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

Awalin Sopan

Problem and Motivation

Online social networks targeted towards 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 Nation of Neighbors (NON) [11], which is a platform for online neighborhood crime-watch portal for 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. 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.

Label	Node Count	activity_type	ActiveMonth	activity_date	Accept	Invites	Post	Reply	Report
Watch-Jefferson-Count	94		7		16	77	60	39	281
jefferson-county-sheriff	74		9		0	25	0	0	283
shannondale	61		7		24	68	12	16	75
mountain-watch	55		8		0	39	0	0	206
stockpolicedepartment	49		5		0	26	0	0	138
shannondale	37		1		0	0	0	0	117
norwoodmeadowbrook	34		5		2	3	0	0	109
laurelbrooke	33		4		28	40	29	7	8
duncanstincluding	26		5		10	10	47	58	5
Blue-Ridge-Acre	19		6		7	65	19	12	59
killianpines	17		4		5	48	23	17	13

Figure 1: ManyNets network table showing 11 communities (one per row) and a selection of the available metrics.

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, 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 aspects. Managers of online communities can take advantage of such visual analytic tool to analyze the activities of the community members, observe the evolution of the communities and make informed decision about their communities.

Background and 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 (Hollaback, SeeClickFix, Crimereports.com, SpotCrime etc.) to community organizers to overcome this shortcoming and new possibilities for neighborhood crime watch programs. Nation of Neighbors, however, has the unique and specific goal of reorganizing neighborhood watch to make it more effective. The system sends real-time email or text message alerts to the community members. 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 and Shneiderman [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 [12]. 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.

Approach and Uniqueness

While most network visualization tools can visualize just one community, ManyNets visualizes features of numbers of communities at a time. Also, its distribution column overview component enables us to explore the distributions of various features which can be very useful in community analysis. We used the visualizations to generate hypotheses and later tested the hypotheses to validate the usefulness of the visualizations. 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. We developed our metrics by close collaboration with social scientists and community manager.

Data preparation: We collected the activity log of NON community members from January 2005 to December 2011, 6370 activities from 230 communities in total. Activities were classified into 5 categories: report (describes an incident which occurred in the community), post (starting a discussion topic), reply (responding to a previously posted discussion), invitation (soliciting a person to join the community via email) and acceptance (new members joining the community following an email invitation). Two members have an edge connecting them (i.e. relationship) if one of them replies to other's post.

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.

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. The metrics are as follows:

<u>Interaction Intensity:</u> This is the total activity divided by total member-months.

$$I = \text{Interaction Intensity} = \sum (A) / \sum (MM)$$

<u>Average Active Months:</u> This is the average number of months the community members have participated.

$$\overline{M} = \text{Average Active Months} = \sum (MM) / \sum (N)$$

Where,

$\sum (A)$ = *Total activity* = $\sum$ *(reports + posts + replies + invitation sent + invitations accepted)*
$\sum (MM)$ = *Total member-months* = $\sum$ *(months since each Member registered)*
$\sum (N)$ = *Number of community members*

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 activity, joining date and our proposed Leadership metric. Analysts can also select particular members and create their ego networks 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).

$$Leadership = M_A - (standard\ deviation(M_A))\times 2 - M_M$$

Where,

M_A= *Total activity of a member* = $\sum$ *all activities by a member*
M_M = *Mean activity of all members*

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 columns 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 activity level remains high), trends and outliers (communities with anomalous activity patterns).

Temporal split of network: ManyNets splits a network into a series of sub-networks, 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 of this table visualizing the changes in activity pattern over time.

Results and Contributions

Our broader goal is to help the community managers understand why some communities succeed and some others die, what are the key factors to keep a community alive. Comparative analysis can answer these questions. Our visualizations and the metrics can produce interesting insights about the activity patterns, growth patterns, and the leadership in the communities. It can be extended to include other metrics based on the activity data. Our visualizations can cluster similar communities based on their activity type and activity distribution over time, it can also show the temporal evolution of a community, and identify key players in the communities and compare their activities and role in the community. The network diagrams show the interaction among the members and highlight the leaders. The following examples illustrate our results from the case study with Nation of Neighbors.

Identifying successful communities: We selected only the communities that have at least 5 Invitation activities and at least 5 active members and then sorted the communities by their Interaction value. Finally we manually reviewed communities that geographically overlapped and kept the largest one, thus having 44 active and independent communities suitable for comparative analysis.

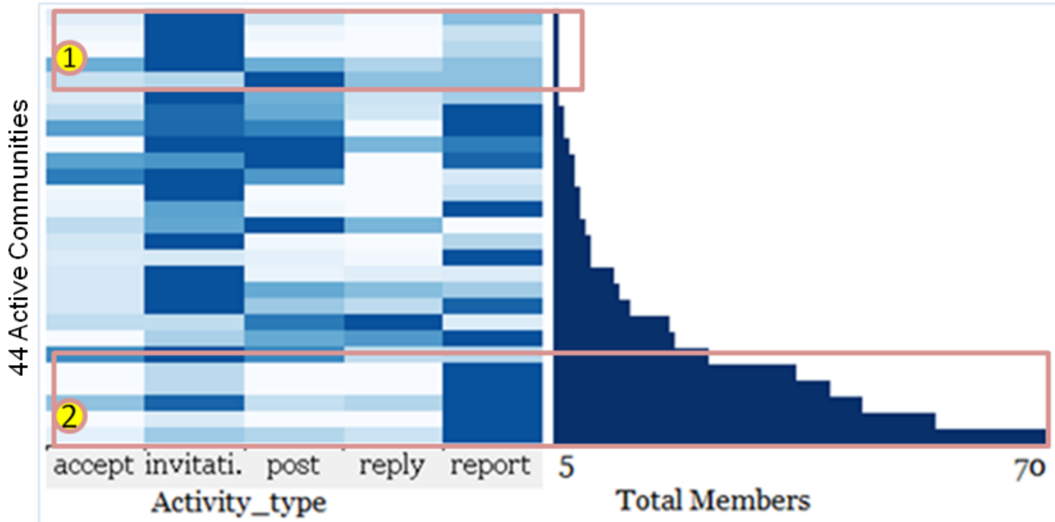

Figure 2: Side by side view of activity type distribution heatmap and total member for the active communities. Each band of heatmap is activity distribution of a community in the Activity_type column and the corresponding member count for that community.

Activity patterns: The histogram view in ManyNets of the activity type for all the 230 communities showed that invitation was the most common activity. But after filtering down to the 44 larger and active communities we observed more reports than invitations which contrast the general activity pattern. 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) 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 have more invitations; the heatmap column overview of 'Activity_type' distribution column [8] rendered each community's activity type distribution as heatmaps 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 and persistent communities we observed that, 16 of them had at least one leader (3 of them had 2 leaders; all others had 1) and none of these successful communities had any people from law enforcement,

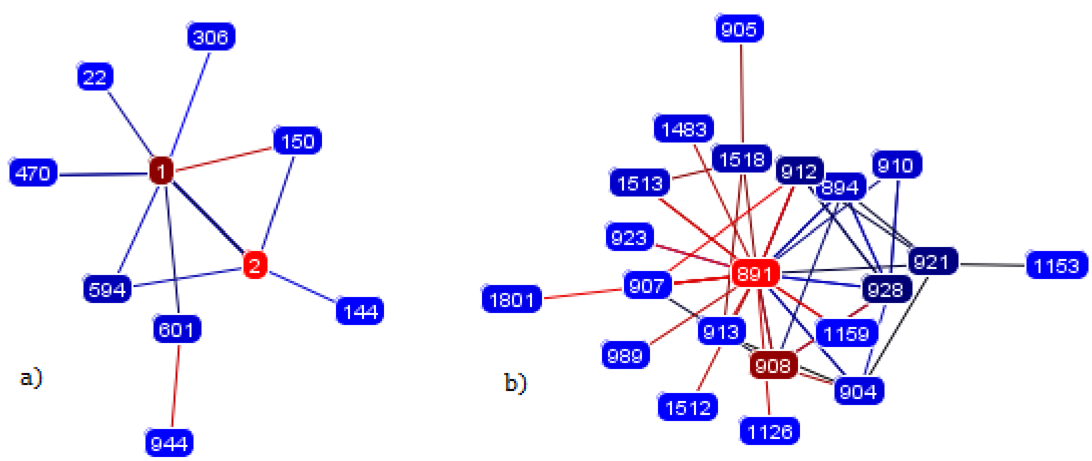

Figure 3: Ego networks of the leaders of a) Watch-Jefferson-County community and b) Duncans community. Dark red nodes are the leaders. Edges showing the post and reply activities.

indicating the involvement of law enforcement people was not a success indicator in those communities. From the distributions of 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. Only a small percentage of community members participated in reporting crime and invitations were mostly sent by the leader.

The 'Activity date' column showed that the temporal patterns in 4 communities 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 (dark red indicting the node with highest value for Leadership, blue being the lowest) (Figure 3). 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 10 nodes even though this was the largest community. In contrast, the "Duncuns" 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. A regression analysis also supported the hypothesis developed with the aid of ManyNets' visualization, i.e. the presence of super-active members strongly correlates with the growth of a community [7]. 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).

Label	Relationship activity_type	Accept	Invites	Post	Reply	Report	TotActivity
2009-07Jul		7	17	12	11	11	58
2009-08Aug		2	3	5	2	9	21
2009-09Sep		1	11	7	3	10	32
2009-10Oct		0	18	8	5	11	42
2009-11Nov		1	5	2	1	16	25
2009-12Dec		0	2	3	2	6	13
2010-01Jan		0	10	3	2	10	25
2010-02Feb		0	0	1	0	6	7
2010-03Mar		0	1	0	0	9	10
2010-04Apr		0	0	0	0	10	10
2010-05May		0	0	3	0	8	11

Figure 4: Temporal splits of the Watch-Jefferson-County community.
After January 2010, more reports are posted whereas number of invitations dropped.

Growth pattern of a community: After sampling the communities, the "Watch-Jefferson-County" community appeared to be the most active one, 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. Initially there were different types of activities but gradually the proportion of reports grew larger while no new 'Invites' and 'Accept' occurred recently. As more people became involved in the community, some would say that it reached critical mass. 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 hypothesis in this case study; however, only 12 communities in the whole dataset had law enforcement officials involved with them so there might just be too few cases for adequate analysis.

Conclusion: This work shows how a visual analytic tool, ManyNets, can help community managers generate hypotheses about community behavior and identify successful cases. When they have thousands of activities from hundreds of communities over several months, a visual analytics tool like ours, can help them understand their communities better without having deep knowledge of statistics. 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".* 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.

REFERENCES

1. Butler, B. S. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research, 12,* 4 (2001), 346-362.
2. Freire, M., Plaisant, C., Shneiderman, B. and Golbeck, J. ManyNets: an interface for multiple network analysis and visualization. In *Proc. Proceeding of the 28th annual SIGCHI conference on Human factors in computing systems* (2010), 213-222.
3. Hansen, D., Shneiderman, B. and Smith, M. Visualizing Threaded Conversation Networks: Mining Message Boards and Email Lists for Actionable Insights. In A. An, P. Lingras, S. Petty and R. Huang (eds.) *Active Media Technology, 6335*, (2010), 47-62.
4. Holloway, K., Bennett, T. and Farrington, D. P. 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. Lindsay, B. and McGillis, D. Citywide community crime prevention: An assessment of the Seattle program. *Community crime prevention: Does it work* (1986), 46-67.
6. Preece, J. Sociability and usability in online communities: determining and measuring success. *Behaviour and Information Technology, 20,* 5 (2001), 347-356.
7. Rey, P. J., Sopan, A., Ahn, J., Kang, J. H., Plaisant, C. and Shneiderman, B. The influence of Leadership on Activity Levels on an Online Crime Watch Community. Technical report, submitted to *Proc. American Sociology Association Annual Meeting* (2012).
8. Sopan, A., Freire, M., Taieb-Maimon, M., Plaisant, C., Golbeck, J., Shneiderman, B.: Exploring Data Distributions: Design and Evaluation. *International Journal of HCI* (2012).
9. Trier, M. Towards Dynamic Visualization for Understanding Evolution of Digital Communication Networks. *Information Systems Research, 19,* 3 (2008), 335-350.
10. Xu, J. and Chen, H. Criminal network analysis and visualization. *Commun. ACM, 48* (June 2005), 100-107.
11. http://www.nationofneighbors.com
12. Ahn, J., Maimon, M., Sopan S., Plaisant C., and Shneiderman B. (2011) Temporal Visualization of Social Network Dynamics: Prototypes for Nation of Neighbors. 2011 *International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction (SBP11).*

Author information: Awalin Sopan is from Department of Computer Science, University of Maryland.

Advisor: Ben Shneiderman.

Acknowledgement: PJ Rey, Catherine Plaisant, Jae-wook Ahn.

Sponsor: NSF.