

Consequences of Content Diversity for Online Public Spaces for Local Communities

Claudia López

University of Pittsburgh
135 North Bellefield Ave, Pittsburgh, PA 15260
cal95 @ pitt.edu

Brian Butler

University of Maryland
2117B Hornbake Bldg, College Park, MD 20742
bsbutler @ umd.edu

ABSTRACT

While there is significant potential for social technologies to strengthen local communities, creating viable online spaces for them remains difficult. Maintaining a reliable content stream is challenging for local communities with their bounded emphases and limited population of potential contributors. Some systems focus on specific information types (e.g. restaurant, events). Others allow many different information types. This paper reports our findings about the consequences of content diversity from a study of neighborhood-oriented Facebook groups. The findings raise questions about the viability of designs for local online communities that focus narrowly on single topics, goals, and audiences.

Author Keywords

Local online groups, design decisions, sustainability

ACM Classification Keywords

H.5.3. Group and Organization Interfaces: Web-based Interaction

INTRODUCTION

Communicative ecology refers to the mix of media that people use to connect to others [31]. Residents of local communities rely on many systems to communicate and propagate information [14]. Traditional media, internet-enabled systems and word of mouth comprise the communicative ecology of a neighborhood. A recent study has shown that local newspapers, radio, and TV continue to be important sources for many kinds of local information [28]. Word of mouth is also identified as a frequent source of community news [28].

Internet-based systems that rely on user-generated content also play a role as local information sources. Specialty websites provide information about restaurants and other local businesses [25]. Search engines enable users to discover locally-constructed information sources [25]. However, mobile applications and online social platforms such as Facebook and Twitter, while widely used in US, are only slowly developing as channels for local information [28]. Listservs

are rarely mentioned as source of local information [28]. In spite of the potential for peer-based online systems to propagate local information, development and deployment of these systems is still in the early stages.

One of the main challenges of online social systems is maintaining a sufficient level of content contribution. Even with a potentially global reach, online discussion communities often fail to maintain a critical mass of activity [13, 24]. Given the much smaller population of potential participants, building sustainable online spaces for local communities is even more challenging.

Systems that provide user-generated information about local communities vary significantly with respect to the diversity of content allowed. Some sites focus on providing information of a single type. Urbanspoon¹ enables users to share reviews about restaurants. Metromix² is a network of local entertainment websites that allows users to comment on the advertised establishments. Eventful³ uses the same mechanisms to advertise events. Each of these sites organizes content and activity in hierarchically-structured sections corresponding to particular regions and communities. Then within each community area, the type of information allowed is focused on a particular type of content.

At the other extreme are sites that allow individuals to post information on a broad range of topics. Websites for traditional media often collect information from individuals on a range of topics (traffic, school closings, sports and entertainment, politics, etc.). Craigslist⁴ allows users to post information ranging from discussion contributions to classified advertisements for all kind of services. e-democracy⁵ supports moderated online forums within a local community. i-neighbors⁶, Neighborland⁷, EveryBlock⁸ and NextDoor⁹ aim to provide a centralized website for each neighborhood where users can meet online, discuss and stay informed about any aspect of their neighborhood. While these information sources also have subsections focused on particular locations, their content

¹<http://www.urbanspoon.com>

²<http://www.metromix.com/>

³<http://eventful.com/>

⁴<http://www.craigslist.org/about/sites>

⁵<http://forums.e-democracy.org/>

⁶<http://i-neighbors.org/>

⁷<https://neighborland.com/>

⁸<http://www.everyblock.com/>

⁹<https://nextdoor.com/>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CSCW '13, February 23–27, 2013, San Antonio, Texas, USA.

Copyright 2013 ACM 978-1-4503-1331-5/13/02...\$15.00.

diversity is much greater than seen in the specialized sources described above.

Our larger research agenda focuses on determining how design decisions and community characteristics affect the viability and impact of locally-focused online social systems. To achieve this goal, we are developing and testing a data collection protocol for both online and offline information systems and information sources that are part of a neighborhood's communicative ecology. As part of this developing stream of research, we have collected data about neighborhood-oriented Facebook groups, websites that advertise local events, local media sites and offline bulletin boards in a sample of urban neighborhoods in Pittsburgh.

The study reported in this paper examines how a single system characteristic (content diversity) affects one aspect of local online system viability (level of activity) in a single type of online spaces for local communities (local Facebook groups). Community networks [30] –information systems that aim to support communication within a neighborhood– have usually tended to increase the diversity of their content over time, as reported in [7]. Therefore, we hypothesize that the diversity of the purposes and content affects future content volume and frequency of information exchanges within a system. In the context of a local community, we expect that the more diverse the communicative purposes and content, the more subsequent posts there will be within the system. However, we expect that this positive relationship will be subject to diminishing effects as higher levels of content diversity leading to problems such as information overload.

RELATED WORK

This work is focused on analyzing the diversity of content and its correlation to the level of activity of online groups for local communities. Our research question is based on three research areas: community networks, public displays and the impact of diversity on online communities. This section reviews the relevant work in these areas.

Community Networks

It has been observed that technology can reshape many aspects of people's lives [26]. Among them, researchers have studied how internet access changes the communication patterns with local social ties [16, 20], and how internet-based systems designed for neighborhoods can affect community involvement and social capital [17, 19]. Although it has been argued that the Internet can diminish the importance of neighborly proximity by enabling people to maintain friendship ties with others regardless of their location [34, 26], place and neighborhoods remain an important focus for many sociotechnical research efforts [1, 2, 7, 15, 18].

Community networks are “computer-based networks created by and for a local community” [7]. They are intended to support communities' efforts to achieve social goals such as greater community awareness and involvement [30].

Community Memory of Berkeley, California [11] is recognized as the first exemplar of community networks. It was known as an “electronic bulletin board” and it was used for

a wide range of topics such as housing, finding people with similar interests (e.g. musicians), organizing groups (e.g. car pooling), buying and selling things, and reviewing restaurants. Additionally, it developed some unexpected uses as users added poems, shared graphics, and started dialogues with others in the system [11]. The Cleveland Free-Net [5] was initially conceived as an online question & answer bulletin board for health issues, but was later extended to incorporate other topics such as law, arts, and government issues.

The PEN project in Santa Monica [27] offered access to public information such as event and bus schedules, city council agendas, public safety tips, and the library catalog. It also included a discussion forum where the neighbors could participate in or initiate discussions about local issues. Two-thirds of the overall activity in the community network took place in these open discussion forums. Generally, community networks tended to offer a broad range of topics with the range of content expanding over time.

Web-based versions of community networks also allowed for high levels of content diversity. The Blacksburg Electronic Village (BEV) [8] offered a wide variety of information and services ranging from health and education information to museums and library services. Particularly, it was also noticed that the BEV-news listserv – that was originally intended to host discussion about BEV policies – was used for a much broader range of discussion topics including event announcements, restaurant reviews, and even personal topics.

Netville [16] offered chat rooms and email lists where neighbors could initiate any kind of conversation. e-Neighborhood [17] offered different forums for reviews of local business and services, classifieds and events. The Carlisle Community Center designers reported that some discussion topics such as politics and local issues related to danger and novelty were more appealing to the neighbors as they generated more participation in the discussion forums [21]. Based on this observation, they suggested that purposefully seeding these sorts of topics might be an effective approach to increasing participation in community networks. e-Democracy [3], a discussion forum seeking to encourage civic engagement within neighborhoods, has specifically hired staff to seed new content in the discussion forums as a mechanism to encourage more activity in the discussion forums. Its studies of content diversity have revealed different proportions of each type of content (announcements, opportunities, requests, news and issues) in different neighborhoods [3].

Although it is often asserted that content diversity is positively related to participation rates, none of these projects have actually tested this hypothesis. Our work is particularly focused on studying the relationship between these two variables in local online groups.

Public Displays

Physical public displays are another prominent form of social information system where neighbors can find local information. Bulletin boards, store windows, light poles, and other features of the physical environment are often used to share local information. These physical objects provide a low

cost mechanism for disseminating information about events, activities, and organizations within a neighborhood [10, 32]. There have also been efforts to develop electronic announcement boards and other forms of publicly accessible, computerized information public displays [10, 22, 29] in order to increase community awareness and involvement.

The content posted in public displays has been previously studied. Based on observation of bulletin boards in three local areas in California and interviews with residents, Churchill et al. [10] concluded that bulletin boards serve as a communication media to seek and advertise (1) viewpoints, (2) activities, (3) events and (4) services. In another observational study, Alt et al. [4] determined that there is an interplay among a bulletin board's location, its stakeholders (the manager, the content providers and the viewers) and its content. Content that has no relation to the location is rarely posted. Besides, each bulletin board develops an "identity" as its managers remove posters related to undesired topics. Munson et al. [22] stated the goals and norms of use of a public display are negotiated among the participants over time. Our work aims to further investigate content in public online displays. We want to analyze how the decisions about the (allowed) diversity of content relate to subsequent content contribution rates.

Diversity in Online Communities

Diversity has been generally associated with participation measures in online communities. Diversity of tenure among editors of Wikipedia groups had a curvilinear association to the group's members withdrawal [9]. 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.

User population diversity in chat channels at their inception stage was a significant predictor of the likelihood of channels' long term sustainability [24]. Channels with more diverse populations were more likely to survive than those with more homogeneous populations. However, the diversity measure used in this study assessed how stable the user population was at an interval of time and did not assess diversity of any characteristic of individual posters.

Regarding content diversity, a study of e-mail-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 [6]. 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) [33]. In contrast, our study focuses on understanding the consequences of content diversity on the creation of content in neighborhood-oriented online groups.

HYPOTHESIS

As reviewed in the previous section, prior work on community networks has usually reported a broad range of topics

being offered or discussed [7]. A diverse spectrum of content can potentially be relevant to more people within the neighborhood, which in turn can lead to more people being motivated to create content in the online group. Given that the potential audience of a local online group is limited (compared to global online groups), we believe that the negative effects of content diversity in larger online groups [33, 6] will not be as detrimental in local online groups.

We hypothesize that the diversity of communicative purposes and content affects subsequent content volume and frequency of information exchanges within a system. In the context of a local community, we expect that the more diverse the purposes and contents, the more posts there will be within the system at later periods of time. However, we expect that this positive relationship will be subject to diminishing effects as high levels of content diversity lead to problems such as information overload and identity confusion.

DATA COLLECTION

Previous work on public displays and community networks has measured and qualitatively analyzed the diversity of content published in these social information systems [3, 4, 10, 21, 22]. This work aims to supplement the prior work by quantitatively examining the relationship between content diversity and subsequent content contributions in online groups related to neighborhoods. To examine the consequences of content diversity, we collected data from Facebook groups related to five urban neighborhoods in Pittsburgh.

Facebook Groups

Examining the consequences of design and contextual factors for local online systems requires a sample of sites that vary with respect to community characteristics, technology capabilities, and system features. This study of Facebook groups is an initial step in an ongoing effort to gather data from a diverse sample of local systems that make up a neighborhood's communicative ecology. Our sample includes local-oriented Facebook groups, websites, e-mail discussion lists, and offline bulletin boards from different communities.

Although Facebook is itself a large-scale, global, online social system, Facebook groups can be used to create online spaces for local communities. We chose this platform for this study because content diversity is not predefined by the technology itself, but rather it is developed and negotiated by the group's members. Facebook groups that are used to discuss local community issues have the potential to vary significantly with respect to content diversity in ways that are not directly determined by design decisions or technology affordances (as every group uses the same technology). Furthermore, as Facebook groups were not designed to be community networks, the groups are more likely to have naturally evolved, without the biases generated by planned interventions, participatory design or purposefully seeded content. Additionally, Facebook is already a very popular site, so many people have the technical skills needed to use it as a medium to share local information. Thus, low levels of participation in groups can be more reliably associated with the

content that was shared in the group than with other factors such as usability or general awareness of the platform.

Neighborhoods

We aim to understand what contextual factors can affect the sustainability of local online groups, therefore we sampled neighborhoods that differ with respect to attributes that might be relevant to the activity in online groups. The sampled neighborhoods vary with respect to population size, age demographics, and median income (Table 1).

Table 1. Neighborhoods' Descriptive Data¹⁰

Neighborhood ¹⁰	Population	% Over 60	Median Income
Bloomfield	9,089	21.60%	15K - 25K
East Liberty	6,871	18%	15K - 25K
Highland Park	6,749	16.90%	35K - 50K
Oakland Central	5,281	7.80%	10K - 15K
Squirrel Hill North	10,408	14%	50K - 75K

Central Oakland has the smallest population (5,281 residents), the lowest proportion of older residents (7.80%) and the lowest median income (10K - 15K) among the sampled neighborhoods. It is also distinctive because two universities (University of Pittsburgh and Carnegie Mellon University) are located there. Squirrel Hill North has the highest median income (50K - 75K) and the largest population. Bloomfield has the highest proportion of older residents (21.6%) and the second lowest median income. Highland Park has an average population size, average proportion of older residents and above average mean income. This small, but heterogeneous sample of urban neighborhoods is a starting point for understanding the design and contextual factors that affect the viability of local online social systems.

The data collection procedure

The following procedure was used to identify Facebook pages and groups associated with each neighborhood:

- A Facebook search was conducted using the neighborhood's name, "Pittsburgh" and "Facebook group" as the search keywords. The top ten search results were recorded.
- A Google search was conducted using the same search keywords and the additional constraint that results be from Facebook.com (site:facebook.com). The first twenty search results were recorded.

Duplicate sites and non-Facebook sites were removed based on manual inspection of the recorded search results. The resulting sample included 16 Facebook pages or groups (2-4 per neighborhood). We collected data about "likers" and posts for each of them using the Facebook developers API. We pulled the page that contained the most recent posts for each site in late October, 2011 (Time 2). We also retrieved their prior

¹⁰Neighborhoods are based on City of Pittsburgh maps. While assessing their accuracy is beyond the scope of this study, these designations are generally consistent with those used by residents, local media, and other organizations.

¹⁰Source: Census 2000 data - http://www.city.pittsburgh.pa.us/cp/assets/census/2000_census_pgh_jan06.pdf

Table 2. Content Coding Scheme

Goal's verb: What is the message's intention?	
Request	Solicit something from the reader, e.g. vote for a campaign
Provide	Provide something to the reader, e.g. inform about an accident.
Goal's noun: What is being requested/provided?	
Information	Objective information
Action	A physical action
Opinion	A subjective opinion
Audience' scale: How many people are needed?	
Peer	Only one reader (e.g. greeting someone for her birthday, selling a car)
Few people	Few (less than 100) readers (e.g. inviting to a meeting in a coffee shop)
Many people	As many readers as possible (e.g. signing an online petition)
Audience' involvement: What should they do?	
Passive Audience	A passive action (e.g. read an online news) Go to a place and be part of a passive audience (e.g. listening a concert)
Active	A physical activity needed to participate in the event (e.g. run a marathon)
Topic: What is the message about?	
Event	A happening that includes date and location and has been organized in advance (e.g. invitations to political meetings)
Activity	Organized actions that perform a function and require a higher commitment than attending an event once (e.g. join a band)
Cause	A perspective or request related to a community project (non-profit) or any community issue (e.g. donate food to a shelter)
Service	An occupation that is being advertised or a product that can be traded (e.g. babysitter wanted, carpool partners sought)
Fact	A happening that includes date and location that has not been organized in advance by the community (e.g. news)

(older) page of posts (Time 1). The number of posts per page varied depending on post frequency and length. In total, 794 posts were collected. This raw data is the basis for the analysis sample that contains measures of each site's communication activity and participants.

CODING LOCAL FACEBOOK POSTS

Facebook posts were automatically parsed to identify the date, type (i.e. status, link, photo, video), author, and number of likes and comments. To measure content diversity we applied a coding scheme to characterize the posts in terms of their communication goal, the scope of the intended audience, and their topic. This categorization scheme, summarized in Table 2, is a modified version of the classification schemes used in [4, 10].

A message's communication goal is indicated by a goal's verb and noun. The goal's verb refers to the intention of the mes-

Table 3. Categorization of Sample Messages

Example 1: “Councilman Patrick Dowd hosts Council to Go in Bloomfield! Councilman Dowd will be at Crazy Mocha in Bloomfield on February 8th to hear your ideas, concerns, and complaints. Please stop by! Monday, February 8, 2010 5:30pm - 7:00pm Crazy Mocha Coffee Company - 4525 Liberty Avenue. We hope to see you there!”	
Goal’s verb:	Solicit
Goal’s noun:	Action
Audience’s scale:	Few people
Audience’s involvement:	Active
Topic:	Event
Example 2: <i>The Allegheny River sometimes seems distant from Highland Park even though we’re really close. What do you think about having riverfront access, perhaps via bike or pedestrian connections (which have been discussed for years)?</i>	
Goal’s verb:	Solicit
Goal’s noun:	Opinion
Audience’s scale:	Many people
Audience’s involvement:	Passive
Topic:	Cause
Example 3: <i>Geotechnical drilling along Penn Avenue between Evaline and Mathilda will occur on 9/19 and 9/20 in the evening hours. This drilling takes samples to determine pavement depth at different locations, in preparation for the Phase 1 reconstruction project due to begin in 2013. It may look like Marcellus shale drilling, but it’s not!</i>	
Goal’s verb:	Provide
Goal’s noun:	Information
Audience’s scale:	Many people
Audience’s involvement:	Passive
Topic:	Fact

sage (request or provide). The goal’s noun indicates what is being communicated or requested (information, action, or opinion). By default every message aims to provide information. The other categories were designed to further capture the more specific goal of a given message. For example, the first message in Table 3 is providing information about a meeting with a city council member, but it is also *soliciting* neighbors to take an *action* by attending and participating in the meeting. In the second example in Table 3, the message poses a question about what others think about an infrastructure project/idea in the neighborhood. In this case, the message was categorized as *soliciting* neighbors’ *opinion*. The last example shows a message that informs about drilling at a particular location in the neighborhood. However, readers are not asked to do anything else. This message was categorized as *providing information*.

The intended audience assesses the scale of the target audience being engaged through the message (only one peer, few people, or many people) and the type of action required for involvement (passive, audience, or active). Table 3 also shows the intended audience codes associated with each example message. Example 1 mentions a meeting that is to be held

in a coffee shop; therefore, the number of people that can attend is limited. This contrasts with other messages that invite larger numbers of people. For example, the following message “*Little Italy Days this weekend! September 23, 24 and 25! Free entertainment (including Live On Liberty bands) free parking and food, lots and lots of delicious food!*” also encourages readers to attend an event, but this event is held in the streets of the neighborhood, so it clearly intends to engage several hundred people. This case was coded as *many-people* in the audience’s scope category. We defined a threshold equal to a hundred people to distinguish between few-people and many-people categories. In general, events that were located in coffee-shops and meeting rooms were coded as *few-people*. Events that were located in theatres or streets and actions that could be performed online (read an article or vote online) were classified as *many-people*. An example of message that was coded as *peer* for the audience’s scope is: “*Anyone eat at Cioppino restaurant in the strip? I have a \$50 gift certificate that will expire next week. I am selling it for \$40.00 or best offer. Hurry up it will expire July 14!*”. Although many people can read the message, only a single reader can participate in the purchase of the gift certificate.

Additionally, the kind of audience involvement describes the level of activity being requested of the audience. Example 1 in Table 3 encourages readers to go to a specific meeting place and express their ideas and concerns at that event. To respond to this message, the audience needs to be *active*. The message’s goal would not be achieved if individuals do not attend and actively participate in the discussion. On the other hand, examples 2 and 3 only require readers to perform an action online (write their opinion or read the information), which is coded as a *passive* action. An intermediate category was defined as *audience*. This category refers to messages that invite people to attend an event in a specific place, but participation in that event is passive. The following message: “*The Reservoir of Jazz concert series will launch on Sunday, August 7, at 5 p.m. with a performance by the BOILERMAKER JAZZ BAND*” was classified as *audience* because readers are encouraged solely to go as spectators at an event that would be performed in the same way whether they attend or not.

Finally, message topic refers to the main subject of the message: event, activity, cause, fact, or service. Events are gatherings that are planned and scheduled in advance that have a location and date. Example 1 in Table 3 refers to an *event*. An activity requires more commitment. It implies that the intended audience commits to perform a certain action in a more sustained fashion. For example, “*ELPC’s performing arts outreach program is looking for some great teens who might be interested in becoming part of HAT Co (Hope Academy Theater Company). If you are one, please apply. If you know one, please forward this on.*” was categorized as an *activity* because it is trying to engage teens in an ongoing theatre program. An sample of *cause* is shown in Example 2. The message sender is sharing their concern about an issue and asking others to share their thoughts about it. Example 3 is a sample of a *fact*. Although it is a happening that has location and date, it does not request people to attend or participate. Other examples of *facts* are accident reports, crime

alerts and any kind of news. Finally, *services* are occupations or products that are being advertised. An example of a service is: “Come by our lemonade stand today to support the library! We’ll be serving lemonade until 1PM.”

Each post was coded by two judges trained with the coding scheme. Inter-coder agreement for each feature was assessed with Cohen’s Kappa coefficients. The Kappa coefficients were: 0.61 (topic), 0.61 (audience’s scale), 0.63 (goal’s noun), 0.67 (audience’s involvement), 0.74 (goal’s verb). These minimally adequate Kappa values suggest that the coding scheme may not be fully refined. In particular, posts containing multiple goals, topics and audiences contributed to lower than expected levels of coder agreement. To evaluate the feasibility of increasing the reliability of the coding, the posts in Time 2 were re-coded with instructions to focus on primary aspects of the message. After this iteration, the coefficients for the messages were: 0.65 (goal’s noun), 0.75 (audience’s involvement), 0.78 (topic), 0.84 (goal’s verb), and 0.89 (audience’s scale). Based on this improvement in the coefficients, the coded sample was determined to be adequate for preliminary analysis. Any remaining disagreements between coders were resolved based on discussion.

DESCRIPTIVE ANALYSIS OF FACEBOOK GROUPS

Table 4 summarizes the number of posts, the time interval, and posting frequency for each of the 16 Facebook sites at Time 1 and Time 2. During Time 2, there were 430 posts in 16 groups with an average of 26.88 posts per group. The average interval of time in which they were posted was 172.16 days (range: 24.55 – 893.32 days). The average post frequency was 0.41 posts per day (range: 0.01 – 1.59 messages per day).

Table 4. Summary of Posts and Audience

site	Time 1			Time 2			
	n	days	freq	n	days	freq	likers
B-1	41	57.5	0.71	39	24.6	1.59	461
B-2	31	98.1	0.32	28	65.8	0.43	219
B-3				12	893.3	0.01	
E-1	6	6.9	0.87	26	99.2	0.26	58
E-2	29	37.9	0.76	25	50.6	0.49	100
E-3	43	115.8	0.37	36	116.3	0.31	126
E-4	30	204	0.15	27	326.6	0.08	124
H-1	26	75.7	0.34	26	118.2	0.22	225
H-2				27	237.5	0.11	93
O-1				11	312.2	0.04	
O-2	7	6.1	1.15	31	102.0	0.30	106
O-3	27	23.9	1.13	29	48.0	0.60	535
O-4	31	49.1	0.63	32	37.0	0.87	329
S-1	31	76.0	0.41	29	41.9	0.69	717
S-2	34	125.1	0.27	27	223.1	0.12	160
S-3	28	35.6	0.79	25	58.2	0.43	287
All	364			430			
Mean	28	70.1	0.61	26.9	172.2	0.41	252.9
Min.	6	6.1	0.15	11	24.6	0.01	58
Max.	43	204	1.15	39	893.3	1.59	717

The Time 1 sample consists of the second most recent batch of posts for each group. There were 364 posts in 13 groups.

Three of the sampled groups (B-3, H-2 and O-1) did not have enough history and activity to populate a prior page of posts in Facebook. Their frequency of posts per day is .01, .11 and .04, which are very low compared to the average frequency of the other groups in Time 1 (.61). As these three groups have existed over a long time (B-3: 893.3 days; H-2: 237.5 days; and O-1:312.2 days), we concluded that these online spaces had not managed to remain viable over time. Among the 13 groups that had sufficient activity, the mean number of posts in Time 1 was 28 and the mean time interval was 70.13 days.

The number of people who “liked” each Facebook site was collected at Time 2 as a measure of the potential audience for each site (Table 4). Two groups (B-3 and O-1) did not report this measure. For the other 15 groups, the average was 252.86 people (range: 58 – 717 people).

Ten groups were associated with nonprofits. Three groups (E-1, O-1, S-3) belong to a public library and two other groups (E-3, E-4) were related to churches. A single group (B-3) was created by an individual.

Posts’ Authors

Groups also varied with respect to the number of posters and the concentration of posts written by the most active authors (Table 5). The average number of people posting in each group is 5.77 at Time 1 and 6.44 at Time 2. Overall the number of posters ranged from 1 to 16.

Table 5. Summary of Audience and Posters

	Time 1		Time 2	
	posters	top 80%	posters	top 80%
Mean	5.77	0.21	6.44	0.45
Min.	1.00	0.00	2.00	0.20
Max.	16.00	0.50	13.00	1.00

Poster concentration was measured by calculating the proportion of authors who accounted for 80% of the posts. This measure ranges from 0 to 1. The lower its value, the higher the level of concentration (i.e. a smaller percentage of the posters account for the majority of the posts). The average poster concentration was 0.21 (Time 1) and 0.45 (Time 2). In Time 1, the sites were more concentrated, in part because of two groups in which only one user posted all the messages. However, in Time 2 there was a shift to lower levels of concentration with a mean of 45% of the posters accounting for 80% of the posts. This change holds true if we removed the most inactive groups (B-3, H-2, O-1). In that case, 37% of the posts’ creators accounted for 80% of the posts on average (range: 30% - 56%).

In general, the Facebook groups’ owners were the most active posters. In Time 1, the owners’ contribution ranged from 32% to 100% of the posts with a mean of 76%. In Time 2, the proportion of posts accounted for by group owners dropped to an average of 64% (range: 4% - 93%). This result was affected by the characteristics of the least active groups in our sample (B-3, H-2, O-1). Their owners contributed very little to their groups (8%, 45%, 4%). After removing them from the sample in Time 2, the owners’ share of posted message in Time 2 (mean: 74%, range: 38% - 93%) is similar to Time 1.

Analyzing the persistence of authors over time, we found that in most groups approximately half of the contributors in Time 1 contributed to the group again in Time 2 (average: 48%, range: 35% - 100%). Changes in group size also varied (with median change percentage: 0%). Five of the groups had fewer members in Time 2 than in Time 1 with member loss ranging from 25% to 50%. Three groups had exactly the same size in both periods. Another five groups increased their size ranging by between 1.23 and 4 times.

Content Diversity

Among the collected posts, 45% were related to events, 24% were about services, 20% are associated with facts, 9% discussed causes and 2% advertised activities (Figure 1a). None of the topic types were present in all of the groups. None of the groups had only one topic type accounting for 80% of posts. On the other hand, seven groups had high topic diversity with three quarters of their topic types accounting for 80% of their posts.

The majority of posts (64%) solicited action. 20% solely provided information, and 11% provided an opinion about an issue. Posts soliciting information (3%) or support (2%) were far less common (Figure 1b). The sampled groups were generally moderate with respect to post goal diversity, with only two groups exhibiting a high concentration (low diversity) and one having low goal concentration (high diversity).

39% of messages sought to engage large audiences in passive activities such as reading specific online news, voting online, or viewing published photos. Another segment of posts (34%) invited small groups of people to get involved in more active happenings such as community meetings. Only 12% of the posts sought to encourage large numbers of people to actively participate in events such as marathons or cleaning campaigns (See Figure 1c). Intended audience diversity among the posts in each group was also moderate.

During the data collection period, there were two street festivals in Bloomfield, which were mentioned in 19% of the posts in the Bloomfield groups (B-1,B-2,B-3). More than 40% of the posts in East Liberty and Squirrel Hill groups announced events or services provided by the Facebook group's owners. Squirrel Hill also had the highest rate of posts offering volunteering opportunities and fundraising campaigns. 23% of the posts were advertising local businesses. Two group owners whose organization aimed to promote local business development and artists largely defined that pattern.

DIVERSITY AND SUSTAINABILITY

To test our hypotheses regarding content diversity and content contribution levels, we used regression analysis to predict the frequency of posts at Time 2 with independent variables from Time 1. Groups that had no activity in Time 1 (B-3, H-2, O-1) were excluded. The number of distinct goals, audiences and topics present in each group in Time 1 were used as measures of different forms of content diversity.

The number of goals (# goals) counts how many different types of message goals were found in a given group in Time 1. A goal type was formed by combining a goal verb and a

goal noun. The possible values for goals were: "solicit action", "solicit information", "solicit opinion", "provide information" and "provide option". If a group has messages that were all coded as "solicit action", then the number of goals would be 1. If a group has some messages that were categorized as "solicit action" and others that were coded as "provide opinion", the number of goals would be assessed as 2. Figure 2 shows a scatter plot between the number of distinct types of goals observed in Time 1 and the frequency of posts in Time 2, illustrating the relationship between one type of content diversity and subsequent post frequency,

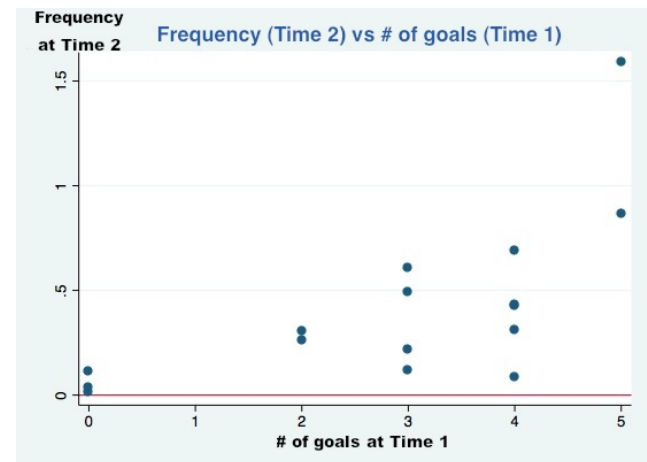


Figure 2. Frequency vs. Number of Goals

The number of audiences (# audience) measured the diversity of message target audiences. Distinct audiences were identified by combining audience scale and audience involvement level codes. Thus, potential audience values included: "peer-active", "peer-audience", "peer-passive", "fewPeople-active", "fewPeople-audience", "fewPeople-passive", "manyPeople-active", "manyPeople-audience" and "manyPeople-passive". The variable # audience was created by calculating the total number of different audience values that were found in each group.

The number of topics (# topics) assessed topic diversity. It was calculated by counting the number of different topic types (i.e. events, facts, causes, services and activities) present in the messages posted in each group.

In the regression models of post frequency, we also controlled for the number of authors and frequency of posts in Time 1. Dummy variables were included to assess neighborhood differences. However, none of the dummy variable were statistically significant in any of the analyses and hence have been excluded from the final results.

Table 6 reports a reduced model with the most significant predictors and the full model containing all the diversity variables. The reduced model was significant ($p = 0.001$) and it explains 81% of the variability in the frequency in Time 2. Diversity of goals was positively associated with high subsequent posting frequency after controlling for the number of authors and the frequency in Time 1. These two control variables were also significant factors to predict frequency of

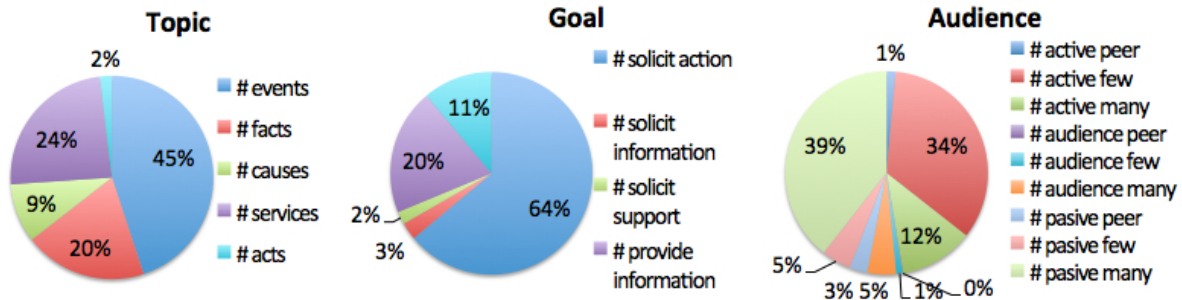


Figure 1. Topic, Goal and Intended Audience

posts in the next period. The number of authors and the frequency in Time 1 are positively associated with the frequency of posts in the next interval of time. A full model that includes all of our content diversity variables was also tested. The full model was significant ($p = 0.015$), however neither topic diversity (# topics) nor intended-audience diversity (# audience) were significant contributors to the full model.

Table 6. Predicting Frequency

Predictor	Reduced Model		Full Model	
	B	SE	B	SE
frequency (T1)	.740***	.196	.743**	.219
# authors (T1)	.042*	.019	.053	.033
# goals (T1)	.243***	.084	.262**	.096
# audience (T1)			-.057	.102
# topics (T1)			.049	.071
_cons			-1.102	.421
Reduced Model:	$R^2=.81, R^2_{adjusted}=.74, p = .0014$			
Full Model:	$R^2=.82, R^2_{adjusted}=.69, p = .0150$			

*: $p < .1$; **: $p < .05$; ***: $p < .01$

Jackknife residuals, leverage and Cook's distance analyses were conducted to assess the robustness of the results. A potential outlier was identified in most of the analyses. The group B-1 had an extremely high frequency of posts compared to the rest of the groups. Both regression models were estimated again after removing this data point and the results were qualitatively the same regarding the diversity of goals. After removing this potential outlier, the p-value of some of the control variables were above the significance level 0.5. However, diversity of goals remained a significant predictor of frequency of posts in the next period of observation. This suggests that the models are robust to outliers.

Multicollinearity tests were conducted using the Collin Stata package. The results show a moderate level of multicollinearity among the independent variables in the reduced model ($VIF = 1.61, CN = 14.64$) and a higher level of collinearity in the full model ($VIF = 3.01, CN = 21.04$). The Pearson correlation score between number of authors and number of goals is .64. The number of authors was highly correlated to the number of audiences ($r = .86$), which is the highest correlation among the independent variables. This may explain the insignificance of the number of authors in the full model.

We also evaluated the possibility that concentration measures

of each category, retention rates of authors and concentration of contributors might be associated with the frequency of posts in the subsequent period. However, none of these variables was a significant predictor in the regression analyses.

DISCUSSION

Our results confirm the hypothesis that content diversity, in terms of goals, is a significant factor in the development and maintenance of content streams in online spaces for local communities. Greater variety of message goals is associated with higher levels of activity in subsequent time periods. This result differs from other analyses of the consequences of content diversity in larger-scale systems, which have found that higher content diversity is associated with negative outcomes [6, 33]. This may mean that that focused streams of information are more desirable in larger-scale social systems, while in smaller, bounded contexts content diversity is more beneficial because it leads to a sufficiently active content stream, which in turn helps the group remain viable over time.

Our results also raise questions about the consequences of the design decisions made when implementing online communities for cities. Single-focus communities such as Eventful and Urbanspoon can engage a considerable number of users globally (their global ranking in Alexa is 5410 and 1699, respectively). However, these goal/topic-limited designs can lead to a segregated distribution of local information online. While this strategy may work in environments where the number of restaurants and events is sufficient to maintain the content stream, our results suggest that a broader content approach may be necessary in many local communities. This may be a factor behind the currently limited role of the online social systems that rely on user-generated content as a source for local information. Even when specialty sites (e.g. restaurant or entertainment event sites) are relevant sources of local information [25], people still rely on traditional media and word-of-mouth to acquire other relevant local information about events and services [28]. Interestingly, local events and services were the more popular topics in our sample of local Facebook groups. Future work should further examine the consequences of content type segregation on use of online local information sources and the role of less constrained platforms, such as Facebook, in propagating information about the local events and services.

We also observed that the majority of the Facebook groups

had highly participative owners. Owners accounted for more than 70% of the posts in the groups with high frequency of posts. Inequality of contribution in online groups is a phenomenon that has been observed in most successful large-scale online communities [23]. Therefore, it is not surprising to also find this behavior in these smaller groups. Although the content in an online community is meant to be provided by many of its members, it seems to be necessary to have a core of very active users that contribute enough content to regularly attract new members and retain existing ones. This can be especially difficult when a local online group is not created by the neighbors, but offered as an external service. However, a community network that aims to serve a neighborhood would need to find and motivate a group of residents to become the core contributor group.

Although the rate of members' retention was not found to be a significant predictor of post frequency, the continuous involvement of 48% of authors (on average) shows that repeated contributions to a local online group is not uncommon. Future work should consider its antecedents and consequences. Who are repeat contributors to local community online groups? What types of content do they post? Is there a relationship between the users' participation in an online local group and their role in their neighborhoods or affiliations with certain organizations? Does the presence of repeat contributor encourage or discourage new participants? As with any group, core participants have the potential to be both critical resources and stifling influences, suggesting that further understanding of these individuals' motivations and behaviors is important when creating local online groups.

Despite the fact that we expected to find an effect of the neighborhoods' characteristics on the online groups, our analysis did not reveal any significant effect. Although some neighborhoods had more groups than others, neighborhoods did not systematically differ with respect to the frequency of posts in their associated Facebook groups. This was especially surprising because one of the neighborhoods (Central Oakland) has a major proportion of transient residents as students who attend nearby universities largely compose its population. Several hypotheses regarding the lack of significant results can be raised. It is possible that our study design has obscured the location effects. For example, we decided to use the official maps to delineate the boundaries of each neighborhood and therefore its characteristics. These boundaries can differ from the boundaries that residents perceive among different sectors in the city, as it is discussed in [12]. If this were the case, we could have been relating local groups to incorrectly defined local communities. It is also possible that there are indeed no neighborhoods' factors that affect the sustainability of online groups. It is feasible that these groups are mainly run for people who are highly committed to the local community (e.g. long-term residents) which are present in any neighborhood, and that more transient (less committed) residents do not significantly impact the viability of these groups as they do not actively participate in them. Further replication of our study in a larger and more heterogeneous sample of online groups needs to be undertaken to fully understand the implications of this result.

As with any study, this research has limitations that must be taken into account when considering its implications. First, our dataset only contains public Facebook groups. Third parties cannot access private Facebook groups; therefore our results only apply to public online spaces. Private groups may function differently because of their association with more cohesive groups. Furthermore, Facebook groups were not particularly designed to support interaction among neighbors. A follow up study should consider the effects of content diversity in neighborhood-oriented network sites such as EveryBlock or NextDoor, in which the design decisions about structure may directly affect the types and consequences of content diversity. Moreover, we looked at the Facebook groups in isolation from other local media. Future work will focus on studying content diversity in a more diverse sample of the local communicative ecology.

Our results depend upon the nature of the coding strategy. Messages that combined different goals, topics, and audiences introduced noise into the measures that may have obscured the relationship between posting frequency and other types of content diversity. Development of coding schemes and models that take into account within-message diversity may provide a clearer characterization of these relationships. Despite these limitations, our findings shed light on aspects of use of online groups for local communities and the consequences of diversity in these groups.

CONCLUSION

This paper reports initial findings regarding the dynamics of local online systems. We collected, coded and analyzed more than 750 posts from 16 Facebook groups associated with five neighborhoods in an US city. The hypothesis that content diversity in local online groups would promote future posting activity was partially supported. Diversity of communication goals was significantly associated with higher frequency of posts in a subsequent time period. More generally, these local Facebook groups all contained a variety of posts with 45% of posts about events and 24% about services; 64% requesting some action and 20% providing information; and, 39% aiming to engage a large audience in passive actions and 34% inviting smaller audiences to join active happenings.

While not conclusive, these results raise questions about the viability of designs for local online systems that focus narrowly on single content and message types. A narrow scope risks the viability of the online group in a local setting, by neglecting one way that a sufficiently active content stream may be maintained. Future work should consider this and other design choices for information systems for local communities with data from a greater variety of communities, technologies, and system designs.

ACKNOWLEDGMENTS

Claudia López's PhD studies are funded by CONICYT, Chile (The National Commission for Science Research and Technology) and a Fulbright Scholarship.

REFERENCES

1. Special issue: Supporting community and building social capital. *Commun. ACM* 45, 4 (2002).

2. Special issue: Urban computing. *IEEE Computer* 39, 9 (2006).
3. Inclusive social media project: Participatory evaluation. Tech. rep., e-democracy.org, 2010.
4. Alt, F., Memarovic, N., Elhart, I., Bial, D., Schmidt, A., Langheinrich, M., Harboe, G., Huang, E., and Scipioni, M. Designing shared public display networks: Implications from today's paper-based notice areas. In *Pervasive Computing* (2011), 258–275.
5. Beamish, A. Communities on-line: Community-based computer networks. Master's thesis, Department of Urban Studies and Planning, Massachusetts Institute of Technology, 1995.
6. Butler, B. S. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Info. Sys. Research* 12, 4 (2001), 346–362.
7. Carroll, J. M. *The Neighborhood in the Internet*. Routledge, 2012, ch. What are community networks?, 23–47.
8. Carroll, J. M., and Rosson, M. B. Developing the Blacksburg electronic village. *Commun. ACM* 39, 12 (1996), 69–74.
9. Chen, J., Ren, Y., and Riedl, J. 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* (2010), 821–830.
10. Churchill, E. F., Nelson, L., and Denoue, L. *Communities and technologies*. Kluwer, B.V., 2003, ch. Multimedia fliers: information sharing with digital community bulletin boards, 97–117.
11. Colstad, K., and Lipkin, E. Community memory: a public information network. *SIGCAS Comput. Soc.* 6, 4 (1975), 6–7.
12. Cranshaw, J., Schwartz, R., Hong, J., and Sadeh, N. The livelihoods project: Utilizing social media to understand the dynamics of a city. In *ICSCW* (2012).
13. Cummings, J. N., Butler, B., and Kraut, R. The quality of online social relationships. *Commun. ACM* 45 (2002), 103–108.
14. Foth, M., and Hearn, G. N. Networked individualism of urban residents: discovering the communicative ecology in inner-city apartment buildings. *Information, Communication & Society* 10, 5 (2007), 749–772.
15. Gordon, E., and de Souza e Silva, A. *Net Locality: Why Location Matters in a Networked World*. Wiley-Blackwell, 2011.
16. Hampton, K., and Wellman, B. Neighboring in Netville: How the internet supports community and social capital in a wired suburb. *City & Community* 2, 4 (2003), 277–311.
17. Hampton, K. N. Neighborhoods in the Network Society the e-Neighbors study. *Information, Communication & Society* 10, 5 (2007), 714–748.
18. Haythornthwaite, C., and Kendall, L. Internet and community. *American Behavioral Scientist* 53, 8 (2010), 1083–1094.
19. Kavanaugh, A. L., and Patterson, S. The impact of community computer networks on social capital and community involvement. *American Behavioral Scientist* 45, 3 (2001), 496–509.
20. Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., and Crawford, A. Internet paradox revisited. *Journal of Social Issues* 58, 1 (2002), 49–74.
21. Millen, D. R., and Patterson, J. F. Stimulating social engagement in a community network. In *CSCW* (2002), 306–313.
22. Munson, S. A., Rosengren, E., and Resnick, P. Thanks and tweets: comparing two public displays. In *CSCW* (2011), 331–340.
23. Ortega, F., Gonzalez-Barahona, J., and Robles, G. On the inequality of contributions to wikipedia. In *HICCS* (2008), 304–308.
24. Raban, D. R., Moldovan, M., and Jones, Q. An empirical study of critical mass and online community survival. In *CSCW* (2010), 71–80.
25. Rainie, L., Purcell, K., Mitchell, A., and Rosenstiel, T. Where people get information about restaurants and other local businesses. Tech. rep., Pew Research Center, 2011.
26. Rainie, L., and Wellman, B. *Networked: The New Social Operating System*. The MIT Press, 2012.
27. Rogers, E. M., Collins-Jarvis, L., and Schmitz, J. The pen project in santa monica: Interactive communication, equality, and political action. *Journal of the American Society for Information Science* 45, 6 (1994), 401–410.
28. Rosenstiel, T., Mitchell, A., Purcell, K., and Rainie, L. How people learn about their local community. Tech. rep., Pew Research Center, 2011.
29. Schroeter, R. Engaging new digital locals with interactive urban screens to collaboratively improve the city. In *CSCW* (2012), 227–236.
30. Schuler, D. Community networks: building a new participatory medium. *Commun. ACM* 37, 1 (Jan. 1994), 38–51.
31. Tacchi, J. A., Slater, D., and Hearn, G. N. *Ethnographic Action Research: A User's Handbook*. UNESCO, India, 2003.
32. Taylor, N., Cheverst, K., and Miller, J. Affordances and signifiers of community noticeboards. In *Workshop on Pervasive Advertising* (2009), 1–4.
33. Wang, Y.-C., and Kraut, R. Twitter and the development of an audience: those who stay on topic thrive! In *CHI* (2012), 1515–1518.
34. Wellman, B. Physical place and cyberplace: The rise of personalized networking. *International Journal of Urban and Regional Research* 25, 2 (2001), 227–252.