

# In search for a virtual settlement: An exploration of weblog community boundaries

*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*

**Abstract.** Although weblogs are perceived as low-threshold tools to publish on-line, 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. 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 a community presence. A pilot study is presented, where a social network analysis of links between weblogs is used to identify a community of knowledge management bloggers.

**Note:** this is an updated version of paper submitted to Communities&Technologies'05, so:

- beware that final version may be different
- please let the authors know what they can improve

## 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 on-line, 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 *on-line identities* of their authors. Most weblogs are not formal, faceless, corporate sites or news sources: they are authored by individuals (known as bloggers or bloggers), and perceived as 'unedited personal voices' (Winer, 2003).

Often a weblog is written as a narration of its author's thoughts and feelings (Walker, 2003; Herring, Scheidt, Bonus, & Wright, 2004), allowing personality and values to emerge from the words. Even weblogs that are little more than collections of links and short commentaries say something about their authors. The selected content a weblog author finds interesting enough to link to and to comment on functions as a public record of personal interest and engagement.

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 et al., 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 a 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 Hodkinson, 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 weblogs analysis with the application of the distinction between *virtual settlement* and *virtual community* as proposed by 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 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 community<sup>1</sup>, 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 is 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 address research challenges of locating a community when there is no single shared space to start an analysis from, 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. Then a pilot study is presented, where social network analysis are used to identify a network of knowledge management bloggers based

---

<sup>1</sup> Although communities emerging around popular weblogs (e.g. Blanchard, 2004, are an interesting target for a research.

on analysis of links between 50 weblogs generated via snowballing sample. We close the paper by conclusions and discussion of further research.

## Finding life between buildings

### Research challenges

In this section we would like to use Jones' archaeology metaphor (Jones, 1997) and to compare locating a virtual settlement where a weblog community resides with archaeological excavation. One has to know where to start digging and what to look for. Then, once artefacts are found, there is a need to interpret their meaning in order to decide more promising directions for extended excavation. One must also know when to stop the exploration.

Thinking about discovering and delineating a weblog community in terms of archaeological excavation we have identified three sets of problems:

- uneven structure of the blogosphere that creates a situation where it is not always clear where to start "excavation";
- "teasing" data (a variety of artefacts is available, but most of them are difficult to get) that complicates decision-making on data collection methods;
- artefacts may be culture-specific, so one must be careful with his or her interpretations.

### Knowing where to start: uneven structure of the blogosphere

The chance of discovering a weblog community depends on a subset of weblogs selected for an 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 (see

below), 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.

#### Data collection methods: "teasing" data

The public nature of weblogs makes them an easy target for a researcher, providing a record of personal interest and engagement in the posts, as well as links that indicate influences and relations with other bloggers. Most weblogs have a simple and well-defined structure (e.g. the weblog post usually has a title, a permalink and a date/time stamp), generate web-feeds (RSS or Atom) representing weblog content in machine-readable format (XML or RDF), or notify centralised weblog tracking tools (e.g. [weblogs.com](http://weblogs.com)) about updates. The relatively simple structure of weblogs and widespread adoption of standards (RSS, XML-RPC, Blogger API) by weblog tool providers enable a variety of tools and services that allow tracking and analysing weblogs. For example, one can visualise a weblog neighbourhood (related weblogs) at [Blogstreet](http://Blogstreet.com) ([blogstreet.com](http://blogstreet.com)), check weblog popularity ranking at [Technorati](http://Technorati.com) ([technorati.com](http://technorati.com)), track ideas contagiously spreading in a weblog community at [Blogdex](http://Blogdex.com) ([blogdex.com](http://blogdex.com)) or read a selected subset of weblogs online at [Bloglines](http://Bloglines.com) ([bloglines.com](http://bloglines.com)).

Publicly available weblog data and a large number of tools to analyse it raise expectations about availability of this data for research purposes, although the practice of weblog research is dramatically different (e.g. Anjewierden, Brussee, & Efimova, 2004; Herring et al., 2005, for explicit indications of challenges of obtaining weblog data). Most weblog tracking and analysis tools index only a subset of weblogs (e.g. those that registered with the system); include partial weblog data usually representing fresh updates (e.g. links from homepages or content from last 45 days); or index only data in machine-readable formats (e.g. RSS/Atom feeds that are not always present or include excerpts of weblog posts instead of full-text).

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. As a result, many weblog researchers have to limit themselves to working with convenient samples (e.g. restricting data collection to a specific weblog platform as in Merelo-Geurvós et al., 2004) or rely on manual work that limits the number of weblogs and weblog characteristics to be included in the analysis. Choices made for data collection in those cases can heavily influence the results of the analysis.

## Clues to define direction: culture-specific artefacts

Weblogging tools are flexible to allow for a variety of uses ("Weblogs simply provide the framework, as haiku imposes order of words", Hourihan, 2002). Like a pen could be used to write a diary, a novel, a letter to a friend, or just a shopping list pinned to a fridge door, weblogging tools can be used to publish a personal diary, to collect and share links, to facilitate a course, as an unfolding novel, a record of an experiment, a recipe book... Each of those uses could require a particular use of weblog functionalities and a particular format in a weblog. In addition, weblog format and features are influenced by weblog tool functionalities available for a blogger, personal preferences and technology skills, as well as norms and practices in social surroundings of a blogger.

Even reading other weblogs often results in observing and "stealing" practices of others (Efimova, Fiedler, Verwijs, & 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 Knitting Bloggers NetRing requires certain frequency of posting and 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. For example, blogrolls (sidebar links to weblogs one reads) could be used to identify relation between weblog authors (e.g. Herring et al., 2005). While this approach can yield valuable results, it is not always applicable for discovering potentially existing communities. Some bloggers do not use blogrolls, not because they do not have connections with others, but because they have adopted news aggregators and read other weblogs via their web-feeds (RSS or Atom), as the following quote illustrates<sup>2</sup>:

Seems to me that blogrolls made sense in a time before RSS aggregators. If you use other blogs and sites as triggers for your own writing, then a blogroll serves as a useful way to organize your surfing. When you shift to an aggregator driven strategy, your subscriptions file becomes the equivalent of your blogroll. Of course, your subscriptions file is invisible while your blogroll was public.

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).

## Virtual settlements, practices and artefacts

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

---

<sup>2</sup> <http://www.mcgeemusings.net/2003/10/04.html#a3711>

of sustained stable membership. Those conditions presuppose that there is a cyber-space with some kind of group interactions that is studied to identify if there is a community there. Figure 1 provides a simplified image of the process.

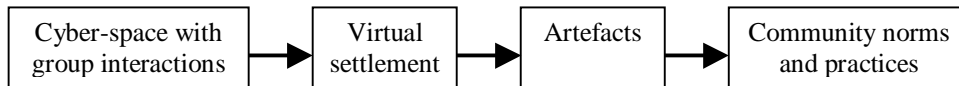


Figure 1. Studying virtual communities

In 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 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.

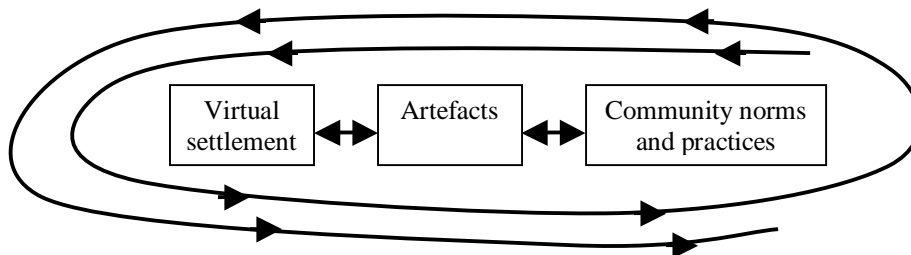


Figure 2. Studying weblog communities

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

### Signs of life between buildings

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

**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, 2004), so analysing weblog reading lists (e.g. in a 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 a 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 necessary 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 necessary 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 a 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 be also maintained via 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 bloggers participation as well driven by an interest of a third party to keep a list of interesting weblogs or aggregate their content.

## Pilot study

In order to try out the suggested approach a pilot study was performed. The purpose of the study was to map a community of knowledge management bloggers where the first author belongs. As lack for appropriate tools for extensive data collection was quickly discovered the study was limited to an examining links between 50 weblogs over 7 months period using social network analysis method. An assumption was made that reciprocal linking indicates a relation between actors (bloggers). This relationship makes the actor part of a network, a network that would indicate participation in a community. This network would need to be analysed further for the community structures outlined by Jones (1997) before one could determine if the network is, in fact, a community as well.

For the pilot study a choice was made to base the analysis on links in weblog text. Although analysing blogrolls and RSS subscriptions is a good way to map the network as they used by bloggers to self identify members of their network, they were not included into the analysis. Blogrolls were excluded as many members of this community rely on RSS readers for reading other weblogs and do not have blogrolls or do not update them regularly. RSS subscriptions are difficult to obtain if they are not shared publicly and have to be collected by asking bloggers directly, so the decision was made to use RSS subscription analysis at a later stage once the core members of the network are identified.

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. Rebecca Blood underscores weblogging self awareness in *The Weblog Handbook*:

After producing Rebecca's Pocket for several months I noticed how frequently I linked to articles on archaeology and scientific discoveries. I chose them because they seemed worth sharing, and in fact they were just the sort of stories I was apt to read aloud to friends when sharing a Sunday *New York Times* over a latte. But reading aloud is even more ephemeral than writing a weblog, and if you had asked me, I would have told you my major interests lay in the arts and humanities, and that I was only marginally interested in science of any kind. Indeed, I had completely forgotten that, for a short while as a child, I had dreamed of becoming an archaeologist. Reading the record of things I deemed worth sharing reminded me of a self that I had overlooked. (Blood, 2002: 30)

## Method

A snowball sample was created from the starting point of Lilia Efimova's blog (blog.mathemagenic.com). Seven months of archives were gathered into one file, and then mined for links. Weblogs referred to more than three times during a seven month archived period were used to generate the point of departure for analysis. This choice was made in order to limit the number of weblogs for further data collection as it was likely to exclude pointers to interesting weblog posts outside of the author's community.

49 links were gathered from the first stage. For each link, the same seven months of archives that were taken from Lilia's blog were manually collected and placed into a separate text file. Each text file was then run through the Ruby script URLScanner, which generated a list of URL's for each file. These links were compiled into master list and cleaned for repetition and non-blogs. For the purpose of this study, a URL is considered a weblog if it contains separate posts, each with enabled comments and post-specific permalinks. The resulting master list included 418 links listed in a matrix. This matrix was compiled in UCINET and exported to Pajek. The network of this data is visualised in Figure 3.

## Results

The visualization of the data yielded many interesting patterns. The centre node in the core cluster represents Lilia's blog, which is where data collection began. The tight cluster of weblogs surrounding her represent those who have the closest ties through instances of reciprocal linking. After identifying the weblogs in the centre cluster, it became clear that these blogs belonged to authors focusing on the topic of knowledge management. When comparing the self-identified network with the network created from analyzing reciprocal linking, over half of those identifying themselves as knowledge management workers and researchers were core members of the reciprocal linking network.

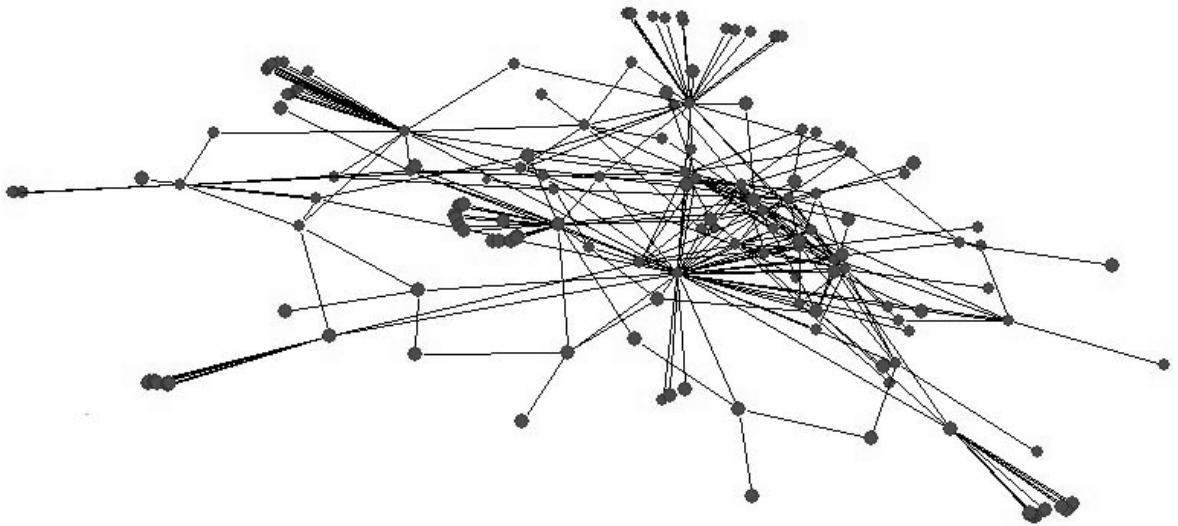


Figure 3. Knowledge Management (KM) Weblog Core and Periphery as measured between March and September, 2004.

Another interesting pattern found in data above concerns the boundaries of weblog networks. There are many instances of mini-clusters on the periphery of the core cluster of weblogs, which have a group of weblogs shooting outwards, not linking back to the core. There are two reasons for these 'non-core linking' clusters. The first represents a portal to different networks, which are of interest to a minority of the members of the core. These blogs often represent networks built around topics other than those which concern the core collective. For example, the bottom most mini-cluster (see Figure 5) begins from the blog Chocolate and Zucchini ([chocolateandzucchini.com](http://chocolateandzucchini.com)). This cooking blog does not link to core knowledge management blogs, yet is linked to by several of them. The blogs linked to by Chocolate and Zucchini are all cooking blogs, none of whom link to the KM core.

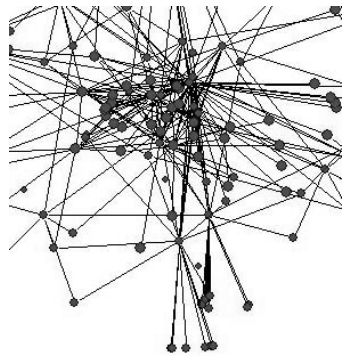


Figure 4 Core cluster of the KM network



Figure 5. Mini-cluster from the KM network

The second type of mini-clusters are those which represent a meme or topic of current interest. For example, during this seven-month period Blogtalk, a conference for and about weblogs, was held. Many of the mini-clusters present in Figure 3 are links to bloggers who presented a paper at this conference. These weblogs held topics of interest to those in KM at a particular moment in time. The weblogs from which these topics originate are often topically similar to the core network, but not so similar that they are core members. One of the presenters from the later example is Lisbeth Klastrup ([www.klastrup.dk](http://www.klastrup.dk)). Klastrup maintains a research blog about persistent online worlds, games, and interactive stories. While these areas are not KM areas, the two do share common interests. These interests are often magnified during events like conferences, or when a shared interest or topic is heavily discussed between blogs. If this network was analysed during another seven-month period, it is very likely that these peripheral mini-clusters would represent different weblogs than those present in Figure 3.

These mini-clusters, however, represent more than a shared topic or interest. They are also an example of the fluidness of the boundaries of weblog networks. The core of a network is built through the exchange of ideas and shared aims between like-minded bloggers. Links are exchanged as social currency for both the recognition of your peers and for acknowledgement of topic. As topic is not limited to your core network, theoretically any weblogger can be brought into a link exchange due to a shared discourse on a certain topic. The boundaries of the network suddenly become fuzzy. It is these fuzzy boundaries that allow for the core to expand and contract. Going back to the virtual settlement and architecture metaphor, suddenly the plot of land next to your house is up for sale. The new neighbours have a lot of interesting thoughts and ideas, and you find yourself going over for coffee more and more often (an exchange of topically related links).

The fuzzy boundaries of the periphery can allow the network to change. It is difficult to exchange members in the core of the network, as these members are related by several attributes: similar belief structures, past exchanges/relationships

and a sense of familiarity. The members of the periphery only need a similar topic exchange to become a temporary member of the network. As time progresses, if a periphery member blogs about core topics and stays in contact with core members through both links of acknowledgement and participation in weblog conversation, that member will become a part of the core.

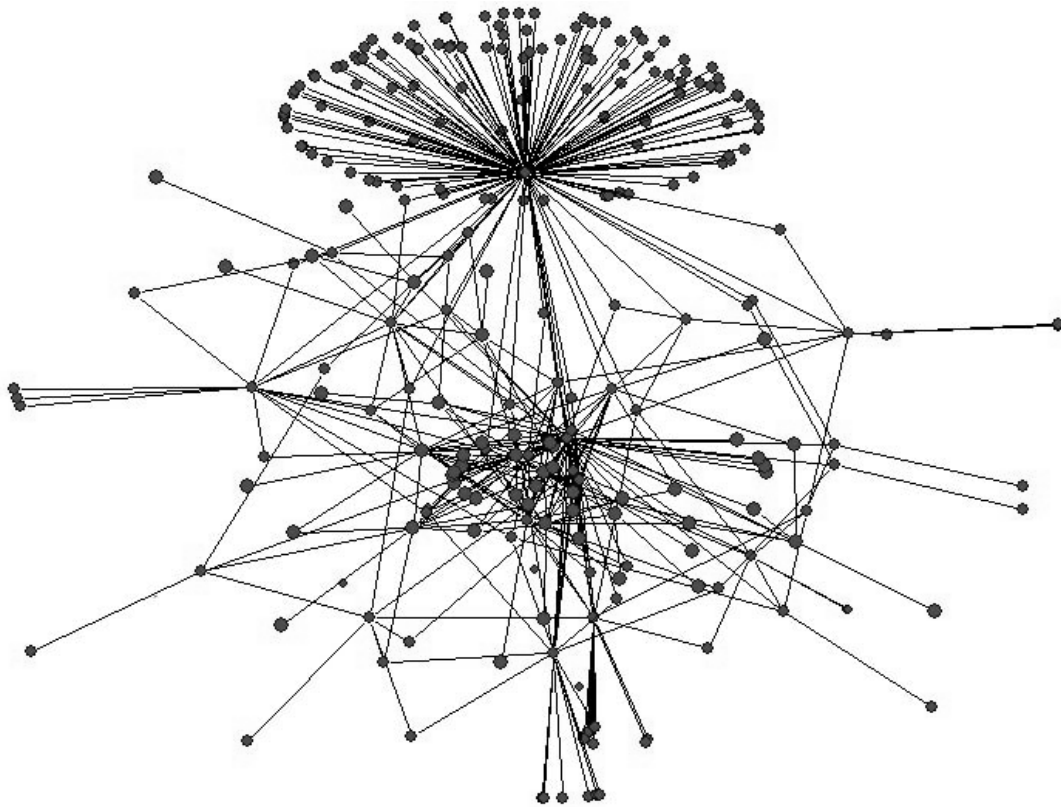


Figure 6. Knowledge Management (KM) Weblog Core and Periphery as measured between March and September, 2004 (with comment anomaly)

Figure 6 represents an anomaly found in the initial data, which ultimately provided an interesting comparison. 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 when the comments were removed from the data indicates that this weblog has a high rate of participation in the comment area. The question remains if this rate of participation is usual in weblogs, and is an interesting topic for further study.

## Conclusions

This paper began as an attempt to get an insight into methods of finding "life between buildings", virtual settlements where weblog communities may reside. Through this process, several research challenges became necessary to address. 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 two authors of this paper are members of slightly different weblog communities. Both 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. 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.

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, which allow for the periphery members to become a core member by both entrance through topic and through reciprocal linking. 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).

## Acknowledgments

The authors wish to thank Samuel Kvarnbrink for his creation of the Ruby script, *URLscanner*, as well as his patience in explaining everything python and ruby related. Also to Alex Halavais, Anjo Anjewierden, Matt Mower and Natalie Glance for their help regarding data collection methods and tools.

## 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., Brussee, R., & Efimova, L. (2004). Shared conceptualisations in weblogs. To be published in *Proceedings of BlogTalk 2.0, Thomas N. Burg (ed.), Vienna, July 2004*.
- 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. (2004). 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.
- 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. (2004a). Linking weblog neighborhoods: between "small pieces" and "winner-take-all". In *Association of Internet Researchers Annual Conference: IR 5.0: Ubiquity, Sussex, September 2004*.
- Halavais, A. C. (2004b). *Tracking Ideas in the Blogosphere*. Unpublished paper.
- 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*. 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*. 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 (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.
- Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
- Lawley, L. (2004). blog research issues. *Many-to-many*. Available: [http://www.corante.com/many/archives/2004/06/24/blog\\_research\\_issues.php](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 (1).
- 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*. Available: <http://radio.weblogs.com/0121664/2004/01/27.html#a356>
- Merelo-Geurvós, 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*. 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.
- Nardi, B. A., Schiano, D. J., & Gumbrecht, M. (2004). Blogging as social activity, or, would you let 900 million people read your diary? In *Proceedings of 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*. Available: <http://blogs.law.harvard.edu/whatMakesAWeblogAWeblog>