Awalin Sopan, PJ Rey,
and Ben Shneiderman

The Dynamics of Web-Based Community Safety Groups: Lessons Learned from the Nation of Neighbors

Online social networks targeted toward national priorities such as disaster planning, crime watch, or food safety are increasingly described in research studies. A central research challenge is to understand the determinants of the successful growth of such communities. The use of a visual analytics tool would guide the community managers to understand their dynamics. To study this, we used data from the Nation of Neighbors (NON) [11], which is a platform for online neighborhood crime watch communities of residents and law enforcement officers to interact over the shared goals of preventing crime and strengthening communities. The members and law enforcement jurisdiction can add their community to NON; report crime or other incidents in their communities; participate in community discussion; share news, photos, or documents; and manage upcoming events. The system sends real-time e-mail or text message alerts to the community members.

During this case study, we worked closely with the executive director and manager of NON, defined metrics to quantify the success and growth of the communities, and refined our visual analytics tool ManyNets [2] to explore and compare communities as well as analyze their temporal dynamics. ManyNets presents network data in three tables: network, node, and edge tables, connected to node-link diagrams of the networks. It also allows users to analyze network growth over time. Our team of computer scientists and sociologists proposed three analyses that community managers can perform to improve their understanding about their communities:

1) *Community-level analysis*: We defined and implemented novel metrics to assess community success, compared communities along those metrics, and developed hypotheses about factors influencing community growth and member participation.

2) *Member-level analysis*: We analyzed the activity of individual community members and defined and implemented metrics to identify leaders and to quantify their impact.

3) *Temporal analysis*: We compared the growth and activity patterns of communities over time.

Our case study shows the capability of ManyNets to deliver insights regarding these three analytical aspects. Managers of online communities can take advantage of such a visual analytic tool to analyze the activities of the community members, observe the evolution of the communities, and make informed decision about their communities.

RELATED WORK

Neighborhood crime-watch initiatives long predate the emergence of the public Internet (coming to national prominence in the 1970s and 1980s). While there is some evidence that traditional neighborhood watch organizations can be effective at lowering crime in local communities, most studies fail to substantiate these claims [4]; moreover, even when demonstrated effective, the positive effects often rapidly dissipate [5]. The Web offers tools (e.g., Hollaback, SeeClickFix, Crimereports.com, SpotCrime) to community organizers to overcome this shortcoming and new possibilities for neighborhood crime-watch programs. NON, however, has the unique and specific goal of reorganizing neighborhood watch to make it more effective. Many studies (e.g., [1]) describe relevant metrics of online community success but, ultimately, success is defined by the unique mission of the community and its organizers [6]. Close collaboration with the community managers can help analysts in identifying these metrics.

Social network analysis and network visualizations are actively used for analyzing the networks of criminals or terrorists [10]. Hansen, Shneiderman, and Smith [3] demonstrated how NodeXL can be used to mine and analyze the conversation networks of such communities. Trier et al. [9] demonstrate the usefulness of using dynamic network visualization tools to understand community development. Two common approaches are 1) plotting summary statistics over time and 2) presenting a separate node-link diagram of the network at each point of time. In contrast, ManyNets uses tabular visualization to compare features of networks. Such visualization techniques can benefit Web-based crime-watch efforts providing insight into what community organizers can do to make the communities more effective.

DATA, METRICS, AND ANALYSIS

DATA PREPARATION

We collected the activity log of NON community members from January 2005 to December 2011—6,370 activities from 230 communities in total. Activities were classified into five categories:

- report: describes an incident that occurred in the community
- post: starting a discussion topic
- reply: responding to a previously posted discussion

Digital Object Identifier 10.1109/MSP.2013.2276513
Date of publication: 15 October 2013

- invitation: soliciting a person to join the community via e-mail
- acceptance: new members joining the community following an e-mail invitation.

Two members have an edge connecting them (i.e., relationship) if one of them replies to other's post.

After loading the data in ManyNets, our initial look at the 230 communities showed dramatic spikes in invitation activity in June 2006. Discussions with NON's manager revealed that it corresponded to a major reorganization of NON and the older data was not usable. The node (i.e., member) table also showed that the member with ID 0 was the most active in most communities, but ID 0 was used whenever a member posts anonymously. We recalculated the leadership metrics (discussed later) by discarding those anonymous contributions. In total, 28.7% of the total activities were posted anonymously.

COMMUNITY-LEVEL ANALYSIS

ManyNets automatically creates a network table where each row is a community (Figure 1) and computes default network metrics for the communities such as node counts (i.e., number of members), edge counts (total count of activities), connected component count, etc. A distribution column shows the distribution of activity type using small color-coded histograms. Here, acceptance, invitation, post, replies, and reports are red, blue, green, purple, and orange, respectively. Separate columns for each type of activity are provided as well. ManyNets allows sorting, filtering, clustering, and selecting communities based on the metrics.

A CENTRAL RESEARCH CHALLENGE IS TO UNDERSTAND THE DETERMINANTS OF THE SUCCESSFUL GROWTH OF SUCH COMMUNITIES.

COMMUNITY-LEVEL HEALTH METRICS

More complex metrics were needed to identify the success of communities [7], so we defined community health metrics to measure the success of the communities in terms of growth and activeness. We added them as new columns in the network table.

Interaction Intensity I

This is the total activity in a community from all its members divided by total member-months

$$I = A/\sum\nolimits_{m\epsilon M_C} U_m,$$

where M_C is the set of community C members, U_m is the number of months member m has been registered, and $A = R + P + T + I_S + I_A$ (R is the number of reports, P is the number of posts, T is the number of replies, I_S is the number of invitations sent, and I_A is the number of invitations accepted).

Average Active Months $\bar{M}$

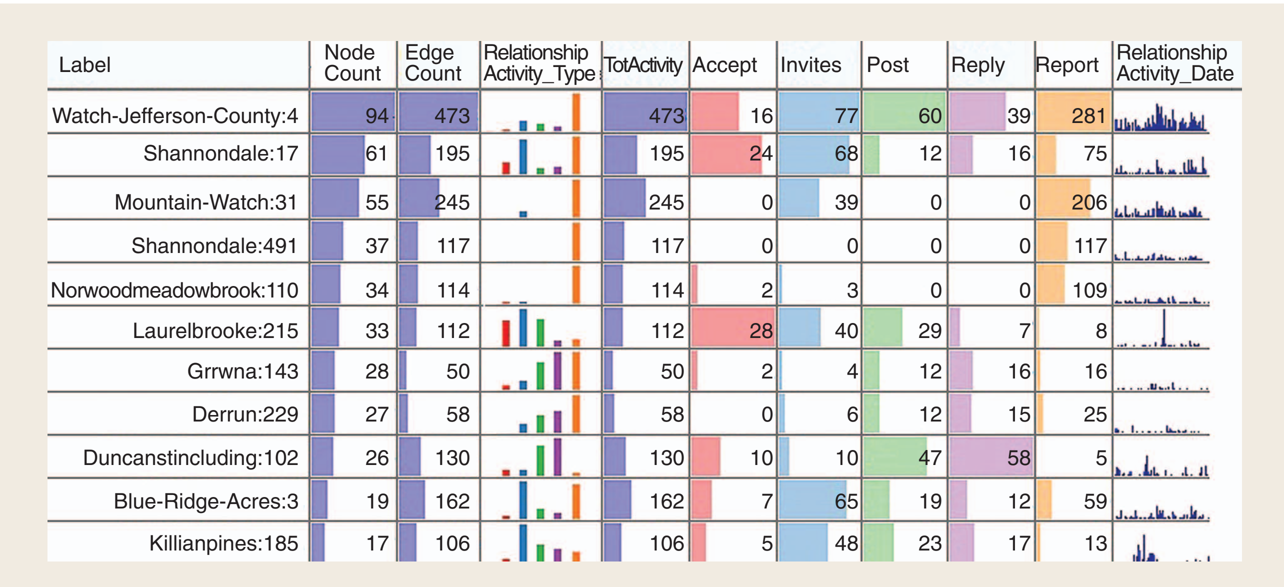
This is the average number of months the community members have participated

$$\bar{M} = \sum\nolimits_{m\epsilon M_C} U_m/|M_C|,$$

where $|M_C|$ is the total number of members in community C.

MEMBER-LEVEL ANALYSIS

We noticed that the most active communities contained one or two members who were far more active than the other community members. Hence, our member-level analysis aimed at finding out leaders and the influence of law enforcement people. In ManyNets, each community has a node table showing each member as a row. The columns are members' activity type distribution, degree, number of total activities, join date, and our proposed leadership metric. Analysts can also select particular members and create their ego networks



Label	Node Count	Edge Count	Relationship Activity_Type	TotActivity	Accept	Invites	Post	Reply	Report	Relationship Activity_Date
Watch-Jefferson-County:4	94	473		473	16	77	60	39	281	
Shannondale:17	61	195		195	24	68	12	16	75	
Mountain-Watch:31	55	245		245	0	39	0	0	206	
Shannondale:491	37	117		117	0	0	0	0	117	
Norwoodmeadowbrook:110	34	114		114	2	3	0	0	109	
Laurelbrooke:215	33	112		112	28	40	29	7	8	
Grrwna:143	28	50		50	2	4	12	16	16	
Derrun:229	27	58		58	0	6	12	15	25	
Duncanstincluding:102	26	130		130	10	10	47	58	5	
Blue-Ridge-Acres:3	19	162		162	7	65	19	12	59	
Killianpines:185	17	106		106	5	48	23	17	13	

[FIG1] The ManyNets network table showing 11 communities (one per row) and a selection of the available metrics.

to visualize the connections of these members with other members.

MEMBER-LEVEL LEADERSHIP METRIC

We decided to look for outliers who participated an extraordinary amount (two standard deviations above the mean activity of all members in that community). We defined *leadership* as follows:

$$L_m = A_m - 2\sigma_A - \mu_A,$$

where $A_m = R_m + P_m + T_m + I_S^{(m)} + I_A^{(m)}$ (R_m is the number of reports submitted by member m, P_m is the number of posts, T_m is the number of replies, $I_S^{(m)}$ is the number of invitations sent, and $I_A^{(m)}$ is the number of invitations accepted); μ_A and σ_A are mean and standard deviation of A over all the members of the community. A positive leadership score indicates a member whose activity level is significantly higher than other members in the community.

TEMPORAL ANALYSIS

Temporal changes include growth patterns, changes in activity levels, and changes in the type of activity over time. To analyze the data in the temporal dimension, we added two new features in ManyNets.

ACTIVITY DISTRIBUTION OVER TIME

In our distribution column "Activity Date" in the network table (Figure 1), each cell is the distribution of activity count per day distributed over time. This column helps identify different activity patterns (e.g., a sudden spike in activities in a community, communities where the activity is diminishing, or persistent communities where the activity level remains high), trends, and outliers (communities with anomalous activity patterns).

TEMPORAL SPLIT OF NETWORK

ManyNets splits a network into a series of subnetworks, each one comprising only the activities of a selected community within a specific time range (a week, month, or year). Activities over a month are shown in each row in Figure 1, visualizing the changes in activity pattern over time.

EXAMPLES OF INSIGHT

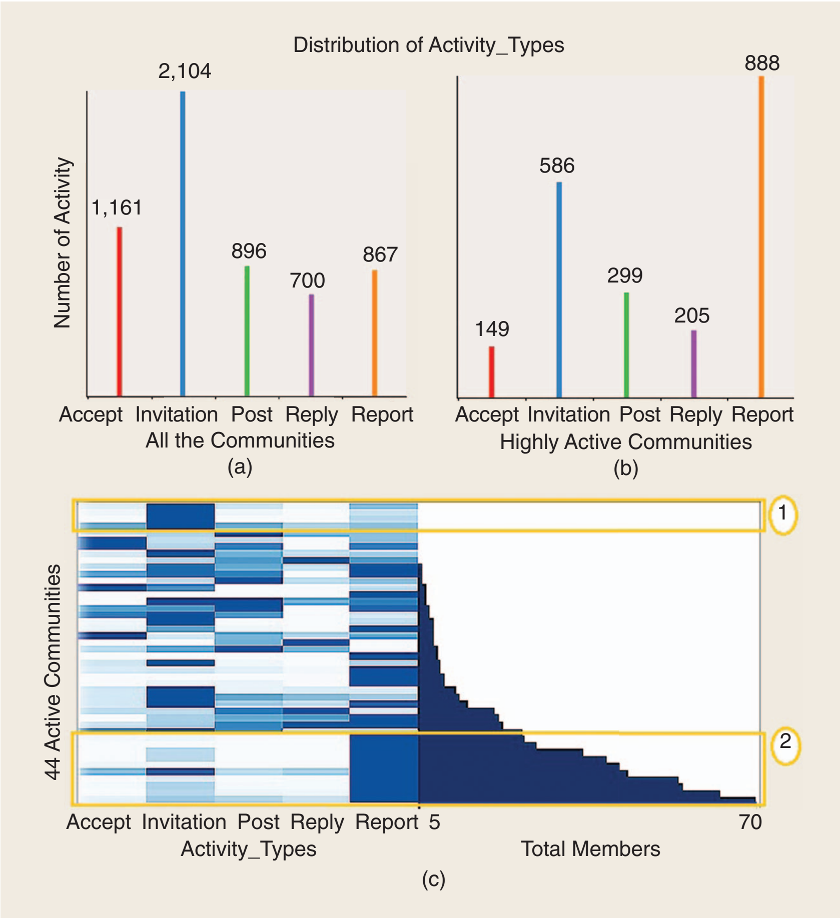
Our analysis of communities, their members, and their temporal dynamics produced interesting insights about the activity patterns, growth patterns, and leadership in the communities.

SELECTING SUCCESSFUL COMMUNITIES

We filtered out the communities with no activities at all, kept only the communities that have at least five invitation activities and at least five active members, and then selected the communities with the highest interaction. Finally, we manually reviewed communities that geographically overlapped and kept the largest one, thus having 44 active and independent communities suitable for comparative analysis.

ACTIVITY PATTERNS OF ALL COMMUNITIES

The histogram of the activity type for all the 230 communities showed that invitation was the most common activity [Figure 2(a)]. But after filtering down to the 44 larger and active communities, we observed more reports than invitations [Figure 2(b)]. To see if there was a correlation between the size of the communities and the activity patterns in the 44 communities, we generated a side-by-side overview of the activity type distribution column and the total member count column [Figure 2(c)] sorting the rows according to the total member count. This showed that the larger communities have comparatively more reports than any other activity, and smaller communities



[FIG2] (a) The relative distribution of activity type in communities for all communities and (b) for active communities. (c) A side-by-side view of activity type distribution heat map and total member for the active communities. Each band of the heat map is activity distribution of a community in the "Activity_Type" column, and the corresponding member count for this community is shown in the "Total Members" column. Cluster 1 shows smaller communities with more invites, while cluster two shows larger communities with more reports.

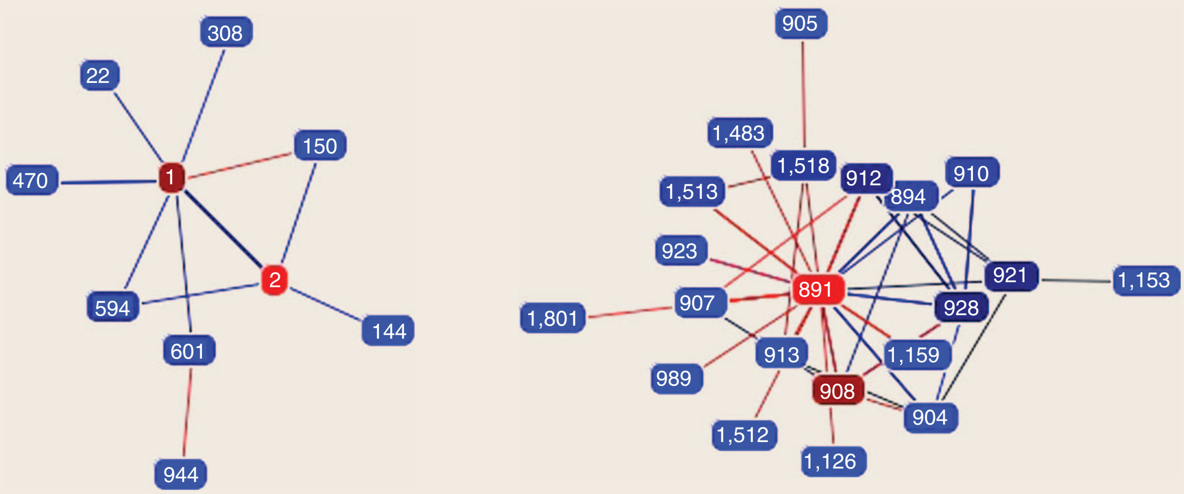
have more invitations. The heat map column overview of Activity_Type distribution column [8] rendered each community's activity type distribution as heat maps stacked one after another (blue being the color for the maximum value).

LEADERSHIP

One important question was whether the successful communities are driven by just a few active members or not and whether those members have any influence over the rest of the community. After selecting the most active communities, we observed that none of the successful communities had any people from law enforcement, indicating the involvement of law enforcement people was not a success indicator in those communities. There were 16 communities that had at least one leader (three of them had two leaders; all others had one). Figure 3(a) shows part of the community table generated by leaders and their activities. From distributions of the activity type of the leaders, we observed that leaders were mostly sending invitations. This indicates their intent to recruit new members to the community, which is vital when the community is still new. Also, most invitations in a community were sent by the leaders—invitations sent by other members were sparse. The "Activity Date" column showed that the

Label	Node Count	Relationship Activity_Type	Relationship Activity_Date	Accept	Invites	Post	Reply	Report	Total Activity
Shannondale:17	2			0	64	4	4	16	88
Harpers-Ferry:16	1			0	39	0	0	1	40
Charles-Town:18	1			0	39	0	0	1	40
Ranson:23	1			0	39	0	0	1	40
Silver-Grove:33	1			0	39	0	0	0	39
Laurelbrooke:215	1			0	38	25	2	6	71
Kendallplace:204	1			0	29	15	2	9	55
Blue-Ridge-Acres:3	1			0	25	18	3	23	69
Keyes-Ferry:80	1			0	25	1	0	19	45
Monrovia-Hollow:51	1			1	22	31	3	0	57
Killianpines:185	1			0	12	16	5	8	41

(a)



(b) (c)

[FIG3] (a) The network table comparing the activity of the leaders in 11 communities. All other members are filtered out from the networks. (b) Ego network of the leaders of the Watch-Jefferson-County community and (c) ego network of the Duncans community. Red nodes are the leaders with higher values of the leadership metric. In the ego networks, two nodes (members) have an edge connecting them if one of them replied to the post of the other.

temporal patterns in the communities in Figure 3(a) from rows two to five were very similar, and they were all from Jefferson County. Although the leaders of these communities were initially very active, their activities decreased over time. To compare the leaders from different communities, we used a node-link diagram to visualize their posts and replies. In the node-link view, the nodes were ranked by leadership value of the members (red indicting the node with maximum leadership value, blue being the lowest) [Figure 3(a) and (b)]. In the Watch-Jefferson-County community, the leader's ego network was small: the leader was connected with only a few other members as the leader's ego network had only ten nodes even though this was the largest community. In contrast, the Duncans community had many posts and replies, and the leader was connected with other members making more posts and replies. This pattern suggests that if the leader engages in a specific type of activity, it may boost the total participation on that type of activity by other members.

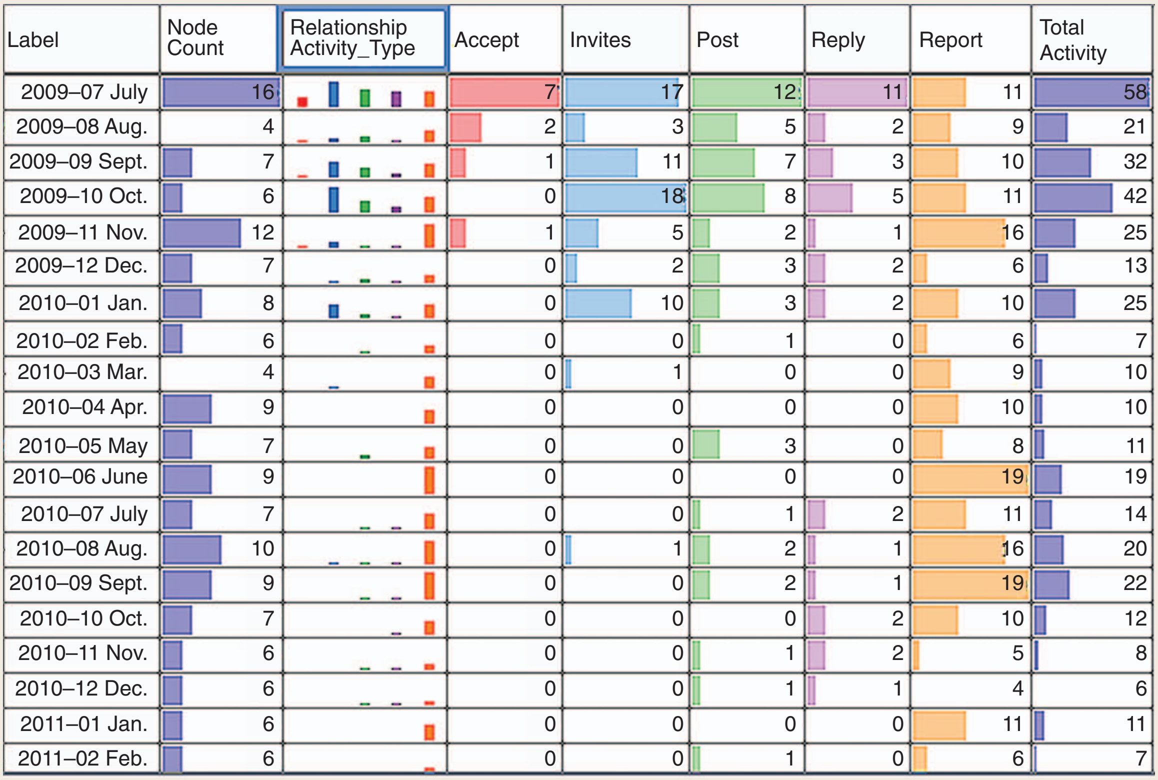
After using ManyNets to identify the leaders, we exported the data to conduct a regression analysis in STATA. The regression analysis supported the hypothesis developed with the aid of ManyNets' visualization, i.e., the presence of superactive members, strongly correlates with the growth of a community [7].

GROWTH PATTERN OF A SINGLE COMMUNITY

After sampling the communities, the Watch-Jefferson-County community appeared to be the most active, so we split it to observe its activity over time. In Figure 4, each row represents the network for a month sorted by time from July 2009 to February 2011. Initially, there were different types of activities, but, gradually, the proportion of reports grew larger while no more invitations and acceptances occurred recently. This community accumulated members first and only after having enough members did they start posting crime reports and having discussions about community safety. As more people became involved in the community, some would say that it reached critical mass, the number of reports increased.

DISCUSSION

In successful and persistent communities, we observed the presence of leaders. As the activities of the leaders can influence the activities of other members, community managers might want to promote such leadership and support their activities online or offline (e.g., encourage them to arrange community safety activities). As the number of members grows, there are fewer invitations sent and more reports posted. Highly active communities appear to have more reports than any other activity. We expected law enforcement involvement to heavily influence activity levels, but we found no evidence to support this



Label	Node Count	Relationship Activity_Type	Accept	Invites	Post	Reply	Report	Total Activity
2009–07 July	16		7	17	12	11	11	58
2009–08 Aug.	4		2	3	5	2	9	21
2009–09 Sept.	7		1	11	7	3	10	32
2009–10 Oct.	6		0	18	8	5	11	42
2009–11 Nov.	12		1	5	2	1	16	25
2009–12 Dec.	7		0	2	3	2	6	13
2010–01 Jan.	8		0	10	3	2	10	25
2010–02 Feb.	6		0	0	1	0	6	7
2010–03 Mar.	4		0	1	0	0	9	10
2010–04 Apr.	9		0	0	0	0	10	10
2010–05 May	7		0	0	3	0	8	11
2010–06 June	9		0	0	0	0	19	19
2010–07 July	7		0	0	1	2	11	14
2010–08 Aug.	10		0	1	2	1	16	20
2010–09 Sept.	9		0	0	2	1	19	22
2010–10 Oct.	7		0	0	0	2	10	12
2010–11 Nov.	6		0	0	1	2	5	8
2010–12 Dec.	6		0	0	1	1	4	6
2011–01 Jan.	6		0	0	0	0	11	11
2011–02 Feb.	6		0	0	1	0	6	7

[FIG4] The temporal splits of the Watch-Jefferson-County community. After January 2010, more reports are posted, whereas the number of invitations dropped.

hypothesis in this case study; however, only 12 communities had law enforcement officials involved with them, so there might just be too few cases for adequate analysis. Nevertheless, law enforcement officers are encouraged to self-identify and use the site as part of their professional duties.

Only a small percentage of community members participated in reporting crime, and invitations were mostly sent by the leader. There were 10% of the posts that were intended to be replies to some previous posts, but the members created a new post instead of replying to the initial post. Lessons for community managers can be to revise the interface so that relevant discussion can be easily performed within the same post and provide more obvious options to send invitations.

CONCLUSIONS AND FUTURE WORK

This article shows how a visual analytic tool, ManyNets, can help community managers generate hypotheses about community behavior and identify successful cases. We started with an overview of the attributes of hundreds of communities and then filtered down to 44 successful and interesting communities, analyzed their member activities, and identified the leaders. Having the capability to generate both statistical and visual insights integrated in the same tool along with its filtering features provided the leverage of rapid reiteration within one tool without going back and forth among several tools. One remark from the NON community manager, Art Hanson was, "Your observations and analysis of what contributes to a 'successful' community will be very helpful going forward—I am hoping to implement some of your measures as built-in tools to help our community managers."

In the future, we want to analyze passive activities as well. An important next step is to suggest possible interventions to the manager that are likely to increase participation, visualization of the topic distribution inside the tool, and observation of their temporal changes.

ACKNOWLEDGMENTS

We cordially thank Catherine Plaisant, Jae-wook Ahn, Alan Neustadtl, Manuel Freire, Jee-hye Kang, Brian Butler, and NON Community Manager Art Hanson for their valuable feedback and development suggestions. This work is supported in part by the National Science Foundation, grant IIS-0968521 "Supporting A Nation of Neighbors with Community Analysis Visualization Environment."

AUTHORS

Awalin Sopan (awalin@cs.umd.edu) is a graduate student in the Department of Computer Science and a graduate research assistant at the Human–Computer Interaction Lab at the University of Maryland. Her research interest encompasses social network analysis, information visualization, and user interaction.

PJ Rey (pjrey@umd.edu) is a Ph.D. student in the Department of Sociology at the University of Maryland. He studies how new forms of digital communication shape and are shaped by America's cultural and economic landscape.

Ben Shneiderman (ben@cs.umd.edu) is a Distinguished University Professor in the Department of Computer Science and is the founding director (1983–2000) of the Human–Computer Interaction Lab at the University of Maryland. He is a Fellow of the IEEE, AAAS, and ACM, and he is also a member of the National Academy of Engineering.

REFERENCES

[1] B. S. Butler, "Membership size, communication activity, and sustainability: A resource-based model of online social structures," *Informat. Syst. Res.*, vol. 12, no. 4, pp. 346–362, 2001.

[2] M. Freire, C. Plaisant, B. Shneiderman, and J. Golbeck, "ManyNets: An interface for multiple network analysis and visualization," in *Proc. 28th Annual SIGCHI Conf. Human Factors in Computing Systems,* 2010, pp. 213–222.

[3] D. Hansen, B. Shneiderman, and M. Smith, "Visualizing threaded conversation networks: Mining message boards and email lists for actionable insights," in *Active Media Technology*, vol. 6335, A. An, P. Lingras, S. Petty, and R. Huang, Eds. Berlin: Springer-Verlag, 2010, pp. 47–62.

[4] K. Holloway, T. Bennett, and D. P. Farrington, *Crime Prevention Research Review No. 3: Does Neighborhood Watch Reduce Crime?* Washington, DC: US Department of Justice Office of Community Oriented Policing Services, 2008.

[5] B. Lindsay and D. McGillis, "Citywide community crime prevention: An assessment of the Seattle program," in *Community Crime Prevention: Does It Work.* Beverly Hills, CA: SAGE Publications, 1986, pp. 46–67.

[6] J. Preece, "Sociability and usability in online communities: Determining and measuring success," *Behav. Informat. Technol.*, vol. 20, no. 5, pp. 347–356, 2001.

[7] P. J. Rey, A. Sopan, J. Ahn, J. H. Kang, C. Plaisant, and B. Shneiderman, "The influence of leadership on activity levels on an online crime watch community," *Tech. Rep*., in *Proc. American Sociology Association Annual Meeting*, 2012.

[8] A. Sopan, M. Freire, M. Taieb-Maimon, C. Plaisant, J. Golbeck, and B. Shneiderman, "Exploring data distributions: Design and evaluation," *Int. J. HCI*, vol. 29, no. 2, pp. 77–95, 2013.

[9] M. Trier, "Towards dynamic visualization for understanding evolution of digital communication networks," *Informat. Syst. Res.*, vol. 19, no. 3, pp. 335–350, 2008.

[10] J. Xu and H. Chen, "Criminal network analysis and visualization," *Commun. ACM,* vol. 48, pp. 100–107, June 2005.

[11] Nation of Neighbors. [Online]. Available: http://www.nationofneighbors.com

[SP]