Measuring and Evaluating the Underlying Social Networks of Visual Blog Communities

Jeffrey Towle

School of Information Management and Systems
University of California, Berkeley
202 South Hall
Berkeley, CA 94720

jtowle@sims.berkeley.edu

ABSTRACT

Measuring and evaluating the social networks that drive new or unique communities is an important step in gaining full understanding of how and why they operate the way they do. This paper proposes a method of collecting and evaluating the social networks of Visual Blog Communities (VBCs). These communities generate large amounts of image data that is difficult to process. However, they also present this image data in a way that embeds social information in the layout. The proposed method for deconstructing this embedded information has two steps. The first is to programmatically collect information about the community and the individuals that comprise its membership. The second is to evaluate the collected data through a lens that utilizes the graphical layout and presentation of the data in order to extract information about social hierarchy, participation levels and individual roles. The end result is a technique that utilizes the visual language of a community web site to inform the development of a social network model of its participants.

Keywords

Visual Blog Communities, Social Network Analysis

1. Introduction

Discovering and investigating social networks in communities where new or novel communication methods are utilized is a crucial step in understanding how these communities work and how these means of communication help to facilitate social ties and reinforce group membership.

In traditional social network analysis, the social network data was collected through ethnographic means [3]. This is in part because traditionally, social network analysis has focused on explicit actions and transactions between people. For example, when one person lends money to another, a rational assumption may be made that the two people involved in the transaction have a strong social tie [Fischer]. Strong social and familial ties are easily discovered using these methods.

Social network information also can be collected from a variety of data sources that are very efficient to use, such as e-mail address books or instant messenger "buddy lists"; however, these sources tend to be private. One way around this is to ask people about their social ties.

The current trend uniting social networking and software has centered on explicit, articulated social networks. Articulated social networking software such as Friendster, Tribe.net or Orkut relies on the user to explicitly express ties in the system [1]. Some systems explicitly ask the user to quantify the strength of a tie based on a scale from one to ten. This is in contrast to the transactional weights that Fischer or Granovetter described in the networks that they studied [2, 4]. Although these systems rely on explicit social information, they are not the only way in which social networking can be utilized by software.

In cases where such explicit information about social ties is not available or easily collectable, implicit social network information derived from a variety of sources can be a helpful tool for studying other types of communication technologies. This paper focuses on the application of this kind of social network analysis to visual blog communities (VBCs).

VBCs are small, web-based communities that are distinguished by their members' ability to communicate visually, through pictures. In addition to a group weblog or discussion forums, each VBC has a "webcam portal," where members are assigned a space to post an image. These pages typically contain around 12 images, with additional pages allocated for overflow. Some VBCs have more than ten pages of images. The images are primarily self-portraits, many with captions (Figure 1) [6].

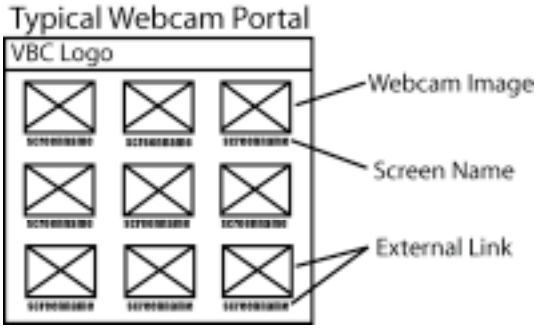

Figure 1 - The typical VBC portal is arranged as shown. Images are arranged from the upper-left to the lower right. Additional images are displayed on additional pages. External links may be activated by clicking on either the image or the screen name, usually sending a user to the image owner's home page.

This information is a prime candidate for automated collection and analysis because it publicly available on the web and is stored in a form that is easily read and processed by machines (HTML). Automated analysis of web sites has been used in the past to evaluate usability based upon site architecture, page layout and formatting [5]. The proposed automated social network discovery and analysis method aims to utilize the same visual and structural cues together with data that uniquely identifies individuals to collect social relationship information that is embedded in the community's web page.

The need for an automated method of collecting and aggregating social data based upon explicit relationships was identified in a pilot study where data was collected involving community composition and the similarities of composition across community.

2. Pilot Study

In pilot research, 752 unique participants were identified across 38 community sites. This analysis covered a four-month period in 2003. For each individual, the user's screen name, any aliases used on different sites, and the user's domain name was recorded. Participation on multiple sites was identified visually, making use of the fact that most participants post pictures of themselves in their images. The process was time consuming, as many participants update their images before the data could be collected, and it was necessary to re-check the data for many of the participants.

The amount of data to process is enormous. Every month, these communities generate gigabytes of images and processing it by hand is not feasible. The amount of social data embedded within the larger data set is quite manageable, but the task of extracting this data is difficult if we do not know how to filter out the irrelevant data.

The major shortcoming of the method used in the pilot study is that while information about community association was collected, no information is collected about the individual's standing in the community. Furthermore, no information is collected regarding changes to the community over time. If a participant were to leave a community, no record would be kept. It is too time consuming, and the amount of data is overwhelming to collect and process by hand. These limitations prevent us from studying or even identifying some types of events, such as change in the social structure or the effects of conflict amongst community members.

A more robust and efficient method is needed in order to accurately model the community composition and its changes over time, as well as the individual's relationships to one another within and across the communities.

3. Proposed Method

The following is a proposal and discussion of a method for automating the collection and evaluation of social network data in the context of VBCs. The ultimate goal is to mine a community's structural and textual content in order to collect information about the underlying community and social network. Though automated collection can often yield more data than can possibly be processed [3], the proposed technique aims to solve this problem by focusing on collecting a targeted subset of the total information generated by the community.

The proposed technique requires two sets of data in order to build a complete model. The first is a set of raw quantitative data that describes the individuals, communities and the explicit links between them. The second is the set of rules that determines how the quantitative data will be used. It is necessary to learn the language of the community and the visual language used by its members so that we can build a model that shows how social hierarchy and relationships are represented by a combination of visual and textual cues. This data must be collected through ethnographic study. In the case of VBCs, this data has been collected through observation of VBCs over time, reading the posted community rules and a content-analysis study [6].

4. Reading the Social Data

There are many ways in which the social structure of the community is reflected in the physical layout of the images across the pages. In the typical VBC, one can learn a lot about the community merely by looking at the "webcam portal." For example, in the typical VBC, the site maintainer will post his or her image in the upper-left of the first page of webcams. Nearby slots are owned by people who have a very high participation level in the community. Participation is very broadly defined; one of the posted rules for the webcams on a VBC reads: "*To get spots on higher pages, you can do several things such as, but not limited to: participation in cam themes, forum posting, fan art, quality of your cams, sending us stuff, links to [this site] you put on your web site, etc*."

The coordinate position of a participant's image in the portal can be recorded as the page, row and column. Although this data is important on its own, it is changes to the layout that provide information about the current state of participation. Position swapping is especially interesting for studying the relationships between individuals. Because the top positions in most portals are assigned to individuals with administrative duties in the community, such as coding or paying the bills, any change in those positions likely indicates a major change in the organizational structure. For example, if someone in a top position moves to a less prestigious slot, it may mean that the person is distancing from the community and moving on to other ventures.

There are levels of data that may be mapped: the community itself, including the individual's roles within that community, and the underlying social network. The first step is to identify the community, its participants, and their roles within that community. Only after those data points have been established will it be possible to dig deeper and uncover the underlying social network.

5. Data Collection

This technique will require constant observation of the community portals. Specifically, an implementation of the technique will need to detect changes to the portal order, community membership (additions, deletions), and individual cam updates. Such a level of observation will require daily crawling of the VBC network, based on the level of resolution determined by McDonald and Towle [6].

The data needed to model VBC social networks can be retrieved from the web by a spider or crawler. Collection should take place in two phases: discovering new communities and monitoring found communities. Once a new community has been found, it will be added to the list of sites to monitor. This crawler will need to run often enough to capture image updates and changes to the page layout.

5.1 Traversing the Network

One of the most difficult tasks in mapping the social network of VBCs is discovering new communities in the greater community of practice. However, based on the preliminary analysis, it is very likely that any given participant hosts (and links to) his or her own web site. Individuals will often place a list of links to other VBCs of which he or is a participant.

It is also common for participants to host a sub-VBC on their own sites. Such a VBC will consist of a small subset of the participants on the larger community. Such sub-communities will be immensely valuable in determining the strength of relationships between participants.

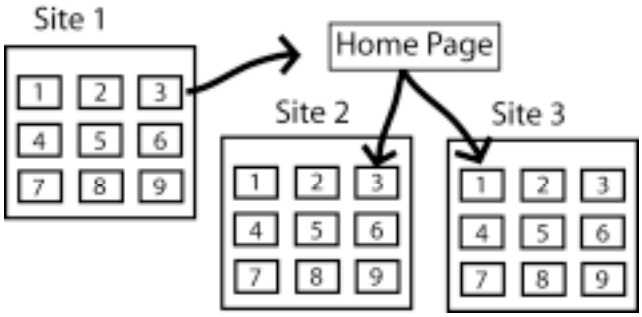

Figure 2 – The procedure for discovering new communities takes advantage of participants' tendency to link their images to their personal sites, which often link to other VBCs

5.2 Uniquely Identifying Participants

Another challenge in data collection is identifying distinct individuals in the communities. In a typical VBC, each participant is allowed three pieces of identifying data: an image URL, a screen name (or username) and a link URL. The image URL is where the participant's webcam image is located. Participants typically host the images on their own servers, and provide the VBC site maintainer with the address. This allows a VBC to link to a single image URL which is updated by the participant on his or her own site, which relieves the site maintainer of the responsibility of setting up a complex file-upload system. This also means that a participant may update his or her image once, and have the change reflected across any number of sites.

The participant's screen name is typically displayed below the image on the webcam grid. Screen names are unique within a site. However this means that when one participates in multiple sites, it may be possible for the screen name to be used by someone else, requiring a change for the newcomer. Screen names are unique to a user within a site, but do not uniquely identify users across sites.

A viewer may activate the link URL provided by the participant when either the image or the screen name is clicked on. URL links typically point to the participant's E-mail address, personal web site, or even to that person's instant messenger name.

In the pilot survey, ample evidence was collected to show that the image URL was the best unique identifier for an individual. In the set of 752 participants, only one participant used multiple image URLs. The proposed method will key individuals on their image URL based on that finding.

5.3 Describing a Community

Several quantitative metrics can be used to describe a community: number of pages, number of participants per page, number of unique participants, distribution of update frequencies, among others. Obviously, these communities have a richer set of descriptors; however descriptors such as community interests and discussion topics are secondary to the task of modeling the community interactions and social networks.

5.4 Describing an Individual

It is important to study both the unique individual and the individual in the context of a community, as individuals typically do not participate in only a single webcam portal. There are a number of different data items that can be collected for any one individual in a community: screen name, frequency of updates, image URL, and link URL (home page or email). These three items describe a unique individual across sites.

One quantifier for describing an individual is the frequency of updates. Some individuals update more than twice a day and some update so infrequently that they can be removed from a community's portal. The frequency of update is a good measure of how active the individual is in VBCs as a whole. Even if an individual is not specifically participating in any one community, a high update frequency appears to be well respected by community members.

An individual may submit his or her image URL to many different community portals. The number of sites on which an individual's image appears is an indicator of how well connected the individual is in the various sub-communities of VBC culture. These ties are by no means strong, but the individual is likely known throughout the VBC culture.

Another metric for computing the likelihood of a social tie is the number of shared contacts in an individual's communities. The more homogeneous the set of shared contacts, the higher the likelihood of a social tie.

An individual's position on a site is an indicator of the person's role on the site, and may provide clues as to which site is considered the "primary" site for that user. An individual whose page, row and column position is greater on one site than another may be a more important member of the first community, indicating a stronger level of overall participation and perhaps an important functional role on the site (administrator, coder, participant).

Personal web sites are also important pieces of data. Not only do they aid in identifying unique individuals, but also frequently provide a view on an individuals preferred links. It becomes a good resource for discovering other aspects of a VBC participant's social network (Figure 2).

6. Interpreting the Data

One of the key challenges in interpreting this data is trying to determine how low-level actions map into high-level social relationships. The types of potentially meaningful relationships are nearly innumerable. The following are some examples of these relationships:

> *Reciprocity* – Some individuals participate in "link-trading," which is an "*I'll put you on my site if you put me on your site*" agreement. In such a situation, a social exchange has clearly taken place. Identifying the link is one thing; interpreting it is another.
>
> *Linking* – Many individuals have a section of links on their personal web sites that includes links that are quite related to their participation in VBCs. Many link to each of the VBCs on which they participate. Others link to the home pages of fellow VBC members. This "behind-the-scenes" linking reveals much about the underlying social structure.
>
> *Neighbors* – Two individuals whose images are proximate in the site layout (such as image "neighbors") may have a

stronger tie than individuals whose images are farther apart in the layout. Page co-presence is another potential tie indicative of a higher-level relationship.

Site Maintainer/Site Participant – A special relationship exists between the site maintainer and the community's participants. In this case, there is clearly a power relationship.

Donations – One way that individuals participate in a community is by donating time or money towards the operation of the community's online presence. Monetary donations are important for keeping the server online, and those who donate typically have the respect of those in their community.

Face-to-Face – It is not uncommon to see husbands and wives, roommates, or face-to-face friends in these communities. However, while such information would be incredibly rich and invaluable for a complete analysis, it is beyond the scope of this project to attempt to collect definitive evidence of such relationships. The system may identify that there is a strong relationship between two members of a community, but it will not attempt to assign a type to the inferred tie.

The end goal is to create a model that translates the low-level, functional relationships into more meaningful social relationships, such as the ones listed above. Such a model or visualization needs to reflect the values that indicate both weak and strong ties within a community. Any interpretation of the results needs to take into account the values of the community from which the data was collected. In a community that highly values link trading, the social weight of a link exchange will be higher and should be reflected as such. It is unlikely that all VBCs have homogeneous values, so this translation is not likely something that can be automated for all communities.

7. Applications

The primary application of this social network collection tool is mapping the communities across the web and the relationships between them. Cross-site comparisons could yield interesting results. Because each site is a small community, we can see how social relations translate across distinct communities. The tool will also be able to detect communities that are merely "mirrors" of other communities, enabling us to determine which communities have the most original content.

The tool may also be used for charting the spread of webcam portals on the Internet. As new users come along, we can use this technique to track the spread of the technology and observe its use as it expands to new communities. Based on preliminary observation, it appears that individuals that join a VBC and actively post images will inevitably start their own VBC. The proposed system would be capable of testing this hypothesis by charting the communities over time.

Another application involves determining the level of interaction that goes on between members of the communities. It also helps to determine which communities are stronger in the various metrics of interaction. For example, link trading may cause a community to appear well linked while the actual participant interaction may be quite low.

The tool will also allow us to identify and describe anomalous communities. Such a community may not follow the norms used by other communities. For example, one group organizes their portal by the age of an image—the newest image posted immediately goes to the top of the portal. If such a community is found, the social network model may help describe how the use of the technology changed in this particular instance.

The data that this tool provides will be a useful resource when designing qualitative studies of the VBC participants and communities. It informs the design of interview protocols and surveys by giving hints as to which questions to ask and who should be asked. Interviewees can be asked to explain certain phenomenon and their answers may be used to validate or correct the model that is produced.

8. Conclusion

The proposed data collection and analysis is designed to answer some high level questions regarding VBCs. Because the data set is so large and rapidly changing, this process needs to be automated. However, the design of an automated data collection tool should not be without the input of ethnographic study, or at minimum, manual observation.

In these communities, the images are the most important language. Even if it were possible to program a computer to interpret the content of these images, without the community context, the meaning is lost. At this stage, the best we can do is to collect every bit of social information that is published by the community. This information is harder to capture when it contains very little text. The communication does not stop at the borders of the images and we can utilize the visual environment that a community has developed for itself as a way of learning about its social structure. Data collection is as simple as interpreting the cues that can be found in communities' online structures, layouts, use patterns and growth patterns—all of which are embedded in the environment that the communities have defined for themselves.

9. References

[1] boyd, d. Friendster and Publicly Articulated Social Networks. *Conference on Human Factors and Computing Systems* (CHI 2004). Vienna: ACM, April 14-29, 2004.

[2] Fischer, C. *To Dwell Among Friends: Personal Networks in Town and City*. University of Chicago, 1982.

[3] Garton, L., Haythornthwaite, C., and Wellman, B. Studying Online Social Networks. In Jones, S., Ed., *Doing Internet Research,* pp. 75-105. Thousand Oaks, CA: Sage, 1999.

[4] Granovetter, M. The strength of weak ties. *American Journal of Sociology*, 78(6), 1360-1380 (1973).

[5] Ivory, M. and Hearst, M. Statistical Profiles of Highly-Rated Web Sites. *Conference on Human Factors and Computing Systems* (CHI 2002). Minneapolis: ACM, April 20-25, 2004.

[6] McDonald, D. and Towle, J. Community through Pictures: Visual Blog Communities. Forthcoming in *Communications of the ACM*, 47, 2004