

Finding "the life between buildings": An approach for defining a weblog community

Lilia Efimova

Telematica Instituut, The Netherlands

lilia.efimova@telin.nl

Stephanie Hendrick

The Department of Modern Languages-Umeå University, Sweden

stephanie.hendrick@engelska.umu.se

Anjo Anjewierden

Human-Computer Studies laboratory, University of Amsterdam, The Netherlands

anjo@science.uva.nl

Abstract. Although weblogs are perceived as low-threshold tools to publish online, empowering individual expression in public, there is growing evidence of social structures evolving around weblogs and their influence on norms and practices of blogging. Emerging from connections between weblogs and their authors, weblog communities often do not have a shared space, clear boundaries, or clear membership, challenging researchers who want to study them. Initially intended to be a study in the delineation of weblog boundaries, the scope of defining these boundaries immediately overwhelmed traditional methods and tools. The problems that arose from using traditional link mining methods led to an exploration of alternative approaches of defining these communities. The purpose of this paper is to get an insight into methods of finding "life between buildings": virtual settlements where weblog communities may reside. We use Jones' (1997) theory of a virtual settlement and archaeological metaphor to address research challenges of locating weblog communities, suggest an iterative approach that includes refinement of research methods based on assumptions about community norms, practices and artefacts, and propose which artefacts could serve as indicators of community presence. Finally we present a pilot study which explores different methods of identifying community membership beginning with a known core member of a group of knowledge management bloggers.

Introduction

Weblogs (also called *blogs*) are not easy to define in one sentence. Most authors agree that a weblog is "a frequently updated website consisting of dated entries arranged in reverse chronological order" (Walker, 2003), but then go on to discuss specific characteristics that make weblogs different from other forms of webpublishing (e.g. Winer, 2003). The difficulty of defining weblogs has something to do with the fact that their authors have different goals, uses, or writing styles with only one thing in common: format. Weblogs are often perceived as low-threshold tools to publish online, empowering individual expression in public. They are seen as a form of individualistic expression, providing "personal protected space" where a weblog author can communicate with others while retaining control (Gumbrecht, 2004).

However, what makes weblogs different is not the publication of content *per se*, but the personalities behind them. Weblogs are increasingly becoming the *online identities* of their authors. Most weblogs are not formal, faceless, corporate sites or news sources: they are authored by individuals (known as webloggers or bloggers), and perceived as 'unedited personal voices' (Winer, 2003).

There is an ongoing debate in the weblog research community over how interactive an average weblog actually is (e.g. Herring et al., 2005). For example, a randomly selected weblog shows limited interactivity and seldom links to other weblogs (Herring, Scheidt, Bonus, & Wright, 2004). Nevertheless, there are indications of the development of weblog conversational practices (e.g. Efimova & de Moor, 2005) and community-forming effects of blogging (e.g. Merelo-Geurvós, Prieto, Rateb, & Tricas, 2004).

Moreover, there is growing evidence of social structures evolving around weblogs. This evidence ranges from voices of bloggers themselves speaking about social effects of blogging (e.g. Mehta, 2004), to studies on specific weblog communities with distinct cultures (e.g. knitting community in Wei, 2004, or goth community in Hodgkinson, 2004), to mathematical analysis of links between weblogs indicating that community formation in the blogosphere is not a random process, but an indication of shared interests binding bloggers together (Kumar, Novak, Raghaven, & Tomkins, 2003)

In comparison with many other types of online communities (e.g. those forming around forums or chats rooms), weblog communities emerge from connections between weblogs and their authors, and not around a single shared space. From one side emergence of weblog communities is paradoxical: in that the existence of highly personal spaces guarded by individuals results in emergence of social structures. From the other side it does not seem strange as human beings are quite used to living in cities where private houses and public areas merge into a whole, providing spaces for both individuals and communities.

The metaphor of the city becomes especially appropriate to weblog analysis with the application of the distinction between *virtual settlement* and *virtual community* as proposed by Jones (Jones, 1997). Jones argues against equating virtual communities with the cyber-places (e.g. IRC channel or web-based forum) they inhabit. He compares virtual community research to archaeology and suggests studying a community through artefacts of its virtual settlement.

According to Jones, a study of a virtual community should start from defining and characterising its virtual settlement. He proposes a set of conditions that a cyber-space and its associated group computer-mediated communication (CMC) must display in order to be considered a virtual settlement. These conditions, however, do not indicate clearly how to define the settlement's boundaries (Liu, 1999). This may not be a problem for studies on online communities forming around a shared space or group communication technology, but it definitely is a challenge for the research on communities supported by weblogs.

An individual weblog is not likely to represent a community¹, while shared social spaces seem to emerge between weblogs, like in a city where life between buildings accounts for many social activities of its inhabitants (Gehl, 2001). As in cities, *blogger communal spaces* are not evenly distributed: some neighbourhoods are full of social activities and conversations, while others look like a random collocation of houses where inhabitants have nothing in common. Blogger communal spaces may have visible boundaries (e.g. NetRing for Knitting Blogger community, Wei, 2004), but more often indicators of a community are subtle and difficult for a non-member to distinguish. Just as a local garden is not likely to have a sign indicating that there is a chess-player community that inhabits it every Sunday, blog communities do not delineate obvious community boundaries. It is here that the researcher faces very specific problems: where to start looking for a weblog community and which indicators to look for.

The purpose of this paper is to get an insight into methods of finding "life between buildings": virtual settlements where weblog communities may reside. We suggest an iterative approach that includes refinement of research methods based on assumptions about community norms, practices and artefacts, and propose which artefacts could serve as indicators of a community presence. This paper goes on to present a pilot study which explores different methods of identifying community membership beginning with a known core member of a self-defined group of knowledge management bloggers. We conclude with a discussion of the benefits and drawbacks of the various methods tested, as well suggestions for further research.

Finding life between buildings

Jones (1997) suggests studying a community through artefacts of its virtual settlement and proposes a set of conditions to delineate it: (1) a minimum level of interactivity; (2) a variety of communicators; (3) a virtual common-public-space where a significant portion of community interactions occur; (4) minimum level of sustained stable membership. Those conditions presuppose that there is a cyber-space with some kind of group interaction that is studied to identify if there is a community. Figure 1 provides a simplified image of the process.

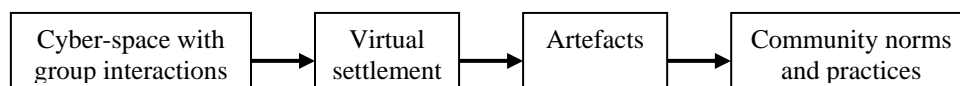


Figure 1. Studying virtual communities

¹ Although communities emerging around popular weblogs, e.g. (Blanchard, 2004), are an interesting target for research.

In the case of weblog communities it is usually difficult to find a single shared space as a point of departure, so we suggest an alternative approach (Figure 2). One can start from weak clues indicating likelihood of community presence and assumptions about norms and practices in it. These would indicate where to start data collection, which weblogs to include in the analysis, what artefacts to look for and how to interpret them in order to define boundaries of the community virtual settlement. While the virtual settlement is being uncovered, researchers could encounter new types of artefacts, develop a better understanding of community norms and practices, and adjust methods for data collection and interpretation.

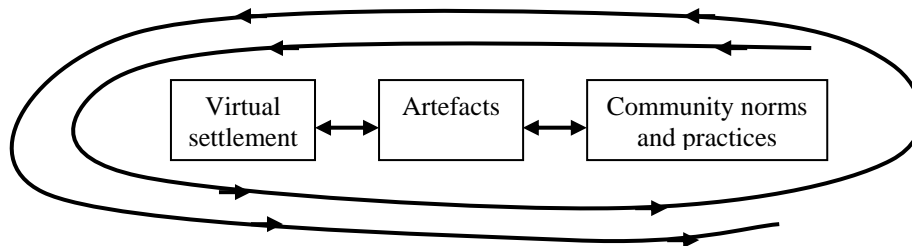


Figure 2. Studying weblog communities

In the following section we identify a set of artefacts that can be used as indicators when defining boundaries of a weblog community.

Community indicators

Although a weblog is a personal writing space, its public nature suggests a need to communicate (Mortensen, 2004) and creates conditions for contact with others by making a blogger present. From this perspective blogging is similar to life between buildings in a real city:

Life between buildings offers an opportunity to be with others in a relaxed and undemanding way. One can take occasional walks, perhaps make a detour along a main street on the way home or pause at an inviting bench near a front door to be among people for a short while. One can take a long bus ride every day, as many retired people have been found to do in large cities. Or one can do daily shopping, even though it practical to do it once a week. Even looking out of the window now and then, if one is fortunate enough to have something to look at, can be rewarding. Being among others, seeing and hearing others, receiving impulses from others, imply positive experiences, alternatives to being alone. One is not necessarily with a specific person, but one is, nevertheless, with others.

As opposed to being a passive observer of other people's experiences on television or video or film, in public spaces the individual himself is present, participating in a modest way, but most definitely participating (Gehl, 2001).

In real life, such low-intensity contact is a starting point for growth of higher intensity interactions or a way of maintaining already established relations (Gehl, 2001). In online communities it is known as *lurking* and rarely considered important, although it is a form of *legitimate peripheral participation* (Lave & Wenger, 1991), a way of becoming a community member through learning about others and the community (Nonnecke & Preece, 2003).

Defining a virtual settlement in a case of shared cyber-space with group interactions do not have to rely on analysing lurking:

A group of "lurkers" who do not communicate can not be called a community. For a group of individuals to qualify as a community, these individuals have to communicate and interact. A lurker may be

considered as part of a virtual community only if the community exists, and the existence of a virtual community is not defined by the presence of lurkers, but by those who do communicate and interact. (Liu, 1999, discussing methods of studying IRC community).

This is not the case of a weblog community, as it is not known in advance what is the "core" interacting group and where are the lurkers. In this case looking for indicators of low-intensity forms of contact (e.g. patterns in weblog reading) may help to locate strong membership and active interactions as finding a crowd of spectators on the street is likely to indicate a social event.

Based on reflection on existing research on weblogs and our personal experiences of blogging we suggest that the following artefacts can be used as indicators of the existence of weblog community.

Meme paths. Ways ideas or references to interesting sources "travel" around weblogs are influenced by social structures behind them – relations between Experts, Link Mavens and Connectors (Hiler, 2002). Mapping those paths (e.g. as in Adar, Zhang, Adamic, & Lukose, 2004; Halavais, 2004b) and especially patterns over multiple meme paths can indicate well connected or influential weblogs as a starting point for an analysis.

Weblog reading patterns. Regular reading of a weblog is a way to start establishing a relation or maintain an existing one (Efimova, 2004b), so analysing weblog reading lists (e.g. in the form of blogrolls, RSS subscriptions, shared Bloglines accounts) and especially relations between reading lists for a group of people (e.g. as in Dennis & Jarret, 2005) can indicate presence of communal ties.

Linking patterns. In many cases, links are not just pointers to additional information, but also "currency of the web" (Walker, 2003) that helps to improve visibility of a page being linked to or, especially in a context of weblogs, signs of value and personal recommendation (see Mortensen & Walker, 2002) for the analogy between linking in weblogs and references in scholarly publications). Links provide an easy way to identify connections between weblogs and used in many studies (e.g. Merelo-Geurvós et al., 2004; Herring et al., 2005), although one should be careful with assigning a particular meaning to a link. For example, linking to someone's weblog in a blogroll does not necessarily mean that links in weblog text would refer to it as well (Marlow, 2004) and even links in the text can have a variety of uses that do not necessarily indicate relations between bloggers (Efimova & de Moor, 2005; Halavais, 2004a).

Weblog conversations. A weblog conversation emerges when a weblog post triggers feedback from others, either using comments to the original post or replies in other weblogs linking to it. Although weblog conversations are difficult to locate intentionally (Efimova & de Moor, 2005) discovering one could be a strong indicator of a shared context and system of relations that enables it (Herring et al., 2005).

Indicators of events. Mentions of face-to-face meetings or participation in online events can indicate existing relations between bloggers. Next to it, participation in events, especially in face-to-face meetings can help establishing new connections, strengthen existing relations and support development of a community sense.

"Tribe" marks, group spaces and blogger directories. Weblogs may display indicators of belonging to a particular community, usually in a form of linking to some kind of blogger directory (e.g. webring of bloggers on a topic as in Wei, 2004). Connection could also be maintained by adding a weblog or some weblog posts to a channel dynamically aggregating weblog content on a specific topic (Paquet & Pearson,

2004). However, topic channels and blogger directories can be maintained without blogger participation as well as driven by an interest of a third party to keep track of a list of interesting weblogs or aggregate their content.

Pilot study

Many factors come into play when discovering a virtual community. Just as an archaeologist must have a hunch about where to begin digging, a starting point must be found in order to begin unfolding the layers of community. Once clues about community membership begin to surface, the metaphor of the archaeologist then becomes one of an ethnographer in search of an appropriate interpretation of the data, vital to uncovering potential directions for further excavation. Equally important, however, when uncovering a virtual community is where to end your exploration. Too close to the starting point and your patterns are too limited in scope to be of any significance, too far and you have ventured into neighbouring communities.

The chance of discovering a weblog community depends on a subset of weblogs selected for analysis, given that weblog social structures can vary between free-floating weblogs, "A-list" weblogs with many incoming connections, or dense clusters of reciprocally linked weblog clusters (Herring et al., 2005). It may also depend on time interval selected for the analysis. For example, the analysis of linking patterns between 25K weblogs over time (Kumar et al., 2003) not only found growth of local-scale community structures, but also bursty periods of linking within those communities. Uneven structure of the blogosphere and different sampling methods may explain differences between studies that found limited interactivity of weblogs and limited linking between them (e.g. Herring et al., 2004; Nardi, Schiano, & Gumbrecht, 2004), and those that indicate close community ties and complex weblog conversations (e.g. Efimova & de Moor, 2005; Nilsson, 2003), as well as differences in representation of gender and age of weblog authors in mainstream press and those characteristics for a random weblog (Herring, Kouper, Scheidt, & Wright, 2004).

Because a randomly selected weblog is not likely to be well connected with other weblogs (Herring et al., 2004) and data collection is resource intensive, researchers looking for weblog communities may need to look for clues that indicate likelihood of the existence of community structures before making decisions on sampling and data collection methods.

For the purposes of this paper the decision was made to focus on the cluster of weblogs focusing on knowledge management and social software. In this case we could safely assume that there is a community around those blogs as we had first-hand experiences with it. The first author of the paper is an active member of the community and explores it in the longitudinal study for her PhD research, other two authors had interactions with the community members on various occasions.

This pilot study represents a dense social network of weblog authors with many bonds and interactions between participants. Many of them are aware of each other: they appear on several KM weblog lists², they link to each other in blogrolls and read each others blogs. Participants engage in multiple weblog conversations over time, pick-up ideas and

² Among many others KM bloggers list at <http://www.voght.com/cgi-bin/pywiki?KMBlogger>

practices from each other (e.g. adoption of Skype), and employ a variety of media to communicate (Efimova & de Moor, 2005). Although most of the contacts between community members were first established via weblogs, many of them participated in a number of events organised by other community members³.

Since we knew that a community existed and had numerous indicators to play with, our goal was to develop a method that would delineate boundaries of the community (or at least to describe it in an objective way). We hoped that by using a known community as a playground we would be able to define a method that could be used to define weblog communities from an outsider perspective. At the moment of writing this paper we can't say that "*the method*" has been found, but we feel that we have discovered relevant components of it and that our experiments and lessons learnt on route would be of interest to others.

Relying on links in the text

Although a number of different indicators of the community presence existed, the decision was made to start from linking patterns as these provide a way to find community members based on their interactions and it is a commonly used approach in weblog studies (e.g. Herring et al., 2005; Kumar, Novak, Raghaven, & Tomkins, 2004; Halavais, 2004a; Merelo-Geurvós et al., 2004). Analysing blogrolls can be a good way to map the network as they are used by bloggers to self identify their close connections, however, we decided against using blogrolls as a starting point as many members of this community rely on RSS readers for reading other weblogs and do not have blogrolls or do not update them regularly⁴. In addition, analysis of links in weblog text lessens the chance that a weblog author will identify him or herself in a network that they do not communicate and/or participate in. Bloggers are often surprised upon rereading their archives that the topics they would associate with their weblog are not necessarily the topics that they link to and about, see (Blood, 2002: 30 for an example).

Analysing links in weblog posts provides a good insight into interactions between bloggers, however using them as a starting point to define a list of weblogs for analysis is a challenge given that this data is not readily available⁵. Our first iteration involved manual data collection that involved copying text from seven months weblog archives and processing them to extract links. The weblogs included in the analysis were identified using a snowball sample starting from blog.mathemagenic.com using three links during the analysis period as a threshold to exclude occasional reference to weblogs that do not involve regular interaction. The resulting data was analysed using the social network analysis program Pajek and documented in the report (Efimova & Hendrick, 2004).

Apart from it being obvious that any serious work of this kind should rely on a bigger data sample (limiting the extent to which manual data collection could be used), the analysis yielded an interesting example of impact that artefacts chosen for the data collection have on the resulting view of the community.

³ For example, BlogWalk series (<http://blogwalk.interdependent.biz>).

⁴ The following post of one of the community members provides an example of this attitude - <http://www.mcgeesmusings.net/2003/10/04.html#a3711>

⁵ More details: <http://blog.mathemagenic.com/2004/11/22.html#a1438> or <http://blog.mathemagenic.com/2005/08/22.html#a1641>

Figure 3 represents an anomaly found in the initial data. The flower shaped formation at the top of the figure is what, at the outset, was thought to be a linklog, a weblog which links heavily to places of interest on the Internet. When going back to the weblog to check if this was so, it was discovered that the blogger in question had very long texts, and very few links; the opposite of what a linklog contains. It turns out that the comments had been captured in the gathering of archives for this particular blogger. On the average, weblogs in this network have about 12-30 links to other weblogs during the seven-month period. This particular blog contained 147 links. When comments were removed from the data, however, it only had 18 links to other blogs in the main posting area. The great number of unique links (139) removed indicates that this weblog has a high rate of participation in the comment area.

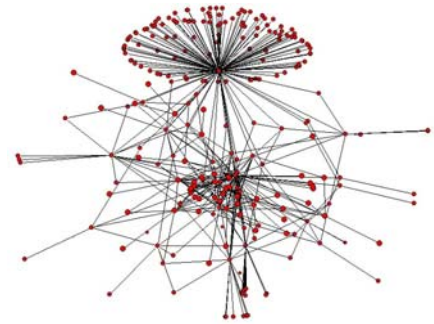


Figure 3 Comment anomaly; mistaken for link log

Automated spidering and weblog discovery

The next stage of our project involved work on the development of tools and algorithms that would allow extracting weblog data in an automatic way. This process has been implemented in BlogTrace (Anjewierden et al., 2005), a general architecture to support weblog research. Although it automated the process of data collection indexing new weblogs was still a time-consuming process. We, thus, worked on the algorithms that would help to identify new weblogs to be spidered based on the likelihood that their authors are connected with the "community" as defined by members of weblogs already indexed.

A partial account of our experiments is provided by Anjewierden (2005). The algorithms we experimented with are based on the idea of "inviting" new members once they are considered socially "acceptable" to the existing members. "Socially acceptability" is operationalised mathematically by considering whether candidate members are linked to by existing members and also whether they link back (reciprocity). The specific operationalisation is based on the idea of resistance: the more a weblog is linked to within the community, the easier it will be found by others.

Looking for reciprocal connections of potential members to an already identified set of weblogs indicated a potentially relevant way to map a weblog community as it would allow for the identification of a number of well-connected blogs starting from one of them. However after some experimentation we discovered that at a certain moment the algorithm was providing data drifting away from the core community before indexing weblogs that were likely to belong there (based on the data we had from earlier analysis stage and personal knowledge of the community).

There are several reasons for this. First, we were not able to automatically index some weblogs identified as strongly connected candidates, which influenced further choices. Moreover, adding a single weblog to the set for each iteration didn't compensate for this by taking into account links from a larger number of weblogs in the community core. Second, we found "black holes", weblogs well connected from the initial set, but directing the

discovery algorithm away from it as they have a stronger connection with "other" communities.

In the case of our analysis, "Many-to-Many"⁶, a group weblog on social software, was an example of a black hole. It's a reliable source of new information and ideas on one of the core topics in the KM blogger community and is written by authors well-connected across multiple networks. Taken that into account it is not very surprising that many people in the community were regularly linking to it and that links from it were heavily directed outwards and not reciprocal. This example corresponds with the data from other studies (Herring et al., 2005, Marlow, 2004) indicating the presence of heavy linking of so-called A-list weblogs by a big number of bloggers, which are not often reciprocal, since A-list blogs tend to link among themselves.

Reciprocal linking within the community

Given our earlier experiences, our latest iteration on the weblog community mapping approach work is based on a number of expansions.

First, weblog indexing was improved. The starting point is a URL to a post by a potential candidate in the community. This URL is used to obtain the RSS feed and the permalinks in the RSS feed are compared to the URL to verify it is indeed a permalink. The pattern is applied to all URLs inside the original URL, picking up additional posts. Applying the spider recursively, all pages with posts are obtained.

Given that all posts for a weblog are now available as HTML pages the principle post attributes have to be extracted: permalink, date, title and body of the post. Unsupervised approaches perform this task with about 85% accuracy (Nanno, Suzuki, Fukiki, & Okumara, 2004). This is clearly unacceptable for research into weblog communities as about 15% of the potential community blogs cannot correctly be analysed. We have opted for a supervised approach that takes advantage of the fact that the HTML structure of a post has a limited (although large) number of variants. With the supervised approach only one potential community member could not be spidered due to an error in the HTML. The supervised approach is slightly more labour intensive than fully automatic approaches, about 15 minutes per new blog to verify correctness and define remedies when necessary. Spidering and analysis itself is complete bound by bandwidth, on average it takes one second per post with a 1Mbit connection.

We also changed our approach for identifying potential members of the community as follows.

1. It starts from the weblog of relatively central community member and identifies blogs linked frequently from it, creating a list of potential weblogs connected by strong ties. The process is performed in two iterations: first indexing all weblogs frequently linked to extract their frequent links and then excluding those that do not link back to a certain number of people in the group.
2. Once a reduced set of weblogs is defined, frequent links from those weblogs are analysed to identify the potential members based on frequency and number of weblogs linking to them from the existing set. Potential members are indexed to

⁶ <http://www.corante.com/many/>

extract their links, and then the set is reduced again to eliminate those that do not link back to others.

3. Step 2 is repeated until no new links are found.

We used this approach as an alternative way to identify a KM community starting from the same weblog as in the first study, the one by Lilia Efimova, and used the following operationalisations:

- The content of all weblogs was indexed to include data from all posts in 2004.
- A connection between weblog A to weblog B was considered a strong tie if A linked to B in at least three posts in 2004. All personal weblogs connected from the seed blog by a strong tie were indexed to extract weblogs they were linking to. Group weblogs were excluded.
- This resulted in the SetOne, which included 42 weblogs. It was analysed for frequency and patterns of linking within. In order to be included as a seed for the following iteration a weblog should link from at least three posts to at least 10% of the set members (four in this case).
- The SetOneReduced included 18 weblogs. The links from those were analysed to define SetTwo which included weblogs linked by the strong tie (at least three posts in 2004) by 10% of members of the SetOneReduced (two in this case). Those weblogs were indexed, resulting in the SetTwo that included 32 weblogs after those that were already indexed as SetOne were excluded.

The resulting set was further analysed qualitatively to identify how far the algorithm was helpful to map the specific community. The weblogs used in the analysis included those in SetOne and SetTwo, as well as Many-to-Many that came up as a member of both sets based on linking, but was excluded as a group weblog.

Qualitative analysis included coding of the weblogs in the resulting set based on their focus. The coding was done by the first author based on her knowledge of the community for the quick check: the categories are roughly defined and would not necessarily be used by bloggers themselves to identify their weblogs. In many cases a blogger would belong to multiple categories; in this case the choice was made to assign the one most strongly associated with the focus of a weblog.

In addition Internet Archive⁷ was used to extract blogrolls of the set members. This choice was made to get blogroll data that corresponds to the linking patterns of 2004. We used a home page of a weblog most close by date to the end of 2004, assuming that it would include relations developed over that year (dates were ranging between September 2004 and November 2004, most of them end of November). Homepages of 61 blogs were accessible, 49 had blogrolls, 45 blogrolls were extracted (4 cases included link to Bloglines to indicate weblogs they read; those were not taken into account as they lead to the current subscription list and can't be easily extracted).

⁷ <http://www.archive.org/>

Figure 4 presents the results of the social network visualisation. It includes weblogs from the extended set (coloured) with all weblogs they linked to during 2004 (grey); number of posts from one weblog to another used as a tie indicator.

Weblog categories are represented by colours:

- Yellow – seed weblog*
- Blue – knowledge management*
- Red – learning*
- Orange – internet research*
- Dark Green – A-list*
- Light green – Many-to-Many*
- White – not decided*

As the data collection has not been finished, it is too early to use it for studying specific patterns in the network, so we used visualisations as a way to check how far our data collection approach corresponds to the subjective feeling of what the community could be. In this process we used different indicators: number of posts between weblogs, strong ties only (3+ posts), reciprocal ties, and blogroll connections.

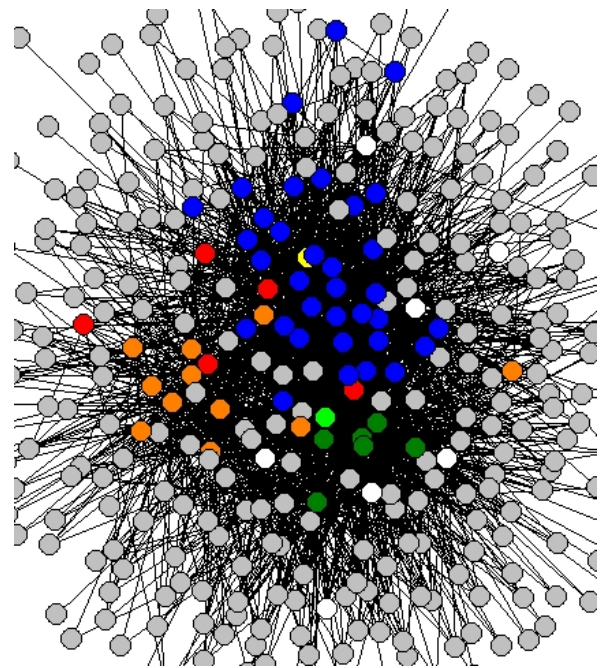


Figure 4. Extended weblog dataset, 2004

Although colour coding allows distinguishing between groups relatively easily, in all cases all weblogs are quite strongly interconnected. This raises a valid question: is there a boundary between groups or do they represent a single community?

Based on our experiences we tend to think that these are separate topical communities. The following quote from the weblog of the first author provides an example:

An observation: while on this trip I met two groups of bloggers ([BlogWalk crowd](#) and [AOIR bloggers](#)) that have some interconnections within they seem not to be connected. I experienced two different sets of references to blogging experiences and other bloggers... Doesn't mean that they do not have anything in common (they do, as bringing [Alex Halavais](#) and [Matt Mower](#) for a dinner shows ;), just seem to live in different parts of the blogosphere (Efimova, 2004a).

If we are right and these are separate communities, what would be a good way to define boundaries between those groups? Is it possible at all based on linking patterns or do we need additional information to do so? What would work in that case: a combination of measures? Text analysis? Could it be that separate communities connected via A-list blogs? Can exclusion of those provide a better view?

In our future work we hope to answer those questions by experimenting with improving spidering algorithm, comparing results starting analysis from different weblogs, as well as complementing link data by text analysis and insight into the relations between bloggers through information provided by the bloggers themselves.

Discussion

In the previous sections we shared our experiences of exploring a way to define boundaries of a weblog community which we know to exist. Although it is too early to say that we have developed a method to do so, we have identified a number of problems and explored directions for solving some of them.

The first problem is about defining what indicators to use to identify a weblog community. Although we used links in weblog text to identify relations between bloggers, this is not applicable to all weblog communities. Our attempts to try this method to identify a community of cooking weblogs did not work: the type of content commonly found in this group of webloggers seems to not require regular linking between weblog posts. Alternatively, lack of linking could be a result of social norms developed in the group. Even reading other weblogs often results in observing and "stealing" practices of others (Efimova, Fiedler, Verwijns, & Boyd, 2004), for example, use of specific plug-ins or particular style of quoting others. Being a member of more defined weblog community may mean complying with a community norms (e.g. being a member of the Knitting Bloggers NetRing requires certain frequency of posting and a focus on knitting, (Wei, 2004). As a result, certain characteristics of a weblog (e.g. language, use of links, weblog structure) can vary significantly between blogging subcultures. Differences between blogging practices and technologies used for blogging suggest that researchers need to understand specific blogging communities they study before interpreting data (Lawley, 2004).

Developing data collection tools for a specific study meets a variety of challenges as well. These include distinguishing a weblog from other types of web-sites and taking into account differences between structure and layout of weblogs due to use of specific functionalities of different weblog platforms, user-modified templates or different practices of using weblog tools.

Conclusions

This paper began as an attempt to gain an insight into methods of finding "life between buildings", virtual settlements where weblog communities may reside. Through this process, it became necessary to address several research challenges. It became clear at an early stage that the linear way of researching online community would not be sufficient as community participation in weblog networks is not located in one place, but distributed both on individual weblogs, as well as in the *space between* the personal weblogs. In order to define community structures, a researcher must not only collect artefacts, but have sufficient prior knowledge in the significance of *reading* the artefact. For example, the authors of this paper are members of slightly different weblog communities. They have more than a year's experience with personal blogging, but only one was sufficiently aware of the subtleties of this community's norms to catch a misinterpretation of data or weblog indexing bugs. On the other hand, the second author's limited knowledge of community norms for this particular community provided an objective counter view in the analysis of the data.

In case of weblog communities, it is usually difficult to find a single shared space (virtual settlement) as a point of departure, so we suggest an iterative approach. One can

start from weak clues indicating the likelihood of a community presence and assumptions about norms and practices in it. These would indicate where to start data collection, which weblogs to include in the analysis, what artefacts to look for and how to interpret them in order to define boundaries of the community virtual settlement. While the virtual settlement is being uncovered, researchers could encounter new types of artefacts, develop a better understanding of community norms and practices, and adjust methods for data collection and interpretation.

In this paper we tried out a suggested iterative approach to map a community broadly defined as "knowledge management bloggers community". We propose a set of indicators of weblog community presence, discuss the alternatives of mapping the community based on those indicators and then select the most promising for further examination.

Through the pilot study, we were able to identify a network structure with, a large periphery membership and a strong core with fuzzy or fluid boundaries. This structure indicates the potential existence of a virtual settlement of a weblog community; however further analysis is needed to identify if fits the conditions suggested by Jones (1997).

References

- Adar, E., Zhang, L., Adamic, L. A., & Lukose, R. M. (2004). *Implicit structure and the dynamics of blogspace* HP Information Dynamics Lab.
- Anjewierden, A. (2005). *Black Holes in the Blogosphere*. Anjo Anjewierden. The source code is the ultimate documentation [On-line]. Available: http://anjo.blogs.com/metis/2005/04/black_holes_in_.html.
- Anjewierden, A., de Hoog, R., Brussee, R. & Efimova, L. Detecting knowledge flows in weblogs, 13th International Conference on Conceptual Structures (ICCS 2005), University of Kassel, Kassel, 1-12, July 2005.
- Blanchard, A. (2004). *Blogs as virtual communities: Identifying a sense of community in the Julie/Julia project*. In L.Gurak, S. Antonijevic, L. Johnson, C. Ratliff, & J. Reyman (Eds.), *Into the Blogosphere; Rhetoric, Community and Culture of Weblogs*. University of Minnesota.
- Blood, R. (2002). *The weblog handbook: Practical advice on creating and maintaining your blog*. Cambridge: Perseus Publishing.
- Dennis, B. & Jarret, A. (2005). NusEye: Visualizing network structure to support navigation of aggregated content. In *Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS'05)*. Los Alamitos: IEEE Press.
- Efimova, L. (2004a). Blog networking, blogwalking and being a boundary subject. Mathemagenic [On-line]. Available: <http://blog.mathemagenic.com/2004/09/28.html#a1363>.
- Efimova, L. (2004b). Discovering the iceberg of knowledge work: A weblog case. In *Proceedings of The Fifth European Conference on Organisational Knowledge, Learning and Capabilities (OKLC 2004), April 2-3, 2004*.
- Efimova, L. & de Moor, A. (2005). Beyond personal webpublishing: An exploratory study of conversational blogging practices. In *Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS'05)*. Los Alamitos: IEEE Press.
- Efimova, L., Fiedler, S., Verwijs, C., & Boyd, A. (2004). Legitimised theft: distributed apprenticeship in weblog networks. In K.Tochterman & H. Maurer (Eds.), *Proceedings of I-KNOW 2004* (pp. 494-502). Graz: Know-Center Austria.

- Efimova, L. & Hendrick, S. (2004). *In search for a virtual settlement: An exploration of weblog community boundaries* Telematica Instituut.
- Gehl, J. (2001). *Life between buildings: using public space*. (5th ed.) Danish Architectural Press.
- Gumbrecht, M. (2004). *Blogs as "protected space"* WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics.
- Halavais, A. C. (2004). Linking weblog neighborhoods: between "small pieces" and "winner-take-all". In *Association of Internet Researchers Annual Conference: IR 5.0: Ubiquity, Sussex, September 2004*.
- Herring, S. C., Kouper, I., Paolillo, J. C., Scheidt, L. A., Tyworth, M., Welsch, P. et al. (2005). Conversations in the blogosphere: An analysis "from the bottom-up". In *Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS'05)*. Los Alamitos: IEEE Press.
- Herring, S. C., Kouper, I., Scheidt, L. A., & Wright, E. (2004). Women and Children Last: The Discursive Construction of Weblogs. In L. Gurak, S. Antonijevic, L. Johnson, C. Ratliff, & J. Reyman (Eds.), *Into the Blogosphere; Rhetoric, Community and Culture of Weblogs*. University of Minnesota.
- Herring, S. C., Scheidt, L. A., Bonus, S., & Wright, E. (2004). Bridging the gap: A genre analysis of weblogs. In *Proceedings of the 37th Hawaii International Conference on System Sciences (HICSS'04)* Los Alamitos: IEEE Press.
- Hiler, J. (2002). The tipping blog: How weblogs can turn an idea into an epidemic. Microcontent news: The on-line magazine for weblogs, webzines, and personal publishing [On-line]. Available: <http://www.microcontentnews.com/articles/tippingblog.htm>.
- Hodkinson, P. (2004). Subcultural blogging? Individual, community and communication. In *Association of Internet Researchers Annual Conference: IR 5.0: Ubiquity, Sussex, September 2004*.
- Hourihan, M. (2002). What we're doing when we blog. O'Reilly Network [On-line]. Available: <http://www.oreillynet.com/pub/a/javascript/2002/06/13/megnut.html>.
- Jones, Q. (1997). Virtual-communities, virtual settlements & cyber-archaeology: A theoretical outline. *Journal of Computer Supported Cooperative Work*, 3.
- Kumar, R., Novak, J., Raghaven, P., & Tomkins, A. (2003). On the bursty evolution of blogspace. In *Proceedings of the twelfth international conference on World Wide Web* (pp. 568-576). Budapest, Hungary.
- Kumar, R., Novak, J., Raghaven, P., & Tomkins, A. (2004). Structure and evolution of blogspace. *Communications of the ACM*, 47, 35-39.
- Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
- Lawley, L. (2004). blog research issues. Many-to-many [On-line]. Available: http://www.corante.com/many/archives/2004/06/24/blog_research_issues.php
- Liu, G. Z. (1999). Virtual Community Presence in Internet Relay Chatting. *Journal of Computer Supported Cooperative Work*, 5.
- Marlow, C. (2004). Audience, structure and authority in the weblog community. In *Presented at the International Communication Association Conference, May, 2004, New Orleans, LA*.
- Mehta, D. (2004). My blog is my social software and my social network. Conversations with Dina [On-line]. Available: <http://radio.weblogs.com/0121664/2004/01/27.html#a356>.
- Merele-Geurvos, J. J., Prieto, B., Rateb, F., & Tricas, F. (2004). Mapping weblog communities. *Submitted to Computer Networks*.

- Mortensen, T. (2004). Dialogue in slow motion: The pleasure of reading and writing across the web. In *Keynote at BlogTalk 2.0*. Vienna, Austria.
- Mortensen, T. & Walker, J. (2002). Blogging thoughts: personal publication as an online research tool. In A. Morrison (Ed.), *Researching ICTs in Context. InterMedia report 3/2002* (pp. 249-278). Oslo.
- Nanno, T., Suzuki, Y., Fukiki, T., & Okumara, M. (2004). *Automatic collection and monitoring of Japanese weblogs* WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics.
- Nardi, B. A., Schiano, D. J., & Gumbrecht, M. (2004). Blogging as social activity, or, would you let 900 million people read your diary? In *Submitted to CSCW 2004*.
- Nilsson, S. (2003). *The function of language to facilitate and maintain social networks in research weblogs*. Umea Universitet, Engelska lingvistik.
- Nonnecke, B. & Preece, J. (2003). Silent participants: Getting to know lurkers better. In C. Lueg & D. Fisher (Eds.), *From Usenet to CoWebs: Interacting with Social Information Spaces* (Springer Verlag).
- Paquet, S. & Pearson, P. (2004). A Topic Sharing Infrastructure for Weblog Networks. In *2nd Annual Conference on Communication Networks and Services Research (CNSR) 2004, Fredericton, May 2004*.
- Walker, J. (2003). Weblog. In *Definition for the Routledge Encyclopedia of Narrative Theory* (Forthcoming from Routledge 2005).
- Wei, C. (2004). Formation of Norms in a Blog Community. In L. Gurak, S. Antonijevic, L. Johnson, C. Ratliff, & J. Reyman (Eds.), *Into the Blogosphere; Rhetoric, Community and Culture of Weblogs* (University of Minnesota).
- Winer, D. (2003). *What makes a weblog a weblog?* Weblogs At Harvard Law [On-line]. Available: <http://blogs.law.harvard.edu/whatMakesAWeblogAWeblog>.