact that the potential users may know (or know about) each other adds another factor that can influence online participation. For example, participation in local online systems may be influenced by the roles that residents play in their urban community. Prior research has found that users who know other users in an "offline context" spend significantly more time on community-building activities in online groups and those users who had leadership roles in offline settings tend to get more involved in cross-posting activities [19]. Participation of well-known people may also encourage participation from other members of the community, similar

to the effect of celebrity endorsements in larger-scale participatory systems [88]. Besides, gender and race significantly differ in their tendencies to join volunteer associations [5], participate in civic activities [136], and the level of contribution to participatory information systems [3, 70, 56]. These kinds of tendencies can have aggregated effects in a system that focuses on urban communities.

4.3 ONLINE ACTIVITY

User tenure and membership overlap are two characteristics of users' online activity in participatory information systems that influence sustainability measures.

Diversity of tenure among editors of Wikipedia groups had a curvilinear association with collective withdrawal [30]. While very low and very high levels of tenure diversity were related to high turnover, moderate levels of diversity of experience in Wikipedia were associated with lower rates of withdrawal. Diversity in chat channels was a significant predictor of the likelihood of a channel's long-term sustainability [121]. Channels with more diverse populations were more likely to survive than those with more homogeneous populations.

Evidence regarding the consequences of *membership overlap* on participatory information systems is mixed. While membership overlap was found to be positively related to the growth of online groups in Wiki Projects [137] and maintenance of an active stream of content in Wikia online communities [153], it was detrimental for the growth of online discussion groups [143].

At the individual level, *user tenure* can affect both the chances of generating online interaction and the probability to continue contributing to an online system. Messages from newcomers with little experience were less likely to receive a response compared to messages sent by old-timers in public newsgroups [4]. On the contrary, new users who eventually became core users in Yahoo groups were always more likely to receive answers [7]. Regarding retention, newcomers were less likely than old-timers to continue posting in public newsgroups [4].

4.4 SOCIAL NETWORKS

At a collective level, performance and viability of offline teams and online communities have been studied through social network analysis [6, 8, 122]. Social structures emerge through the team members' interactions with one another. The characteristics of these social structures can be captured by social network measures. Motivated by prior work, we identified three network measures: size, connectedness and centralization, to be related to attraction, retention and performance.

Network size refers to the number of people involved in the social structure. The size of a group can affect its ability to attract new members. Larger online social structures attracted more new users in listservs [142]. Larger groups can also be more attractive for newcomers - people who enter a large network are more likely to succeed, since the abundance of resources makes it more likely for newcomers to receive informational and social support they need [110]. Availability of resources to support newcomers is particularly crucial for participatory systems for urban communities in which resources are limited to a specific geographical area.

As the community grows, there are more members who can contribute to the community. However, community growth can involve negative consequences as well. As the size of a group increases, members' contributions also grow, but this abundance of contributions can cause information overload [130]. Information overload can overwhelm users and cause their departure from the community [72]. Evidence of a negative effect of size on membership retention has been found in Usenet newsgroups [72] and listservs [142].

Larger groups may also cause social loafing; i.e., individual members of the group feel less responsible to exert their full effort toward the group's goals [75]. In online groups, social loafing can be reflected in the reduction of users' contributions when the user population grows [71, 142]. Indeed, social loafing significantly hindered the performance of groups in online platforms [134].

Network connectedness is an indicator of the interactions among users in a social network. It reflects ties of collaboration or information transfer among members of the network. Connectedness can influence the ability of social structures to attract new members. On one hand, higher connectedness reflects a stronger network that is more likely to provide stronger support to newcomers [110]; therefore, dense networks can be perceived as more attractive by potential new users. On the other hand, highly connected networks can be perceived as a clique or as a less welcom-

ing environment for newcomers and therefore attract fewer new users. In LiveJournal.com, more connected online communities were indeed less likely to grow [6].

Well-connected social structures are more likely to stay together. A meta-analysis of prior literature confirmed that denser teams are associated with team viability [8]. Denser networks provide more social support that leads to more team satisfaction and desire to remain on the team.

Denser social structures can also better provide social support and information to its members, which can in turn help to fulfill the social network's goals. Particularly, more connected offline groups tend to have better performance [8]. A positive relationship between density and performance existed in online communities that discuss the development and use of software tools [80].

Network centralization refers to the inequality of individuals' connections. It is used to distinguish networks with core central, well-connected leaders and many more peripheral, less connected members (e.g., star networks) from more egalitarian networks in which most members are equally connected to others. Network inequality is pervasive in online communities, which has been revealed in studies reporting the preferential attachment phenomena in networks' growth [69] and inequality of contributions in online communities [114]. More connected nodes increase their connectivity faster than less connected nodes [69]. Moreover, the majority of the contributions often can be attributed to a minority of users [114]. In terms of consequences, recent research has shown that higher inequality of contribution in early stages of WikiProjects led to less growth of the group in the long-term [137].

Higher network centralization hints at the presence of central and peripheral members. This uneven structure can affect retention of users. Group members who are more embedded in an online social network are expected to be more likely to remain part of the group; conversely, more peripheral members are expected to be less likely to maintain their online ties [105]. However, there is not yet evidence to support this reasoning.

At the individual level, *user centrality* can positively affect performance. Offline teams with leaders who are central to the group tend to attain higher performance [8]. In professional networks of practice, more central users contribute more and their contributions are perceived as more useful [146]. They were also more likely to sustain their level of contributions [146].

Following the conceptualizations of social capital, user centrality can also positively affect the chance of getting responses from an online group (performance) and retention. The probability of

getting a response to a mobilization request may also be influenced by the social capital available through the online system. As suggested by a model of social capital [92], an individual's ability to mobilize (or capitalize) on the social capital present in the online groups can depend on the individual position in such network. People with more or stronger ties can be more effective at obtaining resources (e.g. advice, emotional support) from their social networks than people with fewer or weaker ties. At the same time, individuals who belong to communities with more collective social capital can also obtain more benefits from their social groups. In online settings, social ties can translate to online interactions. People who have co-participated in the same threads of discussion may feel obligated to reply as a form of reciprocity. Therefore, people who have had interactions may have a better chance of getting a response. Besides, users who have more interactions with other members of the forum can feel more invested in the forum and feel obligated to continue their central role in the community.

Another key individual characteristic is a *brokering position* in the social network. Being a broker among somewhat disconnected sections of a network is associated with individual benefits such as promotions and bonus in teams [16], and it has also been related to facilitate with knowledge transfer among organizational units [140]. These benefits could become better performance measures for the individuals (e.g., getting responses) and for the groups (e.g., knowledge transfer among online groups).

4.5 CONTENT

Prior research has shown that content is related to sustainability at a collective and individual level. A study of email-based online groups found that topic diversity was a significant mediator of the relationship between group size and group membership variation. Size was positively related to content diversity, and both measures had a negative association with the subsequent measure of group size [21]. Larger groups had greater content diversity in their communication and these two factors were associated with losing larger proportions of membership. Another study analyzed the consequences of content diversity in Twitter. It was observed that lower content diversity was associated with a larger and more connected group of followers (i.e. readers) [144].

At the individual level, users who contribute and seek social support instead of informational content are more likely to return to health-related online discussion forums [145].

Related to the notion of social capital, research on Facebook has looked into how content reflects users' attempts to mobilize resources from their ego-centered social networks as additional evidence of the use of this platform for developing and exercising social capital [49, 89]. These studies found that mobilization requests received more comments than non-mobilizations, and specific kinds of mobilizations were more likely than others to get a response and to obtain the response faster [89].

Prior work has also found evidence that receiving online responses in online communities is a factor that influences user retention. In public newsgroups, users were more likely to post again when they have received responses to their prior messages [4, 73]. Similarly, Facebook users who received comments on their pictures were more likely to post pictures again [15]. Other forms of community feedback had comparable positive effects on user retention in large enterprise social sites [13], Slash Dot [90], and other social websites [129].

4.6 PROPOSED FRAMEWORK, MEASURES AND RESEARCH QUESTIONS

Prior evidence from various fields that were revised in this chapter suggests that a number of factors can affect sustainability of participatory information systems for urban communities. We have compiled these factors into an analysis framework that can guide systematic studies of this kind of systems. Figure 4 provides an overview of the proposed framework and the specific measures to operationalize the framework's concepts. This framework helps formulate my research questions:

1. Is sustainability of participatory information systems for urban communities associated with the following aspects?
    a. Collective characteristics of the target urban communities.
    b. Collective features of the online activity, social network and shared content.
    c. Individuals' offline roles and demographics.
    d. Individuals' online activity, network positions and shared content.
    e. Particular system design decisions.

The next chapters report on the studies that were conducted to address these research questions.

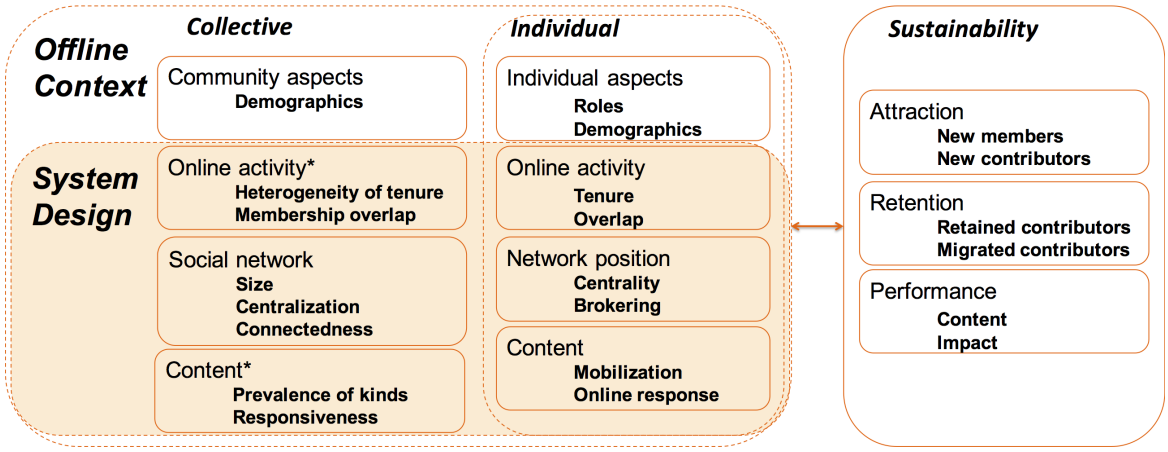


Figure 4: Measures to analyze sustainability of PIS for urban communities

5.0 RESEARCH PLATFORM

To address the research questions, the proposed framework was used to guide an empirical examination of the online platform E-Democracy.[1] The platform hosts and supports online discussion forums for urban communities. This investigation was possible thanks to a partnership with E-Democracy. The platform's founder and senior staff shared with our research group information about the development of the platform, the system's archival data, and the latest user survey.

5.1 E-DEMOCRACY.ORG

E-Democracy is a non-profit organization that has provided online discussion forums for local communities since 1994. E-Democracy's mission is "to harness the power of online tools to support participation in public life, strengthen communities, and build democracy."[2] Originally, E-Democracy was a virtual space (implemented through an email listserv) to obtain and disseminate information about political candidates in the state of Minnesota.[3] [4] The platform was built and maintained by volunteers who contributed in many ways from running servers to moderating the online discussions to transcribing campaign information from other information sources. E-Democracy also hosted pioneer e-debates in which candidates got involved in week-long online conversations (including answers and rebuttals) around topics that residents had raised through the platform.[5] Over time, this virtual space for political discussion during election times became stable

[1] http://www.e-democracy.org

[2] http://forums.e-democracy.org/about/ Last retrieved on October 13th, 2015.

[3] http://www.e-democracy.org/1994/Project_description.html Last retrieved on October 13th, 2015.

[4] http://blog.e-democracy.org/archives Last retrieved on October 13th, 2015.

[5] http://www.e-democracy.org/1994/E-Debates/ Last retrieved on October 13th, 2015.

discussion forums that had attracted more than 1,300 residents. With this user base, E-Democracy later decided to launch their web version and created an announcement-only email list to allow residents to promote their civic events without the need to follow political discussions.[6]

Before 2000, E-Democracy launched online discussion forums for few cities in the state of Minnesota. The platform adopted a metaphor of an online town hall to describe the dynamics and goals of these forums.[7] These city-wide forums had a strong focus on political discussion and deliberation about local issues. For example, the description of the Minneapolis' forum invites users "to discuss local-level Minneapolis civic issues. With over 1600 registered participants, this is a vibrant online space where citizens, elected officials, and community leaders - with diverse ideas and backgrounds - can discuss the important local issues facing our city in a civil and respectful manner."[8] The city-wide forums were meant to bring a more diverse audience to this civic technology, particularly more female voices.[9] While the town hall model has evolved over time according to the lessons learned through the years, several of the key design decisions of the E-Democracy forums were already defined by this time. With the goal of encouraging civil conversations among citizens, users were requested to provide their real names and the platform constrained the daily number of contributions that a single user can provide.[10] [11] [12]

As of 2007, E-Democracy had raised funding from multiple sources including the UK E-Democracy National Project,[13] the Blandin Foundation,[14] the Minneapolis Foundation MSNet Fund,[15] and an Ashoka Fellowship for one of its founders and current leader, Steven Clift.[16] Counting on a budget, E-Democracy established a strategic plan that comprised the deployment and support of city-wide forums in more cities and towns,[17] a focus on increasing the use and re-use of

[6] http://www.e-democracy.org/mn-politics/explain.html Last retrieved on October 13th, 2015.
[7] http://blog.e-democracy.org/posts/624 Last retrieved on October 13th, 2015.
[8] http://forums.e-democracy.org/groups/mpls/charter Last retrieved on October 13th, 2015.
[9] http://blog.e-democracy.org/posts/2647 Last retrieved on November 7th, 2015.
[10] http://blog.e-democracy.org/posts/74 Last retrieved on October 13th, 2015.
[11] http://blog.e-democracy.org/posts/183 Last retrieved on October 13th, 2015.
[12] http://pages.e-democracy.org/Issues_Forums_for_participants Last retrieved on October 13th, 2015.
[13] http://www.e-democracy.org/uk/ Last retrieved on October 13th, 2015.
[14] http://blog.e-democracy.org/posts/115 Last retrieved on October 13th, 2015.
[15] http://blog.e-democracy.org/posts/112 Last retrieved on October 13th, 2015.
[16] http://blog.e-democracy.org/posts/66 Last retrieved on October 13th, 2015.
[17] http://blog.e-democracy.org/posts/105 Last retrieved on October 13th, 2015.

civic information,[18] and a commitment to share their knowledge about online civic engagement and to use and develop open source software.[19] This strategy led to the development of an open source platform that hosts all E-Democracy forums; the expansion of city-wide forums to more than ten city-wide forums in Minnesota, other states in the US, UK and New Zealand; and the development of a number of resources for supporting the creation and maintenance on online groups for civic engagement.[20]

At the same time, the idea of a new kind of forum that would target neighborhoods rather than cities had emerged among the platform's leaders.[21] Different than a focus on city politics, these smaller-scope forums would cover more mundane topics of life in the neighborhoods that could range from community events to garage sales to neighborhood crime. Indeed, some of the funding E-Democracy had obtained at this point was targeted to create these neighborhood forums. The focus on the first neighborhood forums was online citizen engagement in rural[22] and urban communities that were characterized as "high immigrant/low income/communities of color" in Minneapolis and St. Paul.[23]

From 2010 to 2014, and once an initial phase of neighborhood forum creation was over, E-Democracy received additional grants from the Ford Foundation and the Knight Foundation to deepen and expand their efforts on city-wide and neighborhood forums.[24] [25] [26] At this point, the ultimate goal of the neighborhood forums was to encourage inclusive social media that can better engage citizens that might be less likely to get involved in civic participation and use of social media. To achieve that, the focus of the neighborhood forums continued to be on the "neighborhood life". For example, one of the neighborhood forums defines its goal as follows: "Share announcements and discuss neighborhood issues, life, and events specific to the neighborhood."[27] A screen-shot of a neighborhood forum in E-Democracy is shown in Figure 5.

[18] http://blog.e-democracy.org/posts/106 Last retrieved on October 13th, 2015.
[19] http://blog.e-democracy.org/posts/107 Last retrieved on October 13th, 2015.
[20] http://forums.e-democracy.org/projects/know Last retrieved on October 13th, 2015.
[21] http://blog.e-democracy.org/posts/80 Last retrieved on October 13th, 2015.
[22] http://blog.e-democracy.org/posts/112 Last retrieved on October 13th, 2015.
[23] http://pages.e-democracy.org/Neighborhood_Forums Last retrieved on October 13th, 2015.
[24] http://pages.e-democracy.org/Inclusive_Social_Media Last retrieved on October 13th, 2015.
[25] http://forums.e-democracy.org/projects/engage/beneighbors-iceo Last retrieved on October 13th, 2015.
[26] http://blog.e-democracy.org/posts/2647 Last retrieved November 7th, 2015.
[27] http://forums.e-democracy.org/groups/mpls-poho Last retrieved on October 13th, 2015.

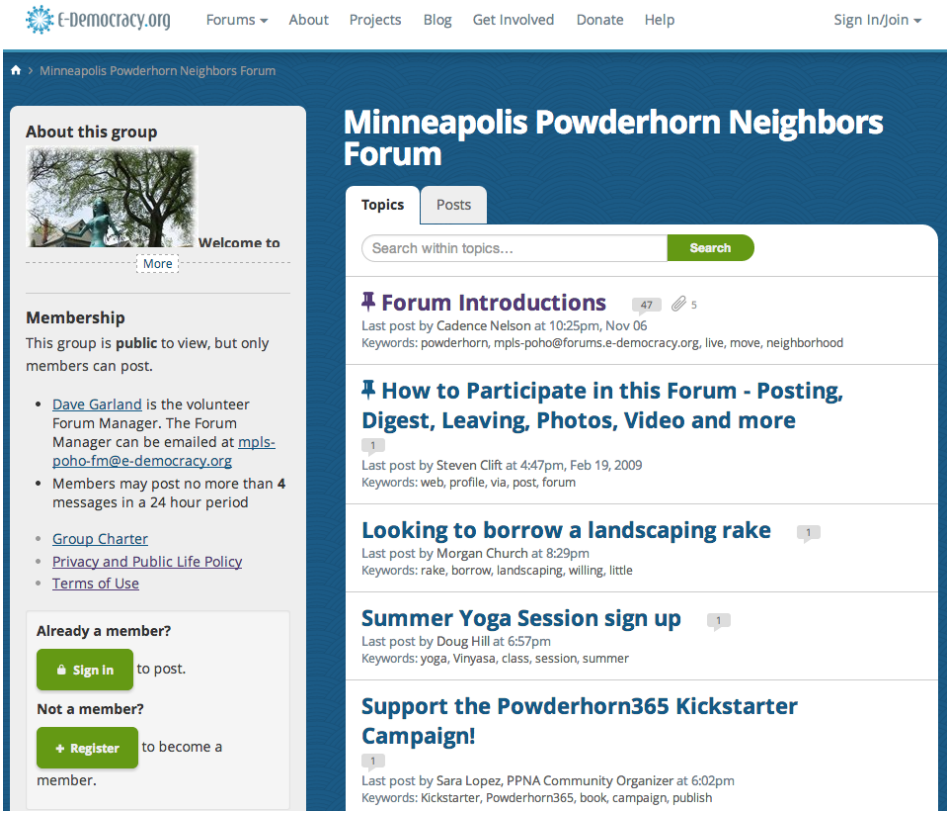


Figure 5: User interface of a local online forum in E-Democracy

Additionally, special offline outreach strategies were designed and undertaken to increase awareness about the local forums and expand the platform's user base. Offline outreach was conducted in highly diverse neighborhoods (those considered "high immigrant/ low income/ communities of color"). The main characteristic of these strategies was to do in-person outreach. The goal was to sign up people in the online local forums through paper-based forms.[28] [29]

Several characteristics of the E-Democracy forums make this platform appropriate for my research purposes. All forums have used the same interface, have followed a similar process of creation, and have enacted similar moderation rules. Creating and maintaining a forum is free to

[28] http://blog.e-democracy.org/posts/639 Last retrieved on October 13th, 2015.
[29] http://blog.e-democracy.org/posts/172 Last retrieved on October 13th, 2015.

do. A new forum is created when at least 100 residents sign up to create their local forum.[30] To post to a forum, users need to register in the platform and subscribe to the particular forum. Although registering and explicitly joining a forum are required to post, the content shared in the forums is public to unregistered users. Some other rules are also relevant for the proposed research. Users are requested to provide their real full name when registering as a way to encourage trust among users (i.e., neighbors) [35]. Every forum has a volunteer forum manager who acts as a moderator when needed, but also encourages participation in several ways.[31] At the same time, all forums constrain the number of daily posts that a user can add in order to avoid a few members dominating the forum activity [39]. The daily maximum number of posts per user varies from two to six.

Nowadays, E-Democracy hosts more than 40 online forums across three countries. Both city-wide and neighborhood forums co-exist in the platform.[32] Aligned with their goal to share their knowledge and lessons learned about online civic engagement, E-Democracy has provided us their archival data of local forums over a period of time and the 2014 user survey data.

5.2 STUDIES

My doctoral dissertation reports on three studies of the E-Democracy's forums that empirically explore the framework proposed in Chapter 4. Study 6 centers on the offline and online collective aspects of sustainability. Study 7 focuses on individual factors that are associated with retention and performance. Study 8 explores the effect of alternative design decisions on measures of attraction and performance. Each of the studies has focused on different aspects of the framework, but their findings complement each other in order to provide a more comprehensive understanding of the sustainability of the E-Democracy forums for urban communities.

Figure 6 illustrates the proposed analysis framework and indicates how the three studies investigate its different aspects. A summary of the data collection and analyses involved in each study is shown in Table 3. These methods will be explained in detail in the corresponding chapters.

[30] http://blog.e-democracy.org/posts/280 Last retrieved on October 13th, 2015

[31] http://pages.e-democracy.org/Forum_manager_position_description Last retrieved on October 13th, 2015.

[32] http://forums.e-democracy.org/about/localforums/ Last retrieved on October 13th, 2015.

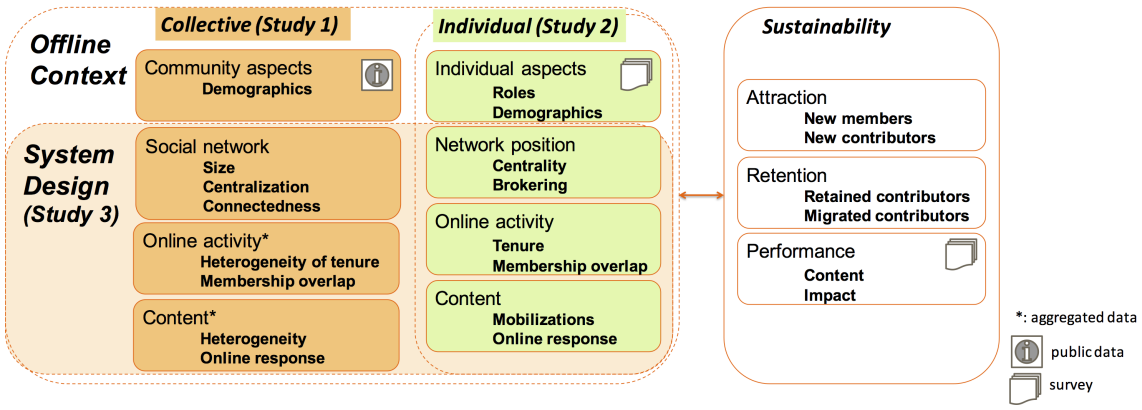


Figure 6: Scope of the three studies on E-Democracy

Table 3: Studies, data collection and analyses of this dissertation work

Study	Data Collection	Analyses
Collective aspects - Ch. 6	a. Archival data of membership/posts b. Manual & automatic coding of posts c. Community public data	Statistical & network analysis Content analysis & classification Factor and survey analysis
Individual aspects - Ch. 7	a. Archival data of membership/posts b. Manual coding of posts c. User survey	Statistical & network analysis Content classification Survey analysis
System design - Ch. 8	a. Archival data of membership/posts b. Community public data c. Race and names public data	Statistical analysis Factor and survey analysis Race & gender prob. estimation

6.0 STUDY 1: COLLECTIVE ASPECTS OF SUSTAINABILITY OF ONLINE FORUMS FOR URBAN COMMUNITIES

While there is growing interest in creating participatory information systems for urban communities, little is known about what makes these kinds of systems thrive and become viable over time. In particular, prior research has argued that characteristics of target urban communities might influence the survival of participatory information systems [77, 149]. However, this remains an open question for researchers and practitioners.

To address this problem, this study aims to investigate whether sustainability of these participatory information systems are associated with (1) collective offline aspects of the target urban communities, and (2) collective characteristics of the online interactions that take place in the participatory information systems.

We conducted a longitudinal observational study of the sustainability of 35 online discussion forums for neighborhoods and districts in the US that are hosted by the E-Democracy platform. Grounded in prior research (see Chapter 4), we assessed how the characteristics of the target urban communities and the forums' social networks, online activity, and content are associated with attraction, retention and performance of the forums. We employed different research methods in this study. First, we used public data about the forums' target urban communities in order to represent their main characteristics. Second, we computed collective measures of the sustainability, online activity and social structures of the forums from archival data of their members and posts. Then, we conducted an automatic classification of the posts in order to characterize the content shared in the forums. Finally, these data were combined to assess the effect of the collective offline and online aspects at a given time period on the sustainability of the forums at a subsequent period of time. This chapter details the research methods that were used to conduct this study, the main findings and their implications.

6.1 DATA COLLECTION AND ANALYSIS

This section explains the measures of sustainability that were used as dependent variables, as well as the measures that were employed as independent variables. It also summarizes the research methods used for data collection. Table 4 sums up the main variables of this study.

Table 4: Independent and dependent variables of Study 1

Independent variables	Dependent variables
1. Community aspects	1. Attraction
a. Size	a. Number of new members
b. Diversity	b. Number of new contributors
c. Instability	c. Proportion of new contributors
d. Full access to Internet	2. Retention
2. Online activity	a. Number of retained members
a. Membership overlap	b. Proportion of retained contributors
b. Heterogeneity of tenure	c. Proportion of migrated contributors
3. Social networks	3. Performance
a. Size	a. Posts
b. Connectedness	b. Productivity
c. Centralization	c. Productivity change
4. Aggregated shared content	
a. Prevalence of mobilization	
b. Responsiveness	

6.1.1 Measuring sustainability

Our dataset includes all the posts that were exchanged in 35 E-Democracy online forums for neighborhoods or districts in the cities of Minneapolis and St. Paul in the US state of Minnesota. We also had data from the two city-wide forums associated with these cities. We segmented the data into calendar quarters as the observation period for our longitudinal analysis. Our panel dataset of posts included data from the first quarter of each forum until the second quarter of 2014. Given that not all forums have been active since the same year, our panel dataset was unbalanced. The oldest forums in our dataset were founded in the first quarter of 2008 and the newest forums were initiated in the second quarter of 2013. Thus, the tenure of the sampled forums ranged from one to six years. Additionally, we had the membership data of the forums. However, this data was only available since last quarter of 2010, not for the whole lifecycle of all forums. Overall, we had 551 observations in our longitudinal data of posts and 402 observations in the membership data. For each quarter in a forum, we computed the following measures to represent attraction, retention and performance:

- *Attraction* is measured using three variables: (a) number of new users who joined the forum in the quarter; (b) number of users who posted for the first time to the forum (i.e., new contributors); and (c) proportion of new contributors in the quarter to the total number of contributors in the previous quarter.

- *Retention* comprises three other measures: (a) number of contributors from the previous quarter who continued to post in the current quarter; (b) ratio of retained contributors (those who kept contributing from the last quarter to this one) to the number of contributors in the prior quarter; and (c) proportion of users who migrated (i.e., stopped contributing to the forum and contributed to another one) to the total number of contributors in the previous quarter.

- *Performance* is represented by three metrics: (a) number of posts in the quarter; (b) volume of posts divided by the number of contributors in the quarter (i.e., productivity); and (c) productivity change between two consecutive periods of time. The change in productivity is the productivity of the current quarter minus the productivity of the prior quarter. This measure could take positive (increase in productivity) or negative values (decrease in productivity).

Figure 7 illustrates the distributions of these dependent variables. While the absolute measures (numbers of new members, new contributors, retained contributors and posts) had a right-skewed distributions, most of the other dependent variables that represent relative measures (percentage of new contributors, percentage of retained contributors, productivity and productivity change) showed roughly normal distributions. The percentage of migrated contributors was right skewed. Therefore, this variable was log-transformed to be able to work with a more symmetrical distribution of this dependent variable. Given these distributions, we used Poisson regressions to model the absolute measures (count variables) and linear regressions to estimate the relative measures. As we have repeated measures of these dependent variables over time, we used xt commands in Stata 14 (tools for analyzing panel data while controlling for the correlation of measures within subjects).

6.1.2 Community aspects: use of public data

The sampled E-Democracy forums target neighborhoods or districts whose demographics vary, according to the data made available by Minnesota Compass.[1] This dataset comprises data collected from the 2010 US Census, the 2009 Local Employment Dynamics data, and the 2005-2009 American Community Survey. The dataset includes a set of 166 variables, each of which can be considered a demographic feature.

To identify the major demographic features, we conducted principal component analysis using Stata 14. The results indicate that 81.15% of the data variance is explained by the first three components. After removing redundant demographic features, the main three components revealed groups of variables that we conceptualized as follows:

- *Size* of the neighborhood includes total population, housing, number of employed residents, and number of jobs in the area.
- *Diversity* of the neighborhood comprises the percentage of the population that is people of color, the percentage of the population that is seventeen or younger, and the percentage of the population that is 25 and over whose education level is less than high school graduate.

[1] Last retrieved on http://www.mncompass.org/profiles/neighborhoods/minneapolis-saint-paul#!areas

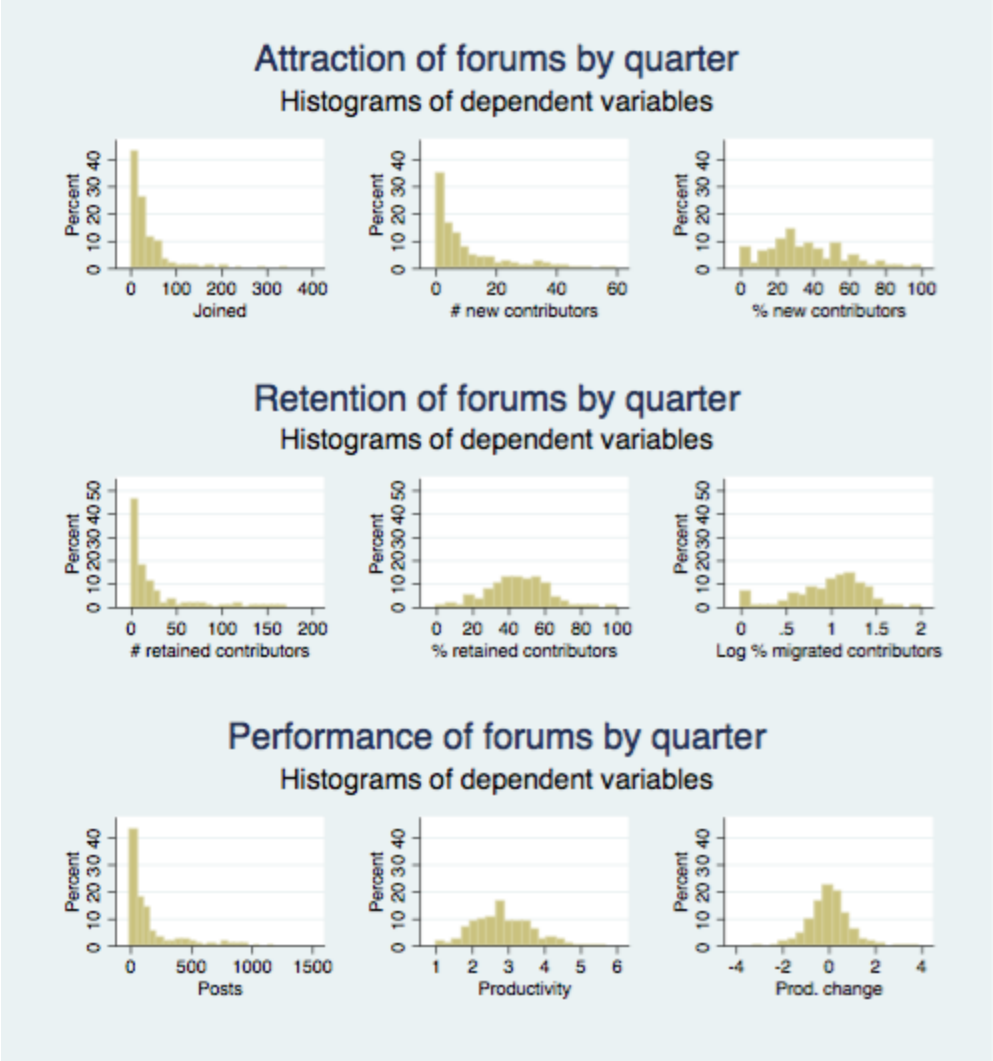


Figure 7: Distribution of the dependent variables that measure sustainability

- *Instability* represents how transient a neighborhood's population is and contains the percentage of rented occupied households, the percentage of occupied households that moved into the neighborhood in 2005 or later, and the percentage of households with an annual income of less than $35,000.

For each component, we chose the most representative factor to characterize the component in our statistical analyses. Size is represented by the total population of the neighborhood. Diversity is measured as the proportion of the population that is people of color. Instability is characterized by the percentage of residents who had moved into the neighborhood in 2005 or later.

We complemented these demographics features with data from the Minneapolis Community Technology Survey.[2] The Information Technology Department of the city of Minneapolis has conducted this annual survey in 2012, 2013 and 2014 to obtain data about the access to and use of computers, mobile technology and Internet among the city residents. The survey is distributed by mail to a random sample of the residents across the eleven communities in the city. We used the raw data of this three-year survey to obtain a measure of Internet access in the E-Democracy forums' target urban communities. We defined the following variable:

- *% with full Internet access*: measures the proportion of survey respondents from each urban community that had access to the Internet both at home and at work.

6.1.3 Online activity: archival data analysis

Using the archival dataset of posts, we measured the characteristics of users' online activity in the forums in terms of the following two measures:

- *Diversity of tenure* is operationalized as the coefficient of variation of the tenure of the users who posted in a forum in a given quarter. For each forum in a given quarter, we computed the tenure of all contributors as the number of minutes since the user's first contribution until the last day of the quarter. Following [30], diversity of tenure was computed as the ratio of the standard deviation of contributor tenure to the mean of the same variable.

[2] http://www.minneapolismn.gov/it/inclusion/WCMS1P-118865 Last retrieved on October 21st, 2015.

- *% membership overlap* represents the ratio of contributors who had posted in two or more forums in the platform at a given quarter, and therefore, had divided their attention into more than one forum. For each contributor of a focal forum, we assessed if the contributor had posted content to another E-Democracy forum in the same quarter. We labeled these contributors as users with membership overlap. To consider this measure at a collective level, we computed the proportion of contributors with membership overlap to the total number of contributors of the forum in the quarter.

6.1.4 Social structures among contributors: social network analysis

We modeled a social network of users according to the following procedure. We created a bipartite network of contributors and threads to describe the online interactions in the online forums. When a user added a message to a thread, a connection between a user node and a thread node was created in the network. When two users participated in the same thread, two user nodes were linked to a single thread node. A projection of this bipartite network over the user nodes illustrates the social structure of users within a forum. Based on this projected user-to-user network (undirected and unweighted), several measures were computed as independent variables.

- *Network size:* The number of user nodes in a forum's social network in the current quarter.
- *Network connectedness:* In order to assess the connectedness of a forum's social network, we first computed individual users' clustering coefficients in the network, and the connectedness was computed as the mean of all users' clustering coefficients in the network. Unlike the network density measure, which is sensitive to the entire size of the network, this measure reflects an average of local connectivity in the social structure; i.e., how embedded each node is in the network. This measure varies from 0 to 1. Larger scores denote more connected social structures.
- *Network centralization:* To capture the centralization or inequality of a forum's social network, we first measured individual users' centrality in the network by degree centrality - the number of connections the users have with other users in the network. Then, following [93], we measured the level of structural concentration using the Gini coefficient of the degree centrality of the forum users in a quarter. As suggested in [93], the Gini coefficient is a measure for iden-

tifying preferential patterns in general networks, as opposed to measures such as power-law exponents that can only apply to networks following power-law distribution. This measure captures the inequality of users' centrality in the social structure. This coefficient can take values from 0 to 1. Larger scores signify more unequal distributions of degree, thus representing more centralized social structures.

6.1.5 Content: manual coding and automatic classification

To further investigate what kinds of content are shared in the online forums for urban communities, we sampled 516 posts that initiated a new thread of conversation in the E-Democracy neighborhood forums and manually annotated them to describe their content. To balance our sample regarding the different rates of responsiveness that they achieve, we randomly sampled roughly equal numbers of posts from three subsets of threads: threads that received no response, threads that received one or two answers, and threads with more than two responses.

Given that the goal of participatory information systems for urban communities is to encourage community involvement and increase social capital, we adopted a coding scheme from prior research [49, 89] that characterizes online posts according to their intention to "mobilize" the social capital available through a participatory information system. Specifically, each of the sampled posts was coded to identify:

- An attempt to mobilize their local community. A post was considered a mobilization when it included an explicit request for action or response. The categories of mobilizations were requests for: (1) Recommendation, (2) Factual knowledge, (3) Opinion/poll, (4) Favor/request/collective action, and (5) Social coordination/invitation/offer.
- The complexity of the attempted mobilization. Complexity was coded in terms of where the requested action was supposed to happen. The options were: (1) In the forum, (2) Somewhere else online (e.g., e-mail, another website), and (3) Offline (e.g., call a phone number, attend a meeting).

Three independent annotators coded the sampled posts after being trained with the coding scheme. A majority vote was used to decide if a post was a mobilization attempt or not. Regarding the other categories, a Cohen's Kappa coefficient was used to assess inter-coder agreement. The

coefficients were 0.67 for kind of mobilization and 0.72 for complexity of mobilization. These inter-coder agreement scores were considered sufficient, and the annotators and researchers discussed further discrepancies until reaching agreement. Table 5 shows examples of local posts and their assigned annotation according to the coding scheme.

To increase the size of coded content, we employed automatic classification algorithms to code all posts that initiated a new thread of conversation in the neighborhood forums. The results of the manual coding were used as a ground truth for the classifiers. We processed the text by running a modified version of a Python code created and made available by Dr. Yu-Ru Lin. The code allowed us to retrieve N-grams (unigrams, bigrams and trigrams) and the count of various linguistic features in the posts, after stemming the text of the posts. Linguistic features, such as pronouns and verbs in past tense, were retrieved by reusing the functions of the Linguistic Inquiry and Word Count (LIWC)[3] package for Python.

N-grams and linguistic data were used as features for the content classification. We used different R[4] packages to run alternative classification algorithms. We used 80% of the coded posts to train the classifiers and compared their performance at classifying the remaining 20% of the content. We assessed the performance of different classification methods such as k-nearest-neighborhoods, decision trees, and support vector machine (SVM). Overall, SVM performed consistently better than the alternative methods. To reduce dimensionality and computation time, we conducted principal component analysis on the N-grams and linguistic features data and explored different strategies to achieve high levels of performance in the automatic classification. Details about the process of classification are provided in Appendix A. In summary, we considered these features in isolation and in conjunction. We also tested different thresholds to filter out very common and uncommon N-grams and to keep the most important components from the principal component analyses. The best results were obtained with 19 components that explain 95% of variance of the linguistic features. Adding the main components of the unigrams and trigrams harmed the performance. The main components from the bigrams performed almost as well as the linguistic features alone. Bigrams generally helped to improve the classification of non-mobilizations, but made the classification of passive mobilizations slightly worse. Therefore, we decided to choose

[3] http://liwc.wpengine.com/

[4] a free software environment for statistical computing

Table 5: Examples of posts by kind of mobilization

Recommendation: "*I am looking for a 'no-Jobs-too-small' handyman, replacing a screen under a porch where I can no longer crawl, etc.. **Anyone have a referral or recommendation?** Thanks*"
Factual knowledge: "*According to the Clean City Minneapolis site on graffiti, we should be able to get graffiti removal wipes from our local community alliance. **Where and when can I access these?** I know that in Whittier the Whittier Alliance has a supply, but I'm not sure the West Bank CDC has the same, especially when their website is a single page with very limited information. There's quite a bit of the unsightly stuff, and I hate to be one who complains that 'someone should do something' without actually doing something.*"
Opinion/poll: "*Do any of you feel invaded when people come to your door, insist on you answering because they continue to ring the bell and then proceed to sell their product or promote their candidate and idea? **Does anyone have any suggestions** for how they deal with this in a positive way? I mentioned this to another person and they thought going door to door would be an effective way to scope out homes and identify vulnerable situations or residents. Does anyone ever ask to see the solicitor's ID?*"
Favor/request/collective action: "*<Name> at <address> has lost his dog 'Georgia' she was last seen yesterday in their yard She has stocky build white with brown spots and has a sweet and shy temperment **please call** <phone number> with any info!*"
Social coordination/invitation/offer: "*NNO event on Hoyt street between Rice and Marion. 4:30-7:30. Block party and everyone is welcome. Free food and fun. **Come join us** and the theme this year is wear blue. Hope to see you as we come together for community*"
Non-Mobilization: "*Please make note of the ramp closures information for our neighborhood. This will have a major impact on commutes and travel in our area.*"

the classifiers that use linguistic components only. The results of this content classification are reported in the next section.

Beyond kinds of content, we were also interested in understanding how the forums reacted to the content that was being posted. Therefore, we measured the rate of online responsiveness in a forum. This variable was computed as the proportion of new threads that were started in a quarter and obtained at least one response to the total number of new threads in the quarter.

6.2 DESCRIPTIVE STATISTICS

The archival data of the sampled 35 neighborhood forums include 75,374 posts that were organized into 32,903 threads of discussion and posted by 5,207 unique users. These posts were collected during the whole lifecycles of the forums until the second quarter of 2014. By the end of our observation period, the neighborhood forums' tenure ranged from one to six years. We also had information from the two city-wide forums since 2005 for Minneapolis and 2006 for St. Paul. These two forums had garnered 92,334 posts that were nested in 25,335 threads and were created by 2,006 unique users.

6.2.1 Sustainability measures

On average, an E-Democracy forum attracts 34 new members and 9 new contributors in a quarter. The maximum number of users who have joined a forum in a quarter is 345 people. The most active neighborhood forum has attracted 60 new contributors in a single quarter. On average, about a third of the forum contributors in a quarter are new contributors who had never posted in the forum before. From one quarter to the next, a neighborhood forum retains 23 of its contributors, which in relative terms make up about half of their contributors. Another 12% of the contributors stop contributing to the forum but do not leave the platform; they instead contribute to other E-Democracy forums (this is what we call migration). In a quarter, the forums garner a mean number of 152 posts. The range of this measure goes from a single post to 1,180 posts in a quarter. On average, the productivity of the forums is 2.8 posts by contributor and the productivity change

is 0.02. This means that on average, the forums increase very slightly their productivity in two consecutive terms. Table 6 shows the descriptive statistics of all dependent variables of this study.

Table 6: Dependent variables that represent sustainability of the neighborhood forums

Dependent variable	Mean	Std. Dev.	Min	Max
# joined (new members)	34.308	39.381	1	345
# new contributors	9.593	11.638	0	60
% new contributors	34.641	23.187	0	100
# retained contributors	23.519	34.932	0	171
% retained contributors	48.932	23.446	0	100
% migrated contributors	12.4805	13.168	0	100
# posts	152.645	215.039	1	1180
Productivity	2.845	1.037	1	9
Productivity change	0.024	1.313	-6.250	6.667

6.2.2 Community aspects

The forums' target neighborhoods and districts vary in their characteristics. Table 7 shows the mean, dispersion and range of the demographic variables and the Internet access measure to be considered in this study.

The sampled E-Democracy forums target areas with population sizes ranging from 2,833 to 36,255 inhabitants. Out of the total population in an area, the percentage of people of color ranged from 10.2% to 86.3%. The percentage of households with new residents (those who moved into the neighborhood in 2005 or later) varied from 16% to 60.2%. The neighborhoods were also heterogeneous in terms of Internet access. On average, 52% of a neighborhood's residents had full

Table 7: Demographics of the forums' target urban communities

Independent variable	Mean	Std. Dev.	Min	Max
Size: Population	12,412	8.527	2,833	36,255
Diversity: % of people of color	42.89	22.31	10.2	86.3
Instability: % moved in 2005 or later	35.53	10.82	16	60.2
% full Internet access	52.09	12.18	24.32	69.27

access to the Internet, but this variable had a range from 24.32% to 69.27%. This variation allowed us to study the impact of these community aspects on the sustainability of their online forums.

6.2.3 Online activity

Among the online measures, there was also a great deal of variation (see Table 8). All of the variables have a wide range of values, which often includes the minimum and maximum possible value of the measures. On average, slightly less than half of the contributors of a forum also contributed to other E-Democracy forums. The average forum had considerable levels of tenure diversity among their contributors, with the standard deviation of tenure being larger than the average (coefficient of variation = 1.12).

Table 8: Online measures of neighborhood forums by quarter

Independent variable	Mean	Std. Dev.	Min	Max
Div. tenure	1.12	0.34	0.13	2.28
% membership overlap	47.60	24.76	0.00	100.00

6.2.4 Social networks

The size of the social networks of contributors ranged from a single user to 284 users in a quarter. Given that the values for centralization and connectedness could go from zero to one by definition, we can say that the average forum had mid-levels of centralization and slightly higher levels of connectedness. Table 9 shows the descriptive statistics of these measures.

Table 9: Social network measures of neighborhood forums by quarter

Independent variable	Mean	Std. Dev.	Min	Max
Network size	53.40	64.06	1.00	284.00
Network connectedness	0.63	0.21	0.00	1.00
Network centralization	0.49	0.26	0.00	0.97

6.2.5 Content

With regard to content, the outcomes of the manual coding indicate that residents mostly use local online forums to mobilize their local communities. Our content analysis of 516 annotated posts shows that a large majority (83%) of the posts that initiated a thread were mobilization requests (Table 10). This can be considered evidence that online forums for local communities are being used for exercising or attempting to obtain benefits from the social capital available through participatory information systems.

Mobilization attempts include explicit requests for different kinds of actions. Requests coded as *social coordination/invitation/offers* are the most frequent. Almost half of the sample (47%) was coded as such. These posts were usually invitations to local events and offers of some sort of service available within the geographical community (see examples in Table 5). The category *Favor/request/collective action* accounts for the second-largest share of annotated messages (20%).

Table 10: Distribution of the kinds of mobilization among the manually-coded posts

Mobilization kind	Frequency	Off-site
Recommendation	35 (.07)	3 (.08)
Factual knowledge	24 (.05)	0 (.00)
Opinion/poll	28 (.05)	6 (.21)
Favor/request/collective action	103 (.20)	72 (.70)
Social coordin./invitation/offer	240 (.47)	229 (.95)
Non-mobilization	86 (.17)	

This group includes requests for information related to crime investigations, lost pets and calls for neighbors to act collectively on a problematic issue. Other kinds of mobilizations, such as requests for *recommendations*, *factual knowledge* and *opinion/poll*, cover much smaller proportions of our sample (5 - 7%). The remaining 17% of the sample was coded as *non-mobilization*. These messages were usually announcements of factual information. They did not explicitly request any actions from the readers, but rather attempted to create awareness about a locally-relevant fact (see Table 5).

Overall, 310 out of 430 mobilization attempts promote off-site actions, which were supposed to happen either offline (e.g. call a phone number, join a party) or in another online system (e.g. by e-mail, another website). Messages annotated as *social coordination/invitation/offer* were almost always classified as off-site actions. On the other hand, *recommendations* and *factual knowledge* were almost always coded as on-site actions.

Given the uneven proportions of the different kinds of mobilization requests, all automatic classification algorithms achieved low levels of performance when we trained them to identify all of these kinds of content. The performance improved considerably when the most infrequent kinds of mobilizations were grouped together. The best results were obtained with SVM classifiers that categorize content into three labels: "non-mobilization"; "active mobilization" which represents the label *social coordination/invitation/offer*; and "passive mobilization" which comprises *recom-*

mendation, *factual knowledge*, *opinion/poll* and *favor/request/collective action*. This distinction is still considered useful for our analysis as it could distinguish *social coordination/invitation/offer* from other kinds of content. *Social coordination/invitation/offer* is the most common kind of content in the neighborhood forums and it is also the kind of content that almost always requested off-site responses instead of online ones.

The automatic classifier with the best performance uses linguistic features only (see Section 6.1.5). Table 11 shows the confusion matrix of predicted and true labels for the testing set of posts. Overall, the classifier has an accuracy level of 0.7037, which is considered acceptable. A closer look at the results of the predicted values by label (see Table 12) reveals that the classifier achieves very high performance at classifying *active mobilizations* (high sensitivity, precision and detection rate). However, it fails more often at classifying the other two labels. Given that we have more confidence in the classification of *active mobilizations*, we focus our data analysis on that kind of content. Nevertheless, we present the results associated with all kinds of content for the completeness of our report.

Given that linguistic features are the only input of our classifiers, we present here a description of these features according to the ground truth. Figure 8 shows the proportions of features by each of the three labels that we used in the automatic content classification. More details about the linguistic differences among the six original content labels are provided in Appendix A. The three-label coding scheme reveals the following linguistic patterns:

Table 11: Confusion matrix of the automatic content classification

	True labels		
Predicted labels	Non-mobilization	Passive mobilization	Active mobilization
Non-mobilization	8	8	3
Passive mobilization	7	25	4
Active mobilization	1	9	43

Table 12: Performance of the automatic content classification by kind of mobilization

Predicted labels	Non-mobilization	Passive mobilization	Active mobilization
Sensitivity (recall)	0.50000	0.5952	0.8600
Specificity	0.88043	0.8333	0.8276
Pos Pred Value (precision)	0.42105	0.6944	0.8113
Neg Pred Value	0.91011	0.7639	0.8727
Prevalence	0.14815	0.3889	0.4630
Detection Rate	0.07407	0.2315	0.3981
Detection Prevalence	0.17593	0.3333	0.4907
Balanced Accuracy	0.69022	0.7143	0.8438

- *Non-mobilizations* tended to be the longest posts and had the smallest proportions of words that represent affective processes, second person pronouns, and leisure-related words. They also had the maximum ratios of first person pronouns and verbs in past tense.
- *Passive mobilizations* had the largest ratio of words associated with home, cognitive processes (especially tentative words), and verbs in present tense.
- *Active mobilizations* had the smallest proportions of function words (particularly first person pronouns) and verbs in past tense. They also had the largest proportions of positive emotions, verbs in future tense, and words related to space, time, work and leisure.

Based on these kinds of linguistic features and after a process of dimension reduction, the automatic classifier was used to predict the kind of content of 32,362 posts that initiated a new thread of conversation in an E-Democracy neighborhood forum. Some posts could not be automatically classified because they had paragraphs written in a language other than English or had no words that were identified by our algorithm. Table 13 shows the number and proportion of posts that were categorized into each kind of content. The predicted values estimate the prevalence of *active mobilizations* to be 47.19% of the content, which is very similar to the ratio found in our ground

Label	observations	sentcnt	wc	dic_wc	funct	social	affect	cogmech	percept
Non mobilizations	119	10.50	206.70	159.37	0.61	0.12	0.05	0.17	0.02
Passive mobilizations	190	8.36	147.56	114.10	0.61	0.14	0.06	0.20	0.02
Active mobilizations	240	8.75	170.60	122.94	0.53	0.14	0.06	0.17	0.02

Label	posemo	negemo	tentat	p1	p2	p3	past	present	future
Non mobilizations	0.04	0.01	0.03	0.05	0.01	0.01	0.04	0.07	0.01
Passive mobilizations	0.04	0.01	0.04	0.04	0.02	0.02	0.03	0.09	0.01
Active mobilizations	0.05	0.01	0.03	0.03	0.02	0.01	0.01	0.07	0.02

Label	motion	space	time	work	achieve	leisure	home	money	death
Non mobilizations	0.03	0.10	0.07	0.04	0.02	0.01	0.02	0.01	0.00
Passive mobilizations	0.02	0.09	0.06	0.03	0.02	0.02	0.03	0.01	0.00
Active mobilizations	0.03	0.11	0.08	0.06	0.03	0.03	0.02	0.01	0.00

Figure 8: Linguistic features by the three kinds of mobilizations

truth (47%, see Table 10). Among the other kinds of content, the proportion of *passive mobilizations* in the full set of tests is 33.76%, slightly smaller than the ratio in the manually-coded posts. The remaining 19.09% are classified as *non-mobilizations*, which is a somewhat higher than the proportion of this kind of content in the set of manually-annotated posts.

The predicted labels were later used to measure the proportion of each kind of content in the neighborhood forums by quarters. Table 14 shows the descriptive statistics of such measures. The average neighborhood forum had 53.89% of posts requesting *active mobilizations*, such as events or offers that require social coordination (range: 14% - 100%). The mean percentage of *passive mobilizations* and *non-mobilizations* are about 30% and 20%, respectively.

Table 14 also includes the average level of responsiveness in the forums. The data reveals that approximately 30% of the posts starting a new thread in a forum at a given quarter received at least one response. Thus, on average, most new threads in a local forum did not receive any response.

6.2.6 Some data considerations

While our goal was to assess the impact of all of these independent variables (community aspects, online participation, social networks and content) on sustainability of the E-Democracy forums, preliminary analysis revealed that this would not be feasible due to some features of our data.

Table 13: Distribution of the kinds of mobilization among the automatically-coded posts

Non-mobilization	Passive mobilization	Active mobilization
6,164	10,926	15,272
(19.05%)	(33.76%)	(47.19%)

First, there was high levels of multi-collinearity among some of the independent variables. The size of the network of contributors were highly correlated to the measure of centralization in the network (larger networks were more centralized) and the total population of the neighborhoods (the more inhabitants, the larger the network). A multi-collinearity analysis revealed that the condition number was higher than 30 when we included all measures of online participation, social networks, and demographics. The results of the analyses also became inconsistent when we included size of the networks. Therefore, we had to drop this variable from our analyses.

Additionally, the social network measures and responsiveness were not independent from each other. Hence, we could not include both kinds of variables in the same regression analysis. The reason behind the interdependency between these variables is that the social networks were created by using data about who has responded to whom in the forums.

The Internet access data was only available for the neighborhoods in Minneapolis. Therefore,

Table 14: Content measures of neighborhood forums by quarter

Independent variable	Mean	Std. Dev.	Min	Max
% active mobilizations	53.89	16.44	14.29	100
% passive mobilizations	29.93	13.20	5.56	100
% non-mobilizations	19.68	11.28	2.70	100
% responded new threads	29.38	15.55	0	100

we ran a separate analysis to assess the impact of Internet access. However, we did not use this measure as an independent variable in all analyses in order to be able to use all data from St. Paul in the remaining analyses.

To deal with these problems but still explore the different aspects of the proposed framework, we decided to present all regression analyses in sets of different types of variables that could be related to sustainability. This approach allowed us to assess which combinations of independent variables are able to better explain the variance of our dependent variables.

Given that the values of the total population had a much larger range than any other variable, we log-transformed them before using it in the analyses. When we used dependent variables that were relative measures (e.g., the proportion of new contributors or proportion of retained users) we constrained our analysis to observation periods where the number of contributors in the prior quarter was four or more. We did this because we reason that these relative measures are meaningless when the number of contributors to a forum in a previous quarter is too small.

6.3 TESTING THE FRAMEWORK: COLLECTIVE ASPECTS

To examine the impact of the neighborhood forums' offline and online characteristics on their future sustainability, we used longitudinal models to estimate the association between all online factors at a given quarter (time *t*) and the measures of attraction, retention and performance at the following quarter (time *t+1*) while controlling for the effect of the offline characteristics of the forums' target geographical areas, the time of creation of the forum, and the tenure of the forum at each observation period.

Given that our analyses involve a large number of independent and dependent variables, they have been organized into subsections that focus on each aspect of sustainability: attraction, retention, and performance. In each subsection, we describe the most consistent patterns of influence of the independent variables on the three specific dependent variables that measure each aspect of sustainability. The results of our analysis of attraction are shown in Tables 15 - 20. Tables 21 - 26 describe the results of the analysis of retention. The output of the analysis of performance are depicted in Tables 27 - 32. Each set of tables are organized according to the three dependent

variables that represent each aspect of sustainability. There are two tables for each dependent variable. The first table describes the results of the regressions that use the following independent variables: (1) demographic community aspects, (2) all community aspects,[5] (3) demographics and online activity, and (4) demographics and both online activity and social network factors. The second table describes the results of adding content variables to the demographic variables. These tables show the results of three alternative analyses that include responsiveness to new threads and (1) the ratio of active mobilizations, (2) the ratio of passive mobilizations, and (3) the ratio of non-mobilizations. In all of these regression analyses, we controlled for the time of creation of the forums and their tenure at the period of observation.

6.3.1 Attraction

Among the community demographics, size and instability of the neighborhoods are related to the absolute values of attraction (number of new members and new contributors), but not the relative metric (ratio of new contributors) (see Tables 15, 17, and 19). While population size is positively related to the number of people who decide to be part of the forums, the instability of the neighborhoods' populations is negatively associated with the same variables. Forums belonging to larger neighborhoods attract more people. Those that belong to neighborhoods with larger proportions of new residents engage fewer people. The relationship between population size and absolute measures of attraction is often, but not always, statistically significant. Generally, the relationship becomes statistically insignificant when accounting for variables that represent or might encourage interaction in the forums (e.g., connectedness of the network, responsiveness, and neighborhood access to the Internet). Population diversity was not a significant factor on any of the measures of forum attraction.

Once we added the measure that comes from the Minneapolis Community Technology Survey, we found that wider access to the Internet in the neighborhoods is significantly related to higher attraction of new members. Forums belonging to neighborhoods with larger proportions of residents who have access to the Internet, both at home and at work, are more likely to attract larger audiences.

[5] These regressions had fewer observations because the Minneapolis Community Technology Survey data was only available for neighborhoods in the city of Minneapolis.

The next analyses of attraction assess the role of online aspects. As we had longitudinal data, we evaluated how the attraction measures in a quarter are related to online measures in the prior quarter (denoted as L. in the Tables). We also controlled for the effect of the same variable as measured in the prior quarter. We made this decision as a way to control for the effect of hidden variables that could affect the decision to join a forum in every quarter, but that we were not able to measure. An analogous approach was taken in the regression analyses conducted for all aspects of sustainability.

Regarding online participation, the analyses indicate that diversity of tenure among the forums contributors is positively related to all variables of attraction, but is only consistently significant for the absolute measures of sustainability. Neighborhood forums made up of a combination of old-timers and new contributors are better at attracting new members and encouraging active participation in the subsequent period.

The ratio of members who had participated in other E-Democracy forums has a somewhat mixed influence on attraction. While larger membership overlap helps to bring new members to a forum in the next quarter, it has a negative relationship with the relative measure of attraction of new contributors. This reveals that membership overlap helps to broaden the audience of the forums, but can harm contribution levels in the long-term.

The results are less conclusive for the role of connectedness and centralization of the social network. Connectedness is negatively related to attraction of new members and new contributors to the forums; however, it is only statistically significant with regard to attraction of new members. Forums that exhibit more connected social networks attract fewer people as new members in a subsequent quarter. However, the impact of this variable on the proportion of new contributors is positive but not significant. In turn, even though only the positive relationships between centralization and attraction are statistically significant, the sign of the association varies across different variables of attraction. As this association is not supported by prior research, we conclude that further research is needed to confirm it.

Finally, we report the results of alternative regressions that assess the impact of responsiveness and the proportions of different kinds of content on attraction (Tables 16, 18 and 20). While the ratio of posts that received at least one response is negatively related to attraction of new members, it is positively associated with the number and proportion of new contributors. It is possible that

responsiveness is perceived negatively (e.g., controversy, fights) by outsiders and thus prevents people from joining the forums. However, once people are part of the forums, responsiveness might be understood as feedback in a more positive way and it encourages people to start posting.

The kinds of content are related to the absolute measures of attraction, but not the relative ones. Larger proportions of passive mobilizations are significantly associated with more new members and new contributors in the next quarter. The positive relationship between the proportion of active mobilization and attraction is positive but not statistically significant. Together, these results provide support for a positive effect of mobilization requests on absolute measures of attraction.

Across all of the reported regressions, we controlled for the time of creation of each forum and the tenure of the forum at the observation period. Consistently, we found that these two variables are significantly associated with the levels of attraction of the forums. Forums that were created in 2011 or later are less likely to engage new users and new contributors. These newer forums also attract smaller proportions of new contributors. This indicates that forums created more recently struggle more than older forums to attract residents. This can be due to the larger number of alternative participatory information systems that have become available for urban communities in the past few years. Newer forums might have more trouble attracting an initial base of active contributors that makes the forums thrive. Forums that were created earlier might have had the chance to create such a user base, which in turn has helped to attract more people.

Tenure has a more nuanced relationship with attraction. There is partial evidence that forums attract larger number of new contributors over time, but they engage fewer new members and smaller proportions of new contributors. This behavior might be reflecting that more new members are engaged in early phases of the forums' lifecycle and, over time, more of these members contribute content to the systems. However, the rate to which members become contributors (transitioning from readers of content to producers of content) decreases over time.

6.3.2 Retention

While the neighborhoods' population size has a significant role in some measures of attraction, it has a less relevant impact on retention measures. Even though the relationships are often positive, they were only significant in very few cases (See Tables 21, 23 and 25).

Table 15: Attraction: Number of new members

	(1) Joined	(2) Joined	(3) Joined	(4) Joined
Log population	2.762***	1.931	2.796*	1.354
% of color	1.001	1.003	0.999	0.997
% moved 2005+	0.977**	0.981*	0.963*	0.963**
% full int. access		1.024*		
L.Joined			0.998***	0.998***
L.Div. user tenure			2.517***	2.666***
L.% overlap			1.021***	1.020***
L.Connectedness				0.645***
L.Centralization				1.288**
Created 2011+	0.514***	0.306***	0.240***	0.319***
F. tenure	0.976***	0.988***	0.970***	0.982***
Observations	435	253	335	305
AIC	11447.8	2874.7	8109.0	6562.2
BIC	11476.3	2902.9	8147.1	6606.9

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 16: Attraction: Number of new members using content variables

	(1) Joined	(2) Joined	(3) Joined
Log population	3.056**	3.897***	2.489**
% of color	0.999	1.003	0.999
% moved 2005+	0.974**	0.967***	0.978**
L.Joined	0.999***	0.998***	0.998***
L.% active mobilizations	1.001		
L.% passive mobilizations		1.011***	
L.% non-mobilization			0.977***
L.% responded	0.995***	0.993***	0.997***
Created 2011+	0.379***	0.378***	0.338***
F. tenure	0.957***	0.957***	0.955***
Observations	407	383	365
AIC	10251.0	9576.2	8067.0
BIC	10291.0	9615.7	8106.0

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 17: Attraction: Number of new contributors

	(1)	(2)	(3)	(4)
	# new	# new	# new	# new
Log population	3.782*	4.893	2.585*	1.193
% of color	1.004	1.016	1.001	1.001
% moved 2005+	0.934***	0.927***	0.955**	0.959**
% full int. access		1.037		
L.# new			1.008***	1.008***
L.Div. user tenure			1.596***	1.627***
L.% overlap			1.001	0.998
L.Connectedness				0.886
L.Centralization				0.956
Created 2011+	0.308**	0.139***	0.314***	0.425***
F. tenure	0.999	1.001	1.012***	1.015***
Observations	481	291	375	344
AIC	3548.0	1972.2	2856.2	2600.3
BIC	3577.2	2001.6	2895.4	2646.4

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 18: Attraction: Number of new contributors using content variables

	(1)	(2)	(3)
	# new	# new	# new
Log population	2.805*	2.967*	2.399*
% of color	1.002	1.003	1.000
% moved 2005+	0.958**	0.959**	0.961**
L.# new	1.010***	1.009***	1.009***
L.% active mobilizations	1.001		
L.% passive mobilizations		1.006*	
L.% non-mobilization			0.989***
L.% responded	1.006**	1.005*	1.009***
Created 2011+	0.363***	0.367***	0.345***
F. tenure	0.999	0.998	0.997
Observations	375	373	366
AIC	2887.1	2875.3	2744.0
BIC	2926.3	2914.5	2783.0

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 19: Relative attraction: Proportion of new contributors

	(1)	(2)	(3)	(4)
	% new	% new	% new	% new
Log population	0.156	−0.067	0.069	0.043
% of color	0.053	−0.245	0.068	0.094
% moved 2005+	−0.046	0.293	−0.026	−0.061
% full int. access		−0.021		
L.Div. user tenure			0.238*	0.213
L.% overlap			−0.201**	−0.089
L.Connectedness				0.008
L.Centralization				0.128*
Created 2011+	−0.188	−0.306	−0.250**	−0.211*
F. tenure	−0.498***	−0.565***	−0.358***	−0.443***
Observations	392	235	382	351
R^2	0.137	0.142	0.275	0.277

Standardized beta coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 20: Relative attraction: Proportion of new contributors using content

	(1)	(2)	(3)
	% new	% new	% new
Log population	0.084	0.124*	0.077
% of color	0.102	0.123*	0.085
% moved 2005+	−0.023	−0.048	−0.052
L.% new	0.273***	0.312***	0.288***
L.% active mobilizations	−0.036		
L.% passive mobilizations		0.120	
L.% non-mobilization			−0.004
L.% responded	0.157**	0.152**	0.195***
Created 2011+	−0.129*	−0.114	−0.134*
F. tenure	−0.292***	−0.257***	−0.275***
Observations	374	372	366
R^2	0.284	0.316	0.292

Standardized beta coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

On the contrary, diversity and instability of the neighborhoods' populations are more consistently related to absolute and relative measures of contributor retention. Forums belonging to more diverse neighborhoods retain significantly more of their contributors and at larger proportions. In turn, forums of neighborhoods with larger proportions of new residents (who moved into the neighborhood in 2005 or later) keep fewer contributors and at smaller rates.

Similar to the results about attraction of new members, higher access to the Internet is significantly associated with higher absolute and relative measures of retention. This confirms the important role of Internet availability in sustainable participation of residents as content providers.

Regarding online participation, there is weak evidence that diversity of tenure plays a role in retention. Membership overlap has a more consistent impact on retention. Neighborhood forums that have larger proportions of users who also contribute to other forums retain fewer of their contributors from one quarter to the next. From those who leave these forums, larger proportions do not leave the platform but migrate to other E-Democracy forums. Together with the results of attraction, the data analysis reveals that membership overlap helps to attract new members to a forum, but also brings the risk of compromising the number of contributors in the long-term.

The social structure of contributors has a milder effect on retention. Centralization is not a significant factor on the measures of this aspect of sustainability. Besides, connectedness is only significantly associated with the proportion of people that migrate to other E-Democracy forums in a subsequent quarter. Forums that have more connected social structures have larger proportions of members moving to another forum within the E-Democracy platform. This means that while connectedness might not help retain users in the forums, it positively influences retention within the E-Democracy platform.

In regard to content, both the proportion of the kinds of content and responsiveness have a significant impact on the retention variables (see Tables 22, 24, and 26). While larger shares of active mobilizations in a forum are related to smaller proportions of retained users, larger proportions of requests for passive mobilizations are expected to increase the number of contributors that remain active from one quarter to the next. In turn, the impact of responsiveness on the other measures of retention is more dubious. More responsiveness in a forum positively affects the number of retained contributors in the following quarter, but this effect becomes negative and not significant when considering a relative measure of retention. Responsiveness is also significantly associated

with smaller rates of contributors migrating to other E-Democracy forums.

Time is a significant aspect on retention measures. Forums that were created in 2011 or later are expected to retain significantly fewer contributors; however, this effect is not significant when considering a relative measure of contributor retention. The impact of forum tenure has a less consistent pattern. Over time, and after controlling for other measures, the neighborhoods forums are expected to maintain more of their contributors from one quarter to the next. However, this effect is negative but not significant in a model to estimate the proportion of retained contributors.

Overall, the count and relative measures of the retention of contributors are strongly related to several community and content aspects, but have very weak relationships with the measures based on online interaction among users in the forums. Contrarily, the proportion of contributors who migrate from one E-Democracy forum to another is not influenced by any of the community aspects. The behavioral measures that describe the online interactions among users are the only factors that show significant associations with this dependent variable.

Table 21: Retention: Number of retained users

	(1)	(2)	(3)	(4)
	# retained	# retained	# retained	# retained
Log population	1.909	4.557*	1.416	1.063
% of color	1.016*	1.036***	1.011*	1.009
% moved 2005+	0.932***	0.922***	0.964**	0.962**
% full int. access		1.046**		
L.# retained			1.007***	1.007***
L.Div. user tenure			1.042	1.038
L.% overlap			0.991***	0.991***
L.Connectedness				0.867
L.Centralization				1.217
Created 2011+	0.340***	0.204***	0.433***	0.478***
F. tenure	1.046***	1.043***	1.022***	1.021***
Observations	457	278	342	314
AIC	3757.8	2557.8	2604.0	2491.1
BIC	3786.7	2586.8	2642.3	2536.1

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 22: Retention: Number of retained users using content

	(1)	(2)	(3)
	# retained	# retained	# retained
Log population	1.422	1.469	1.306
% of color	1.013*	1.014*	1.012*
% moved 2005+	0.960**	0.961**	0.962**
L.# retained	1.006***	1.006***	1.006***
L.% active mobilizations	0.999		
L.% passive mobilizations		1.008***	
L.% non-mobilization			0.987***
L.% responded	1.014***	1.012***	1.016***
Created 2011+	0.424***	0.430***	0.410***
F. tenure	1.023***	1.023***	1.021***
Observations	342	340	335
AIC	2552.5	2525.2	2481.8
BIC	2590.9	2563.5	2519.9

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 23: Relative retention: Proportion of retained contributors

	(1)	(2)	(3)	(4)
	% retained	% retained	% retained	% retained
Log population	0.001	0.195*	−0.030	−0.086
% of color	0.183*	0.564***	0.172*	0.134
% moved 2005+	−0.267**	−0.542***	−0.187	−0.110
% full int. access		0.326***		
L.Div. user tenure			−0.132	−0.066
L.% overlap			−0.101	−0.116
L.Connectedness				−0.034
L.Centralization				−0.100
Created 2011+	−0.319***	−0.276***	−0.273***	−0.299**
F. tenure	−0.115	−0.088	−0.179**	−0.160*
Observations	373	226	365	336
R^2	0.140	0.265	0.167	0.159

Standardized beta coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 24: Relative retention: Proportion of retained contributors using content

	(1)	(2)	(3)
	% retained	% retained	% retained
Log population	0.045	0.027	0.005
% of color	0.236**	0.225*	0.163
% moved 2005+	−0.281***	−0.289***	−0.264**
L.% retained	0.007	0.005	−0.017
L.% active mobilizations	−0.163**		
L.% passive mobilizations		0.082	
L.% non-mobilization			0.080
L.% responded	−0.138	−0.097	−0.096
Created 2011+	−0.341***	−0.371***	−0.354***
F. tenure	−0.134**	−0.151***	−0.147**
Observations	342	340	335
R^2	0.198	0.181	0.140

Standardized beta coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 25: Relative retention: Proportion of migrated contributors

	(1)	(2)	(3)	(4)
	Log % migrated	Log % migrated	Log % migrated	Log % migrated
Log population	−0.034	−0.113	0.080	0.108
% of color	0.046	0.061	0.005	0.079
% moved 2005+	0.083	0.070	0.017	−0.101
% full int. access		−0.194		
L.Div. user tenure			−0.227*	−0.189*
L.% overlap			0.392***	0.429***
L.Connectedness				0.101*
L.Centralization				−0.079
Created 2011+	0.084	0.051	0.071	−0.000
F. tenure	0.199***	0.174**	0.051	0.070
Observations	373	226	365	336
R^2	0.017	0.075	0.314	0.371

Standardized beta coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 26: Relative retention: Proportion of migrated contributors using content

	(1)	(2)	(3)
	Log % migrated	Log % migrated	Log % migrated
Log population	0.023	−0.013	0.060
% of color	−0.051	−0.104	0.021
% moved 2005+	0.154	0.089	0.065
L.Log % migrated	−0.054	−0.004	−0.042
L.% active mobilizations	−0.016		
L.% passive mobilizations		−0.099	
L.% non-mobilization			0.065
L.% responded	−0.176**	−0.220***	−0.161*
Created 2011+	0.074	0.052	0.063
F. tenure	0.180**	0.149*	0.187**
Observations	342	340	335
R^2	0.081	0.151	0.063

Standardized beta coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6.3.3 Performance

Similar to the case of retention, diversity and instability of the neighborhoods' populations have consistently significant relationships with the performance of their forums. Compared to forums of homogeneous neighborhoods, forums of more diverse neighborhoods are expected to garner more posts, be more productive as a group, and have more positive productivity changes from one quarter to the next. On the contrary, forums belonging to neighborhoods with less stable populations have worse performance at generating content. The more newly-arrived residents that a neighborhood has, the less content their forums garner in a quarter. More unstable neighborhoods are also associated with less productive forums and less positive changes in productivity between two consecutive quarters. There is weak evidence that the neighborhoods' population size has significant relationships with their forums' performance measures.

We also found that the forums serving neighborhoods with wider Internet access are expected to receive a larger number of posts. These forums are also expected to be more productive. We conclude that when more neighborhood residents have access to Internet from home and work, their local forums become more likely to have active streams of content.

The number of posts in a quarter is significantly related to measures of the online characteristics of the forums in the previous quarter. Larger diversity of tenure and smaller percentages of membership overlap are associated with more posts in a subsequent calendar quarter. More centralized and more connected social networks are expected to generate a smaller number of posts in the next period of observation. Higher prevalence of active mobilizations and more responsiveness are also associated with a larger number of posts in the following quarter.

Time, again, plays a significant role. Newer forums are expected to collect fewer posts than older forums. Similar to the results of attraction, this data analysis shows that newer forums have more difficulty becoming sustainable in terms of performance measures. Furthermore, tenure has a mixed impact. Over time, the forums are likely to garner more posts; however, they are also expected to be less productive and have less positive changes in productivity.

In general, the results indicate that the number of posts is influenced by time and offline and online aspects. However, we found no evidence that productivity and change in productivity are influenced by any online measure that we considered in the regressions. Neither online participa-

tion nor social network nor content aspects are significantly related to productivity or productivity change, after controlling for time variables and demographics. We interpret this as evidence that there are other hidden variables, which we are not measuring, that more reliably determine both productivity and productivity change.

Table 27: Performance: Number of posts

	(1) Posts	(2) Posts	(3) Posts	(4) Posts
Log population	2.347	5.598*	1.812	1.180
% of color	1.018*	1.036***	1.013*	1.012*
% moved 2005+	0.928***	0.915***	0.955***	0.951***
% full int. access		1.049**		
L.Posts			1.001***	1.001***
L.Div. user tenure			1.080***	1.152***
L.% overlap			0.996***	0.996***
L.Connectedness				0.823***
L.Centralization				0.898*
Created 2011+	0.354***	0.198***	0.428***	0.494**
F. tenure	1.039***	1.036***	1.022***	1.023***
Observations	481	291	383	351
AIC	17068.4	11119.6	12681.6	11959.6
BIC	17097.6	11149.0	12721.1	12005.9

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 28: Performance: Number of posts using content

	(1) Posts	(2) Posts	(3) Posts
Log population	1.782	1.834	1.670
% of color	1.012*	1.013*	1.012*
% moved 2005+	0.952***	0.953***	0.957**
L.Posts	1.001***	1.001***	1.001***
L.% active mobilizations	1.006***		
L.% passive mobilizations		0.999	
L.% non-mobilization			0.990***
L.% responded	1.007***	1.005***	1.006***
Created 2011+	0.404***	0.420***	0.409***
F. tenure	1.020***	1.021***	1.019***
Observations	383	381	374
AIC	12605.8	12688.0	12327.2
BIC	12645.3	12727.5	12366.5

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 29: Relative performance: Productivity

	(1) Productivity	(2) Productivity	(3) Productivity	(4) Productivity
Log population	-0.014	0.200	-0.031	-0.013
% of color	0.446^{***}	0.766^{***}	0.463^{***}	0.489^{***}
% moved 2005+	-0.349^{**}	-0.700^{***}	-0.288^{*}	-0.242
% full int. access		0.259^{*}		
L.Div. user tenure			-0.061	-0.038
L.% overlap			-0.083	-0.049
L.Connectedness				-0.086
L.Centralization				0.106
Created 2011+	0.011	0.066	0.050	0.019
F. tenure	-0.089	-0.113	-0.149	-0.179
Observations	393	236	383	351
R^2	0.138	0.238	0.161	0.201

Standardized beta coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 30: Relative performance: Productivity using content

	(1)	(2)	(3)
	Productivity	Productivity	Productivity
Log population	−0.037	−0.057	−0.056
% of color	0.326***	0.334***	0.344**
% moved 2005+	−0.225	−0.270*	−0.225
L.Productivity	0.265***	0.234**	0.263***
L.% active mobilizations	−0.031		
L.% passive mobilizations		0.041	
L.% non-mobilization			−0.019
L.% responded	−0.017	−0.023	−0.001
Created 2011+	0.018	0.043	0.003
F. tenure	−0.128*	−0.135*	−0.129*
Observations	383	381	374
R^2	0.258	0.237	0.251

Standardized beta coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 31: Relative performance: Change in Productivity

	(1)	(2)	(3)	(4)
	Prod. change	Prod. change	Prod. change	Prod. change
Log population	-0.076	-0.054	-0.102	-0.073
% of color	-0.056	-0.069	-0.069	0.010
% moved 2005+	0.007	-0.116	0.001	0.032
% full int. access		-0.080		
L.Div. user tenure			0.180^{*}	0.212^{**}
L.% overlap			0.034	0.066
L.Connectedness				0.211^{**}
L.Centralization				0.039
Created 2011+	0.015	0.083	-0.042	-0.072
F. tenure	-0.138^{***}	-0.136^{**}	-0.065	-0.032
Observations	383	231	383	351
R^2	0.019	0.020	0.027	0.082

Standardized beta coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 32: Relative performance: Change in Productivity using content

	(1)	(2)	(3)
	Prod. change	Prod. change	Prod. change
Log population	-0.036	-0.055	-0.054
% of color	0.316^{***}	0.323^{***}	0.331^{**}
% moved 2005+	-0.218	-0.261^{*}	-0.216
L.Productivity	-0.711^{***}	-0.737^{***}	-0.706^{***}
L.% active mobilizations	-0.030		
L.% passive mobilizations		0.040	
L.% non-mobilization			-0.018
L.% responded	-0.016	-0.022	-0.001
Created 2011+	0.017	0.041	0.003
F. tenure	-0.124^{*}	-0.131^{*}	-0.124^{*}
Observations	383	381	374
R^2	0.314	0.307	0.320

Standardized beta coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

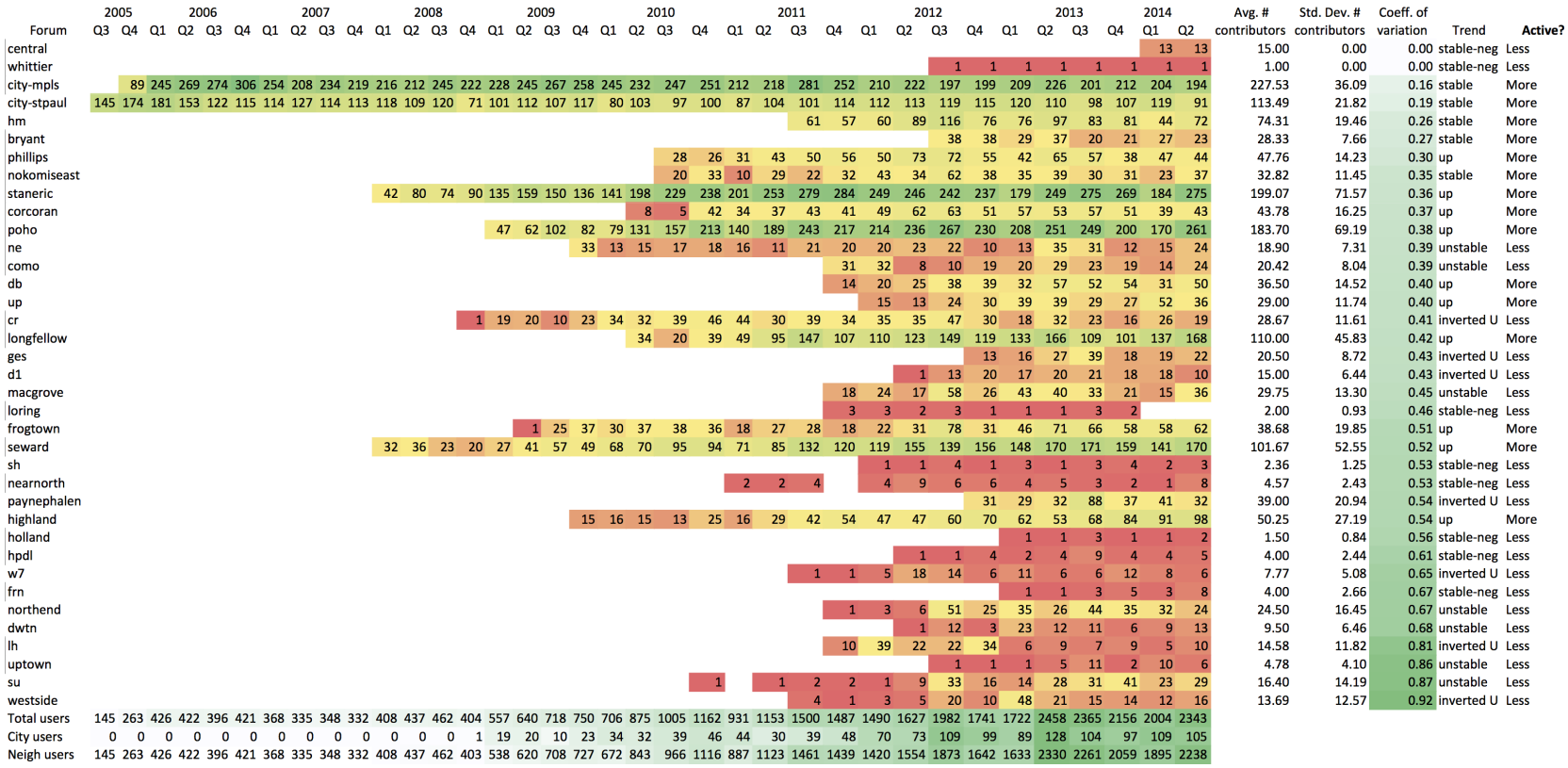
6.4 SUSTAINABILITY, NUMBER OF CONTRIBUTORS, AND SCALE

Given that our preliminary data analysis revealed that network size (i.e., the number of contributors) is highly correlated to other independent variables, we had to drop this variable from our previous analyses. Unfortunately, such a decision leaves us with open questions about how the number of contributors relates to sustainability. To address this issue, we provide here an analysis that explores the relationship between sustainability, number of contributors and kind of forums.

6.4.1 More active vs. less active neighborhood forums

On average, a neighborhood forum has 53 users who post at least once within a calendar quarter (st. dev. = 60). Similar to our absolute measures of sustainability, the distribution of the number of contributors is right-skewed, with a few neighborhood forums that have many contributors and many more forums that have much smaller numbers of contributors. This variable ranges from a single contributor to 284 contributors in a neighborhood forum at a given quarter.

The count of contributors at different quarters allows us to visualize the trends of number of contributors over time. In Figure 9, we color-coded the largest numbers of contributors with green background and the smallest numbers with a red background. The yellow background represents a mid-range number of contributors. We also computed the coefficient of variation of the number of contributors by quarter. Using these two kinds of data (trends and coefficients of variation), we classified forums into five subcategories: *up* includes forums with increasing numbers of contributors over time; *stable* comprises forums with stable numbers of contributors over time, always fourteen people or greater; *inverted U* contains forums that show curvilinear trends with an increasing trend at the beginning followed by a consistent decreasing pattern in more recent quarters; *unstable* covers forums with high variability on the numbers of contributors over time; and *stable-neg* consists of forums that have stable numbers of contributors fewer than fourteen people. While the categories *up* and *stable* tend to have low coefficients of variance, the categories *unstable* and *inverted U* tend to present higher coefficients of variance. To simplify our analysis, we grouped these categories into two larger groups: *more active* that includes the categories *up* and *stable*, and *less active* that comprises the remaining categories.



Forum	2005 Q3	Q4	2006 Q1	Q2	Q3	Q4	2007 Q1	Q2	Q3	Q4	2008 Q1	Q2	Q3	Q4	2009 Q1	Q2	Q3	Q4	2010 Q1	Q2	Q3	Q4	2011 Q1	Q2	Q3	Q4	2012 Q1	Q2	Q3	Q4	2013 Q1	Q2	Q3	Q4	2014 Q1	Q2	Avg. # contributors	Std. Dev. # contributors	Coeff. of variation	Trend	Active?
central																																			13	13	15.00	0.00	0.00	stable-neg	Less
whittier																											1	1	1	1	1	1	1	1	1.00	0.00	0.00	stable-neg	Less
city-mpls		89	245	269	274	306	254	208	234	219	216	212	245	222	228	245	267	258	245	232	247	251	212	218	281	252	210	222	197	199	209	226	201	212	204	194	227.53	36.09	0.16	stable	More
city-stpaul	145	174	181	153	122	115	114	127	114	113	118	109	120	71	101	112	107	117	80	103	97	100	87	104	101	114	112	113	119	115	120	110	98	107	119	91	113.49	21.82	0.19	stable	More
hm																									61	57	60	89	116	76	76	97	83	81	44	72	74.31	19.46	0.26	stable	More
bryant																													38	38	29	37	20	21	27	23	28.33	7.66	0.27	stable	More
phillips																					28	26	31	43	50	56	50	73	72	55	42	65	57	38	47	44	47.76	14.23	0.30	up	More
nokomiseast																					20	33	10	29	22	32	43	34	62	38	35	39	30	31	23	37	32.82	11.45	0.35	stable	More
staneric											42	80	74	90	135	159	150	136	141	198	229	238	201	253	279	284	249	246	242	237	179	249	275	269	184	275	199.07	71.57	0.36	up	More
corcoran																				8	5	42	34	37	43	41	49	62	63	51	57	53	57	51	39	43	43.78	16.25	0.37	up	More
poho															47	62	102	82	79	131	157	213	140	189	243	217	214	236	267	230	208	251	249	200	170	261	183.70	69.19	0.38	up	More
ne																		33	13	15	17	18	16	11	21	20	20	23	22	10	13	35	31	12	15	24	18.90	7.31	0.39	unstable	Less
como																										31	32	8	10	19	20	29	23	19	14	24	20.42	8.04	0.39	unstable	Less
db																										14	20	25	38	39	32	57	52	54	31	50	36.50	14.52	0.40	up	More
up																											15	13	24	30	39	39	29	27	52	36	29.00	11.74	0.40	up	More
cr														1	19	20	10	23	34	32	39	46	44	30	39	34	35	35	47	30	18	32	23	16	26	19	28.67	11.61	0.41	inverted U	Less
longfellow																				34	20	39	49	95	147	107	110	123	149	119	133	166	109	101	137	168	110.00	45.83	0.42	up	More
ges																													13	16	27	39	18	19	22	20.50	8.72	0.43	inverted U	Less
d1																												1	13	20	17	20	21	18	18	10	15.00	6.44	0.43	inverted U	Less
macgrove																										18	24	17	58	26	43	40	33	21	15	36	29.75	13.30	0.45	unstable	Less
loring																										3	3	2	3	1	1	1	3	2			2.00	0.93	0.46	stable-neg	Less
frogtown															1	25	37	30	37	38	36	18	27	28	18	22	31	78	31	46	71	66	58	58	62	38.68	19.85	0.51	up	More
seward											32	36	23	20	27	41	57	49	68	70	95	94	71	85	132	120	119	155	139	156	148	170	171	159	141	170	101.67	52.55	0.52	up	More
sh																											1	1	4	1	3	1	3	4	2	3	2.36	1.25	0.53	stable-neg	Less
nearnorth																						2	2	4			4	9	6	6	4	5	3	2	1	8	4.57	2.43	0.53	stable-neg	Less
paynephalen																													31	29	32	88	37	41	32	39.00	20.94	0.54	inverted U	Less
highland																		15	16	15	13	25	16	29	42	54	47	47	60	70	62	53	68	84	91	98	50.25	27.19	0.54	up	More
holland																													1	1	3	1	1	2	1.50	0.84	0.56	stable-neg	Less
hpdl																											1	1	4	2	4	9	4	4	5	4.00	2.44	0.61	stable-neg	Less
w7																									1	1	5	18	14	6	11	6	6	12	8	6	7.77	5.08	0.65	inverted U	Less
frn																													1	1	3	5	3	8	4.00	2.66	0.67	stable-neg	Less
northend																										1	3	6	51	25	35	26	44	35	32	24	24.50	16.45	0.67	unstable	Less
dwtn																												1	12	3	23	12	11	6	9	13	9.50	6.46	0.68	unstable	Less
lh																										10	39	22	22	34	6	9	7	9	5	10	14.58	11.82	0.81	inverted U	Less
uptown																													1	1	1	5	11	2	10	6	4.78	4.10	0.86	unstable	Less
su																						1		1	2	2	1	9	33	16	14	28	31	41	23	29	16.40	14.19	0.87	unstable	Less
westside																									4	1	3	5	20	10	48	21	15	14	12	16	13.69	12.57	0.92	inverted U	Less
Total users	145	263	426	422	396	421	368	335	348	332	408	437	462	404	557	640	718	750	706	875	1005	1162	931	1153	1500	1487	1490	1627	1982	1741	1722	2458	2365	2156	2004	2343					
City users	0	0	0	0	0	0	0	0	0	0	0	0	0	1	19	20	10	23	34	32	39	46	44	30	39	48	70	73	109	99	89	128	104	97	109	105					
Neigh users	145	263	426	422	396	421	368	335	348	332	408	437	462	403	538	620	708	727	672	843	966	1116	887	1123	1461	1439	1420	1554	1873	1642	1633	2330	2261	2059	1895	2238					

Figure 9: Trends of the number of contributors in the forums by quarters

This classification allows us to assess the connection between the number of contributors of a forum to other measures of the forum's sustainability. Tables 33 and 34 show the mean and standard deviation (in parentheses) of each measure of sustainability. Beyond having a larger number of contributors, the forums that were categorized as *more active* have on average more new members, more new contributors, more retained contributors, and more posts. All of these differences are statistically significant.

Table 33: Absolute measures of sustainability by activity level

Level	# contributors	# joined	# new	# retained	# posts
Less active	14.842	26.519	3.607	5.395	42.008
	(13.612)	(40.910)	(5.475)	(5.351)	(49.681)
More active	86.825	43.631	15.910	42.371	269.427
	(73.580)	(35.373)	(13.012)	(42.014)	(256.873)

Table 34: Relative measures of sustainability by activity level

Level	% new	% retained	% migrated	Prod.	Δ prod.
Less active	31.676	48.302	15.539	2.783	.0128
	(27.013)	(28.893)	(16.678)	(1.269)	(1.731)
More active	37.403	49.588	9.299	2.910	.0355
	(18.582)	(15.960)	(6.725)	(.710)	(.641)

The relative measures also show better trends for *more active* forums. On average, *more active* forums have larger proportions of new contributors, better ratios of retained contributors, smaller proportions of contributors who migrate to other forums, better productivity and more positive changes in productivity. However, among the relative measures, only the difference of ratio of retained contributors is statistically significant.

Together, this data analysis suggests that our classification of active forums according to the number of contributors is consistent with better measures of sustainability at the collective level, especially in absolute measures.

Additional analysis (see Tables 35 - 37) was conducted in regard to the independent variables associated with sustainability, according to our framework and data analysis. In summary, *more active* forums are significantly more diverse in terms of tenure of their contributors, have lower ratios of contributors with membership overlap, and have more hierarchical social structures. They also have smaller proportions of active mobilizations and higher ratios of responsiveness to posts that create a new thread of conversation.

Table 35: Community aspects by activity level

Sustainability	Size	Diversity	Instability	% full access
Less sustainable	13,298	40.855	37.282	53.130
	(9,597)	(23.148)	(11.671)	(11.802)
Active	10,912	46.331	32.554	50.521
	(7,210)	(21.263)	(8.836)	(13.390)

Table 36: Online activity and social networks measures by activity level

Sustainability	Div. tenure	% overlap	Centralization	Connectedness
Less active	1.236	62.426	0.316	0.660
	(0.429)	(28.403)	(0.278)	(0.258)
More active	1.074	37.381	0.644	0.613
	(0.294)	(16.982)	(0.142)	(0.178)

Table 37: Content measures by activity level

Sustainability	% active mob.	% passive mob.	% non mob.	% responded
Less active	58.434	28.620	20.830	25.462
	(18.119)	(17.041)	(14.118)	(17.937)
More active	47.181	32.517	19.280	34.338
	(10.645)	(10.911)	(6.170)	(12.734)

Similar to our analysis of different measures of sustainability, we conducted regressions to estimate the effect of collective online and offline characteristics of the forums on the number of contributors. It is important to note that these models put together the effect of attraction of new contributors and retention of old-timers.

The results (see Tables 38 and 39) show that the impact of neighborhood size and diversity is often not statistically significant. This can be explained by the opposite effects that these variables have on attraction and retention of contributors. The opposite effects compensate each other when we use the overall number of contributors as the dependent variable. This result confirms that it is important to study attraction and retention of contributors as separate phenomena.

When the influence of the independent variables on attraction and retention had the same direction, the influence on the overall number of contributors was always statistically significant. Aligned with the results about attraction and retention, the effect of instability of the neighborhoods on the total number of contributors is negative and significant. In prior analyses, the effect of the level of Internet access was not significant for the attraction of new contributors, but significant for the retention of old-timers. Summing up these two influences, the ratio of residents that have Internet access both at home and at work has a significant and positive association with the number of contributors in a neighborhood forum. In analogous ways, the positive effects of diversity of tenure, responsiveness, active mobilizations and tenure of the forums were further confirmed. The negative influences of membership overlap, connectedness and more recent creation time of a forum were also further confirmed. The influence of the centralization of the networks continues to be not statistically significant.

6.4.2 Neighborhood vs. city forums

Even though the focus of this study is the neighborhood forums, we present here some analyses that show the difference in sustainability between neighborhood-oriented and city-wide forums for informative purposes. These two kinds of forums not only target geographical areas of different size, but they also appeal to different motivations. City-wide forums aim to encourage political discussion about local issues. On the other hand, neighborhood forums focus on community engagement and information-sharing among neighbors. These distinctions clearly lead to different

Table 38: Number of contributors

	(1) Contributors	(2) Contributors	(3) Contributors	(4) Contributors
Log population	2.186	4.446*	1.712	1.115
% of color	1.010	1.025**	1.005	1.004
% moved 2005+	0.939***	0.931***	0.967**	0.965**
% full int. access		1.042*		
L.Contributors			1.004***	1.004***
L.Div. user tenure			1.142***	1.194***
L.% overlap			0.998*	0.998*
L.Connectedness				0.837***
L.Centralization				0.908
Created 2011+	0.383***	0.212***	0.451***	0.516***
F. tenure	1.047***	1.045***	1.025***	1.025***
Observations	481	291	383	351
AIC	5474.6	3395.3	4171.0	3862.7
BIC	5503.8	3424.7	4210.5	3909.0

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 39: Number of contributors using content variables

	(1)	(2)	(3)
	Contributors	Contributors	Contributors
Log population	1.716	1.750	1.574
% of color	1.006	1.006	1.004
% moved 2005+	0.965**	0.966**	0.969*
L.Contributors	1.004***	1.004***	1.003***
L.% active mobilizations	1.003**		
L.% passive mobilizations		1.001	
L.% non-mobilization			0.991***
L.% responded	1.006***	1.004***	1.006***
Created 2011+	0.451***	0.460***	0.451***
F. tenure	1.022***	1.023***	1.022***
Observations	383	381	374
AIC	4163.5	4162.1	4034.1
BIC	4203.0	4201.6	4073.3

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

kinds of online groups according to the data. Tables 40 and 41 show the results of regression analyses that model the effect of the kind of forum on our sustainability measures.

Neighborhood forums achieve lower levels of sustainability than city-wide forums. Neighborhood forums are expected to have significantly fewer contributors, less success in the attraction of new contributors, less retention and poorer performance both in absolute and relative numbers. For example, compared to a city-wide forum, the number of contributors is expected to drop by 18% in neighborhood forums. These results provide evidence that neighborhood and city-wide forums work on different scales of online participation and could not be merged together in our analysis. This is one of the reasons that we focused our analysis on neighborhood forums only.

Table 40: Attraction, retention and performance by kind of forum

	(1) Contribs	(2) Joined	(3) # new	(4) # retained	(5) Posts
Neighb.	0.177**	0.506	0.196*	0.135**	0.097***
F. tenure	1.015***	0.979***	0.970***	1.011***	0.998***
Obs.	572	487	572	546	572
AIC	7974.4	12178.1	4879.4	5815.9	34018.9
BIC	7991.8	12194.8	4896.8	5833.2	34036.3

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 41: Relative attraction, retention and performance by kind of forum

	(1) % new	(2) % retained	(3) Log % migrated	(4) Productivity	(5) Prod. change
Neighb.	-0.243^{***}	-0.483^{***}	0.438^{***}	-0.675^{**}	-0.056^{*}
F. tenure	-0.590^{***}	-0.148^{**}	0.340^{***}	-0.166	-0.118^{**}
R^2	0.165	0.152	0.106	0.612	0.011
Obs.	482	461	461	483	473

Standardized beta coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6.5 SUMMARY OF RESULTS AND DISCUSSION

A summary of the results are depicted in Table 42. The positive and negative signs indicate the kinds of relationships between the dependent (rows) and independent (columns) variables. The colored cells in the table indicate the significant results.

Together, the results indicate that community aspects are significant factors on all three dimensions of forum sustainability. While the population size plays a positive role in the attraction of new members and new contributors, its impact on the measures of retention and performance is not statistically significant. This result is aligned with prior research that has found that larger communities tend to have larger volunteer associations [103]. Nevertheless, we found no evidence that this effect further reflects sustained participation as contributors in the forums. Our results indicate that population size is relevant for attracting new people but is not significantly related to the production of content or retention in the long-term. This finding has implications for technology designers. While focusing on small neighborhoods might not be considered a good decision because it might be hard to achieve critical mass, we found that the size of the neighborhoods only affects the attraction of new people, but not retention or performance. Other aspects of the target community might be more critical for sustainability than population size.

On the other hand, racial diversity of a neighborhood's population does not significantly relate to attraction to the forums, but it is significantly and positively associated with retention and performance of the forums. Forums of more diverse neighborhoods keep more of their contributors from one quarter to the next (they also do so at larger rates) and they have more active streams of content (more content, more productivity, and with more positive changes in productivity between consecutive calendar quarters). The finding of a positive relationship between diversity and retention is unexpected. Prior research suggests that diversity at the collective and individual levels tends to harm retention in volunteer associations [128, 119, 104]. Heterogeneity usually results in members' departure from volunteer associations. At the individual level, departure is explained by the opportunities of people to join alternative associations that might offer less hostile environments (with people more similar to the individual). There are several potential reasons that might explain this unexpected result. It is possible that there are no alternative online spaces for the sampled diverse neighborhoods, and therefore people might not have other options to post their information. It is also possible that although the forum targets a racially diverse neighborhood, the users of the forums are not as diverse as the neighborhoods and hence, the consequences of heterogeneity within the forum do not develop. Another feasible rationale is that people who live in the same neighborhood and decide to use the forums are already homogeneous enough regarding other aspects, so racial heterogeneity becomes less relevant.

Perhaps the most influential and consistent demographic factor for sustainability of the E-Democracy forums is residential instability in the neighborhood. Neighborhoods that have larger proportions of residents who had moved in 2005 or later are expected to have less sustainable forums. They engage fewer new members, fewer new contributors and smaller proportions of new contributors. Forums of more unstable neighborhoods also retain fewer contributors both in absolute and relative numbers. Furthermore, these forums also tend to receive less content, be less productive and have less positive changes in productivity from one quarter to the next. This demographic factor emerged from our attempt to reduce dimensionality of demographic factors that characterize neighborhoods. However, we did not have a specific hypothesis to test regarding this variable. The concept of social capital can offer a feasible explanation for the impact of residential instability on sustainability. Prior research provides evidence that residential stability is highly and positively connected to social capital in geographical communities [74]. Neighborhoods

Table 42: Summary of results: Collective online and offline aspects

	Attraction			Retention			Performance		
	#join	#new	%new	#ret	%ret	%migr	#posts	prod	Δprod
Log population	+	+		+			+		-
% of color		+		+	+		+	+	
% moved 2005+	-	-		-	-		-	-	
Created 2011+	-	-	-	-	-	+	-	+	
F. tenure	-		-	+	-	+	+	-	-
% full int. access	+	+		+	+		+	+	
L.Div. user tenure	+	+	+	+	-	-	+	+	+
L.% overlap	+		-	-	-	+	-	-	-
L.Centralization	+	-	+	+	-	-	-	+	+
L.Connectedness	-	-	-	-	-	+	-	+	+
L.% active mobiliz.	+	+	-	-	-	-	+	-	-
L.% passive mobiliz.	+	+	+	+	+	-	-	+	+
L.% responded	-	+	+	+	-	-	+	-	-

with more stable populations are characterized by higher levels of social capital among residents. This can be viewed as a product of people having the time to invest over and over again in the social relationships with those who live nearby. We reason that the path of social capital can explain the impact of residential instability on sustainability. Unstable populations have less social capital, which in turn negatively affects the neighborhood's ability to make their participatory information systems function. This reasoning is aligned with prior research that speculates about the need for social capital as a key antecedent of the success of community networks [77, 149]. Although our results cannot confirm such a hypothesis yet, they provide additional evidence in favor of such a direction.

Using data from the Minneapolis Community Technology Survey, we were also able to assess the impact of the level of Internet access in the neighborhoods on the sustainability of their participatory information systems. Prior research has documented that Internet access both at home and at work, as opposed to only one or none, is a significant predictor of the production of online content at the individual level [131]. Our results confirm this relationship at the collective level in online forums for urban communities. Neighborhoods with higher ratios of residents with Internet access both at home and at work are able to maintain forums that not only generate more online content, but also attract more new members and have better retention. It seems obvious that wider Internet access leads to more online participation. However, this relationship might be forgotten in the age of the Internet, mobile technologies and big data, especially in developed countries. While Internet access seems pervasive, there are still differences in the availability and quality of Internet access. Content production requires digital skills that can be better developed when people have full access to the Internet. At the neighborhood level, this becomes a critical factor for the sustainability of participatory information systems.

Among the variables of online participation, our data shows that diversity of tenure of the forums' contributors has a positive effect on attraction and generation of content. However, its impact on retention is not statistically significant. While diversity of tenure has a curvilinear effect on growth rates in Wiki Projects [30], the relationship in the E-Democracy forums is linear. It is possible that these local forums are not as large as the Wiki projects and therefore the turning point is not yet achieved. The results indicate that a mix of old-timers and newcomers is a healthy combination for local online forums. They make the forums look more attractive for residents to

join and they also bring larger numbers of new contributors every quarter. More diversity of tenure also helps achieve larger volumes of content. It is possible that the combination of old-timers and newcomers brings more skills and topics to the forums, thus making them more engaging and active. Based on this data, we conclude that online forums for urban communities should try to develop old-timer contributor loyalty and attraction of new contributors every quarter, so the forums maintain enough tenure diversity to thrive.

Membership overlap is more of a double-edged sword for the sustainability of online forums for urban communities. While it creates mobility of users within the E-Democracy platform by bringing more new members to the forums and having larger proportions of people migrating to other forums instead of abandoning the platform, larger membership overlap is also related to smaller levels of contribution in several ways. Forums with larger proportions of contributors who add content to other E-Democracy forums are likely to retain fewer of their own contributors and garner fewer posts in a subsequent period of time. This means that membership overlap increases exposure of the forums, but does not contribute to building sustainable streams of content in the long-term.

After controlling for demographics and online participation, the impact of the social network of contributors is somewhat dubious. The average level of embeddedness of contributors in a social network has a negative relationship with the attraction of new members and performance in absolute numbers, but a positive association with migration. This data reflects that more connected social networks are perceived as more closed groups, and are less appealing to new people who might not feel welcomed to join the group. Nevertheless, from those people who decide to leave these groups, a larger proportion do not leave the platform completely but decide to contribute elsewhere. This can be seen as a positive impact of connectedness on the sustainability of the platform, even though it does not help the sustainability of the focal forum with high connectedness. Beyond that, more connected groups discourage content creation in the future. Again, this can be an effect of the group being perceived as closed. In more connected groups, there is less attraction and not necessarily more retention, which results in groups with fewer people that are able and wanting to generate content.

Centralization in the networks has a positive impact on the attraction of new members and proportion of new contributors, but it is negatively associated with the number of posts that a forum

garners. Centralization is a measure of inequality of degree centrality among the contributors. Unlike what has been reported in Wiki Projects, where inequality is associated with less growth [137], our results indicate that forums with clear leaders who are particularly more connected than the majority become more attractive for newcomers. Given the offline shared context that functions as a background of local online forums, it is possible that the existence of clear leaders raises the expectations about the usefulness of these forums, and therefore drives attraction. Nevertheless, inequality does not drive the retention of contributors. Centralization helps to broaden the forums' audiences, but does not increase the commitment to actively participate in the forums in the long-term. A less desirable consequence of centralization is that it harms performance in absolute terms. Centralized networks seem to attract new people as readers, but generate less content as a group. Additional mechanisms need to be developed to deal with this negative impact of centralization on content creation.

To assess the influence of content on sustainability, we conducted an automatic content classification of all posts that started a new thread in the forums. We found that a large percentage (about 80%) of the content shared in local online forums are requests for mobilizations of the social capital available in target urban communities. This provides evidence of the use of these participatory information systems to exercise or use the social capital that is embedded in the urban communities. This finding confirms the results of a preliminary analysis in which we explored a sample of posts that was manually coded as different kinds of mobilizations [97].

The proportion of mobilization requests influences sustainability. Specifically, passive mobilization requests (such as requests for recommendations and factual knowledge) attract people to the forums as members and new contributors. Beyond that, active and passive mobilizations have different impact on other aspects of sustainability. Active mobilizations (such as invitations to events) are related to more content in the next quarter, but less retention of contributors. On the other hand, passive mobilizations are associated with more retained contributors. We interpret these results as evidence that mobilizations of social capital are a key factor in these forums. However, exceeding the amount of active mobilization requests might lead to an overwhelming flow of information that discourages online conversation. A forum with too many active mobilizations might convert the forums into one-way communication channels which people only use when they need to disseminate information. This can result in a lack of two-way online interactions, which

can then drive away considerable proportions of future contributors.

Once we account for kinds of content, the level of overall responsiveness in the forums has mixed effects. Responsiveness is associated with more new contributors, larger proportions of new contributors and more posts. This means that once people are involved in the forums, responsiveness helps sustainability. However, responsiveness reduces the attraction of new members. The positive effect of making people contribute do not extend to broaden the forums' audience. It is possible that outsiders do not perceive responsiveness as a major benefit of the forums, but as a burden. Responsiveness might be seen as information overload that can discourage people from joining. The influence of responsiveness on retention is more problematic. While it positively affects the number of retained users, this is not reflected in larger proportions of retention. Furthermore, among people who leave, smaller proportions of people migrate to other forums. Instead, larger proportions of people leave the platform for good.

Overall, our results are also informative about whether online or offline factors influence sustainability variables. The absolute measures are usually significantly influenced by both offline and online aspects, but most of the relative variables of attraction and retention are only affected by one of them, not both. The proportion of new contributors and the ratio of contributors who migrate to other forums are affected only by online measures. The percentage of retained users is associated with offline and content measures. Productivity and productivity change are only significantly affected by offline measures.

The pseudo R-squared and BIC and AIC scores allow us to compare the models regarding their ability to explain the variance across observations. For several dependent variables, the models with higher performance are those that include a collective measure of full Internet access. Productivity, the proportion of retained contributors, the number of new members, the number of new contributors and the number of posts in a quarter were better estimated when we accounted for the effect of the neighborhood's demographics and level of full access to the Internet. As documented by research on the digital divide [131], full Internet access can strongly influence individuals' tendencies to create online content. Our work complements this finding at the collective level. Several measures associated with content creation and attraction in participatory information systems were better explained when accounting for a collective measure of full Internet access.

Our relative measure of attraction (% of new contributors) is better estimated when we include

content measures in the models. This finding suggests that the kinds of content and responsiveness play important roles in encouraging people to transition from reader to creators of content in these forums. While content was not a significant factor of productivity and productivity change, it did increase the predictive power of the models. Further research is needed to better understand this result. The number of retained contributors and the proportion of migration are better explained by the models that include measures of online interaction in the forums.

Together, these results indicate that attraction and performance depend more on offline aspects of the target communities and, in specific cases, on content aspects. Beyond the demographics, retention also depends on the characteristics of the online participation and social structures that organically develop in the online forums.

Our analysis also reveals that the volume of contributors and posts in the E-Democracy forums for local communities is rather low. On average, a neighborhood forum engages 53 contributors and gathers 152 posts in a calendar quarter. The most active neighborhood forum in E-Democracy has attracted 284 contributors and received 1,180 posts in a quarter. Responsiveness in these forums is also low. Out of the new conversation threads that are initiated in a quarter, less than a third , on average, obtain at least one response. This confirms prior research that argues that virtual spaces for local communities hardly generate new content on a daily basis and have difficulty encouraging online discussion [25]. Nevertheless, some of the sampled forums have been active for six years. This suggests that users perceive value in the forums, even though the stream of content is not as active as other social media. This data raises questions about the suitability of using volume as a measure of success of participatory systems for urban communities. Although these concerns have been discussed in prior literature [27, 125], further research is needed to explore alternative measures of performance that might better represent the effectiveness of participatory information systems in urban settings.

7.0 STUDY 2: INDIVIDUAL ASPECTS OF SUSTAINABILITY OF ONLINE FORUMS FOR URBAN COMMUNITIES

It is known that participatory information systems for urban communities face the challenge of low content volume and modest levels of online discussions [25]. This has also been confirmed in the E-Democracy platform, where on average a neighborhood forum accumulates 152 posts in a calendar quarter (See Chapter 6). Nevertheless, the local forums continue to be active in the long-term, with older forums having been active for six years. This sparks the research goal of this study: why do individuals continue participating in apparently dormant information systems?

To answer this question, we employ the proposed framework (see Chapter 4) to conduct a study of the offline and online characteristics of individuals associated with sustainable participation in the forums. We explore the interplay of online and offline characteristics of individual users that can influence measures of attraction, retention and performance of participatory information systems for urban communities. Similar to the previous study (Chapter 6), this analysis is based on archival data of all posts exchanged in 35 online forums for urban communities in the US. However, this study focuses on individual posts as the unit of analysis, as opposed to collective measures of participation. An analysis of behavioral measures is then supplemented with cross-sectional user survey data in order to model the users' perspectives about the forums' impact on their involvement with the community. This chapter reports on the research methods used in this study, the results and their implications.

7.1 TESTING THE FRAMEWORK: ONLINE INDIVIDUAL ASPECTS

This study is composed of two parts: an analysis of the online behaviors of individual users on the E-Democracy platform and an analysis of user perception about the platform.

7.1.1 Archival data analysis

The goal of this study is to assess the relationship between individual aspects of a user's online behavior and sustainability measures. Following our proposed framework, we focused on measures of online activity, social network position, and content at the individual level. To assess sustainability, we defined measures of performance and retention at the individual level. To compute these measures, we used archival data of the posts that were exchanged in 35 E-Democracy online forums for neighborhoods and districts in the cities of Minneapolis and St. Paul from 2008 until the second quarter of 2014.

We focused our analysis on posts that initiated new threads in the forums; therefore, we could reasonably assume that other posts in the same thread were responses (a measure of performance of generating content). For each initial post, online *response* was represented as a binary variable with values of one if the initial post had received at least one answer, and zero otherwise.

Retention was measured as other binary variables that represent whether the user posted again in the forum within a timeframe of 25 days (the average time between two consecutive posts of a user) and of a quarter (to discuss these results in comparison with the prior study - Chapter 6). These variables were called *within avg.* and *within quarter*. We also considered another binary measure of retention that indicates whether the user never posted again in the forum and *migrated* to another forum within the E-Democracy platform.

We measured the independent variables at the time of posting. For each post that initiated a new thread in the neighborhood forums, we computed several measures that describe the characteristics of the post's author at the time of the post. In regard to the user's online activity, we computed the *user's level of contribution* in the current forum as an ordinal variable with three values: (1) *newcomers* with zero prior posts; (2) *old-timers* with a number of prior posts between one and 37; and, (3) *power-users* with more than 37 prior posts. We chose the threshold of 37 posts (the 95th

percentile) to have a balanced number of observations in the old-timer and power-user groups. We also defined a binary indicator of *membership overlap* that represents if the user had posted to another E-Democracy forum in the same quarter.

Additionally, we used the results of our social network analysis (see Sections 6.1.4 and 6.2.4) and automatic classification of content (see Sections 6.1.5 and 6.2.5) to add these aspects to our individual-level analysis. Two network position measures were considered at the time of posting: the *degree centrality* of the user and to what extent they were a *broker* in the forum's social network. The broker position was determined by the network constraint index, as suggested by prior literature [16, 17, 18]. Regarding content, we included the predicted *kind of content* of the post, including three possible values: non-mobilization, passive mobilization and active mobilization.

To estimate retention measures, we included a binary variable that describes whether the last post of a user received a *response*. To control for the level of activity in the forum (a collective measure), we included the number of users who had contributed to the forum in the past. Table 43 summarizes the independent and dependent variables of this analysis.

7.1.2 Descriptive statistics

Our analysis of the archival data indicates that 3,466 individuals have initiated a new thread of conversation at least once in a neighborhood forum. Together, these users created 32,903 new threads. In response to these initial posts, a total of 42,471 other posts were generated by the online groups.

The analysis also reveals that about a third (35.4%) of the new threads received a response in the neighborhood forum. Nevertheless, 67% of the authors of those initial posts contributed again within 25 days and 80% within a calendar quarter. Only 4.7% of people who posted in a neighborhood forum stopped posting in the forum and migrated to another E-Democracy forum (see Table 44).

The distribution of posts by contributors is right-skewed, as is the case in most participatory information systems [69, 114]. On average, users created 9.5 posts that initiated new threads in a forum, ranging from 1 to 1,168 and with a median value of two posts. Ten percent of the posts were created by *newcomers*, slightly more than half were posted by *old-timers*, and the remaining

Table 43: Independent and dependent variables of archival analysis of study 2

Independent variables	Dependent variables
1. Users' online activity in the forums	1. Performance
a. Prior level of contribution	a. Got a response?
b. Membership overlap	2. Retention
2. Network position in the forum	a. Posted again within average time?
a. Centrality	b. Posted again within a quarter?
b. Brokering	c. Migrated to another forum?
3. Content	
a. Kind of shared content	
b. Responded	

were added by *power-users*. Among the 3,466 unique contributors who had started at least one thread of conversation in the neighborhood forums, 68% (2,359) had posted exclusively in one neighborhood forum, and another third of the contributors had posted to more than one forum. People who posted in more than one forum were often power-users.

Regarding the network position of the users, the degree centrality ranges from 0 to 624 with an average of 65. The measure network constraint that represents a user brokerage position varies from 0 to 1.281 and its average is 0.166 (st. dev. = 0.221). This indicates that there is little variation on the brokerage position of the users in the forums' networks.

With respect to the kinds of content, 47% of the posts that initiated a thread were *active mobilizations*, 33% of them were *passive mobilizations*, and the remaining were *non-mobilizations*.

Table 44: Descriptive statistics of online measures of sustainability at the individual level

Dependent variable	Mean	Std. Dev.
Received a response	0.354	0.478
Posted again within avg. time	0.670	0.47
Posted again within a quarter	0.803	0.398
Migrated to another forum	0.047	0.211

7.1.3 Performance and retention

To assess the impact of individual online characteristics on performance and retention, we used xt models in Stata to estimate logistic regressions that model the relationship between the individual online behavior and (1) the probability of receiving online response as a measure of performance, and (2) the probability of posting again within average time, within a quarter, or posting to another E-Democracy forum (i.e., migration) as alternative measures of retention. The dependent variables are binary indicators. The independent variables are the prior level of contribution of the user, whether the user had posted to other forums in the past, the user's degree centrality and brokering position in the forum's social network, and the kind of content and length of the post. A control variable that represents the forum's number of contributors was also included to account for the forum's collective effect on the individuals' behavior. These multilevel models considered the posts nested within forums and users. Table 45 describes the results of these regression analyses.

Unlike what has been observed in related research [4], compared to newcomers, users with more experience as contributors were less likely to receive an online response. However, even though old-timers are less likely to receive a response, they are significantly more likely to post again in the forums. These results have two possible explanations: (1) there are differences in online interactions that take place in systems for urban communities as opposed to systems with global reach; and (2) users with different levels of experience in the local forums tend to use the system in different ways. For example, more experienced users might tend to use the forums as a

tool to propagate information on a periodic basis, rather than to acquire information. On the other hand, newcomers might tend to engage in more back-and-forth conversations in the forum. These results also reveal that power-users are much less likely to migrate to other E-Democracy forums, thus confirming their commitment to the focal forums.

Membership overlap is correlated with lower likelihood to receive a response and more likelihood to migrate to another E-Democracy forum. These results can signify that membership overlap is often associated with the need to cross-post information for information dissemination across several geographical areas in the city. Such information dissemination need is often a one-way path in which an online response is not necessarily required.

Aligned with the expectations from prior research on social capital [92], central users are more likely to receive a social capital benefit available through the local information system. Users with higher degree centrality (more prior online interactions with other users in the forums) are more likely to receive a response to their requests. These users are also more likely to continue posting in the same forums in the future, and much less likely to migrate to other forums. This result indicates that stronger connections with other members enable users to benefit more from the forum (by getting responses). Beyond that, stronger connections among users contribute to make the system more viable, as they tend to keep posting in the forum.

While it was expected that brokers in the forums' social networks were more likely to get benefits [16], we found no support for an effect on the probability of receiving a response. The brokering position is not significantly related to retention either. It is possible that these results are associated with the rather low and homogeneous values of the network constraint measure in our data.

Our analysis of the kinds of content suggests that they matter for performance, but not for retention. Compared to non-mobilizations, posts that are coded as *active mobilizations* are significantly less likely to obtain an online response. Compared to non-mobilizations, the probability of receiving a response drops by a factor of 0.529 when the post is coded as an *active mobilization*. This means that while the largest share of posts are *active mobilizations*, these posts lead to little content generation. Thus, they have a negative effect on the performance of the forums. On the other hand, *passive mobilizations* were not significantly different than non-mobilizations in their impact on performance and retention. Furthermore, post length of the post is also significantly

Table 45: Performance and retention: online response and new post at the individual level

	(1) Response	(2) Within avg.	(3) Within quarter	(4) Migrated
Old-timer	0.802^{***}	1.463^{***}	1.443^{***}	1.306
Power user	0.585^{***}	1.856^{***}	2.046^{***}	0.315^{*}
Membership overlap	0.711^{***}	1.008	0.906	17.28^{***}
Degree cent. (log)	1.455^{***}	1.675^{***}	2.114^{***}	0.173^{***}
Brokering	1.022	1.112	1.145	0.407^{***}
Passive mob.	1.071	0.998	0.932	1.001
Active mob.	0.529^{***}	1.079	1.105	1.081
Network size (log)	0.939	0.457^{***}	0.334^{***}	8.728^{***}
# words (log)	0.840^{***}			
Responded		2.559^{***}	2.299^{***}	0.320^{***}
Observations	23006	21605	21605	21605
AIC	28102.4	24160.9	18819.1	6088.4
BIC	28198.9	24256.7	18914.9	6184.2

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

related to the chance of obtaining online responses from the neighborhood forums. As suggested by prior work [73], shorter posts are significantly more likely to receive a response.

Our analysis also provides support for a significant relationship between online feedback and user retention in E-Democracy online forums for local communities. Users who received responses to their prior posts are more likely to post again in their local forums than those who did not receive any answer. This result confirms what has been consistently found in different studies of participatory information systems with global reach. Online response is a significant predictor of future participation in public newsgroups [4, 73], Facebook [15], large enterprise social sites [13], and other user-generated content sites [90, 129]. We were not able to find support for this relationship in a preliminary study [97] that explored a smaller dataset of posts from a subset of the same forums. In such analysis, the effect size of the user's prior level of contribution was much larger than the estimation in the current study, and online response was not a significant factor. To better understand this difference, we conducted further analysis on the data and found that there is a significant interaction between prior level of contribution and online response ($p < 0.001$). The positive effect of responsiveness is significantly larger for newcomers than other contributors, as illustrated in Figure 10. This offers a feasible explanation for the result discrepancies in the two studies with two different sample sizes.

7.2 TESTING THE FRAMEWORK: OFFLINE INDIVIDUAL ASPECTS

While the archival data helps us answer questions about online response and retention of contributors at an individual level, other aspects of the proposed framework such as the offline characteristics of the users and the impact of the forums remained unexplored. To address this gap, we investigated data gathered through the 2014 E-Democracy user survey. This online survey was designed and conducted by E-Democracy. The survey was open to responses for a period of one week in December 2014. It was advertised through the platform and collected a total of 1,373 responses from users.

To maximize response rates, E-Democracy decided to keep the questionnaire as short as possible. This resulted in a questionnaire with a single question for the majority of the concepts that

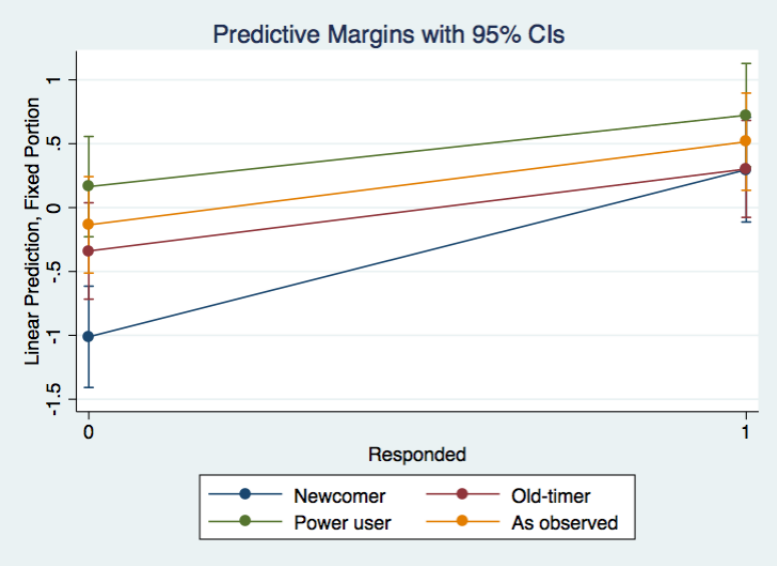


Figure 10: Interaction effect between prior level of contribution and online response

it measured. While this might have reduced the reliability of the measures that describe subjective evaluations, it encouraged a large number of responses. E-Democracy estimates that the number of respondents is larger than 10% of its user base in the cities of Minneapolis and St. Paul.[1]

This section describes our analysis of the E-Democracy user survey data. We first introduce our approach to study the survey data. Then, we present the descriptive statistics of the dependent and independent variables of the study. Finally, we explain the main findings of the regression analyses of the cross-sectional survey data.

7.2.1 Survey data analysis

The questionnaire included 30 questions about various topics, including which was the respondents' primary forum and whether they had contributed content to it. It also requested information about the respondents' expectations of local forums, to what extent these expectations had been met, the respondents' use of the system, and the types of impact the forums have had on the

[1] http://blog.e-democracy.org/posts/2610 Last retrieved on October 28th, 2015.

respondents' engagement with their communities. It also had questions about the respondents' demographics and roles in the local community. The questionnaire is available in Appendix B.

Guided by our proposed framework, we organized the questions into components that represent similar aspects of sustainability and their potential antecedents. Table 46 summarizes the dependent and independent variables that were explored through the survey data.

The dependent variables were the respondents' *goals*, their *activity* in the forums, their *satisfaction* levels, and their *involvement* with their communities as a result of being a user of the local forums. Tables 47, 48 and 49 describe the specific questions in the survey that were used to compute our dependent variables.

Table 46: Independent and dependent variables of survey analysis of study 2

Independent variables	Dependent variables
1. Offline aspects	1. Attraction
a. Role in the community	a. Goals
b. Women	2. Performance
c. People of color	a. User's activity
2. System	b. Satisfaction and involvement
a. Kind of forum	
b. Offline outreach	
c. Forum's activity	

Three kinds of goals were considered: (1) to *get* local information such as news, local events, businesses, resources and services; (2) to *exchange* information such as points of view and ideas regarding community issues; and (3) to *engage* with the community via local initiatives, meeting

neighbors, or helping neighbors in need. Each of these goals was represented as a binary variable that was set to one if the respondents had rated any of the corresponding survey questions "very important" (see Table 47), and zero otherwise.

We also included two binary measures of the respondent's activity in relation with the local forums: (1) *communicated on-site* was set to one if the user had either posted at least once to their forum or received a reply on their forum in the last year; and (2) *communicated off-site* was set to one if the respondent had either received a response to her posts or replied to someone else's posts in person, by email, phone or other media.

Regarding the outcomes of using local forums, we defined three variables of satisfaction that match the kinds of goals that were defined above: *get* information, *exchange* information, and *engage* with the community. Given that the questions about satisfaction were phrased as a 5-point Likert scale, we coded the answers from zero (strongly disagree) to four (strongly agree). Our dependent variables were then defined as binary indicators that are set to one if the average satisfaction among the corresponding survey questions was higher than two.

To better understand the kinds of impact that the local forums might have on the users' engagement with the community, the survey also included 16 questions about very specific outcomes. We considered each of these questions separately. The questions are presented here in an order that starts with feelings of community involvement and finishes with active/physical actions reflecting involvement with or participation in the community. The possible answers to these questions were (1) "Yes, I did this AND it increased because of the forum"; (2) "I did this, but did not increase this activity because of the forum"; and, (3) "No, I didn't do this / not applicable". We re-coded the answers to these questions as binary variables that were defined as one if the respondent chose the first option, and zero otherwise.

With regard to the independent variables, we included offline aspects of the users and characteristics of the forums the user participates in. The users' offline aspects were represented by the user's gender, race, and role in the community. Gender was coded as a binary variable *Women* that is set to one for females and zero for males. Race was represented by a binary variable *People of color* that was defined as zero if the respondent was a white person and one if the respondent belonged to any other race or ethnicity. The respondents' roles in the community were operationalized into three binary variables that indicated if the user was a *Resident*, a *Parent*, and/or a

Table 47: Mapping between survey questions and goals and activities

Dependent variables	Questions in the survey
Goals	How important to you are the following things you can do on your Neighbors Forum? (3-point Likert scale)
Get	Get community news and local event announcements
	Learn about local businesses, resources and services
Exchange	Discuss or understand others' views on community issues and happenings
	Share information or ideas
Engage	Get involved in local initiatives or causes
	Meet neighbors and other community members
	Help neighbors in need (sharing, lost pets, etc.)
Activity	Check all that apply ...
Communicated on-site	I posted at least once to my forum in the last year
	I received on-forum replies to my forum post(s)
Communicated off-site	I received private email replies to my forum post(s)
	I received a response or acknowledgement of my forum posts in person or via the telephone
	I have never posted publicly on my Neighbors Forum, but I have replied privately to posts

Worker in the community. The variable worker was set as one if the respondent had any of the following roles: student, business owner, employee or volunteer in any kind of organization within the community, or an elected official. The specific questions used to set a value for these variables are shown in Table 50.

Finally, we also included three variables that describe the kind of forum the user is part of.

Table 48: Mapping between survey questions and satisfaction

Dependent variables	Questions in the survey
Satisfaction	To what extent is your forum meeting your needs? How *satisfied* are you with the opportunity that your forum has provided in the last 12 months to... (5-point Likert scale)
Get	Get community news and local event announcements
	Learn about local businesses, services, and resources
Exchange	Discuss or understand others' views on community issues and happenings
	Share information or ideas
Engage	Get involved in local initiatives or causes
	Meet neighbors and other community members
	Help neighbors in need (sharing, lost pets, etc.)

Kind of forum can take three values: zero if the user participated in a city-wide forum only, one if the user was affiliated to a neighborhood forum only, and two if the user was part of both a neighborhood and a city-wide forum. For neighborhood forums, we also defined two additional variables. *Offline outreach* indicates whether the user's neighborhood forum had enacted offline outreach strategies such as door-to-door advertisement to encourage residents to join the forums. *Forum's activity* denotes whether the user's neighborhood forum had been classified as more active than others according to our archival data analysis (see subsection 6.4.1).

7.2.2 Descriptive statistics

All of the dependent variables are binary indicators; therefore, the average value also represents the percentage of respondents that have given a positive answer to the question at hand. The descriptive statistics (Table 51) show that the majority of respondents (more than 60%) consider at least one

of the goals (get, exchange, and engage) to be very important. Among them, the most appreciated goal is to get information, which is valued as very important by 78.3% of the respondents.

Slightly more than half of the respondents have engaged in communication on the forums (on-site) and more than 40% report to have been involved in some kind of off-site communication about topics that were raised on the online forums.

The data also reveals that there are high levels of satisfaction regarding goals. 77% of the respondents are satisfied with the system's opportunities to get information, 72% are satisfied with regard to engaging with the community, and 69% are satisfied with respect to exchanging information.

Nevertheless, when the respondents were asked about specific outcomes that might reveal the impact of the forums on their involvement in the community, the rates of agreement dropped considerably. People agreed the most with the claims that they had become better informed (78%), and learned from others' points of views (67%) thanks to the online local forums. However, only about half of the respondents agreed with other claims, such as having became more satisfied with their communities, felt welcomed in the forums, and learned how to influence decision-making in their communities. A minority (31.6%) indicated that they had learned about other residents' races or ethnicities through the forums.

The levels of agreement about outcomes that represent actions rather than feelings are generally lower. The most common actions were attending a community event and using a local service such as to visit a local business. These actions were carried out by slightly more than 40% of the respondents. Other outcomes were even less likely. Between 20% and 30% of respondents had done favors for others in the community and attended meetings to discuss local affairs as a result of their forum participation. Less than 20% had signed a petition, contacted government officials, donated money to a cause, performed volunteer work, collaborated with others to encourage a change in the community, or met a neighbor. It is important to note that this does not mean that the respondents did not engage with these actions. It means they do not perceive or attribute an increase of their involvement with these actions to their participation in the local online forums.

Table 52 shows the descriptive statistics of the offline characteristics of the E-Democracy users that took the survey. These data also allow us to characterize our sample. Regarding the roles the users play in the community, a large majority of users (84.7%) are residents of their forum's target

urban community. About 40% of the respondents were workers in the local community and only 26.7% were parents. Most of the survey respondents were women (64.7%) and a small minority were people of color (7.8%).

The respondents were asked to identify their primary forum in E-Democracy. If they chose a forum associated with a neighborhood in Minnesota or St. Paul, they were also asked if they participated in the corresponding city-wide forum. Based on this data, we were able to code the kinds of forums in which a respondent participated. The majority (64.8%) was part of a neighborhood forum only. Slightly more than a quarter of the respondents were affiliated with both a neighborhood and city-wide forum, and less than 7% were users of a city-wide forum only.

The respondents were affiliated with 37 different neighborhoods in the cities of St. Paul (16) and Minneapolis (21). E-Democracy had implemented offline outreach strategies in ten of these neighborhoods. Our archival data analysis concludes that 13 of these neighborhoods were active (see Section 6.4.1). Overall, we had more than 140 observations in all combinations of the kinds of neighborhood forums, and more than 700 observations for the most common combination (active forum with no offline outreach).

The survey collected other data about the respondents that help to further describe our sample. Beyond a bias towards women and white users, the respondents tend to be highly-educated adults (average: 45 - 54 years old) who own property and whose income is above the poverty levels in the US. The majority of respondents (52.64%) were 45 years old or older at the time of data collection. Less than 2% of respondents were younger than 24 years old. More than 80% of respondents had a bachelor's, graduate or professional degree. Almost the same proportion of people owned the place where they live. Seventy percent had a total household income higher than $30,000, and another 16% preferred not to say their range of income.

Regarding the use of the system, about 40% reported learning about the E-Democracy online and another 40% had heard about it through offline media (e.g., word of mouth, offline outreach). Almost 60% were regular email subscribers that use email to read and post in the forums, a quarter received daily digests, and 12% accessed the system via web.

Approximately one out of five respondents reported that they had neither posted in the forums nor communicated privately with others about the content of the forums. The respondents also used other social media. Eighty-seven percent used Facebook, 41% used Twitter, 65% had a LinkedIn

account, and 19% participated in NextDoor. Slightly less than a third of the users also participated in small, private online groups for their block or nearest neighbors (29%) or other public or large online groups for the neighborhood or part of the town (26%).

Based on the data that this sample of E-Democracy users provides, we conducted various analyses to better understand the relationship between the offline characteristics of the users and their goals, activity and community involvement in regard to the local forums. For each dependent variable, we conducted two regression analyses. First, we assessed the effect of the kind of forum the respondent participates in (city-wide only, neighborhood only, or both). The second regression took into account only residents that participate in neighborhood forums. For them, we estimated the effect of being part of a forum that was subject to offline outreach engagement and if the forum was considered an active forum, given their number of contributors over time. The rest of this section summarizes the results of such analyses.

7.2.3 Performance: Goals

Women are more likely than men to believe that getting local information and engaging with their local communities is a very important thing they can do in the forums (see (1) and (2) in Table 53). This result is consistent after controlling for other user characteristics, the kind of forums in which the user participates, and the level of activity of the forum. For example, the estimation indicates that the probability of valuing as "very important" the ability to get local information increases by factor of 1.696 when the respondent is a woman (as opposed to a man). The effect size of gender on the importance of engagement ranges from an increase factor of 1.570 to 1.766. The role that a user plays in the community and her race do not significantly relate to their appreciation of the forum goals.

Compared to users who solely participate in city-wide forums, people who participate in neighborhood forums (only or in conjunction with a city-wide forum) are more likely to value the following two goals as "very important": getting local information and engaging with the community. On the contrary, these users are significantly less likely to consider help with exchanging views or ideas "very important". These results are aligned to the neighborhood forums' goal to encourage community involvement, as opposed to the city-wide forums' goal of fostering political discussion.

Considering the neighborhood forums, users who belong to active forums and those who belong to forums with offline outreach are more likely to perceive the goal of getting information as very important. Being part of an active neighborhood forum is also significantly related to considering engaging with other people and initiatives in the community "very important".

7.2.4 Performance: Activity

As users differ in their appreciation of the diverse goals they could achieve through the local forums, they also differ in their probability of engaging in different activities associated with their forums. We measured two kinds of activities: whether the user communicates with other people on-site (by posting or receiving answers online), and whether they communicated with other people off-site in regard to content that was exchanged in the forum. Table 54 shows the results of the analyses about these two measures.

Users who value the goals of exchanging views and engaging with the community are more likely to have communicated on-site with fellow residents than those who did not consider these goals "very important". Those who value engaging with the community are also more likely to communicate off-site with other people about issues raised in the forums. These results confirm that the goals of the users (things they consider to be very important) are significant factors on the kind of activities with which they engage. This also reveals that there is some level of communication happening outside the site that is related to the content shared on the platform, and this is more commonly reported by people who value the goal of engagement with the local community.

People who are workers in the community are more likely to communicate on-site that those who are not workers. This phenomenon can be associated with these workers' need to use the system to disseminate or obtain information as part of their work-related activities. None of the other users' offline characteristics have consistently significant relationships with the kinds of activities they undertake in regard to the system.

Users who belong to neighborhood forums, as opposed to those who participate in city-wide forums only, are less likely to engage in offline communication. Among neighborhood forums, active forums are significantly associated with both more on-site and more off-site communication among residents of the local communities. This leads us to believe that a thriving local forum is

associated with a community that uses several media to communicate. However, we cannot yet distinguish whether there is a causal relationship and in which direction.

7.2.5 Performance: Satisfaction

Once we have taken into consideration goals and activity, we can now assess which factors affect users' satisfaction with their forums in regard to their goals. Table 55 shows the results of this new set of analyses. To simplify the interpretation of the data, the measures of satisfaction were classified into the same categories as the users' goals: get, exchange, and engage.

Both on-site and off-site communication among residents have positive relationships with satisfaction. The relationship between on-site communication and satisfaction in regard to exchanging views and engaging with the community becomes statistically insignificant when we restricted our analysis to users of neighborhood forums (and controlled for the level of sustainability of the forum and whether it had offline outreach). A potential explanation for this result is that passive (not posting) and active participation in a neighborhood forum leads to the same levels of satisfaction in terms of exchanging views and engaging with others. There are other aspects of the forums (e.g. level of activity) that more critically influence the satisfaction regarding these goals. On the other hand, there is a consistently significant relationship between being involved in on-site communication and feeling satisfied with getting local information. To meet this goal, active and direct communication on-site is crucial.

The role of off-site communication is significantly associated with being satisfied regarding the goals of exchanging views and engaging with the community, but not getting information. Keeping all other independent variables constant, users who communicated off-site with other residents regarding topics raised in E-Democracy are 1.4 times more likely to feel satisfied about exchanging views than users who did not communicate off-site. Off-site communication also increases the chance of satisfaction of engaging with the local community by a factor larger than 1.5. This signifies that the off-site communication generated by the online forums is significantly related to satisfaction with the forums' possibilities to exchange views and engage with the community.

The respondents' goals are also significantly associated with their satisfaction levels. There is a match between goals and satisfaction. Those who consider getting local information to be

important are more likely to feel satisfied with the opportunities their forums have given them to get local information; users who value exchanging views are more likely to feel satisfied about it; and respondents that appreciate engaging with their community are more likely to be satisfied with this goal. Besides, users who consider important getting local information to be important are also more likely to be satisfied regarding the goal of exchanging views. Those that expect the forum to be useful for engagement are also more likely to be satisfied with getting local information. Thus, beyond the perfect match between goals and satisfaction, the results show that even though a given goal might not be valued as very important, users can perceive some level of satisfaction regarding the ability of the system to meet such a goal.

None of the offline characteristics of the users are significantly associated with satisfaction after controlling for the effect of their goals, activities related to the forums, and characteristics of the forums. While this is a consistent result, it is relevant to note that some of the offline characteristics were already significantly associated with goals and kinds of activity. Besides, these characteristics might also affect the decision to participate in different kinds of forums (city-wide, neighborhood or both). Therefore, this result should not be read as evidence that offline aspects do not matter, but as evidence of a potentially fully-mediated relationship among offline characteristics and satisfaction. Further analysis is required to confirm the mediation.

The results also indicate that users who participate in neighborhood forums only are more likely to be satisfied in regard to the goal of engagement. This suggests that neighborhood forums are playing a role in community engagement that city-wide forums are not. This is not a criticism of city-wide forums, but rather a confirmation that these two kinds of E-Democracy forums in E-Democracy are not only targeting different goals and geographical scopes, but are also achieving different kinds of impacts on their users.

Among the neighborhood forums, people who participate in forums that with offline outreach are less likely to be satisfied with the goal of exchanging views. Offline outreach was undertaken in racially diverse neighborhoods and a great deal of its motivation was to encourage the digital inclusion of minorities in the neighborhoods. Regardless of the impact of offline outreach on bringing more diverse audiences to the forums, it is possible that the offline outreach also raised expectations regarding the goal of exchanging views. Our results are an indication that these expectations have not necessarily been met.

Finally, the forums that were categorized as more active are related to survey respondents that are more satisfied regarding all goals. Compared to users who participate in the less active forums, users of more active neighborhood forums are likelier to be satisfied with getting local information and engaging with the community by a factor above 2.4. The impact of an active forum is slightly smaller in regard to satisfaction with exchanging views (factor=1.9). This last piece of evidence continues to support the idea that meeting the expectations of exchanging views might be the most difficult goal for the E-Democracy forums to achieve.

7.2.6 Performance: Specific outcomes

With the goal of further exploring the potential outcomes of using E-Democracy, the survey assessed whether respondents have engaged in various actions as a result of participating in the forums. We analyzed the data in an analogous way to the study of measures of satisfaction. Tables 56 to 63 summarize the main results.

Survey respondents who had engaged in on-site communication with other residents are more likely to feel welcomed in the forums than respondents who had not communicated online. However, they are less likely to feel that they have learned about other races through the platform. Besides, actively participating in communications with others on-site is significantly related to having donated money to a cause or charity, done favors for fellow residents, collaborated with other people to encourage a change in the community, participated in community discussion meetings, met a neighbor, and used local services such as visiting a business. These results indicate that on-site communication has an impact on the actions people undertake, rather than on people's feelings.

In turn, off-site communication is related to a larger range of outcomes. It is positively and significantly associated with a larger probability of feeling that participation is welcomed in the forums and higher satisfaction with their local community as a place to live or work. Beyond that, people who had communicated off-site with other residents are more likely to have been introduced to new ideas, learned about other races, and learned how to influence decisions in their communities. Off-site communication is also positively associated with actions such as having contacted elected officials, donated money to causes, done favors for others, collaborated with

other community members to trigger a change in the community, and used a local service. Thus, those who engage in off-site communication about topics initiated in the platform are more likely to perceive an impact on their feelings and actions related to their community.

There are several statistically significant positive relationships among the users' goals and the kinds of outcomes. The goal of engagement is associated with the broadest range of outcomes, including feeling welcomed in the forums, becoming more satisfied with the community, learning about other views, and learning how to influence decisions in their communities. This goal is also positively associated with action-related outcomes such as contacting government officials, donating money, doing favors for others, meeting neighbors and attending local events.

The other goals are related to fewer outcomes. All goals are related to feeling welcomed and becoming more satisfied with the community. The goal of getting informed is the only one associated with the outcomes of feeling more informed about what is happening in the community and signing a petition. None of the goals are consistently associated with having learned about races and performed volunteer work as a result of the forum participation. These results suggest that there is an intricate relationship between the diverse things that people value from a local civic technology and the actual outcomes that can be perceived as a result of using those tools. Further research is needed to continue understanding these relationships at this level of granularity.

After controlling for the users' goals, activities and kinds of forums in which they participate, there are some relationships between offline users' characteristics and specific outcomes of the use of local forums. Users who are workers in the community are more likely than non-workers to have collaborated with others to make a change in the local community. Compared with users who have no children, parents are more likely to have attended community events as a consequence of using local forums.

Besides the role that people play in the community, gender and race also play a role on some specific outcomes. Women are less likely to have contacted a government official and done favors for others. People of color are more likely to perceive that they have learned new ideas or points of views. These significant and consistent relationships add to the associations that already exist between offline characteristics of the users, their goals and their kinds of participation in the system. Together, these results provide support for the existence of a relationship between offline aspects of the individuals and sustainability, as proposed in our framework.

The kind of forum that a user participates in and the characteristics of this forum also relate in some specific outcomes. Compared to users who participate in city-wide forums only, those who participate in neighborhood forums are more likely to have become more satisfied with their local communities and done favors for other members of their communities. Users who participate in both kinds of forums are more prone to contact a government official than users who participate in the city-wide forums only.

We found evidence of a negative effect for users who participate in neighborhood forums instead of only city-wide forums. Participating in a neighborhood forum is negatively related to have learned from other people's perspectives. Together, these findings provide further evidence that the two kinds of forums are different and, therefore, lead to different outcomes. While it can be expected that neighborhood forums make people more likely to do favors for other members of their community, they are less effective than city-wide forums to help their members learn new ideas or points of view. It is possible that the smaller scope of neighborhood forums and their focus on community constraint the diversity of the topics or ideas presented in the online conversation among neighbors.

Among users who participate in neighborhood forums, those who belong to forums with offline outreach are more likely to report that they have learned about other people's races. Users who participate in more active neighborhood forums tend to have felt welcomed in the forums, become more satisfied with their local communities, and learned how to influence decisions in their communities. These users were also more likely to have been involved in actions such as collaboration with others to encourage a change in the community, attending a community event and visiting a business. Again, these results provide evidence that the level of activity in the forum has a significant effect on the consequences of using the forums that the users perceive. Beyond and above the effect of the offline characteristics of the users, their goals and activities in the forums, thriving neighborhood forums are positively related to achieve desired outcomes.

7.3 SUMMARY OF RESULTS AND DISCUSSION

Together, the results of these two studies shed light on the online and offline aspects that have an effect on sustainability at an individual level. Based on user behavior and user survey data, we attempt to characterize (1) who uses the E-Democracy forums, (2) how the users' online behavior can explain whether they receive a response from the community and if they decide to continue posting in a local forum, and (3) how the users' offline characteristics relate to their goals, activities and perceived impact of the forums.

Data from the 2014 E-Democracy user survey reveals that the platform user-base is largely comprised of highly-educated adults and property owners, with household incomes above the poverty line in the US. While the platform strives to attract racially diverse audiences, a broad majority (92%) of the survey respondents self-reported to be white. Nevertheless, and somewhat unexpectedly, more than 60% of the respondents were women, which is an encouraging sign of the inclusion of female voices in online public discourse.

Even though about two-thirds of the new threads in the neighborhood forums do not receive a single answer, most of their authors (80%) continue to post in the forums within a timeframe of a calendar quarter. Receiving a response to the previous post, especially for newcomers, positively influences the probability of posting again. However, several other characteristics of the posts' authors significantly affect this decision. Particularly, the characteristics of the users' prior experience in the local forums shape the probability of user retention. The number of prior contributions and the level of interactions with other users on-site positively relate to the probability of posting again in the forum. These two variables are also highly correlated to a lower chance of migrating to other forums in the E-Democracy platform. Thus, central users and power-users are very likely to continue being active in the same forums in the future. This hints at the existence of core groups of committed members that generate content constantly in the neighborhood forums.

Although responsiveness is low, there are individual characteristics that significantly associated with the chance of getting an on-site response from the local community. First, local communities are more likely to provide answers for newcomers than old-timers or power-users. This result is contrary to prior research in other social media [4]. We speculate that this can be related to different posting styles between newcomers and more experienced users. It is possible that those users who

post more often have an agenda related to information dissemination, while newcomers bring a more conversational style to the forums. However, further research is needed to confirm this.

More central users in the social structure of the forums are also more likely to receive responses. This result connects to the notion of social capital and reciprocity. Those users who have been involved in on-site interactions might have been developed a need for reciprocity from other forum members. Therefore, when central users post, they are more likely to receive a response as a result of these reciprocity obligations. This can be interpreted as a way of capitalizing on the social capital that is available to them through the online forums.

Post characteristics that a user adds to a neighborhood forum are also relevant. Requests for active mobilization, such as event announcements and longer posts, are much less likely than other kinds of content and shorter posts to receive an answer. Overall, these results indicate that who posts and what is being posted are influential aspects on the probability of receiving on-site responses from the local communities.

While the archival data provides evidence that on-site responsiveness is important for user retention, the survey data allows us to complement and extend the understanding of the implications of on-site communication among users. We found that on-site communication also has a significant impact on the users' satisfaction with the local forums' possibilities to get local information.

Besides, there is also evidence that the topics raised in local forums also encourage off-site communication among people in the neighborhoods. This off-site communication not only exists, but is significantly associated with users' positive perceptions of satisfaction in regard to the forum's role in information exchange and community involvement. This finding provides evidence that the local forums are succeeding at creating opportunities, not only for information awareness, but also for more reciprocal information exchanges and community engagement. However, off-site communication also poses a challenge for the evaluation of the systems. If local systems were only analyzed through the examination of the users' online behavior, then an influential aspect of the users' experiences in regard to the system would be overlooked. Mechanisms to deal with this challenge should be developed.

We also found partial support for a significant relationship between offline user characteristics and sustainability. Women are more likely to expect that a local information system helps them to get local information and engage with others in the community. Besides, people who work in

the community are more likely to be involved in on-site communication. This reveals that gender and the role that people play in the community, at least, are characteristics that could be taken into account when designing personalization strategies of participatory information systems for urban communities. Different user profiles have different goals and might want to engage with the local systems in diverse ways.

The survey data analysis also sheds additional light on understanding the impact of participatory information systems on their users. First, we learned that satisfaction with the forums is strongly influenced by people's goals, on-site and off-site communication, and the kinds of forums in which they participate. Among the three goals we studied, perhaps the most elusive goal to meet is encouraging the exchange of views and ideas among users.

While most E-Democracy users agree that they are satisfied in regard to getting local information and engaging with other people or initiatives in the community, when they are asked about particular feelings or actions that might have been influenced by the use of the forums, they seem to perceive lower levels of impact. Even though a majority of people self-report high levels of satisfaction with the local forums, a minority of people report changes on their feelings or actions related to involvement with their community. We conclude that assessing the impact of these local information systems at this level of granularity requires the development of more robust metrics that can lead to more conclusive results.

The characteristics of the forums in which the users participate are also influential. The data provides evidence that users who are part of neighborhood forums are more interested in getting local information and engaging with their communities, and less interested in information exchange than users who participate in city-wide forums only. These differences translate to perceptions of satisfaction as well. Neighborhood forum users are much more likely to perceive satisfaction with the goal of community engagement, but less likely to report that they have learned from other people's perspectives. This result confirms that these two kinds of forums appeal to different motivations and achieve different kinds of impact.

Among neighborhood forums, the level of activity in the forums has a major role on user perception. Compared to users of less active neighborhood forums, users of more active neighborhood forums are more likely to perceive the goals of getting information and engaging with the community as very important. They are also much more likely to be involved in both on-site and off-site

communication. Furthermore, they are more prone to be satisfied in regard to these three goals. In active neighborhood forums, the relationship between on-site communication and satisfaction with the goals of information exchange and community involvement becomes not statistically significant. This suggests that in active forums, both passive and active online participation leads to satisfaction. All of these results seem to indicate that achieving higher levels of activity in the local forums is crucial for users to perceive that local information systems make a difference.

Table 49: Mapping between survey questions and other outcomes

Dependent variables	Questions in the survey
Outcomes	As a result of information or discussions on your Neighbors Forum, in the last 12 months,
Felt welcomed	I feel that my participation is welcomed or valued by others on the forum
Became more satisfied	I am more satisfied with my local community as a place to live or work
Learned new views	I have been introduced to new ideas or points of view
Learned about races	I have learned more about my neighbors of races, ethnicities, or home languages different from my own
Got better informed	I am more informed about issues that affect my community
Learned how to influence	I have learned more about how to influence decisions in my community
Signed	Sign a petition
Contacted gvmt.	Contact an elected official or government office
Donated	Donate money to a local charity or cause
Volunteered	Perform local volunteer work
Did favors	Do favors for or share goods (e.g., lend tools, give away items) with neighbors or local community members
Collaborated for change	Work with other residents to make change in the local community
Attended discussion mtgs.	Attend a community meeting in which local issues were discussed
Met neighbor	Meet other community members in person
Attended event	Attend community events such as a festival, picnic, or parade
Used local services	Visit a business, restaurant, or hire someone recommended on the forum

Table 50: Mapping between survey questions and independent variables

Independent variables	Questions in the survey
Roles in the community	What are your roles in the local community of your Neighborhood forums? (Please check all that apply):
Resident	Resident
Parent	Parent or guardian of children 18 or under living at home
Worker	Student
	Local business owner
	Employee of a business in the forum area
	Employee or volunteer with a nonprofit, community/cultural organizations, or place of worship serving the forum area
	Employee of local government or school
	Elected official
Women	Your gender: male, female (coded as 1), other
People of color	Your race/ethnitic (please check all that apply):
	White or European American (coded as 0)
	Black, African American, or African immigrant OR Asian OR American Indian or Alaska Native OR Native Hawaiian or Other Pacific Islander OR Hispanic/Latino OR Other (coded as 1)

Table 51: Descriptive statistics of goals, activity, satisfaction and impact

Dependent variable	Mean	Std. Dev.
Goal: get information	0.783	0.412
Goal: exchange information	0.623	0.485
Goal: engage with community	0.668	0.471
Communicated on-site	0.565	0.496
Communicated off-site	0.420	0.494
Satisfaction: get information	0.774	0.418
Satisfaction: exchange information	0.694	0.461
Satisfaction: engage with community	0.722	0.448
Felt welcomed	0.545	0.498
Became more satisfied	0.555	0.497
Learned new views	0.670	0.471
Learned about races	0.316	0.465
Got better informed	0.787	0.410
Learned how to influence	0.448	0.498
Signed	0.160	0.367
Contacted government	0.183	0.387
Donate	0.109	0.312
Volunteered	0.084	0.278
Did favors	0.220	0.415
Collaborated for change	0.153	0.360
Attended discussion mtgs.	0.278	0.448
Met neighbor	0.171	0.377
Attended event	0.414	0.493
Used local services	0.425	0.495

Table 52: Descriptive statistics of offline characteristics of the users

Independent variable	Mean	Std. Dev.	Min	Max
Resident	0.847	0.36	0	1
Worker	0.406	0.491	0	1
Parent	0.267	0.443	0	1
Women	0.642	0.48	0	1
People of color	0.078	0.268	0	1

Table 53: Performance: Goals

	Get		Exchange		Engage	
	(1)	(2)	(3)	(4)	(5)	(6)
Resident	1.112	1.139	0.643	0.765	1.020	0.821
Worker	0.879	0.876	1.153	1.103	0.974	0.932
Parent	1.237	1.149	1.040	1.084	1.123	1.190
Women	1.740***	1.696***	1.085	1.040	1.766***	1.570**
People of color	1.197	1.183	1.353	1.435	1.343	1.339
Neighborhood	1.836*		0.414**		2.132**	
Both	1.922*		0.505*		2.140**	
Outreach		1.768**		1.202		1.278
Active		1.946***		1.318		1.856***
Observations	1139	1070	1135	1068	1128	1060
Pseudo R^2	0.022	0.032	0.013	0.007	0.023	0.023

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 54: Performance: Activity

	Communicated on-site		Communicated off-site	
	(1)	(2)	(3)	(4)
Goal: get	1.174	1.043	1.429*	1.371
Goal: exchange	1.499**	1.505**	1.109	1.042
Goal: engage	1.580**	1.603**	1.471**	1.631**
Resident	1.761*	1.630	1.115	1.140
Worker	1.554***	1.714***	1.118	1.249
Parent	0.942	0.922	1.170	1.147
Women	0.794	0.808	1.096	1.147
People of color	0.567*	0.709	0.660	0.933
Neighborhood	0.682		0.432**	
Both	0.818		0.490**	
Outreach		0.772		0.923
Active		3.025***		1.964***
Observations	1121	1054	1121	1054
Pseudo R^2	0.038	0.083	0.024	0.038

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 55: Performance: Satisfaction

	Satisfaction: Get		Satisfaction: Exchange		Satisfaction: Engage	
	(1)	(2)	(3)	(4)	(5)	(6)
Comm. on-site	1.844***	1.552*	1.382*	1.184	1.705***	1.367
Comm. off-site	1.375	1.094	1.466**	1.407*	1.750***	1.587**
Goal: get	3.126***	3.131***	1.521*	1.712**	1.462*	1.339
Goal: exchange	0.869	0.890	2.389***	2.414***	1.190	1.204
Goal: engage	1.636**	1.829**	1.278	1.241	2.421***	2.772***
Resident	1.184	0.975	0.877	0.829	0.746	0.798
Worker	0.796	0.755	0.825	0.867	0.863	0.946
Parent	1.009	0.992	1.329	1.285	1.136	1.038
Women	1.104	1.182	1.170	1.255	1.287	1.252
People of color	1.090	1.056	0.641	0.855	0.919	0.987
Neighborhood	1.660		1.204		2.093**	
Both	1.186		1.075		1.677	
Outreach		1.021		0.637**		0.904
Active		2.495***		1.961***		2.818***
Observations	1109	1044	1108	1042	1098	1035
Pseudo R^2	0.098	0.124	0.076	0.101	0.097	0.132

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 56: Performance: Felt welcomed and satisfied

	Felt welcome		Became more satisfied	
	(1)	(2)	(3)	(4)
Comm. on-site	2.217***	2.040***	0.973	0.883
Comm. off-site	1.720***	1.622***	1.409**	1.364*
Goal: get	1.533*	1.512*	2.122***	2.165***
Goal: exchange	1.855***	1.846***	1.625***	1.586**
Goal: engage	1.353*	1.395*	1.445*	1.527**
Resident	0.841	0.938	1.226	1.069
Employee	0.951	0.986	0.887	0.934
Parent	1.116	1.016	1.146	1.109
Women	0.868	0.973	0.849	0.788
People of color	0.830	0.854	0.811	0.832
Neighborhood	1.603		2.385**	
Both	1.278		2.692***	
Outreach		0.789		0.922
Active		1.377*		1.885***
Observations	1117	1051	1113	1048
Pseudo R^2	0.089	0.093	0.065	0.073

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 57: Performance: Learned new perspectives

	Learned new views		Learned about races	
	(1)	(2)	(3)	(4)
Comm. on-site	1.152	1.049	0.628**	0.684*
Comm. off-site	1.743***	1.661***	1.467**	1.401*
Goal: get	1.145	1.235	1.164	1.145
Goal: exchange	1.536**	1.519**	1.351*	1.316
Goal: engage	1.483*	1.597**	1.138	1.270
Resident	0.726	0.765	0.730	0.689
Employee	1.210	1.279	1.005	0.981
Parent	1.347	1.342	0.916	0.995
Women	0.984	1.031	0.853	0.812
People of color	0.712	0.768	1.646*	1.528
Neighborhood	0.493*		0.402***	
Both	0.597		0.580*	
Outreach		0.864		1.421*
Active		1.496*		0.904
Observations	1118	1052	1117	1051
Pseudo R^2	0.054	0.059	0.036	0.029

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 58: Performance: Got information

	Got better informed		Learned how to influence	
	(1)	(2)	(3)	(4)
Comm. on-site	1.313	1.193	1.166	1.110
Comm. off-site	1.489*	1.286	1.616***	1.483**
Goal: get	2.206***	2.227***	1.031	1.091
Goal: exchange	1.251	1.170	1.578***	1.586***
Goal: engage	1.149	1.234	1.536**	1.568**
Resident	1.177	1.211	1.031	0.948
Employee	1.066	1.176	0.841	0.855
Parent	1.124	1.151	1.162	1.133
Women	0.810	0.790	0.902	0.909
People of color	0.776	0.864	1.222	1.309
Neighborhood	0.772		1.055	
Both	1.043		1.165	
Outreach		1.147		0.771
Active		1.985***		1.240
Observations	1118	1053	1114	1048
Pseudo R^2	0.050	0.064	0.039	0.042

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 59: Performance: Engaged in activism

	Signed		Contacted gvmt.	
	(1)	(2)	(3)	(4)
Comm. on-site	1.406	1.345	1.248	1.314
Comm. off-site	1.269	1.269	1.692**	1.776***
Goal: get	2.299**	2.295**	1.232	1.374
Goal: exchange	1.143	1.213	1.249	1.246
Goal: engage	1.230	1.268	1.816**	1.770**
Resident	0.721	0.997	0.793	0.722
Employee	0.868	1.001	1.017	1.000
Parent	1.195	1.135	0.766	0.811
Women	1.196	1.264	0.830	0.711*
People of color	1.475	1.499	0.499	0.587
Neighborhood	1.198		0.729	
Both	1.231		1.014	
Outreach		0.833		1.423
Active		1.186		1.026
Observations	1116	1049	1116	1051
Pseudo R^2	0.035	0.038	0.046	0.049

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 60: Performance: Donated resources or time

	Donated		Volunteered	
	(1)	(2)	(3)	(4)
Comm. on-site	1.728*	1.639*	1.181	1.145
Comm. off-site	2.511***	2.398***	1.396	1.323
Goal: get	1.583	1.521	1.384	1.621
Goal: exchange	1.979**	2.208**	1.119	1.116
Goal: engage	2.093*	2.385**	1.820*	1.792
Resident	0.757	0.673	0.871	0.774
Employee	0.800	0.730	1.168	1.091
Parent	0.797	0.808	0.680	0.704
Women	0.847	0.869	0.767	0.761
People of color	1.091	1.131	1.591	1.683
Neighborhood	2.697		3.311	
Both	2.343		4.361*	
Outreach		1.222		1.216
Active		1.441		1.374
Observations	1115	1049	1113	1047
Pseudo R^2	0.091	0.100	0.039	0.035

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 61: Performance: Collaborated

	Did favors		Collaborated for change	
	(1)	(2)	(3)	(4)
Comm. on-site	2.831***	2.589***	2.122***	2.226***
Comm. off-site	2.273***	2.081***	2.024***	1.995***
Goal: get	1.458	1.320	1.007	0.869
Goal: exchange	1.039	1.003	1.399	1.489*
Goal: engage	1.880**	1.979***	1.365	1.480
Resident	1.328	1.357	0.603	0.663
Employee	0.963	1.048	1.375	1.441*
Parent	0.990	0.948	0.869	0.913
Women	1.186	1.287	0.698*	0.704
People of color	0.908	1.006	1.230	1.397
Neighborhood	3.941**		1.079	
Both	3.392**		1.212	
Outreach		1.037		1.409
Active		2.713***		1.139
Observations	1115	1050	1114	1048
Pseudo R^2	0.108	0.124	0.066	0.076

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 62: Performance: Engaged with local affairs and people

	Attended discussion mtgs.		Met neighbor	
	(1)	(2)	(3)	(4)
Comm. on-site	1.384*	1.478*	2.003***	2.036***
Comm. off-site	1.256	1.212	2.378***	2.228***
Goal: get	1.758**	1.976**	1.018	1.109
Goal: exchange	1.337	1.220	1.198	1.221
Goa: engage	1.179	1.341	2.171***	2.307***
Resident	0.702	0.852	0.704	0.698
Employee	1.176	1.275	1.135	1.177
Parent	0.870	0.901	0.877	0.888
Women	0.832	0.791	0.799	0.851
People of color	0.587	0.613	1.414	1.630
Neighborhood	0.655		1.421	
Both	0.804		1.527	
Outreach		1.338		0.949
Active		0.916		1.115
Observations	1113	1047	1114	1048
Pseudo R^2	0.036	0.040	0.079	0.084

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 63: Performance: Participated in local activities and economy

	Attended event		Used local services	
	(1)	(2)	(3)	(4)
Comm. on-site	1.195	1.119	1.854***	1.598**
Comm. off-site	1.160	1.064	1.742***	1.551**
Goal: get	1.819***	1.902***	1.973***	2.052***
Goal: exchange	0.912	0.847	0.912	0.943
Goal: engage	1.755***	1.969***	1.338	1.427*
Resident	0.870	1.050	1.026	1.153
Employee	1.130	1.304*	0.825	0.909
Parent	1.541**	1.492**	0.971	0.911
Women	1.010	1.033	1.037	1.092
People of color	0.699	0.734	0.998	1.140
Neighborhood	1.211		1.334	
Both	0.995		1.065	
Outreach		0.867		0.747
Active		1.600**		2.249***
Observations	1118	1052	1111	1046
Pseudo R^2	0.043	0.058	0.060	0.085

Exponentiated coefficients

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

8.0 STUDY 3: SYSTEM DESIGN AND ITS IMPACT ON SUSTAINABILITY OF ONLINE FORUMS FOR URBAN COMMUNITIES

The previous two studies allowed us to explore how collective and individual aspects influence sustainability of participatory information systems for urban communities. Our findings reveal that urban communities with larger proportions of people of color are related to local information systems with higher retention and performance. However, while the forums' target urban communities have more than 40% of people of color on average, the survey data reveals that less than 8% of the user survey respondents were people of color. This suggests that the E-Democracy user base is not representative of their target neighborhoods' racial diversity. Beyond that, the survey data also shows that more than 60% of the respondents are women, which is unexpected given that only a minority of the civic technologies' users in the UK are women [56]. These findings motivate this study, in which we explore whether the design of E-Democracy has influenced the participation of people of color and women, who are often underrepresented in public civic online discourse.

One of goals of the E-Democracy platform is to foster digital inclusion by increasing the online participation of women and people of color.[1] These two specific populations were particularly reluctant to join the earliest forums of the platform. This pattern is aligned with a broader phenomenon of systematic inequalities on Internet access and use. Early research on the digital divide often reports that women and people of color had less access to and usage of the Internet, compared to men and white people respectively [70, 112]. However, more recent research reports that the access divide is disappearing [117]. Besides, the differences of Internet use have become more nuanced [70, 131] and might be linked to other variables such as social class [131], digital skills [63], and self-efficacy [65].

[1] http://pages.e-democracy.org/Inclusive_Social_Media Last retrieved on November 7th, 2015.

In the US, the most recent national surveys conclude that women are as likely as men to participate in social media in general. However, there are significant differences among specific social media platforms [3]. While women are more likely than men to engage in Pinterest, Facebook, and Instagram, they are significantly less likely to participate in online forums [3], which is the technology used by E-Democracy.

In the particular context of civic participation, the divide remains in the US. Beyond social class differences, there is significant inequality of participation across races. White people are more likely than African-Americans and Hispanics to engage in civic communications, including political communications, both online and offline [136].

To deal with these challenges of digital inclusion, E-Democracy has enacted design decisions to encourage the online participation of women and people of color on the platform. Based on (1) interviews with the founder and senior staff members of E-Democracy, and (2) content analysis of the E-Democracy blog[2] that describes the major events in the lifecycle of the platform, we identified three major system design decisions that aim to foster digital inclusion:

- City vs. neighborhood forums: Early on, E-Democracy had only city-level forums that focused on creating a virtual space for political discussion. These kinds of forums were successful at attracting residents to provide content for the discussions. However, white males were over-represented among the contributors. As an attempt to reach a more diverse audience, E-Democracy created neighborhood-oriented forums that aim to increase community involvement as opposed to political discussion. The intention was to create a less polarized virtual space that might be more welcoming for newcomers, especially those who might be less accustomed to participating in online conversations.

- Offline outreach: To further encourage the participation of under-represented populations, E-Democracy implemented a special user engagement strategy in neighborhoods with high racial diversity.[3] This strategy focused on reaching the neighborhoods' residents through offline channels in order to increase awareness about the forums and to sign up new users. This strat-

[2] http://blog.e-democracy.org/ Last retrieved on October 13th, 2015.

[3] http://forums.e-democracy.org/projects/engage/inclusive-social-media/ Last retrieved on November 7th, 2015.

egy was implemented through several mechanisms,[4] [5] including hosting workshops about how to use the forums, attending neighborhood festivals to publicize the forums, and conducting door-to-door advertisement. When residents were interested in the forums, they were asked to sign up using paper forms that were later used to manually register them as members of the corresponding forum. Due to budgetary issues, the offline outreach was only conducted in some, not all, highly-diverse neighborhoods.

- Maximum daily number of posts: All E-Democracy forums constrain the number of daily messages a user can post in order to prevent a few members from dominating the activity in the forum [39]. This rule also aims to diversify the voices that are represented in the forums.[6] The majority of the forums restrict the daily number of posts to two per user, while others allow between three to six posts per user per day.

To assess the impact of these design decisions on digital inclusion, we conducted longitudinal analyses of the participation of women and people of color in the E-Democracy forums.

8.1 RESEARCH METHODS

For this study, we first had to determine the gender and race of the E-Democracy users. Then, we used longitudinal regression analysis to model the influence of the design decisions on digital inclusion. The details of these methods are explained in this section.

8.1.1 Determining gender and race of users

Data about the gender and race of the E-Democracy users were not available in the platform: therefore, we used an alternative mechanism to determine these demographics. Given that the users are strongly encouraged to provide real names to register in the system, we matched the users' first and last names to aggregated data from The United States Social Security Administration (SSA) and the US Census in order to estimate the probability of a name belonging to a specific gender

[4] http://blog.e-democracy.org/posts/639 Last retrieved on October 13th, 2015.
[5] http://blog.e-democracy.org/posts/172 Last retrieved on October 13th, 2015.
[6] http://blog.e-democracy.org/posts/51 Last retrieved on November 7th, 2015.

and race, respectively. We adopted these approaches from prior research's efforts as a reasonable estimation of gender [108] and race [47] from people's names.

We used an iterative process to retrieve the probability of each user's first name belonging to a male or female person. In this process, we employed data of the most frequent baby names by gender in the US,[7] according to the SSA. The data includes the frequency of all first names that have more than five occurrences for the same gender in a year. We used the list of names corresponding to the beginning of each decade since 1920 until 2010. For all E-Democracy users' first names, we initially examined the 1960 Census that corresponds to the average birth year of E-Democracy users.[8] If the name was found, the corresponding frequencies were assigned to the user's name in order to compute the probability of each gender. Otherwise, we followed the same procedure of looking for the name in other Censuses from 1920 to 2010 in chronological order. Once the process was over, all found names had been associated with a probability of belonging to a male or female person.

To compute the users' probability of belonging to a racial group, we used two kinds of data: the list of most-frequently occurring surnames from the 2000 US Census,[9] and the racial distribution of the neighborhoods in the cities of Minneapolis and St. Paul according to the 2010 US Census, as compiled by Minnesota Compass.[10] The first dataset includes all last names with more than 100 occurrences in the 2000 Census and their distribution across people of different races. For each user's last name, we retrieved its probability of belonging to a person of each of the races and ethnicities coded in the US Census: White, Black, Asian and Pacific Islander (API), American Indian and Alaska Native (AIAN), Two or More Races, and Hispanic. As the E-Democracy users are affiliated with either a neighborhood or a city forum, we also gathered the racial distribution of these geographical areas from our second source of data. Finally, we computed the joint probability of each race considering both the probability of such race given the user's last name and the user's geographical area.

To determine gender and race, we employed two main mechanisms: a pseudo-random sam-

[7] http://www.ssa.gov/oact/babynames/limits.html Last retrieved on May 18th, 2015.

[8] According to the 2014 E-Democracy's user survey.

[9] http://www.census.gov/topics/population/genealogy/data/2000_surnames.html Last retrieved on May 18th, 2015.

[10] http://www.mncompass.org/profiles/neighborhoods/minneapolis-saint-paul#!areas Last retrieved on September 14th, 2015.

pling according to the probability distributions of each user's name, and a threshold-based approach in which a threshold had to be met to assign the top category as the user's gender and race. For the latter, we used three different thresholds to assess sensitivity of our results. The thresholds were 0.66, 0.75 and 0.90 for gender, and 0.50, 0.75, and 0.90 for race.

8.1.2 Longitudinal analysis of digital inclusion

We merged the gender and race data to the archival data of 35 neighborhood forums and two city forums in Minnesota. Using these data, we computed the proportion of women and people of color who joined and contributed to the forums in different quarters. We also calculated the proportion of posts provided by women and people of color in each forum by quarter. These variables were used as dependent variables to measure the effect of the E-Democracy design decisions on digital inclusion. To estimate this effect, we conducted longitudinal linear regression analyses for nested data to control for repeated measures of the forums over time. The independent variables of these analyses were the design decisions enacted in each forum. We also controlled for other variables such as the year that the forum was created, the forum's tenure at each observation period, and the demographics of each forum's target urban community.

All of the dependent variables associated with female participation in the forums were normally distributed. There were a few potential outliers; however, they did not significantly change the results so we kept them in the dataset. The distribution of the variables related to participation of people of color were right-skewed; therefore, we log-transformed these data in order to employ a more symmetrical dependent variable for the analysis. These variables had no outliers. Figures 11 and 12 show the distribution of our dependent variables as estimated by the pseudo-random mechanism, which are qualitatively similar to the results generated by the other mechanisms.

To ensure that the proportion of women and people of color in each measure of participation is meaningful, we limited our analysis to observation periods in which this proportion was computed in regard to more than ten users whose gender and race had been determined by our algorithms. For this reason, the number of observations in the regression analyses varied from 124 to 327.

Three design decisions were the focus of the study: the kind of forums (city vs. neighborhoods), if the forums have had offline outreach, and the daily maximum number of posts. Given

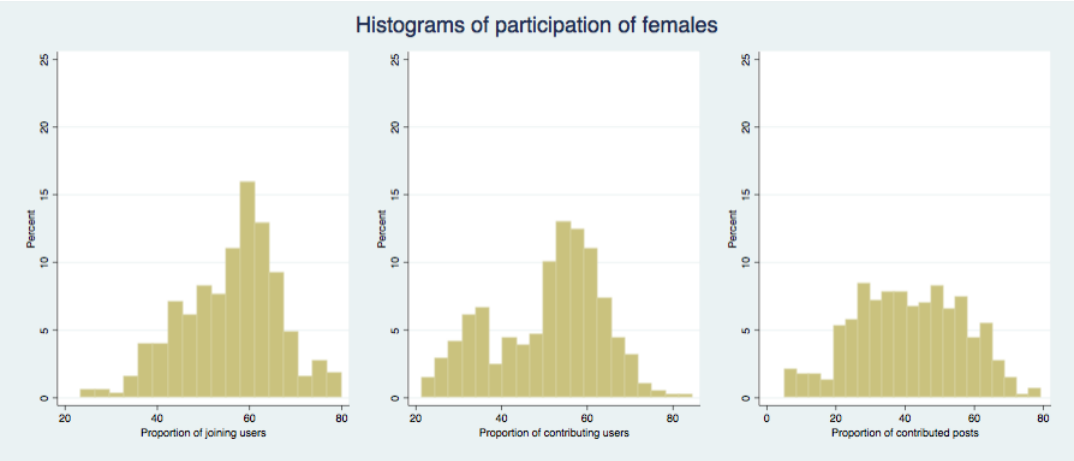


Figure 11: Distribution of the dependent variables of participation of women

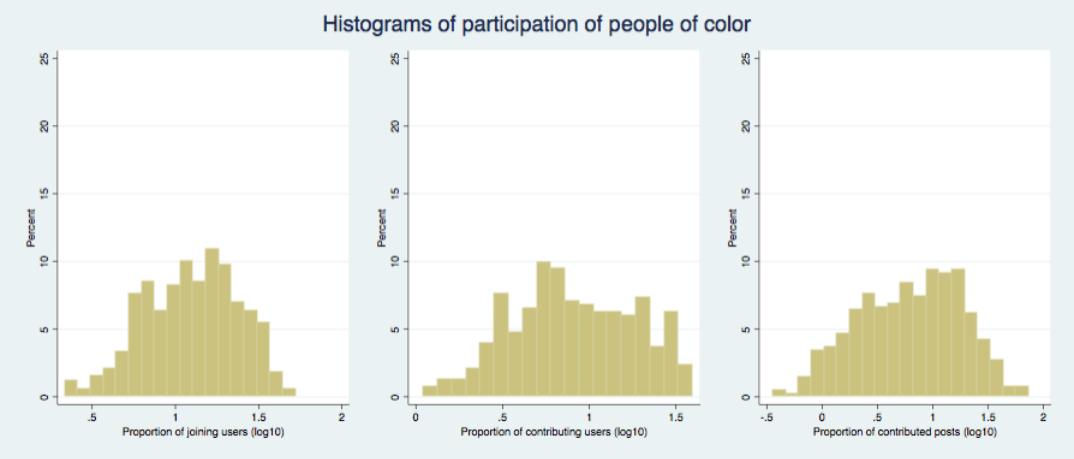


Figure 12: Distribution of the dependent variables of participation of people of color

that none of the city forums have had offline outreach and their maximum number of posts is the same, we limited our analysis of these two design decisions to the 35 neighborhood forums. The assignment of forums to these decisions was neither balanced nor random. Therefore, our analyses had to deal with unbalanced samples and potential co-linearity among the variables that denote the design decisions and other neighborhood demographics. To address some of these issues, we created a binary variable to indicate if the maximum number of daily posts was above three. The correlation between this variable and the offline outreach variable was not significant (Spearman's $\rho = 0.22$, $p = 0.18$). Therefore, we included both of them in the same regression analyses. Table 64 describes the number of forums that enacted each of these design decisions. The forums were also aggregated regarding their year of creation in order to avoid very unbalanced samples. This variable had two levels: if the forum was created before 2011 or not.

Table 64: Neighborhood forums by design decisions

Daily max. # of posts	Without outreach	With outreach	Total
Two or three	20	6	26
Four or five or six	6	5	11

The neighborhood demographics were limited to three main components following the results of a dimension reduction procedure (see subsection 6.1.2). These components are: size, diversity and instability of the neighborhoods. Each was measured by the most representative variable in the US census: population size, % of people of color, and % of people who moved into the neighborhood in 2005 or later. We used a log-transformation of the population size in our analyses.

8.2 PARTICIPATION OF WOMEN AND PEOPLE OF COLOR

Overall, we have data on 13,705 E-Democracy users (members and/or contributors). Our approach was able to retrieve the probabilities of gender for 10,105 of them (74.39%) and the probabilities of race for 11,120 users (81%). The first names of 23.21% of the users were not found in the SSA data. A smaller proportion of user names (2.4%) did not include a clear first name. These cases included organization names, users that registered as families instead of as individuals, and users who reported their initials only. On the other hand, the last names of 15% of the users did not match the surnames from the US Census, and another 4% of the users' names did not include a clear last name.

Our results will report on the analysis of users who were associated with gender and/or race probabilities. Although these results will not describe the whole user base of the system, we believe that they are representative of the actual trends of participation of women and people of color on E-Democracy. First, we reason that the SSA data is unlikely to be biased towards any gender. Hence, our approach to determine gender should lead to representative patterns regarding this aspect.

Second, the US census data that was used to determine race includes information about surnames with more than 100 occurrences. This dataset might be less likely to have information about people of color given that they are minorities in the US. This bias might result in an underestimation of participation of people of color on E-Democracy. Hence, we believe that the results we obtain about participation of people of color in the forums will represent a conservative estimate that could only be higher if this bias indeed exists.

Another source of potential bias could be related to the people who decide to use their initials instead of their full names in E-Democracy. However, we do not have strong evidence to argue that there are systematic biases towards gender or race regarding this behavior. Anecdotal evidence could suggest that women might be less willing to share their identity in this kind of geographically-bounded online forums. If this were the case, this would again result in an underestimation of one of our dependent variables, which would not change the trends of our results and could only increase the effect sizes.

8.2.1 Descriptive statistics about gender

Consistently across the different mechanisms to determine gender, the results indicate that there are more female than male users on E-Democracy. Table 65 details the results of each of the alternative approaches. The pseudo-random approach shows that while women account for 40% of the user base, men comprise 35% of it. The other mechanisms result in lower percentages for both genders, but they all reveal a gap in a range of 4% - 6% between the two genders.

Table 65: Users and posts by gender

		Pseudo-random	t = 0.66	t = 0.75	t = 0.90
Users	Female	5,445 (0.40)	5,406 (0.39)	5,204 (0.38)	4,948 (0.36)
	Male	4,750 (0.35)	4,771 (0.35)	4,537 (0.33)	4,170 (0.30)
	Not classified	0 (0.00)	18 (0.00)	454(0.03)	1,077 (0.08)
Posts	Female	53,267 (0.32)	52,378 (0.32)	50,239 (0.31)	49,013 (0.30)
	Male	88,220 (0.54)	89,077 (0.54)	85,979 (0.52)	83,368 (0.51)
	Not classified	0 (0.00)	32 (0.00)	5,185 (0.03)	9,106 (0.06)
Posts/Users	Female	9.78	9.68	9.65	9.90
	Male	18.57	18.67	18.92	19.99

While women make up the largest proportion of the E-Democracy users, they contribute a significantly smaller share of the posts. Table 65 denotes the estimated amount of posts provided by women and men, according to our four approaches. The pseudo-random mechanism indicates that men have provided 54% of posts and women have contributed 32% of posts. This pattern of difference is similar to the results from threshold-based approaches. The gap in number of posts across genders is estimated to be slightly higher than 20% of the posts.

These opposing trends are due to considerably different contribution patterns of women and

men on E-Democracy. On average, women contribute about half the number of posts that men provide. Our pseudo-random estimation indicates that while women have added 9.78 posts on average, men have generated an average of 18.57 posts. Together, these results reveal that E-Democracy has been able to engage women as users. However, this has not yet resulted in a balanced representation of both female and male voices on the platform.

8.2.2 Descriptive statistics about race

All of the mechanisms to determine race reveal a great deal of participation inequality between white people and people of color in the E-Democracy forums. White people comprise the majority of the users of the platform and the other races and ethnicities together constitute a small minority. According to our pseudo-random approach, 68% of the user base are white people and only 13% of the users belong to any other race or ethnicity. Among the minorities, people whose race is Black and Asian Pacific Islander comprise larger shares of the user base than Hispanics or the remaining races. According to the threshold-based mechanisms, the white majority is in a range of 56% - 71%, and the remaining races are in a range of 6% - 10%. Table 66 details the results of the four approaches.[11]

Table 66 also shows that the different mechanisms are consistent in estimating the proportions of most races, except for white and black races. Compared to the pseudo-random approach, the least conservative threshold mechanism generates a larger proportion of white people and a smaller proportion of black people. This reveals that a considerable amount of last names have comparable probabilities of belonging to either of these two races. This observation makes it necessary to further consider the results provided by all of our alternative mechanisms to determine race in our upcoming analyses.

The white majority in the user base is confirmed in the number of posts. The pseudo-random approach indicates that white people (68% of the user base) contributed 83% of the content in the forums. In turn, users who belong to other races (13% of the user base) added only 7% of the posts. The threshold-based approaches calculate that the share of posts provided by minorities is even lower (see Table 66).

[11] This table uses a gray background color in some cells to indicate that a particular regression was not significant.

Table 66: Users and posts by race

		Pseudo-random	t = 0.50	t = 0.75	t = 0.90
Users	White	9,316 (0.68)	9,715 (0.71)	9,051 (0.66)	7,613 (0.56)
	Black	698 (0.05)	280 (0.02)	119 (0.01)	69 (0.01)
	API	675 (0.05)	628 (0.05)	558 (0.04)	481 (0.04)
	AIAN	8 (0.00)	2 (0.00)	1 (0.00)	0 (0.00)
	2+ Races	37 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
	Hispanic	386 (0.03)	366 (0.03)	251 (0.02)	97 (0.01)
	Undefined	0 (0.00)	129 (0.01)	1140 (0.08)	2,860 (0.21)
Posts	White	136,748 (0.83)	144,003 (0.87)	138,277 (0.84)	121,022 (0.73)
	Black	5,938 (0.04)	650 (0.00)	134 (0.00)	61 (0.00)
	API	1,208 (0.01)	832 (0.01)	553 (0.00)	345 (0.00)
	AIAN	14 (0.00)	13 (0.00)	13 (0.00)	0 (0.00)
	2+ Races	83 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
	Hispanic	2,529 (0.02)	1,686 (0.01)	1,230 (0.01)	597 (0.00)
	Undefined	0 (0.00)	580 (0.00)	7,557 (0.05)	25,739 (0.16)
Posts/Users	White	14.68	14.82	15.28	15.90
	Black	8.51	2.32	1.13	0.88
	API	1.79	1.32	0.99	0.72
	AIAN	1.75	6.5	13.00	-
	2+ Races	2.24	-	-	-
	Hispanic	6.55	4.61	4.90	6.15

On average, white users add more posts than users from any other race. While all mechanisms put the mean contribution of white people at more than 14 posts, the average contribution of any other race is, at most, 8.51 posts. The results are not conclusive regarding the average contribution levels of black people. The pseudo-random approach leads to a mean contribution of more than eight posts. However, the threshold-based mechanisms lower this measure to a quarter of this value. Consistently across all approaches, the results indicate that Hispanics contribute more than the other minorities.

Overall, these results signify that the platform still attracts a white majority who are also more likely than users from other races to express their voices in the platform. The rest of our analyses aims to unveil if any of the design decisions enacted in E-Democracy have helped to re-shape the contribution patterns of women and people of color at the forum level.

8.3 MODELING THE IMPACT OF DESIGN DECISIONS ON DIGITAL INCLUSION

We report the results of our longitudinal regression analyses to model the impact of three design decisions on fostering digital inclusion in E-Democracy. Particularly, we study the participation rates of women and people of color in the online forums. Given that people of color participate at low rates in the platform, we aggregated the participation levels of all people of color together in the following analyses.

8.3.1 Neighborhood vs city forums

Our first set of regression analyses consider longitudinal data (in quarters) of all forums to estimate whether city-wide and neighborhood forums vary in terms of the participation of women and people of color, while controlling for the tenure of the forum at each observation period. Tables 67 and 68 show the results of the population-averaged GEE (generalized estimating equation) models to estimate such effects.

Compared to city-wide forums, neighborhood forums have significantly larger proportions of female users who join and contribute. Furthermore, the neighborhood forums have significantly

larger proportions of posts provided by women. While the sign of the impact of the forum type remains constant across different mechanisms to determine user gender, the effect sizes vary. Specifically, the percentage of female joining users is expected to be from 5.92 to 10.38 units higher in neighborhood forums. The effect size is even larger when analyzing the proportion of female post contributors. All of the alternative mechanisms indicate that the proportion of female contributors is predicted to be more than 25 units higher in neighborhood forums than in city-wide forums. Women also contribute significantly larger shares of content in neighborhood forums than in city-wide forums. Compared to city-wide forums, neighborhood forums are predicted to have a share of women-provided posts that is between 10.55 to 12.63 units larger.

Beyond the forum type, time turned out to be a significant factor on the proportion of woman. Over time (i.e. longer tenure), the proportion of female contributors grows. The effect size is predicted to be in a range of 0.30 to 0.37 units by every quarter of tenure. This result can be related to a more general trend of wider female participation in social media [3].

Regarding race, the analyses show that, as opposed to city-wide forums, neighborhood forums have larger proportions of members and contributors who are people of color. Furthermore, these neighborhood forums also gather broader shares of content provided by people of color. The effect of the kind of forum on these three dependent variables is significant across all of the mechanisms to determine race. Given that we used a log transformation of the dependent variable, the interpretation of the results considers changes in percent (instead of units) of user proportion. In particular, neighborhood forums are expected to increase in proportion of new members who are people of color by a factor of 2.81 to 9.28%. The proportion of contributors who are people of color is predicted to grow by a factor of 4.96% (using the most conservative approach) to a factor of 7.52% (using the pseudo-random mechanism). The positive effect of neighborhood forums on the attraction of new users and contributors is further extended by a positive association with production of content. The proportion of content generated by people of color is between 3.89 and 5.75% larger in neighborhood forums than in city-wide forums.

Forum tenure was found to be a significant factor only when using the pseudo-random approach. Therefore, our data does not provide conclusive evidence regarding the role of time on the participation of people of color in the platform.

Together, these findings provide evidence that the decision to create neighborhood-based fo-

Table 67: Participation of women: City vs neighborhood forums

DVs	IVs	pseudo-random	t = 0.66	t = 0.75	t = 0.90
% of joining users	Neighborhood	10.38^{***}	6.26^{*}	6.40^{*}	5.92^{*}
	Tenure	0.20^{*}	0.11	0.11	0.12
% of contributing users	Neighborhood	27.03^{***}	27.71^{***}	27.15^{*}	25.60^{***}
	Tenure	0.37^{***}	0.37^{***}	0.34^{***}	0.30^{***}
% of posts	Neighborhood	12.60^{***}	12.63^{***}	11.45^{***}	10.55^{**}
	Tenure	0.04	0.01	-0.03	-0.07

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 68: Participation of people of color: City vs neighborhood forums

DVs	IVs	pseudo-random	t = 0.66	t = 0.75	t = 0.90
% of joining users	Neighborhood	9.28^{***}	6.87^{**}	5.52^{**}	2.81^{*}
	Tenure	0.26^{*}	0.09	0.07	-0.05
% of contributing users	Neighborhood	7.52^{***}	6.48^{***}	5.89^{***}	4.96^{***}
	Tenure	$0.07*$	0.07	0.04	0.01
% of posts	Neighborhood	5.75^{*}	5.58^{***}	3.95^{***}	3.89^{***}
	Tenure	0.06	0.09	0.01	-0.01

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Gray background: non-significant regression

rums is associated with attracting larger proportions of women and people of color as users and contributors in E-Democracy. Furthermore, neighborhood forums also gather larger proportions of content provided by people in these populations. Therefore, in the case of E-Democracy, we found that limiting the geographical scope from cities to neighborhoods and shifting the focus from political discussion to community engagement is indeed related to attracting populations that had been elusive in the platform's early days. These results are encouraging with regard to the impact of particular design decisions on fostering digital inclusion in the context of civic technologies.

8.3.2 Offline outreach and daily maximum number of posts

The next set of regression analyses aim to model the effect of the other two design decisions, offline outreach and daily maximum number of posts, on digital inclusion. To conduct these analyses, we constrained our data to the neighborhood forums because city-wide forums never enacted offline outreach strategies and always allowed the same daily maximum number of posts. In the analyses, we also assess and control for the effect of each forum's year of creation, the tenure of the forum at the observation period, and the demographics of each forum's target neighborhood. We estimated the effects of these variables using xt models in Stata 14 to account for correlations of repeated measures of the same forums over time.

8.3.2.1 Modeling the impact of the design decisions Table 69 describes the results of assessing the effect of the design decisions on digital inclusion, controlling for time measures. The table reports the results for participation of women and people of color, including the four mechanisms to determine gender and race.

The results reveal that while there is consistent evidence for a positive impact of the offline outreach on the participation of people of color, there is no evidence in regard to a positive impact of this decision on the participation of women. We found consistent evidence that enacting offline outreach strategies is expected to increase percentages of new members and contributors who are people of color by approximately 0.30%. On the other hand, this design decision is hardly related to female participation in the forums. The only exception is that the two most conservative approaches lead our models to predict that offline outreach is significantly related to smaller proportions of

women joining the forums. Thus, even though offline outreach has a positive impact on attracting more racially diverse participation, it has a null to a negative effect on female participation.

The effect of the policies regarding the daily maximum number of posts is more focused on contribution rather than membership in the forums. However, these results differ in their significance depending on the mechanisms to determine gender and race. Table 69 denotes that forums that had more relaxed constraints of daily maximum posts (four or more posts daily) were associated with a broader proportion of female contributors. This relationship was significant according to the most conservative approaches to determine gender. In these cases, a more relaxed policy is expected to increase the proportion of female contributors by a factor ranging from 3.8 to 4.35 units. On the other hand, the most conservative approach leads our models to predict that a more relaxed policy about the daily maximum number of posts is also associated with larger proportions of contributors who are people of color and larger shares of content provided by them. These findings provide partial support for a positive effect of increasing the maximum number of daily posts.

The year of creation of the forum also has an effect on digital inclusion in the E-Democracy neighborhood forums. Compared to newer forums, older forums (those created in 2010 or earlier) have larger shares of content provided by women, but they have smaller proportions of female contributors. Compared to forums founded in 2011 or later, older forums are expected to exhibit larger shares of women-generated content by a factor above 5.5 units. However, these forums are also expected to have a share of women contributors between 6.83 and 8.12 units lower than forums that were created after 2010. A potential explanation for this phenomenon is that the older forums attracted female early adopters, who might be much more active contributors than average. Over time, more women became active users of the system. However, these women who are late adopters might contribute posts at a smaller rate, thus lowering the overall proportion of women-generated posts in the forums.

Older forums also have significantly less participation of people of color than newer forums. Older forums have engaged fewer new members and fewer contributors who are people of color. This pattern is also extended to content production. Forums that were created before 2011 are expected to have smaller proportions of content created by people of color by a factor that ranges from 0.28 to 0.42%.

Table 69: Participation of women and people of color: Outreach and max number of posts

	p-rand	t=0.66	t=0.75	t=0.90	p-rand	t=0.66	t=0.75	t=0.90
	% of women joining				% of people of color joining			
Created 2010-	.81	0.19	−0.32	−0.32	−0.13*	−0.14*	−0.16*	−0.14
Tenure	0.11	0.01	0.06	0.10	0.01**	0.01**	0.00	0.00
Outreach	1.09	−3.30	−4.21*	−3.7*	0.27***	0.29***	0.32***	0.29***
4+ max posts	−0.24	2.18	2.73	1.74	0.02	−0.03	−0.03	−0.00
R-square	0.01	0.03	0.04	0.04	0.21	0.20	0.25	0.22
	% of women contributing				% of people of color contributing			
Created 2010–	−7.48**	−8.12**	−7.56**	−6.83**	−0.15	−0.32**	−0.40**	−0.38
Tenure	0.43***	0.41**	0.36**	0.32**	0.00	0.01*	0.01*	0.01
Outreach	2.98	2.66	0.42	0.78	0.39***	0.37***	0.39**	0.29***
4+ max posts	2.58	3.27	4.35*	3.8*	0.17	0.11	0.11	0.21*
R-square	0.16	0.17	0.18	0.16	0.42	0.37	0.38	0.24
	% of posts added by women				% of posts added by people of color			
Created 2010-	5.57	5.61	7.82*	7.89*	−0.19	−0.28*	−0.42**	−0.40*
Tenure	−0.04	−0.10	−0.19	−0.23	0.01	0.02*	0.02	0.02
Outreach	3.68	3.24	−0.42	0.02	0.31*	0.37**	0.44*	0.26
4+ max posts	−0.77	−0.47	1.07	0.98	0.29	0.10	0.13	0.27*
R-square	0.03	0.03	0.07	0.07	0.22	0.24	0.24	0.12

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Gray background: non-significant regression

Over time, some measures of inclusion also change significantly. Particularly, the proportion of female contributors increases along with the tenure of the forums by a factor between 0.32 and 0.43 units. However, tenure is not significantly associated with the other dependent variables of female participation. There are positive associations between tenure and all variables of participation of people of color, but their significance level varies across the different approaches to determine race. Nevertheless, the effect size of this impact is generally very low (0.01 to 0.02%).

In general, the models of female participation achieve lower explanatory power than the models of participation of people of color (see R-squared rows in Table 69). Therefore, we conclude that design decisions and time variables better explain the variability of the data about racial digital inequality than gender digital inequality.

8.3.2.2 Controlling for demographics of the target neighborhoods Given that the forums target neighborhoods with different demographics, here, we aim to assess whether the neighborhood demographics have any impact on the measures of digital inclusion and/or the effectiveness of the design decisions. Tables 70 and 71 show the results of the regression analyses to model participation of women and people of color after controlling for time and demographics.[12]

After controlling for neighborhood measures, our analyses point out that the design decisions are not associated with female participation in neighborhood forums (see Table 70). Compared to our prior analyses, the few significant relationships between the design decisions and dependent variables became insignificant when we accounted for the effect of the neighborhood demographics. Neither the offline outreach nor the alternative policies about the daily maximum number of posts per user were found to be significant factors on the proportion of women who join and contribute to the forum or the proportion of content they provide. As the offline outreach was primarily focused on engaging a racially diverse audience, the nonexistent effect of offline outreach on female participation is not surprising. On the other hand, the result regarding the maximum number of posts is more unexpected. The goal of the latter strategy is to diversify the voices that are expressed in the forums. However, we found no support for a significant difference on the proportion of female voices between the two alternative policies of the daily maximum number of posts that are present in E-Democracy.

[12]These tables use a gray background color in some cells to indicate that a particular regression was not significant.

Table 70: Participation of women: Controlling for demographics

DVs	IVs	pseudo-random	t = 0.66	t = 0.75	t = 0.90
% of joining users	Log population	−0.44	1.01	−0.54	−2.14
	% of color	−0.01	0.01	−0.01	−0.04
	% moved 2005+	−0.13	−0.23**	−0.22**	−0.20**
	Created 2010-	1.21	1.06	0.41	0.36
	Tenure	0.08	−0.03	0.01	0.05
	Outreach	2.01	−3.17	−2.98	−1.00
	4+ max posts	−0.03	2.59	2.99	1.96
	R-Square	0.03	0.07	0.08	0.09
% of contributing users	Log population	3.71	2.77	0.65	0.23
	% of color	0.10	0.11	0.07	0.09
	% moved 2005+	−0.23	−0.25	−0.13	−0.19
	Created 2010-	−6.93**	−7.70**	−7.46**	−6.78***
	Tenure	0.43***	0.41**	0.36**	0.32**
	Outreach	0.47	0.45	−0.77	−0.55
	4+ max posts	2.42	3.01	3.87	3.16
	R-Square	0.17	0.18	0.18	0.18
% of posts	Log population	13.25	12.79	12.03	11.68
	% of color	0.13	0.14	0.12	0.13
	% moved 2005+	−0.56*	−0.57*	−0.52*	−0.52*
	Created 2010-	7.99*	7.99*	10.03**	9.99**
	Tenure	−0.05	−0.09	−0.19	−0.23
	Outreach	−2.82	−3.17	−6.28	−6.03
	4+ max posts	0.76	0.98	2.49	2.17
	R-Square	0.15	0.14	0.15	0.15

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Gray background: non-significant regression

After controlling for demographics, time continues to be a relevant factor on the proportion of female contributors and the percentage of posts generated by women. Both measures are significantly related to the year of forum's creation. Compared to forums that were created after 2010, older neighborhood forums are expected to have smaller proportion of female contributors by a factor in the range of 6.78 - 7.70 units. Nevertheless, older forums are also predicted to have larger proportions of posts provided by women (effect size: 7.99 - 10.03 units).

Aligned with the prior analyses, the tenure of the forums is also a significant factor on engaging female contributors. The proportion of women adding content to the forums increases along with the tenure of the neighborhood forums. An additional quarter in a forum's tenure is predicted to increase this proportion by a factor between 0.32 and 0.43 units.

Among the demographic factors, our findings reveal that neighborhoods with more residential instability (i.e. larger proportions of people moving in since 2005) engage smaller proportions of new female members and gather smaller proportions of content provided by women. An increase of 1% in the proportion of people that moved recently into the neighborhood is expected to reduce the proportion of new female members by a range of 0.13 - 0.23 units. Such an increase in neighborhood instability is also significantly related to a drop of slightly more than 0.5 units in the proportion of content provided by women.

Accounting for the effect of demographics improves the ability of our models to explain variability of participation of women in neighborhood forums (see R-squared in Tables 70 and 69). Furthermore, the effect of a particular design decision (daily maximum number of posts) became statistically insignificant after accounting for one of the demographics (residential instability). We interpret these findings as evidence that the neighborhood characteristics have a significant role on the impact that a civic technology (such as E-Democracy) and its design decisions might have on their target communities.

Regarding the participation of people of color, and after controlling for demographic variables, we found that there are very weak signs of the impact of the design decisions on the participation levels of people of color. Unlike the results of the design decisions alone (in the Subsection 8.3.2.1), we found only one significant relationship between a design decision and our dependent variables about racial digital inclusion. All the other relationships maintain their direction, but they become statistically insignificant.

The pseudo-random approach indicates that forums that implemented offline outreach are significantly associated with larger proportions of people of color who join the forums as members. According to this method, offline outreach increases the proportion of people of color by a factor of 0.12%. This effect size is smaller than the effect size that was estimated by the analysis of design decisions alone. Although all of the other mechanisms reveal a positive impact of offline outreach on the proportion of people of color joining the forums, these associations are not significant at the 0.05 level. Besides, offline outreach is no longer a significant factor on either the percentage of contributors who are people of color or the proportion of content provided by them.

The models also indicate that the two different policies regarding the daily maximum number of posts do not differ significantly from each other on their impact on the participation of people of color. Although some of these relationships were significant in the prior analyses (Subsection 8.3.2.1), they became statistically insignificant when accounting for demographics. Again, these results confirm the importance of accounting for the demographics of target urban communities when evaluating civic technologies.

The effect of both time variables (the forum's year of creation and tenure at the observation period) remain the same after controlling for demographics. Forums that were created in 2010 or earlier are significantly and negatively related to all measures of participation of people of color. In turn, we found some evidence that an additional quarter of forum tenure positively relates to the participation of people of color. However, the effect size is rather small and the significance level is only achieved under some mechanisms to determine the user race.

Regarding neighborhood demographics, we found that racial diversity in the neighborhood is significantly associated with larger proportions of contributors who are people of color. A unit of increase in a neighborhood's percentage of people of color is expected to increase the percentage of contributors who are people of color by a factor of 0.01 percent. This positive but small impact is also significant in other measures of participation, according to the pseudo-random and the least conservative approaches to determine race. Beyond that, in some cases, residential instability is significantly related to the participation of people of color. However, these results are very sensitive to the mechanism by which we estimated the race of users.

The inclusion of neighborhood demographics in the models improves the models' ability to explain the variability of participation of people of color in the neighborhood forums (see R-Squared

Table 71: Participation of people of color: Controlling for demographics

DVs	IVs	pseudo-random	t = 0.66	t = 0.75	t = 0.90
Log % of joining users	Log population	0.01	−0.03	0.14	0.05
	% of color	0.01*	0.01	0.00	0.01
	% moved 2005+	0.01*	0.00	0.01*	0.00
	Created 2010-	−0.16***	−0.19**	−0.19**	−0.18*
	Tenure	0.01*	0.01**	0.00	0.00
	Outreach	0.12*	0.17	0.14	0.12
	4+ max posts	−0.04	−0.07	−0.06	−0.04
	R-Square	0.35	0.28	0.34	0.29
Log % of contributing users	Log population	0.08	−0.05	0.07	−0.04
	% of color	0.01***	0.01	0.01	0.00
	% moved 2005+	0.01**	0.01*	0.02	0.01
	Created 2010-	−0.17**	−0.35***	−0.41**	−0.41**
	Tenure	0.00	0.01*	0.01*	0.01
	Outreach	0.09	0.20	0.14	0.24
	4+ max posts	−0.08	−0.02	−0.06	0.12
	R-Square	0.67	0.53	0.50	0.29
Log % of posts	Log population	−0.10	−0.21	−0.08	−0.28
	% of color	0.01***	0.01*	0.01***	0.01**
	% moved 2005+	0.01	0.01**	0.02	−0.00
	Created 2010-	−0.32**	−0.38**	−0.49**	−0.52**
	Tenure	0.01	0.02*	0.02	0.02
	Outreach	0.04	0.22	0.09	0.09
	4+ max posts	0.04	−0.09	−0.10	0.16
	R-Square	0.54	0.46	0.38	0.24

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

rows in Tables 71 and 69). Similar to the study of female participation, the predictive power of all models grew. The magnitude of variability explained by the models is even larger for racial digital inclusion than for gender digital inclusion.

While the offline outreach was undoubtedly a significant factor on fostering racial digital inclusion in our initial analysis, this evidence becomes weaker when accounting for demographics (compare Tables 71 and 69). Particularly, a neighborhood's racial diversity rises to be the most influential demographic in regard to digital participation of people of color. Therefore, we hypothesize that the effectiveness of offline outreach might depend on the racial diversity of the target neighborhood. In the next subsection, we further analyze the data to explore this hypothesis.

8.3.2.3 Zooming in racially-diverse neighborhoods To test a potential interaction effect between racial diversity in the neighborhoods and offline outreach on digital inclusion, we classified the neighborhoods in two ways: above and below median racial diversity. Data analysis revealed that offline outreach strategies were only implemented in neighborhoods that had above-median racial diversity, but not to all of them. We focused our analysis on these highly diverse neighborhoods to better assess the impact of offline outreach in similarly diverse neighborhoods. The results are shown in Table 72.

Among neighborhoods with above-median racial diversity, we found some evidence that offline outreach is significantly associated with broader participation of people of color in E-Democracy. However, this evidence is not conclusive across the different mechanisms to determine race. Particularly, the least conservative approach (t = 0.50) leads our model to provide support for a significant effect of offline outreach strategies on all the variables of participation of people of color. Offline outreach is predicted to increase the percentage of people of color who join the forums by 0.34%, compared to forums that did not enact this design decision. Offline outreach is also expected to increase the proportion of contributors who are people of color by a factor of 0.66%. The share of posts generated by people of color is expected to increase by a factor of 0.91% percent in forums with offline outreach. Among these relationships, only the relationship between offline outreach and percentage of people of color who join a forum was significant according to the pseudo-random approach. Other approaches to determine race did not result in statistically significant relationships between offline outreach and participation of people of color.

Table 72: Participation of people: Zooming in racially diverse neighborhoods

DVs	IVs	pseudo-random	t = 0.66	t = 0.75	t = 0.90
Log % of joining users	Log population	−0.24	−0.29	−0.15	−0.23
	% moved 2005+	0.01**	0.01*	0.01	0.01
	Created 2010-	−0.26***	−0.29**	−0.36*	−0.34*
	Tenure	0.00	0.00	0.01	0.00
	Outreach	0.24*	0.34*	0.31	0.30
	4+ max posts	0.04	−0.08	−0.04	−0.01
	R-Square	0.17	0.19	0.22	0.20
Log % of contributing users	Log population	−0.20	−0.81	−0.66	−0.44
	% moved 2005+	0.01	0.03	0.01	0.00
	Created 2010-	−0.25	−0.63*	−0.62*	−0.53*
	Tenure	0.01	0.01*	0.01	0.01
	Outreach	0.24	0.66*	0.47	0.41
	4+ max posts	0.09	0.04	0.20	0.16
	R-Square	0.24	0.25	0.19	0.14
Log % of posts	Log population	−0.64	−1.26*	−1.08*	−0.80
	% moved 2005+	0.01	0.05**	0.02	0.01
	Created 2010-	−0.47	−0.82*	−0.77*	−0.69
	Tenure	0.01	0.02*	0.02	0.01
	Outreach	0.20	0.91*	0.55	0.38
	4+ max posts	0.22	−0.01	0.26*	0.21
	R-Square	0.14	0.24	0.15	0.07

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Gray background: non-significant regression

These analyses among highly diverse-neighborhoods also confirm the impact of all time variables. Overall, these regressions have a lower predictive power than the ones reported in Table 71. This might be related to the smaller size of the dataset that now includes only half of neighborhoods (those with above median racial diversity).

Together, all of these results indicate that the impact of these two design decisions is highly sensitive to the demographics of the forums' target urban communities. While an analysis of the design decisions alone show strong support for the impact of the offline outreach, the evidence becomes weaker when accounting for neighborhood demographics. Overall, we can confidently argue that the alternative policies regarding the maximum number of daily posts per person (four or more vs. three or less) do not significantly differ from each other when it comes to fostering digital inclusion. Aligned with the goals of the offline outreach strategy, we found partial evidence that this design decision positively affects the forum participation of people of color. On the other hand, offline outreach did not influence female participation. While the relationships between offline outreach and participation of people of color is consistently positive, their significance value is sensitive to the mechanisms by which we determined user race. Therefore, we believe that our results are encouraging but not yet conclusive.

8.4 SUMMARY OF RESULTS AND DISCUSSION

Summing up, this chapter has explored the impact of three design decisions on gender and racial digital inclusion. To assess such effects, we adopted an approach to automatically estimate people's gender and race using their first and last names. We reported the results generated by four alternative approaches to determine both gender and race. Thus, we can assess the sensitivity of our results to the mechanisms we used to determine race and gender. Our main findings are:

- While E-Democracy attracts more women than men as users and contributors, women account for a considerably smaller proportion of posts in the platform. On average, a woman provides about half the amount of posts that a man contributes. This finding complements the current assessment of the gender digital divide. Aligned with current surveys in the US, it shows that women use participatory information systems at similar rates as men. At the same time, and

unlike what has been observed in the UK [56], our results indicate that women make up the majority of the user base of this civic technology. Nevertheless, the online discourse in this civic technology is still highly dominated by male voices.

- The broad majority of the E-Democracy users are white people and this is further reflected by the amount of content generated by these users. Although the majority of white users is estimated to be in a range of 56% - 71%, our results show that their share of posts is between 73% and 87% of all posts on the platform. Correspondingly, a small minority of users are people of color. Our results estimate this proportion in a range of 5% - 13% of the user base. Beyond that, our preliminary analysis showed that the proportion of users who are people of color is right-skewed across all forums. This pattern reveals that most of the forums in the platform have a very low proportion of users who are people of color. This overall pattern is particularly discouraging, as one of the main goals of E-Democracy is engaging a racially-diverse audience. Nevertheless, these patterns are aligned to the US national patterns of higher levels of civic engagement online and offline among white people.
- Compared to city-wide forums with a focus on political discussion, neighborhood forums with a focus on community involvement had a significant and positive impact on digital inclusion. Neighborhood forums have broader participation of both women and people of color in terms of user proportion and amount of content. These findings are encouraging regarding the positive outcomes of creating virtual spaces for small geographical communities that aim to create a friendly environment to discuss local issues, without creating the potential entry barriers commonly associated with political discussion.
- E-Democracy has always enacted rules to constrain the maximum number of posts a user can provide in a single day. The initial threshold was two posts each day. The goal of such a rule is to avoid monopolization of the online discussion, thus encouraging more people to contribute their opinions to the online conversation. This rule has created controversy among the users, and some forums have made this rule more flexible. We compared two policies regarding this threshold: two or three posts max vs. four to six post max. Overall, we found no evidence that these two policies differ significantly from each other regarding their ability to influence the participation of women and people of color. Although these results do not support or challenge the belief that constraining the number of posts helps to diversify public discourse, it does

provide evidence that a small change in the threshold has a negligible effect on diversifying the voices represented in civic technology.

- We found partial support for a positive relationship between offline outreach strategies in the target neighborhoods and broader participation of people of color in the online forums. While there was strong evidence of a statistically significant positive relationship when assessing the design decisions alone, the significant levels were not consistently achieved when controlling for neighborhood demographics. Although not all of the relationships became insignificant, the results seem less compelling and more sensitive to alternative mechanisms to determine the race of the users. We conclude that these results are encouraging in regard to the effectiveness of this design decision, but remain inconclusive.
- Methodologically, we found evidence that accounting for demographics when assessing the impact of participatory information systems for urban communities is critical. In several cases, the impact of a design decision was considerably changed when accounting for demographics. Furthermore, the predictive power of all models grew when the neighborhood demographics were added. We conclude that this finding calls for more systematic analysis of community and urban informatics projects that can be deployed in several geographical communities. This can lead to better isolate the impact of the technology while controlling for the communities' characteristics.
- Regarding the characteristics of the neighborhoods, our results show that a population's racial diversity and residential stability can affect the measures of digital inclusion in a civic platform. First, more racially-diverse neighborhoods are associated with broader online participation of people of color. Even though the size effect is small, this variable might be critical when assessing the success of particular design decisions that aim to foster racial digital inclusion (such as the case of offline outreach). On the other hand, we found evidence that neighborhoods with more residential instability (i.e. more people moving in during 2005 or later) are associated with less female participation in the forums. We believe that a better understanding of the connections between demographics and digital participation is fundamental to better design civic technologies that reach diverse audiences.
- We also found that both the forum's time of creation and tenure were significantly related to digital inclusion. Although these effects might be specific to the state of the technology and

technology adoption at particular periods of time, our results suggest that controlling for time measures in the analysis of civic technologies is necessary.

- As with any other study, this research has limitations that need to be taken into consideration when assessing the implications of the results. First of all, we did not have a ground truth about the gender and race of the E-Democracy users. We use alternative algorithms to reliably determine these attributes and we presented the results generated by all of them. Although we believe that our results regarding gender are likely to be representative of the user population in E-Democracy, our results may under-represent the population of people of color. Therefore, it is possible that some of our non-significant relationships could become significant if some of these limitations were solved. Another important limitation of our work is the fact that the design decisions were enacted for reasons beyond our research purposes. Therefore, the assignment was neither random nor balanced. When conducting the data analyses, we strove to control extremely unbalanced samples by each design decision and potential multi-collinearity among the independent variables. Therefore, we are confident that the results reported here are robust in terms of statistical associations. However, we cannot claim causality among the variables we have studied.

9.0 DISCUSSION

My dissertation work proposes an analysis framework to study participatory information systems for urban communities. It also reports on three studies that, guided by the framework, investigate 35 online forums for neighborhoods and districts in the cities of Minneapolis and St. Paul, Minnesota, US. Beyond the specific results of each of the three studies previously discussed (see sections 6.5, 7.3, and 8.4), we believe that taken together, the results provide key lessons to inform the design and study of local information systems. These lessons are as follows.

Local vs. hyper-local: Participatory information systems need to engage people as content contributors. However, engaging enough contributors is challenging. It is estimated that about 10% of social media users are active providers of content. Therefore, targeting very restricted user audiences, such as residents of a specific neighborhood, might seem to be an even bigger challenge. Nevertheless, our data analyses suggest that while population size of a neighborhood is important for constantly attracting new people to neighborhood forums, it is not critical for performance and retention. Besides, while city-wide forums attract more people, neighborhood forums engage more diverse audiences to online civic conversations. Proportionally, more females and people of color join and contribute content to neighborhood forums. The neighborhood forums are also associated with more satisfaction with regard to the goal of engaging with fellow residents and local initiatives in the community. Similar to paper-based bulletin boards (exemplars of offline participatory information systems) located in neighborhood public spaces [96], online forums for small geographically-defined communities can maintain active streams of content and be significant for the communities. In a range from 2,800 to 36,000 inhabitants, a very active E-Democracy neighborhood forum targets an area of about 11,000 people on average (std. dev. 7,210). Thus, we provide evidence here that focusing on hyper-local communities such as neighborhoods can create active-enough streams of content to positively impact community involvement and increase

inclusion of populations that are often underrepresented in online civic discourse.

Content that mobilizes the local community: As many information systems for local communities aim to increase community involvement, they have often been assessed regarding their impact on the residents' social capital. Based on self-reporting research methods, some projects have found evidence that participatory information systems positively influences social capital. To complement this evidence, we have taken another research approach. We analyzed the content shared in local information systems to find out whether it reveals a connection with social capital. Our approach has shown that more than 80% of the posts that initiate a new thread of conversation in a neighborhood forum can be categorized as mobilization requests. These represent user attempts to mobilize social connections and their resources that are available through the local information system. This proportion is significantly larger than the 4.4% of posts that were categorized as mobilization requests in a study of Facebook [49], where users draw resources from their ego-centric social networks. We interpret this high occurrence rate of mobilization requests as additional evidence that local online systems are closely related to accumulating and exercising social capital within the neighborhoods. Further research can consolidate the connection between different kinds of mobilization requests and self-reported assessments of social capital.

Beyond the high prevalence of mobilization requests, different kinds of mobilizations have divergent effects on sustainability. Considering the two kinds of mobilizations that we were able to distinguish using an automatic content classification, active mobilizations such as event announcements or requests to meet up account for a majority of the posts in the neighborhood forums. This evidence is aligned with our prior research that has found that local events made up for a considerable amount of content in other participatory information systems for local communities [94, 95, 96]. Nevertheless, active mobilizations were less likely than non-mobilizations to receive on-site response from others. At a collective level, larger proportions of these kind of mobilizations were positively associated with more posts, but negatively related to retention of users. This might indicate that certain users constantly use the forums to post active mobilizations (thus influencing performance over time), but these mobilizations do not create online interaction in the short term, which might be driving away other users from the forums. In turn, passive mobilizations positively affect attraction and retention, but have no role on performance at a collective level. We conclude that a good balance between active and passive mobilization requests is needed to maintain an ac-

tive stream of content in the forums without allowing the forums to become a one-way information dissemination tool. The more active forums had on average 47% active mobilizations and 32% of passive mobilizations (std. dev. of 10% for both). These proportions are indicators of a trade-off that works well for E-Democracy neighborhood forums. Finding that a balance between different kinds of requests is healthy for sustainability complements our prior results that content diversity in Facebook groups is associated with more viability of the groups at subsequent times [95].

Off-site communication matters: Local information systems exist in a context where there is a high chance that users communicate through other media. Residents of the same neighborhood can talk to each other on the streets, attend the same local events or meetings, receive information through intermediaries, or communicate by more personal media (e.g. emails, phones). Therefore, communication through a participatory information system is more likely to be interwoven with other kinds of communication that cannot be logged in the system archival data. In our studies, we found evidence that off-site communication about topics discussed in the neighborhood forums happen at a considerable scale (more than 40% of the survey respondents). Although our data cannot confirm such a thing, it is possible that some off-site communication is related to the almost two-thirds of unanswered new threads that exist in the neighborhood forums. Given that our manual content analysis indicates that almost all active mobilizations requested off-site responses instead of on-site ones, it is possible that some off-site communication is associated with the reduced probability of active mobilizations to get on-site responses. Future research might further clarify what drives the low levels of on-site responsiveness in these information systems.

Our data analysis indicated that off-site communication not only exists, it is influential for participatory information systems to achieve their expected impact. Being involved in off-site communication is significantly associated with higher satisfaction in regard to the forum's opportunities for information exchange and community involvement, even after controlling for the effect of on-site communication. Thus, off-site communication matters as part of the user experience. A side effect of this phenomenon is that there may be a considerable part of the users' interactions that is not be available in the archival data. This could affect, for example, the interpretation of some online behavioral measures, such as volume of content or responsiveness. A key implication is that off-site communication should not be overlooked when assessing the success of participatory information systems for local communities and their impact on said communities. Further

research might craft strategies to facilitate feedback mechanisms that allow the system designers or evaluators to have indicators of this off-site communication.

Design for digital inclusion, always: Digital inclusion continues to be a challenge in the context of civic participation. While E-Democracy has succeeded in attracting women to their platform, women still add significantly less content than men in the forums. Compared to city-wide forums, neighborhood forums are also associated with larger proportions of people of color as users and contributors. However, even the most optimistic estimation predicts that the proportion of new members who are people of color is less than 15%. Other measures are expected to be even lower. Confirming this pattern, less than 8% of the user survey respondents self-reported to be a person of color. Even though more racially diverse neighborhoods are more likely to have larger proportions of users who are people of color, these numbers are still far from the average ratio of people of color in the neighborhoods (42%).

Nevertheless, racial diversity is associated with participatory systems that are more sustainable in other aspects. More diverse neighborhoods are expected to have better retention and performance. Indeed, more active neighborhoods forums tend to have a more racially diverse composition of contributors (see Appendix C). The problem remains to be associated with attraction of a diverse user base. Our analysis has shown that design decisions such as narrowing the geographical scope of the information systems and appealing to community involvement help to diversify the voices in a civic platform. Our evidence regarding the positive effect of offline outreach on the attraction of more people of color to the system is not conclusive, but it is encouraging. It is also notable that the sample of survey respondents is considerably more homogeneous than the user base. Given that no special outreach was undertaken to get survey answers, we interpret this as more evidence that constant effort to achieve more racially diverse audiences is required in this context. Design decisions can make a difference, but it seems that the challenge is big enough to warrant multiple and consistent design strategies.

Address the challenge of residential instability: Residential instability is a solid threat to sustainability. It negatively affects attraction, retention and performance. It is also negatively associated with female participation in the neighborhood forums. Residential instability has been connected to the level of social capital in a neighborhood [74]. More stable populations tend to have more social capital, because there are more people that have shared the same geographical context

for longer and have had more time to develop social ties with fellow residents. Considering this line of reasoning, our results regarding residential instability can be evidence that prior levels of social capital in the neighborhood are necessary to make a local information system sustainable. This finding is aligned with related work on the importance of initial social capital in the community [77, 149] and individuals [45] for technology to be have a positive effect on social capital.

We also speculate that the reason behind this influential aspect might be connected to the forums' strong focus on community involvement. Survey data shows that community involvement appeals more to women. Possibly, it is also more appealing to those who have lived longer in, and are more attached to, the neighborhoods. However, not everyone seems to be motivated by the same goal. Therefore, designers of participatory information systems for local communities need to develop new mechanisms to engage newer residents and transient populations that do not plan to stay long in the neighborhoods. Being able to satisfy goals that different residents can engage with has been found to be a key aspect of larger volunteer associations in urban communities [5]. In prior work, using adaptive approaches to engage users with different roles, personality traits, and online experience is useful for attraction and performance in other audience-bounded participatory information systems [98, 99, 111]. Further research can explore the adaptation to different users' goals that can be achieved in local information systems in order to engage populations that might not be interested in community involvement.

Keep a mix of core group and newcomers: Several of our results indicate that the most sustainable forums have a mix of experienced contributors and newcomers. Collectively, diversity of tenure leads to better attraction of people to the forums and better performance. At the individual level, experienced users are more likely to contribute again in the future, less likely receive on-site responses, and much less prone to migrate to other E-Democracy forums. Beyond and above this effect, those users who have been involved in on-site direct communication with more people (high degree centrality) are also significantly more likely to remain active in the forums and less likely to migrate. These results indicate that the neighborhood forums tend to have an experienced and central core of users that generates content constantly, even though at least some of them might not encourage as much online conversations as other users in the platform. The core users might be strongly influenced by the 100 people that are needed to start an E-Democracy forum.[1] In turn,

[1] http://blog.e-democracy.org/posts/280 Last retrieved on November 7th, 2015.

newcomers, along with more central users, in the forums are significantly more likely to receive an online response and therefore generate a more interactive conversation in the forums.

Thus, we conclude that both core central and peripheral users are needed for local information systems to generate different kinds of content to make them relevant to a diverse audience. Not all effort needs to be focused on sustaining participation of current contributors. It is also important to constantly engage newcomers that bring a more conversational style to the local discourse.

Diverse effects of membership overlap: Local information systems can focus on different geographical areas. They can target neighborhoods, cities or other small geographical areas. Residents might want to participate in more than a single system, such as adjacent neighborhoods or a city along with neighborhood forums. Using the both archival and survey data, we found evidence that about a third of the users in E-Democracy participate in more than a local forum. The existence of this membership overlap has divergent influences on sustainability. Collectively, membership overlap increases attraction: more people feel that they can be part of the forums. However, membership overlap negatively influences retention and performance. At the individual level, these results are confirmed: people with membership overlap are less likely to receive a response and more likely to migrate. It seems that people who divide their attention between two or more local forums are less likely to be part of the core central users of the forum. Thus, membership overlap increases the audience of a focal forum and encourages mobility across forums in the platform, but it does not contribute to the other aspects of the focal forum's sustainability in the long term.

Different than global outreach: A concern that motivated this research was to understand what makes the participatory information systems for urban communities different from participatory information systems with global reach. Beyond the role of the demographic context and its critical influence on sustainability, the impact of collective and individual online features on sustainability can be compared to the findings of larger-scope participatory information systems. In comparison, E-Democracy neighborhood forums have a smaller volume of content, fewer contributors, and larger retention of contributors. These local systems are also especially more responsive to newcomers than systems with global reach.

These differences can be associated with the special conditions of sharing a limited offline context. While the narrow scope of local information systems does not seem to affect their ability to have an impact on the users, our data complements prior evidence that local communities do

not generate much new information daily [25]. We argue that this should not be interpreted as a failure, but as part of the context that system designers need to deal with. High levels of retention can be associated with the finding that some users are very likely to contribute again, even after controlling for their level of interaction with others. These users might be residents with very strong commitments for community involvement, or people who work in the neighborhoods and use the forums as part of their roles. These people can become the core group users that partially ensures the sustainability of a local information system. E-Democracy's policy of starting forums with at least 100 people might be a way to ensure that this group exists in every new forum. The fact that newcomers are more likely to receive responses could be also explained by the core group of users. Core users can feel particularly motivated to help new people in the forums. It is also possible that newcomers are recognized by other users who know them offline, and these social ties make it more likely that responses will be provided. We believe that it is important for the designers of these systems to capitalize on the kind of peculiarities that an offline shared context provides in order to better tackle the challenges of sustainability.

Context matters, a lot!: Finally, our last implication is that context matters considerably when assessing the use and impact of participatory information systems for urban communities. We found that the year in which a local forum was created is a significant factor on almost all sustainability measures. Generally, more recently created forums were associated with lower measures of sustainability. We speculate that this effect is related to the larger amount of social media sites that have been available more recently. Newer forums had to compete with more sites for the residents' attention. Tenure is often significant as well. As time goes by, some aspects of sustainability become better and others worse. Not only are demographics of the target neighborhoods significant, but the level of full Internet access in the local community is also an influential factor on the contribution levels of a local forum. Our lesson from these findings is that the study and design of participatory information systems should take into account the context in which the information system is being deployed.

My dissertation work has limitations proper of any research endeavor. We have used the archival data of a sample of forums in a specific state in the US, which might not be representative of information forums for urban communities in different cultural and social contexts. While we strove to control for neighborhood demographics, we could not control for other neighborhood

variables that could have affected the results, such as the average level of civic engagement in the neighborhood. Additionally, we were not able to control for the number of contributors in our analysis of collective aspects due to multicollinearity among the independent variables. Our social network measures were based on data about who responded to whom and not in an underlying social structure, as in the studies on Facebook and Twitter. This might explain why we were not able to replicate results from prior work in regard to social network measures. Besides, users who answered the survey were self-selected and not randomly chosen: therefore, the collected data might be biased towards the perspectives of people who wanted to give their opinion about the local forums. Additionally, in our last study, we used public data to be able to automatically determine the gender and race of the E-Democracy users. The ability of our algorithm to determine race and gender could be biased as well. We have reasoned that our approach might have underestimated the number of females and people of color that participated in the forums, thus underestimating the effect of the design decision under study. Nevertheless, we strove for triangulating data across the results of our different research methods whenever possible in order to achieve more confidence on our interpretation of the results we report here.

10.0 CONCLUSIONS

My dissertation work has focused on conducting three empirical studies on a sample of 35 long-tenure online forums for neighborhoods and districts in the US. These studies aim to explore the different factors that can influence sustainability according to our proposed framework. The results provide evidence that all of these factors affect at least one aspect of sustainability of participatory information systems for urban communities. Thus, these studies provide evidence that the proposed framework is useful to analyze the sustainability of local information systems. Overall, the contributions of my dissertation work cntributes to research on social computing and community/urban informatics by:

1. Providing a conceptual framework to investigate both online and offline aspects of urban communities and their residents that can affect sustainability of their participatory information systems;
2. Reporting on a longitudinal empirical analysis that uses the proposed framework to study the sustainability of long-tenure online forums across multiple urban communities in the US;
3. Conducting a mixed-method approach to explore, from different perspectives, the various factors that can affect the sustainability of participatory information systems for urban communities;
4. Compiling a set of evidence-based design guidelines that capitalize on lessons learned in order to better tackle the challenge of these systems' sustainability.

APPENDIX A

DETAILS ABOUT THE AUTOMATIC CONTENT CLASSIFICATION

A.1 IMPROVING PERFORMANCE OF THE AUTOMATIC CLASSIFIERS

We tested different mechanisms that could improve the classification performance. We explored the use of different features in the classification. The features include unigrams, bigrams, trigrams and linguistic features. We considered using these features in isolation and in conjunction. We also grouped some of the kinds of mobilization into broader categories to have more balanced samples of each label in the classification. We coded additional posts to increase the sample size of one of the categories (non-mobilizations). To reduce dimensionality and computation time, we conducted principal component analysis on the N-grams and linguistic features data. We also tested different thresholds to filter out very common and uncommon N-grams and to keep the most important components from the principal component analyses. We assessed the effectiveness of using N-grams in isolation and in conjunction with the count of linguistic features. We also tried downsizing the content categories that had more observations. We ran different classifiers to distinguish between two content labels and then used a voting process to choose the most common assigned label to a post. We used the tuning methods with 10-fold cross-validation in order to find the most appropriate hyper-parameters for running SVM on our data. To compare results, we use accuracy, precision and recall measures. The best results were obtained with 19 components that explain 95% of variance of the linguistic features. Adding the main components of the unigrams and trigrams harmed the performance. The main components from the bigrams performed almost as well as the linguistic features alone. Bigrams generally helped to improve the classification of

non-mobilizations, but made the classification of passive mobilizations slightly worse. Therefore, we decided to choose the classifiers that use the linguistic components only.

A.2 FEATURES AND MOBILIZATION REQUESTS

Given that the linguistic features are the only input of our classifiers, here, we present a description of these features according to our ground truth (see Figure 13) as a way to better understand the linguistic differences among the different kinds of content in the neighborhood forums. On average, a *non-mobilization* had more sentences and words than all other kinds of mobilizations. The mean number of sentences was ten. The average *non-mobilization* had slightly more than 200 words, out of which more than 150 words were recognized as a linguistic feature by our algorithm. *Recommendations* and *factual knowledge* tended to have the shortest text among the posts.

Among the words that were recognized as linguistic features, there were differences in the prevalence of particular kinds of words among the kinds of content. Compared to other kinds of content, *social coordination/invitation/offer* had smaller proportions of function words, which includes pronouns, articles, verbs and adverbs, among others. The six different kinds of content did not vary considerably in terms of the number of words that represent psychological processes. They had between 10% and 15% of words that characterize social processes related to family, friends and humans. A smaller proportion of words (around 6%) were related to affective process such as positive and negative emotions. The proportion of words that reflect cognitive processes ranged from 17% to 20% of the words. Perceptual processes accounted for a much more reduced share of words (about 2%). Perhaps the most distinguishable pattern is that of the requests for factual knowledge that included smaller proportions of words related to social and affective processes and larger ratios of cognitive and perception processes, as opposed to other kinds of content.

These major word categories have further classifications (sub-categories) that provide more details about the linguistic characteristics of the text in the posts. A closer look at some of these sub-categories is shown in the middle part of Figure 13. We show here the aspects that vary to a larger extent among all the features we could consider. There were larger proportions of positive than negative emotions across all kinds of content. *Recommendations* and *factual knowledge*

Label	observations	sentcnt	wc	dic_wc	funct	social	affect	cogmech	percept
Non mobilizations	119	10.50	206.70	159.37	0.61	0.12	0.05	0.17	0.02
Recommendation	35	6.97	97.23	76.03	0.61	0.14	0.07	0.20	0.02
Factual knowledge	24	6.00	97.25	74.67	0.63	0.10	0.04	0.21	0.03
Opinion/poll	28	8.14	142.14	117.86	0.63	0.15	0.05	0.20	0.02
Favor/request/collective action	103	9.45	177.85	135.20	0.60	0.14	0.06	0.19	0.02
Social coordination/invitation/offer	240	8.75	170.60	122.94	0.53	0.14	0.06	0.17	0.02

Label	posemo	negemo	tentat	p1	p2	p3	past	present	future
Non mobilizations	0.04	0.01	0.03	0.05	0.01	0.01	0.04	0.07	0.01
Recommendation	0.06	0.01	0.06	0.06	0.02	0.01	0.02	0.12	0.01
Factual knowledge	0.03	0.01	0.06	0.06	0.00	0.02	0.04	0.10	0.01
Opinion/poll	0.04	0.01	0.03	0.06	0.01	0.02	0.03	0.09	0.01
Favor/request/collective action	0.04	0.01	0.04	0.04	0.02	0.02	0.03	0.09	0.01
Social coordination/invitation/offer	0.05	0.01	0.03	0.03	0.02	0.01	0.01	0.07	0.02

Label	motion	space	time	work	achieve	leisure	home	money	death
Non mobilizations	0.03	0.10	0.07	0.04	0.02	0.01	0.02	0.01	0.00
Recommendation	0.02	0.09	0.05	0.04	0.02	0.01	0.02	0.01	0.00
Factual knowledge	0.04	0.12	0.07	0.02	0.01	0.02	0.02	0.01	0.00
Opinion/poll	0.02	0.09	0.05	0.03	0.03	0.02	0.02	0.01	0.00
Favor/request/collective action	0.02	0.09	0.06	0.04	0.02	0.02	0.03	0.01	0.00
Social coordination/invitation/offer	0.03	0.11	0.08	0.06	0.03	0.03	0.02	0.01	0.00

Figure 13: Linguistic features by the six kinds of mobilizations

tended to include more tentative words such as maybe, perhaps and guess. They had larger proportions of verbs in present tense. These kinds of content and *opinion/poll* used more first person pronouns. Meanwhile, *social coordination/invitation/offer* had a slightly larger ratio of verbs in future tense and smaller percentage of verbs in past tense than any other kind of content. Other features were related to motion, space, time and personal concerns such as work and leisure (see Table 13). Among these aspects, space, time and work were the features with more prevalence across all kinds of content in the neighborhood forums. The posts that were classified as *Social coordination/invitation/offer* had narrowly larger ratios of time, work and leisure words compared to other kinds of content.

Once we collapsed all infrequent kinds of mobilizations into a single kind of content called *passive mobilizations*, the majority of the proportion trends of these features were preserved. One exception was that requests for *factual knowledge* had a more distinctive pattern of features than other passive mobilizations, and this distinction was lost (averaged up/down) when we included *factual knowledge* in a larger category. This is a limitation of our automatic classification.

APPENDIX B

2014 E-DEMOCRACY USER QUESTIONNAIRE

The following pages show the questions included in the 2014 E-Democracy user questionnaire.

Introduction

As a member of an E-Democracy Neighbors Forum, thank you for helping improve your forum and community by completing this short survey.

We promise to keep your identity absolutely private.

To help you get started:

• Questions in this survey are about your experience with your primary online "Neighbors Forum" -- the forum that you use most (where you live or work).

• “Neighbors,” means people who live or work in the area covered by your primary online Neighbors Forum - not just nearest neighbors.

• Your “community” or “neighborhood,” means the place that corresponds to your primary online Neighbors Forum area.

1. Please enter the primary email address you use on E-Democracy. (Optional)

*2. Where is your primary online Neighbors Forum hosted on E-Democracy?

- Saint Paul
- Minneapolis
- Other

Saint Paul Forums

*3. What is your primary Saint Paul forum?

- Como Neighbors
- Dayton's Bluff Neighbors Forum
- District 1 Neighbors (Eastview-Conway-Battle Creek-Highwood Hills)
- Capitol River / Downtown Community Forum
- Frogtown Neighbors Forum
- Greater East Side Neighbors (D2)
- Hamline Midway Neighbors Forum
- Highland Park Neighbors Forum
- Macalester Groveland Neighbors
- North End Neighbors Forum (North End-South Como)
- Payne Phalen (D5) Neighbors Forum
- Summit Hill Neighbors Forum
- Summit University Rondo Neighbors Forum
- Union Park Neighbors Forum
- West Seventh Fort Road Neighbors
- West Side Neighbors
- Saint Paul Citywide Issues Forum

Other (please specify)

Minneapolis Forums

*4. What is your primary Minneapolis forum?

- Audubon Park Neighbors
- Bryant Neighbors
- Cedar Riverside Neighbors Forum
- Central Neighbors Forum
- Cleveland Neighbors
- Corcoran Neighbors Forum
- East Harriet Neighbors Forum
- Field Regina Northrop Neighbors Forum
- Hale Page Diamond Lake Neighbors
- Holland Neighbors Forum
- Kingfield Neighbors Forum
- Linden Hills & Fulton Neighbors Forum
- Longfellow Community Neighbors Forum
- Loring Park Neighbors
- Near North and Heritage Park Neighbors Forum
- Nokomis East Neighbors Forum
- Northeast Neighbors Forum
- Phillips Community Forum
- Powderhorn Neighbors Forum
- Seward Neighbors Forum
- Standish Ericsson Neighbors Forum
- Uptown Neighbors Forum
- Whittier Neighbors Forum
- Minneapolis Citywide Issues Forum

Other (please specify)

Other Forums/Communities of Practice

*5. Other Forums/Communities of Practice

- Framingham Neighbors-FramBors
- Framingham Nobscot Neighbors Forum
- Framingham Government Forum
- Roseville Community Forum
- Minnesota Politics and Issues Forum
- United States Issues Forum
- UK-based forums
- NZ-based forums
- ComGar - Community Gardeners in Minnesota
- Minneapolis Youth Workers
- Block Connectors
- CityCamp Exchange
- Community Builders
- Crime Prevention and Social Media
- Democracies Online Newswire
- Digital Inclusion Network
- E-Democracy Exchange
- E-Democracy Projects
- LocalLabs
- Locals Online
- Open Government and Civic Technology (Facebook)
- Open Twin Cities Google Group

Other (please specify)

*6. How did you first learn about your primary forum?

- Online
- Not online (in person, print, etc.)
- Don't recall

7. Please tell us more about where you learned about your forum.

- [ ] Email message or invite – from someone
- [ ] Email newsletter
- [ ] Facebook
- [ ] Twitter
- [ ] Web search
- [ ] Web link from another site
- [ ] Online ad
- [ ] Online news story
- [ ] Don't recall
- [ ] Other (please specify)

8. Please tell us more about where you learned about your forum.

- [ ] From someone who came to my door
- [ ] At a community gathering, festival, or meeting
- [ ] From someone I know (via word of mouth, be it in person or telephone)
- [ ] Print flyer, poster, or sign
- [ ] In a community organization newsletter or publication
- [ ] Local media (radio, TV, print newspaper)
- [ ] Don't recall

Other (please specify)

Forum Value and Impact

9. How important to you are the following things you can do on your Neighbors Forum?

	Very important	Somewhat important	Not important
Get community news and local event announcements	○	○	○
Learn about local businesses, resources and services	○	○	○
Discuss or understand others' views on community issues and happenings	○	○	○
Share information or ideas	○	○	○
Get involved in local initiatives or causes	○	○	○
Meet neighbors and other community members	○	○	○
Help neighbors in need (sharing, lost pets, etc.)	○	○	○

Other (please describe)

Forum Value and Impact

10. To what extent is your forum meeting your needs? How *satisfied* are you with the opportunity that your forum has provided in the last 12 months to...

	Very satisfied	Satisfied	Neutral	Dissatisfied	Very dissatisfied
Get community news and local event announcements	○	○	○	○	○
Learn about local businesses, services, and resources	○	○	○	○	○
Discuss or understand others' views on community issues and happenings	○	○	○	○	○
Share information or ideas	○	○	○	○	○
Get involved in local initiatives or causes	○	○	○	○	○
Meet neighbors and other community members	○	○	○	○	○
Help neighbors in need (sharing, lost pets, etc.)	○	○	○	○	○

Other (please describe)

Forum Value and Impact

11. As a result of information or discussions on your Neighbors Forum, in the last 12 months...

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
I feel that my participation is welcomed or valued by others on the forum	○	○	○	○	○
I have been introduced to new ideas or points of view	○	○	○	○	○
I have learned more about my neighbors of races, ethnicities, or home languages different from my own	○	○	○	○	○
I am more informed about issues that affect my community	○	○	○	○	○
I have learned more about how to influence decisions in my community	○	○	○	○	○
I am more satisfied with my local community as a place to live or work	○	○	○	○	○

Actions and Ideas

12. In the last 12 months, did something on your Neighbors Forum lead you to do or increase any of the following? (Please take the time to review column choices carefully. Very important question.)

	Yes, I did this AND it increased because of the Forum	I did this, but -did not increase- this activity because of the Forum	No, I didn't do this / not applicable
Do favors for or share goods (e.g., lend tools, give away items) with neighbors or local community members	○	○	○
Perform local volunteer work	○	○	○
Attend a community meeting in which local issues were discussed	○	○	○
Contact an elected official or government office	○	○	○
Work with other residents to make change in the local community	○	○	○
Attend community events such as a festival, picnic, or parade	○	○	○
Donate money to a local charity or cause	○	○	○
Sign a petition	○	○	○
Meet other community members in person	○	○	○
Visit a business, restaurant, or hire someone recommended on the forum	○	○	○

Other (please specify)

Actions and Ideas

Tip: You can adjust the size of the comment boxes by selecting the bottom right corner and dragging it down and to the right.

13. Please share a story about when your Neighbors Forum made a difference to you or your community.

14. In what ways could your Neighbors Forum be improved?

15. Additional comments or input you'd like to share:

Forum Use

16. On your primary Neighbors Forum are you a ...

- ○ Regular email subscriber - read and post via email
- ○ Daily digest subscriber - read and post via web
- ○ Web-only member
- ○ Web visitor (no E-Democracy account)

17. Check all that apply ...

- ☐ I posted a personal introduction to my primary Neighbors Forum at some point
- ☐ I posted at least once to my forum in the last year
- ☐ I received on-forum replies to my forum post(s)
- ☐ I received private email replies to my forum post(s)
- ☐ I received a response or acknowledgement of my forum posts in person or via the telephone
- ☐ I have never posted publicly on my Neighbors Forum, but I have replied privately to posts
- ☐ I have never posted publicly on my Neighbors Forum and have not communicated privately from its content

18. What are your roles in the local community of your Neighbors Forum? (Please check all that apply.)

- [ ] Resident
- [ ] Parent or guardian of children 18 or under living at home
- [ ] Student
- [ ] Local business owner
- [ ] Employee of a business in the forum area
- [ ] Employee or volunteer with a nonprofit, community/cultural organization, or place of worship serving the forum area
- [ ] Employee of local government or school
- [ ] Elected official

Other (please describe)

19. Check all the online services you are a member of:

- [ ] Citywide Issues Forum (e.g., St. Paul Issues, Minneapolis Issues on E-Democracy)
- [ ] Facebook
- [ ] Twitter
- [ ] LinkedIn
- [ ] NextDoor.com
- [ ] A small private online group (Facebook Group or email list/cc: group) for your block or nearest neighbors
- [ ] A public or large online group (Facebook Group or email list) for your larger neighborhood or part of town

Comments - As it relates to community engagement, what do you value about these services and why? Any comparisons with your Neighbors Forum you would like to share positive or negative?:

20. I would like more information on: (if checking these options, be sure to leave your best email at the very end)

- [ ] Volunteering with my local Neighbors Forum
- [ ] Volunteering with E-Democracy overall
- [ ] Donating to support forum outreach and online hosting
- [ ] Getting involved with efforts to promote more open government and the use of technology to benefit the community

Other (please specify)

A little about you...

We promise to keep your identity absolutely private. This information is crucial to evaluating our inclusion efforts.

21. What is your age:

- Under 18
- 18 to 24 years
- 25 to 34 years
- 35 to 44 years
- 45 to 54 years
- 55 to 64 years
- 65 to 74 years
- 75 and over

22. Your gender:

- Male
- Female
- Other identification

*23. Were you or your parents born outside the United States?

- Yes
- No

24. I was born in ...

- The United States

Another country (please specify):

25. My parents were born in:

- The United States

Another country (please specify):

26. Your race/ethnicity (please check all that apply):

- [ ] White or European American
- [ ] Black, African American, or African immigrant
- [ ] Asian
- [ ] American Indian or Alaska Native
- [ ] Native Hawaiian or Other Pacific Islander
- [ ] Hispanic / Latino
- [ ] Other (please specify)

27. What is the highest level of education you have completed?

- ○ Less than a high school diploma
- ○ High school diploma or GED
- ○ Some college
- ○ Two-year college degree
- ○ Bachelor's degree
- ○ Graduate or professional degree

28. Do you own or rent your residence?

- ○ Own
- ○ Rent
- ○ Other (e.g., rent a room, live as a lodger, etc.)

29. What is your total household income? (Please check one.)

- ○ Less than $10,000
- ○ $10,000 to $29,999
- ○ $30,000 to $59,999
- ○ $60,000 to $99,999
- ○ $100,000 or more
- ○ Prefer not to say

Thank You!

Thank you for taking the time to complete this survey!

• Special thanks to the Knight Foundation for supporting this survey, the iPad mini prize (entries closed) and our inclusive BeNeighbors.org project from 2012-2014. See our blog for more details: http://blog.e-democracy.org

* Like us on Facebook! *

As you complete this survey, please Like us on Facebook. You will be taken to that page when you are done.

30. Best email address to contact you if you requested more information, offered to volunteer, donate, etc.

APPENDIX C

MORE ACTIVE FORUMS AND RACIAL DIVERSITY

Table 73: Participation of people of color: New members by forums' activity level

Sustainability	pseudo-random	t = 0.50	t = 0.75	t = 0.90
Less sustainable	5.577	5.586	5.976	5.113
	14.769	13.715	14.639	11.771
Active	4.911	3.578	3.008	2.558
	9.118	7.125	6.531	5.787

Table 74: Participation of people of color: Contributors by forums' activity level

Sustainability	pseudo-random	t = 0.50	t = 0.75	t = 0.90
Less sustainable	2.987	2.414	2.051	1.638
	2.777	1.924	1.449	1.098
Active	6.049	3.800	3.287	2.814
	5.749	4.101	3.503	2.485

Table 75: Participation of people of color: Posts by forums' activity level

Sustainability	pseudo-random	t = 0.50	t = 0.75	t = 0.90
Less sustainable	7.400	5.346	4.013	3.275
	10.085	6.474	4.533	3.674
Active	16.676	10.166	9.697	8.651
	19.771	15.984	14.932	10.565

BIBLIOGRAPHY

[1] P. S. Adler and S.-w. Kwon. Social Capital: Prospects for a New Concept. *The Academy of Management Review*, 27(1):17 – 40, 2002.

[2] M. Ananny and C. Strohecker. Textales: Creating interactive forums with urban publics. *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City, IGI Global, Hershey, PA*, 2009.

[3] M. Anderson. Americans internet access: 2000-2015. year = 2015. Technical report, Pew Research Center.

[4] J. Arguello, B. S. Butler, E. Joyce, R. Kraut, K. S. Ling, C. Rosé, and X. Wang. Talk to me: Foundations for successful individual-group interactions in online communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 959–968, New York, NY, USA, 2006. ACM.

[5] N. Babchuk and A. Booth. Voluntary association membership: A longitudinal analysis. *American sociological review*, pages 31–45, 1969.

[6] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. 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*, KDD '06, pages 44–54, New York, NY, USA, 2006. ACM.

[7] L. Backstrom, R. Kumar, C. Marlow, J. Novak, and A. Tomkins. Preferential behavior in online groups. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, WSDM '08, pages 117–128, New York, NY, USA, 2008. ACM.

[8] P. Balkundi and D. A. Harrison. Ties, leaders, and time in teams: Strong inference about network structure's effects on team viability and performance. *Academy of Management Journal*, 49(1):49–68, 2006.

[9] D. J. Beal, R. R. Cohen, M. J. Burke, and C. L. McLendon. Cohesion and performance in groups: a meta-analytic clarification of construct relations. *Journal of applied psychology*, 88(6):989, 2003.

[10] A. Beamish. Communities on-line: Community-based computer networks. Master's thesis, Department of Urban Studies and Planning. Massachusetts Institute of Technology, 1995.

[11] G. Beenen, K. Ling, X. Wang, K. Chang, D. Frankowski, P. Resnick, and R. E. Kraut. Using social psychology to motivate contributions to online communities. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, CSCW '04, pages 212–221, New York, NY, USA, 2004. ACM.

[12] W. Beyea, C. Geith, and C. McKeown. Place making through participatory planning. *Marcus Foth (Hg.): Urban Informatics. The Practice and Promise of the Real-Time City. Hershey PA: IGI Global*, pages 55–67, 2009.

[13] M. J. Brzozowski, T. Sandholm, and T. Hogg. Effects of feedback and peer pressure on contributions to enterprise social media. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work*, GROUP '09, pages 61–70, New York, NY, USA, 2009. ACM.

[14] M. Burke and R. Kraut. Social capital on facebook: Differentiating uses and users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 571–580, 2011.

[15] M. Burke, C. Marlow, and T. Lento. Feed me: Motivating newcomer contribution in social network sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 945–954, New York, NY, USA, 2009. ACM.

[16] R. S. Burt. The social capital of opinion leaders. *The Annals of the American Academy of Political and Social Science*, 566(1):37–54, 1999.

[17] R. S. Burt. The network structure of social capital. *Research in organizational behavior*, 22:345–423, 2000.

[18] R. S. Burt. Structural Holes versus Network Closure as Social Capital. In *Social Capital: Theory and Research*. Transaction Publishers, 2001.

[19] B. Butler, L. Sproull, S. Kiesler, and R. Kraut. Community effort in online groups: Who does the work and why. In *Leadership at a Distance: Research in Technologically-Supported Work.*, pages 171–194. Lawrence Erlbaum Associates/Taylor Francis, 2007.

[20] B. S. Butler. *The dynamics of cyberspace: Examining and modeling online social structure*. PhD thesis, Carnegie Mellon University, 1999. Chapter: When is a group not a group: An empirical examination of metaphors for online social structure.

[21] B. S. Butler. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Info. Sys. Research*, 12(4):346–362, Dec. 2001.

[22] B. S. Butler, P. J. Bateman, P. H. Gray, and E. I. Diamant. An attraction–selection–attrition theory of online community size and resilience. *MIS Quarterly*, 2015, To Appear.

[23] A. Button and H. Partridge. Community networks today: Analysing new media for local social networking and community engagement. In *Communities and Action: Prato CIRN Conference*, pages 5–7, 2007.

[24] G. Capece and R. Costa. The new neighbourhood in the internet era: network communities serving local communities. *Behaviour & Information Technology*, 32(5):438–448, 2013.

[25] J. M. Carroll. *The Neighborhood in the Internet*, chapter What are community networks?, pages 23–47. Routledge, 2012.

[26] J. M. Carroll and M. B. Rosson. Developing the Blacksburg electronic village. *Commun. ACM*, 39(12):69–74, Dec. 1996.

[27] J. M. Carroll and M. B. Rosson. Better home shopping or new democracy?: evaluating community network outcomes. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 372–379. ACM, 2001.

[28] J. M. Carroll and M. B. Rosson. Wild at home: The neighborhood as a living laboratory for hci. *ACM Trans. Comput.-Hum. Interact.*, 20(3):16:1–16:28, July 2013.

[29] M. Castells. The rise of the network society, volume i of the information age: Economy, society and culture. *Massachusetts: Blackwell Publishing*, 1996.

[30] J. Chen, Y. Ren, and J. Riedl. The effects of diversity on group productivity and member withdrawal in online volunteer groups. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 821–830, 2010.

[31] Y. Chen and J. Xie. Online consumer review: word-of-mouth as a new element of marketing communication mix. *Management science*, 54(3):477–491, 2008.

[32] C.-M. Chiu, M.-H. Hsu, and E. T. Wang. Understanding knowledge sharing in virtual communities: an integration of social capital and social cognitive theories. *Decision support systems*, 42(3):1872–1888, 2006.

[33] B. Choi, K. Alexander, R. E. Kraut, and J. M. Levine. Socialization tactics in wikipedia and their effects. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, CSCW '10, pages 107–116, New York, NY, USA, 2010. ACM.

[34] J. H.-j. Choi and A. Greenfield. To connect and flow in seoul: Ubiquitous technologies, urban infrastructure and everyday life in the contemporary korean city. *Handbook of research on urban informatics: The practice and promise of the real-time city*, pages 21–36, 2009.

[35] S. Clift. What is a local community forum? http://forums.e-democracy.org/about/localforums/, (accessed October 28, 2014).

[36] J. S. Coleman. Social capital in the creation of human capital. *American Journal of Sociology*, 94:95–120, 1988.

[37] K. Colstad and E. Lipkin. Community memory: a public information network. *SIGCAS Comput. Soc.*, 6(4):6–7, Dec. 1975.

[38] D. Cosley, D. Frankowski, L. Terveen, and J. Riedl. Suggestbot: using intelligent task routing to help people find work in wikipedia. In *Proceedings of the 12th international conference on Intelligent user interfaces*, IUI '07, pages 32–41, New York, NY, USA, 2007. ACM.

[39] L. Dahlberg. Extending the public sphere through cyberspace: The case of minnesota e-democracy. *First Monday*, 6(3), 2001.

[40] L. Dahlberg. The internet and democratic discourse: Exploring the prospects of online deliberative forums extending the public sphere. *Information, Communication & Society*, 4(4):615–633, 2001.

[41] F. De Cindio and C. Peraboni. Building digital participation hives: Toward a local public sphere. *From Social Butterfly to Engaged Citizen: Urban Informatics, Social Media, Ubiquitous Computing, and Mobile Technology to Support Citizen Engagement*, page 93, 2011.

[42] A. de Moor. Moving community informatics research forward. *The Journal of Community Informatics*, 5(1), 2009.

[43] M. de Waal. The ideas and ideals in urban media theory and design. In *From Social Butterfly to Engaged Citizen*, pages 1–14. MIT Press, 2011.

[44] L. A. DeChurch and J. R. Mesmer-Magnus. The cognitive underpinnings of effective teamwork: a meta-analysis. *Journal of Applied Psychology*, 95(1):32, 2010.

[45] T. R. Dillahunt. Fostering social capital in economically distressed communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 531–540, New York, NY, USA, 2014. ACM.

[46] S. Drenner, S. Sen, and L. Terveen. Crafting the initial user experience to achieve community goals. In *Proceedings of the 2008 ACM conference on Recommender systems*, RecSys '08, pages 187–194, New York, NY, USA, 2008. ACM.

[47] M. N. Elliott, P. A. Morrison, A. Fremont, D. F. McCaffrey, P. Pantoja, and N. Lurie. Using the census bureaus surname list to improve estimates of race/ethnicity and associated disparities. *Health Services and Outcomes Research Methodology*, 9(2):69–83, 2009.

[48] N. B. Ellison, R. Gray, J. Vitak, C. Lampe, and A. Fiore. Cultivating social resources on social network sites: Facebook relationship maintenance behaviors and their role in social capital processes. *Journal of Computer-Mediated Communication*, In press.

[49] N. B. Ellison, R. Gray, J. Vitak, C. Lampe, and A. T. Fiore. Calling all facebook friends: Exploring requests for help on facebook. *Proc. of ICWSM 2013*, pages 155–164, 2013.

[50] R. Farzan, R. Kraut, A. Pal, and J. Konstan. Socializing volunteers in an online community: A field experiment. In *Proceedings of the 2012 ACM conference on Computer supported cooperative work*, page To appear, 2012.

[51] M. Foth. *Handbook of research on urban informatics: The practice and promise of the real-time city*. Information Science Reference, IGI Global, 2009.

[52] M. Foth, J. H.-j. Choi, and C. Satchell. Urban informatics. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, CSCW '11, pages 1–8, New York, NY, USA, 2011. ACM.

[53] M. Foth, L. Forlano, and C. Satchell. *From social butterfly to engaged citizen: urban informatics, social media, ubiquitous computing, and mobile technology to support citizen engagement*. The MIT Press, 2011.

[54] S. Gad, N. Ramakrishnan, K. N. Hampton, and A. Kavanaugh. Bridging the divide in democratic engagement: Studying conversation patterns in advantaged and disadvantaged communities. In *Social Informatics (SocialInformatics), 2012 International Conference on*, pages 165–176. IEEE, 2012.

[55] P. Gerbaudo. *Tweets and the streets: Social media and contemporary activism*. Pluto Press, 2012.

[56] R. Gibson, M. Cantijoch, and S. Galandini. The third sector and online citizen empowerment: the case of mysociety. Technical report, mySociety.

[57] H. Gil de Zúñiga, N. Jung, and S. Valenzuela. Social media use for news and individuals' social capital, civic engagement and political participation. *Journal of Computer-Mediated Communication*, 17(3):319–336, 2012.

[58] M. Granovetter. The strength of weak ties: A network theory revisited. *Sociological Theory*, 1(1983):201–233, 2012.

[59] K. Hampton and B. Wellman. Neighboring in Netville: How the internet supports community and social capital in a wired suburb. *City & Community*, 2(4):277–311, Dec. 2003.

[60] K. N. Hampton. Place-based and it mediated "community". *Planning Theory & Practice*, 3(2):228–231, 2002.

[61] K. N. Hampton. Neighborhoods in the Network Society the e-Neighbors study. *Information, Communication & Society*, 10(5):714–748, 2007.

[62] K. N. Hampton, C.-j. Lee, and E. J. Her. How new media affords network diversity: Direct and mediated access to social capital through participation in local social settings. *New Media & Society*, 13(7):1031–1049, Feb. 2011.

[63] E. Hargittai and A. Shaw. Mind the skills gap: the role of internet know-how and gender in differentiated contributions to wikipedia. *Information, Communication & Society*, 18(4):424–442, 2015.

[64] F. M. Harper, D. Frankowski, S. Drenner, Y. Ren, S. Kiesler, L. Terveen, R. Kraut, and J. Riedl. Talk amongst yourselves: inviting users to participate in online conversations. In *Proceedings of the 12th international conference on Intelligent user interfaces*, IUI '07, pages 62–71, New York, NY, USA, 2007. ACM.

[65] C. P. Hoffmann, C. Lutz, and M. Meckel. Content creation on the internet: A social cognitive perspective on the participation divide. *Information, Communication & Society*, 18(6):696–716, 2015.

[66] M. A. Hogg. Group cohesiveness: A critical review and some new directions. *European review of social psychology*, 4(1):85–111, 1993.

[67] Y. Hu, S. D. Farnham, and A. Monroy-Hernández. Whoo.ly: Facilitating information seeking for hyperlocal communities using social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 3481–3490, New York, NY, USA, 2013. ACM.

[68] A. L. Hughes and L. Palen. Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management*, 6(3):248–260, 2009.

[69] H. Jeong, Z. Néda, and A.-L. Barabási. Measuring preferential attachment in evolving networks. *EPL (Europhysics Letters)*, 61(4):567, 2003.

[70] R. Joiner, C. Stewart, and C. Beaney. Gender digital divide. *The Wiley Blackwell Handbook of Psychology, Technology and Society*, page 74, 2015.

[71] Q. Jones and S. Rafaeli. User population and user contributions to virtual publics: a systems model. In *Proceedings of the international ACM SIGGROUP conference on Supporting group work*, pages 239–248. ACM, 1999.

[72] Q. Jones, G. Ravid, and S. Rafaeli. Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Info. Sys. Research*, 15:194–210, June 2004.

[73] E. Joyce and R. E. Kraut. Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communication*, 11(3):723–747, 2006.

[74] N. Kang and N. Kwak. A multilevel approach to civic participation individual length of residence, neighborhood residential stability, and their interactive effects with media use. *Communication Research*, 30(1):80–106, 2003.

[75] S. J. Karau and K. D. Williams. Social loafing: A meta-analytic review and theoretical integration. *Journal of personality and social psychology*, 65(4):681, 1993.

[76] J. E. Katz, R. E. Rice, and P. Aspden. The internet, 1995-2000 access, civic involvement, and social interaction. *American behavioral scientist*, 45(3):405–419, 2001.

[77] A. L. Kavanaugh and S. Patterson. The impact of community computer networks on social capital and community involvement. *American Behavioral Scientist*, 45(3):496–509, 2001.

[78] A. L. Kavanaugh, T. T. Zin, M. B. Rosson, J. M. Carroll, J. Schmitz, and B. J. Kim. Local Groups Online: Political Learning and Participation. *Computer Supported Cooperative Work (CSCW)*, 16(4-5):375–395, Nov. 2006.

[79] B. Keegan, D. Gergle, and N. Contractor. Hot off the wiki: Structures and dynamics of wikipedia?s coverage of breaking news events. *American Behavioral Scientist*, page 0002764212469367, 2013.

[80] Y. H. Kidane and P. A. Gloor. Correlating temporal communication patterns of the eclipse open source community with performance and creativity. *Computational and mathematical organization theory*, 13(1):17–27, 2007.

[81] C. Klein, D. DiazGranados, E. Salas, H. Le, C. S. Burke, R. Lyons, and G. F. Goodwin. Does team building work? *Small Group Research*, 2009.

[82] A. Kleingeld, H. van Mierlo, and L. Arends. The effect of goal setting on group performance: a meta-analysis. *Journal of Applied Psychology*, 96(6):1289, 2011.

[83] S. W. Kozlowski and D. R. Ilgen. Enhancing the effectiveness of work groups and teams. *Psychological science in the public interest*, 7(3):77–124, 2006.

[84] R. Kraut, M. L. Maher, J. Olson, T. W. Malone, P. Pirolli, and J. C. Thomas. Scientific foundations: A case for technology-mediated social-participation theory. 2010.

[85] R. Kraut, M. Patterson, V. Lundmark, S. Kiesler, T. Mukophadhyay, and W. Scherlis. Internet paradox: A social technology that reduces social involvement and psychological well-being? *American psychologist*, 53(9):1017, 1998.

[86] R. E. Kraut and A. T. Fiore. The role of founders in building online groups. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 722–732. ACM, 2014.

[87] R. E. Kraut and P. Resnick. *Building Successful Online Communities: Evidence-based Social Design*, chapter Encouraging Contribution to Online Communities. Cambridge, MA: MIT Press., 2012.

[88] R. E. Kraut and P. Resnick. Encouraging contribution to online communities. In *Building successful online communities: Evidence-based social design*, pages 21–76. MIT Press Cambridge, MA, 2012.

[89] C. Lampe, R. Gray, A. T. Fiore, and N. Ellison. Help is on the way: Patterns of responses to resource requests on facebook. In *Proceedings of the 17th ACM Conference on Computer*

Supported Cooperative Work & Social Computing, CSCW '14, pages 3–15, New York, NY, USA, 2014. ACM.

[90] C. Lampe and E. Johnston. Follow the (slash) dot: Effects of feedback on new members in an online community. In *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work*, GROUP '05, pages 11–20, New York, NY, USA, 2005. ACM.

[91] J. M. Levine and R. L. Moreland. A history of small group research. *Handbook of the history of social psychology*, page 383, 2012.

[92] N. Lin. Building a network theory of social capital. *Connections*, 22(1):28–51, 1999.

[93] Y.-R. Lin, B. Keegan, D. Margolin, and D. Lazer. Rising tides or rising stars?: Dynamics of shared attention on twitter during media events. *PloS one*, 9(5):e94093, 2014.

[94] C. López, B. Butler, and P. Brusilovsky. Does anything ever happen around here? assessing the online information landscape for local events. *Journal of Urban Technology*, 21(4):95–123, 2014.

[95] C. López and B. S. Butler. Consequences of content diversity for online public spaces for local communities. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 673–682. ACM, 2013.

[96] C. López and R. Farzan. Designing for neighborhoods: lessons learned from paper-based bulletin boards. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 1963–1968. ACM, 2014.

[97] C. López and R. Farzan. Lend me sugar, i am your neighbor!: A content analysis of online forums for local communities. In *Proceedings of the 7th International Conference on Communities and Technologies*, C&T '15, pages 59–67, New York, NY, USA, 2015. ACM.

[98] C. López, R. Farzan, and P. Brusilovsky. Personalized incremental users' engagement: driving contributions one step forward. In *Proceedings of the 17th ACM international conference on Supporting group work*, pages 189–198. ACM, 2012.

[99] C. López, R. Farzan, S. S. Sahebi, and P. Brusilovsky. What influences the decision to participate in audience-bounded online communities? 2013.

[100] C. López, Y.-r. Lin, and R. Farzan. What makes hyper-local online discussion forums sustainable? In *Hawaii International Conference on System Sciences, Proceedings of the 48th Annual*, page To appear. IEEE, 2015.

[101] P. J. Ludford, D. Cosley, D. Frankowski, and L. Terveen. Think different: increasing online community participation using uniqueness and group dissimilarity. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 631–638. ACM, 2004.

[102] C. A. Masden, C. Grevet, R. E. Grinter, E. Gilbert, and W. K. Edwards. Tensions in scaling-up community social media: a multi-neighborhood study of nextdoor. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 3239–3248. ACM, 2014.

[103] M. Mc Pherson. The size of voluntary organizations. *Social Forces*, 61(4):1044–1064, 1983.

[104] J. M. McPherson and T. Rotolo. Testing a dynamic model of social composition: Diversity and change in voluntary groups. *American Sociological Review*, pages 179–202, 1996.

[105] I. Mergel and T. Langenberg. What makes online ties sustainable? a research design proposal to analyze online social networks. *Harvard Program on networked governance*, 6:60–69, 2006.

[106] J. R. Mesmer-Magnus and L. A. DeChurch. Information sharing and team performance: a meta-analysis. *Journal of Applied Psychology*, 94(2):535, 2009.

[107] D. R. Millen and J. F. Patterson. Stimulating social engagement in a community network. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, pages 306–313, 2002.

[108] A. Mislove, S. Lehmann, Y.-Y. Ahn, J.-P. Onnela, and J. N. Rosenquist. Understanding the demographics of twitter users. *ICWSM*, 11:5th, 2011.

[109] F. P. Morgeson, D. S. DeRue, and E. P. Karam. Leadership in teams: A functional approach to understanding leadership structures and processes. *Journal of Management*, 2009.

[110] E. W. Morrison. Newcomers' relationships: The role of social network ties during socialization. *Academy of management Journal*, 45(6):1149–1160, 2002.

[111] O. Nov, O. Arazy, C. López, and P. Brusilovsky. Exploring personality-targeted ui design in online social participation systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 361–370. ACM, 2013.

[112] T. P. Novak. Bridging the racial divide on the internet. *Science-AAAS-Weekly Paper Edition*, 280(5362):390–391, 1998.

[113] D. O'Neil. Assessing community informatics: a review of methodological approaches for evaluating community networks and community technology centers. *Internet Research*, 12(1):76–102, 2002.

[114] F. Ortega, J. Gonzalez-Barahona, and G. Robles. On the inequality of contributions to wikipedia. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, pages 304–308, 2008.

[115] Z. Papacharissi and A. Mendelson. Toward a New(er) Sociability: Uses, Gratifications, and Social Capital on Facebook. *Media perspectives for the 21st century*, page 212, 2010.

[116] E. Paulos, R. Honicky, and B. Hooker. Citizen science: Enabling participatory urbanism. *Handbook of Research on Urban Informatics*, pages 414–436, 2008.

[117] A. Perrin and M. Duggan. Men catch up with women on overall social media use. Technical report, Pew Research Center.

[118] K. E. Pettigrew, J. C. Durrance, and K. T. Unruh. Facilitating community information seeking using the internet: Findings from three public library–community network systems. *Journal of the American Society for Information Science and Technology*, 53(11):894–903, 2002.

[119] P. A. Popielarz and J. M. McPherson. On the edge or in between: Niche position, niche overlap, and the duration of voluntary association memberships. *American Journal of Sociology*, pages 698–720, 1995.

[120] R. D. Putnam. *Bowling alone: The collapse and revival of American community*. Simon and Schuster, 2001.

[121] D. R. Raban, M. Moldovan, and Q. Jones. An empirical study of critical mass and online community survival. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 71–80, 2010.

[122] S. Rafaeli, G. Ravid, and V. Soroka. De-lurking in virtual communities: a social communication network approach to measuring the effects of social and cultural capital. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, pages 10–pp. IEEE, 2004.

[123] H. Rainie, L. Rainie, and B. Wellman. *Networked: The new social operating system*. The MIT Press, 2012.

[124] A. M. Rashid, K. Ling, R. D. Tassone, P. Resnick, R. Kraut, and J. Riedl. Motivating participation by displaying the value of contribution. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, CHI '06, pages 955–958, New York, NY, USA, 2006. ACM.

[125] P. Resnick. Beyond bowling together: Sociotechnical capital. *HCI in the New Millennium*, (77):247–272, 2001.

[126] P. Resnick, J. Konstan, Y. Chen, and R. E. Kraut. Starting new online communities. In *Building successful online communities: Evidence-based social design*, pages 231–280. MIT Press Cambridge, MA, 2012.

[127] E. M. Rogers, L. Collins-Jarvis, and J. Schmitz. The pen project in santa monica: Interactive communication, equality, and political action. *Journal of the American Society for Information Science*, 45(6):401–410, 1994.

[128] T. Rotolo. Town heterogeneity and affiliation: A multilevel analysis of voluntary association membership. *Sociological perspectives*, 43(2):271–289, 2000.

[129] C. Sarkar, D. Y. Wohn, and C. Lampe. Predicting length of membership in online community "everything2" using feedback. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion*, CSCW '12, pages 207–210, New York, NY, USA, 2012. ACM.

[130] T. Schoberth, J. Preece, and A. Heinzl. Online communities: A longitudinal analysis of communication activities. In *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on*, pages 10–pp. IEEE, 2003.

[131] J. Schradie. The trend of class, race, and ethnicity in social media inequality: Who still cannot afford to blog? *Information, Communication & Society*, 15(4):555–571, 2012.

[132] D. Schuler. Community networks: building a new participatory medium. *Commun. ACM*, 37(1):38–51, Jan. 1994.

[133] J. C. Scott. Membership and participation in voluntary associations. *American Sociological Review*, pages 315–326, 1957.

[134] Y.-C. Shiue, C.-M. Chiu, and C.-C. Chang. Exploring and mitigating social loafing in online communities. *Computers in Human Behavior*, 26(4):768–777, 2010.

[135] L. Simpson. Community informatics and sustainability : Why social capital matters. *The Journal of Community Informatics*, 1(2):102–119, 2005.

[136] A. Smith. Civic engagement in the digital age. Technical report, Pew Research Center, 2013.

[137] J. Solomon and R. Wash. Critical mass of what? exploring community growth in wikiprojects. In *ICWSM*, 2014.

[138] C. Steinfield, N. B. Ellison, and C. Lampe. Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology*, 29(6):434–445, Nov. 2008.

[139] T. Thielmann. Navigation becomes travel scouting: The augmented spaces of car navigation systems. *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City*, pages 230–243, 2009.

[140] M. Tortoriello, R. Reagans, and B. McEvily. Bridging the knowledge gap: The influence of strong ties, network cohesion, and network range on the transfer of knowledge between organizational units. *Organization Science*, 23(4):1024–1039, 2012.

[141] S. Valenzuela, N. Park, and K. F. Kee. Is There Social Capital in a Social Network Site?: Facebook Use and College Students' Life Satisfaction, Trust, and Participation. *Journal of Computer-Mediated Communication*, 14(4):875–901, July 2009.

[142] X. Wang. *An ecological perspective on online communities*. PhD thesis, University of Pittsburgh, 2007.

[143] X. Wang, B. S. Butler, and Y. Ren. The impact of membership overlap on growth: An ecological competition view of online groups. *Organization Science*, 24(2):414–431, 2013.

[144] Y.-C. Wang and R. Kraut. Twitter and the development of an audience: those who stay on topic thrive! In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 1515–1518, 2012.

[145] Y.-C. Wang, R. Kraut, and J. M. Levine. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, pages 833–842, New York, NY, USA, 2012. ACM.

[146] M. M. Wasko and S. Faraj. Why should i share? examining social capital and knowledge contribution in electronic networks of practice. *MIS quarterly*, pages 35–57, 2005.

[147] B. Wellman, a. Q. Haase, J. Witte, and K. Hampton. Does the Internet Increase, Decrease, or Supplement Social Capital?: Social Networks, Participation, and Community Commitment. *American Behavioral Scientist*, 45(3):436–455, Nov. 2001.

[148] M. Wen and C. P. Rose. Understanding participant behavior trajectories in online health support groups using automatic extraction methods. In *Proceedings of the 17th ACM International Conference on Supporting Group Work*, GROUP '12, pages 179–188, New York, NY, USA, 2012. ACM.

[149] K. Williams and J. C. Durrance. Social networks and social capital : Rethinking theory in community informatics. *The Journal of Community Informatics*, 4(3):1–20, 2008.

[150] C. Xia, R. Schwartz, K. Xie, A. Krebs, A. Langdon, J. Ting, and M. Naaman. Citybeat: Real-time social media visualization of hyper-local city data. In *Demo in International World Wide Web Conference*, WWW '14, 2014.

[151] J. Yang, X. Wei, M. S. Ackerman, and L. A. Adamic. Activity lifespan: An analysis of user survival patterns in online knowledge sharing communities. In *ICWSM*, 2010.

[152] H. Zhu, J. Chen, T. Matthews, A. Pal, H. Badenes, and R. E. Kraut. Selecting an effective niche: an ecological view of the success of online communities. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 301–310. ACM, 2014.

[153] H. Zhu, R. E. Kraut, and A. Kittur. The impact of membership overlap on the survival of online communities. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 281–290. ACM, 2014.