

# The Blog Network in America: Blogs as Indicators of Relationships among US Cities

**Jia Lin**<sup>1</sup>

*Department of Communication; State University of New York at Buffalo*

**Alexander Halavais**<sup>2</sup>

*Department of Communication; State University of New York at Buffalo*

**Bin Zhang**<sup>3</sup>

*School of Medicine; University of California Los Angeles*

*An analysis of links among U.S. weblogs is done to examine the interpersonal social network and social connections among U.S. cities. Drawing 4,241 weblogs from the NITTE census dataset that are identified as being located in the United States, this project extracts the outward links of these weblogs and uses them to analyze the relationship between cities. A total of 632 U.S. city/region units, represented by the first three-digits of US postal codes, are taken as nodes of the network. In total, 41,212 permanent links from blogs of each of the city units are counted as weighted arcs in the network. Inlinks and outlinks of each city unit are recorded for analysis. The study finds that the city units whose bloggers attract most inlinks are Manhattan, San Francisco and Bay Area, Washington, D.C. and its western suburbs, Boston and its suburbs, Los Angeles and Seattle. The study discovers a super-metropolitan cluster, transcending geographical boundaries, within which the cities traditionally associated with cultural elites are closely connected. For other less metropolitan areas, blogs are most heavily connected at a geographically local level, and then extend to a national network.*

## INTRODUCTION

Weblogs, or *blogs*, are self-published websites that have burgeoned since the late 1990s and by December of 2004, the number of blogs had grown to 7 million (Technorati.com). Spontaneous, self-reported expressions made conveniently available online provided great opportunities for social-science research. The blogosphere, the totality of interconnected blogs, provides two layers of information: content and relationships. Writings in weblog entries archive people's everyday experience, while hyperlinks among individual blogs trace some form of social structure. Bloggers are not only noting down their experiences and thoughts, but also trying to reach out to broader audiences, share opinions and to manage their personal knowledge base.

The digital revolution has profoundly redefined the dynamics between space and place. Though people may remain physically stationary, their identity, social capital, and flows of communication often exist in a spatial form. On the other hand, what people bring to online communications is inevitably shaped by "their gender, stage in life-cycle, cultural milieu, socioeconomic status, and offline connections with others" (Wellman & Gulia, 1999). While blogging is an on-line activity that transcends geographical boundaries, the self-images presented to the public and the hyperlinks used are shaped by who the bloggers are in real life, including their physical location.

This project explores social connections among American cities

---

<sup>1</sup> [jialin@gmail.com](mailto:jialin@gmail.com)

<sup>2</sup> [alex@halavais.net](mailto:alex@halavais.net)

<sup>3</sup> [binzhang.ucla@gmail.com](mailto:binzhang.ucla@gmail.com)

by examining weblog hyperlinks among them. We postulate that the range and density of networks of social relations reflect the strength of social connections among cities. By studying the network of cities linked through hyperlinks on blogs, this project will identify the most influential cities in the country, as well as the geographical distribution of city clusters connected by blog links.

## LITERATURE REVIEW

### Weblog networks as indicators of social connection

The blogosphere provides two layers of information: content and relationships. Compared to personal journals, blogs engage a public readership through posting online and linking to other sites. According to Rebecca Blood (2000), blogs started as link-driven sites and each was a mixture of links to news, commentary and thoughts. Blogging makes use of a set of practices and protocols (pinging, subscribing, commenting, trackbacks, and hyperlinking) through which bloggers co-construct a set of affiliation networks, within which individuals with the same interests or close social ties interact with and refer to one another. MP

Tens of millions of personal weblogs have become an indispensable part of World Wide Web. On-line hosting services like Blogger, TypePad, and Livejournal, have made blogging easier for internet users and have stimulated a wider adoption of blogging. According to National Institute for Technology & Liberal Education (NITLE) census data, from July to November 2003 the blogs on *blog-hosting sites* such as *Blogspot*, *Livejournal* and *Diaryland* increased more than 4 times from 176 thousand to 720 thousand. It is estimated that there are at least as many *stand-alone blogs*, blogs that are self-hosted or at least have their own domain names. In either case, bloggers continue to update and post new entries regularly, leading to a large and quickly-growing network of hyperlinked sites that form what has been termed the *blogosphere*.

Broadly, a weblog is any “hierarchy of text, images, media objects and data, arranged chronologically, that can be viewed in an HTML browser” (Winer, 2003). This loose definition can be applied to almost any sort of regularly updated website, but what makes weblogs special is that they are the “unedited voice of a person.” Weblogs can be categorized into one or more of three types: first, a personal diary recording daily activity and thoughts; second, a collection of links to other websites worthy of recommendations, usually with a few words of comments; and third, a forum devoted to specific topics. Beyond personal blogs, there are also a significant number of community weblogs that are coauthored by a number of contributors; this includes sites like Slashdot and Boing Boing, among others. Some corporations and other organizations—from Microsoft to Ford to Boeing—also maintain weblogs.

The depth and size of the content, as well as its relatively easy accessibility, make weblogs a potentially valuable resource for

social-psychological studies. Blogging technology breaks down traditional centralized authorship, enabling everyone with internet access to become a potential author. The blogosphere provides a reflection of ongoing distributed discourse (Halavias, 2002). Political blogs — among the most popular — have been the subject of several studies related to political agendas and public opinions in the US (Lin & Halavias, 2005; Adamic & Glance, 2005; Cornfield et al., 2005). In the business world, personal blogs have become an useful source for marketers and advertisers to study word of mouth communication and detect new trends and consumer behaviors.

The blogosphere provides two layers of information: content and relationships. Compared to personal journals, blogs engage a public readership through posting online and linking to other sites. According to Rebecca Blood (2000), blogs started as link-driven sites and each was a mixture of links to news, commentary and thoughts. Blogging makes use of a set of practices and protocols (pinging, subscribing, commenting, trackbacks, and hyperlinking) through which bloggers co-construct a set of affiliation networks, within which individuals with the same interests or close social ties interact with and refer to one another.

For the purpose of regular reading, many bloggers place links of other blogs on their index pages (permanent links, also sometimes called “blogrolls”) or subscribe to various blogs through syndicated feeds. In this sense, one cluster of hyperlinks could be viewed a virtual neighborhood or community (see Kumar et al, 1999). Such groups may be driven by political preference, culture or academic interests, health or spiritual support, or by familial and friendship ties. According to Wellman (2001), networks online or offline represent social network of relationships that provide sociability, information and sense of belonging. Hyperlink networks reflect the structure of social relationships online, and can be used in the research of international communication, interpersonal communication and ecommerce (Park, 2003). Blog networks, as an increasingly important component of Internet, can be viewed as another effective indicator of virtual community on line. Because the hyperlinks on blogs are frequently created by individual authors, they represent a more fine-grained view of social structure. The collective linking behavior among geographical units can reflect the overall social connections: do bloggers link to each other regardless of geographical boundaries? Are there stronger bonds among certain cities than others?

### Geography in the digital age

The prevalence of telecommunication technologies has generated some popular notions of the fading impact of geographical location and physical distance. From the early prophecy of the “global village” by Marshal McLuhan (1968), to the expansion of the “global city” (King, 1990), from the claim of the “death of distance” by Cairncross (1997), to the design of the “City of Bits” and the “E-topia” by Mitchell (1999), scholars suggest that in the digital age, where people work, live and socialize is not limited by where they physically are anymore. The rise of

“world cities” has been widely accepted in urban studies. Cities like New York, Los Angeles, London or Paris have virtually exceeded the national boundaries and become the world’s “headquarters” because of their prominent place in corporate control, communication networks, and cultural production (Logan & Molotch, 1987; King, 1991). King (1991) introduced the concept of the world urban system, an interdependent system, consisting of people, knowledge, images and ideas. Such “world cities”, according to King, bring together the highly-paid international elites and a transnational producer service class to form a spatial center for global transmission of news, information and culture. In all, these scholars believe information technology enables a time-space compression and the decrease of unidimensional spatial patterns.

Meanwhile, the notion that “geography is dead” is criticized as one of the greatest of the new economy myths. Opposing voices arise to reiterate the importance of geography in a new urban landscape, and suggest that information technology only reinforces community ties. Wellman and colleagues’ research (2001) on Canadians’ use of the Internet finds that people use email mainly to enhance communications with acquaintance such as kin and neighbors, and communication is lower with distant than nearby friends. Kotkin (2001), in his book on the geographical distribution of the new information industry in the United States, declares that the appeal of a place is an increasing priority when people choose where to work and live, since the communication can be achieved over a long distance anyway. Richard Florida’s (2002) study of the creative economy finds that human capital today is more selective when deciding work-sites. In the international setting, Halavias (2000) finds that the distribution of the hyperlinks is far from matching the worldwide distribution of websites, and there are clear national borders on the Internet. Barnett and Choi’s (1995, 1999) examination of the international telephone network suggests that international telecommunication network is to a large degree determined by the factor of physical location and the network can be clearly differentiated into three subgroups (Latin American, Europe and Asia sub groups), with United States acting as a “liaison” in the center. Zook’s (2001) track of internet domain registrations concludes that “the Internet is a more selective network that parallels physical geography and economic development” (p. 3). The growth of the Internet is concentrated in the big cities and urban areas; rather than destroying geography, the internet is selectively connecting a small group of people into highly interactive networks.

The United States is one of the most connected countries in the world with the largest internet population. According to the Pew Internet and American Life Project, 2% to 7% of adult Internet users in the United States are keeping weblogs, of which 10% update them daily. Blogs have become an alternative, grass-roots form of media. Especially when it comes to political news, the blogosphere has been viewed as an indispensable source by mainstream media and internet users. The occurrence of decentralized publishing not only redefines community space, but also redefines the role of authorship and

readership. The blog attracting many in-links can be viewed as a credible news source, a popular opinion leader in a certain field or a platform for good writing, while out-links from a blog indicate the seeking of such news, opinion or writings. By relating blogging to the question of physical geography versus cyberspace, this project will examine where the popular authorship located; that is, the degree to which the geographical location of the blogger is related to his or her prominence within the blogosphere. At the same time, we may ask, do such linking behaviors create an integrated network or a fragmented one with subgroup clusters?

Our society is not a collection of random units, but an interactive and hierarchical network consisting of specific geographical relationships that help to define cultural relationships within and between cities. A city is “a state of mind, a body of customs and traditions and of the organized attitudes and sentiments that inhere in these customs and are transmitted with this tradition” (Park, 1984, p. 1). Accordingly, we assumed that people of same place will demonstrate more or less similar patterns in selecting and attaching to social networks. Social network analysis is widely used in examining the structure of a social entity where geographical location is often seen as a node in the network (Brunn and Dudge, 2001; Zook, 2001; Barnett & Park, 2005; Barnett, 1999; Barnett & Choi, 1995; Halavias, 2000). This study postulates that the range and density of social relation networks reflect the strength of social connections among city units.

In sum, this project seeks to answer the following two research questions:

1. To what degree are the hypertextual expressions of blog authors related to the geographical locations from which they blog?
2. Where are the centers of opinion leaders? Are there clusters of opinion congregation, and if so, how are they identified?

## METHODOLOGY

### Data

Blog samples for this study are retrieved from NITLE census data from June and July of 2003. The database consists of index pages from about 120,000 blogs. Permanent links are retrieved from the blogs that are identified as being authored in the US. In most cases, the permanent links appear on the index pages; but for many blogs hosted by *Livejournal* and *Diaryland*, such links appear instead on member information pages. To exclude hyperlinks appearing in periodic postings, which are less likely to be to weblogs and are therefore less indicative of interpersonal affiliation, only URLs placed on sidebars and grouped together are extracted. The links to such popular web services as hosting sites, providers of visitor statistics, commenting services, technology assistance, together with popular mainstream news sites are excluded from the data.

The geographical locations of blogs and the targets of their outlinks are extracted. Computer-assisted automatic retrieval is realized by a custom set of software tools. A crawler searches all relevant web pages, retrieving the source code (including html code and plain text) of every page, and extracts keywords with geographical information. The crawler searched following patterns that contain geographical information:

*Geotags*: When present, explicit meta-tags pinpointing the geographical location of a site were the most unambiguous indicator of a given site's location. After extracting the values of longitude and latitude from meta-tags, they were mapped to zip code, if located in the United States. Unfortunately, only very small proportion of weblogs provide such meta-data.

*Local weather*: City location can be inferred by weather-related links, since the more exact location the blogger provides to the weather service, the more precise weather forecast they can receive. *Weather.com* and *weatherpixie.com* are the two dominant weather services used by bloggers. In each case, there is an indicator of the geographic location in the text of the URL; namely, either a zip code or local airport code.

*Blogchalk profile*: Blogchalk represented personal information about the blogger in machine- and-human- readable forms. Included among these keywords were the home city and country of the blogger. While now largely defunct, the service was still occasionally used among the weblogs surveyed during this period.

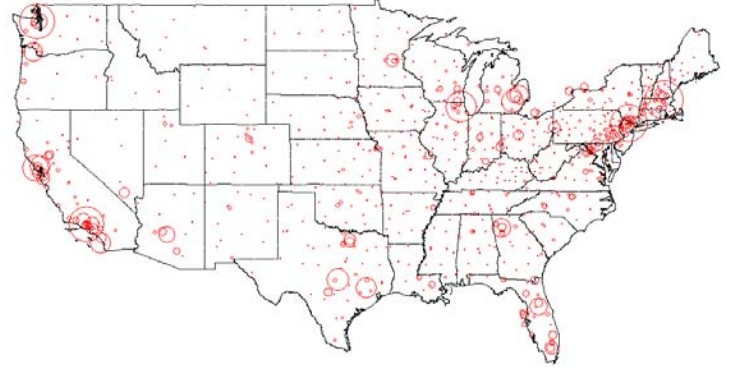
*Blogger profiles at hosted weblogs*: *Blogger*, *Livejournal* and *Diaryland*, the three major blogging host services during the period, provided web pages for user profiles where users were able to list their location.

Using the technologies described above, a previous study on mapping the distribution of weblogs in America generate the map in Figure 1 (Lin & Halavais, 2005).

This analysis made use of data from 2003, which, while slightly dated, provides a good snapshot of the web. The approach taken here may not be as appropriate when applied to the current or future blogosphere, because of the increased use of syndicated feeds (RSS) and aggregators, which reduce the apparent links between blogs.

A total of 4,241 weblogs and 41,212 permanent links are identified for their location in the US. Blogs and links are plotted into their corresponding three-digit zip codes. A total of 632 U.S. city/region units have at least one blog present. The geographical location of blogs has been shown to be consistent with the population distribution and concentrations of high socio-economic status (Lin & Halavais, 2005). Park (1984) notes residential homogeneity as an important indicator in sorting neighborhoods within a city space. Compared to a partition of population in greater metropolitan area which is too general, or into 5-digit zip code corresponding to streets or blocks (Weiss, 1989), which overstates the population variance,

3-digit zip code units represent a middle approach that defines a geographical unit in a way that is widely used in marketing and political targeting strategies.



**Figure 1.** Distribution of bloggers in America. Circle radius is proportional to the number of webbloggers in a region.

## Data analysis

After collecting all weblogs with links and geographical locations, each weblog  $i$  and its  $n_i$  out-linked weblogs are transformed into a list of 3-digit zipcodes, represented by

$$l_i = \langle z_0^i, z_1^i, z_2^i, \dots, z_{n_i}^i \rangle \quad (1)$$

where  $z_0^i$  is the 3-digit zipcode for the weblog  $i$  and the remaining elements are the 3-digit zipcodes of the weblogs referred by the weblog  $i$ .

Given  $N$  weblogs, there will be  $N$  lists encoded as the formula (1). From these lists of three digit zipcodes, we can construct a linkage matrix  $A = [a_{ij}]$ , where each entry  $a_{ij}$  represents the number of outgoing links from a region represented by a 3-digit zipcode  $i$  to a region encoded by a 3-digit zipcode  $j$ . The linkage matrix  $A$  represents a directed network. Several important network-based measures are the number of incoming links  $k_{in}$ , the number of outgoing links  $k_{out}$ , and the number of local links  $k_{local}$  (the number links to the blogs of the same zip code) and they are formally expressed as follows:

$$k_{in}^i = \sum_{j, j \neq i} a_{ji}, \quad (2)$$

$$k_{out}^i = \sum_{j, j \neq i} a_{ij}, \quad (3)$$

$$k_{local}^i = \sum_{j, j \neq i} a_{ji}, \quad (4)$$

where  $i$  and  $j$  represent different regions encoded by 3-digit zipcodes.

Notice that the local links are excluded from  $k_{in}$  and  $k_{out}$ .  $k_{in}$ , the total number of inlinks from outside zip codes will be taken as the prestige score.

In order to later perform a cluster analysis, we derive an undirected adjacency matrix  $A = [a_{ij}]$  from the linkage matrix  $A$ :

$$\begin{cases} u_{ij} = a_{ij} + a_{ji}, & i \neq j \\ u_{ii} = 0 \end{cases} \quad (5)$$

The adjacency matrix  $U$  is symmetric and each entry represents the strength of accumulative interactions (total links) between the two corresponding regions. The connectivity of a region to adjacent nodes is indicated by the total number of inlinks and outlinks (excluding the links within the region), given by

$$k^i = \sum_j u_{ij} \quad (5)$$

Additionally, we cluster regions within the network to discover linking relationships. Clusters are subsets of nodes that are tightly linked with each other. For this purpose, we perform an *agglomerative* hierarchical clustering using an average distance clustering algorithm, in which the distance between two clusters is defined as the average number of links between all pairs of nodes in each pair of clusters (Duda et.al, 2000). The hierarchical clustering initially assigns each node to a cluster and then iteratively merges the closest pair of clusters until all nodes are clustered into a single cluster.

The network analyses presented above were implemented in R (Ihaka & Gentleman, 1996). To better visualize the weblog network, we also use the Fruchterman Rheingold algorithm implemented in Pajek to represent the network in 3-dimensional space.

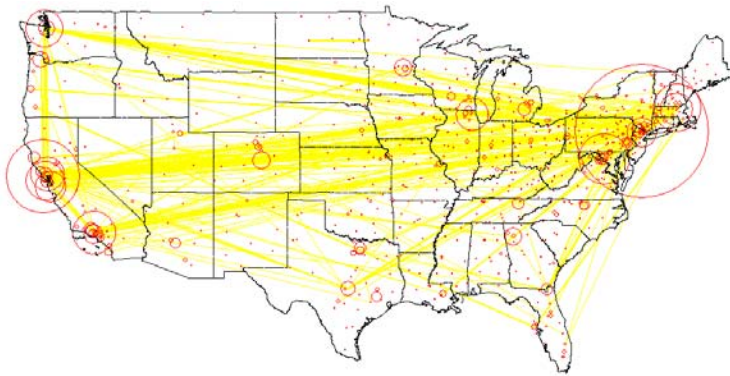


Figure 2. Blog network plot on American map. Circle radius is proportional to the number of inbound links to a region.

**FINDINGS**

**Geographical map of blog links**

Blog links are plotted into the American map shown in Figure 2. There are clearly heavy links is across the continent from coast to coast. The bigger circles in areas indicate larger numbers of in-links from other city units.

The ranking of city units by the number of in-links are shown in Table 1. The city units whose bloggers attracted most inlinks

include Manhattan, San Francisco and the San Francisco Bay Area, Washington, D.C. and its western suburbs, Boston and its suburbs, Los Angeles and Seattle.

Table 1. Prestige score-City units with the most inlinks (excluding links from local blogs)

Prestige ranks	3-digit zip code	City unit	No. of inlinks
1	100	Manhattan	2657
2	941	San Francisco	1446
3	200	D.C	1017
4	21	Boston	942
5	201	Northern Virginia	885
6	900	Los Angeles	845
7	940	South Bay of SF	778
8	981	Seattle	728
9	945	West Bay of SF	662
10	606	Chicago	634
11	24	Suburb of Boston	634
12	950	Santa Cruz	628
13	112	Brooklyn	573
14	902	West Los Angeles	361
15	554	MINNEAPOLIS	347
16	809	Colorado Springs	345
17	303	Atlanta	342
18	980	West Seattle	322
19	972	Portland	301
20	787	Austin	288

To examine if the number of inlinks are determined by the number of bloggers, we weight inlinks by total number of blogs in each city, resulting in Table 2. Compared to the indegree ranking (Table 1), 7 out of the top ten cities remain the same. This indicates that the blog ties are only partially determined by population size.

Table 2. Ranking of cities with the highest degree of normalized inlinks.

Prestige ranks	3-digit zip code	City unit	Normalized inlinks
1	200	D.C	16.67213
2	940	South San Francisco	15.56
3	100	NewYork City	13.08867
4	941	San Francisco	11.568
5	24	Suburb of Boston	11.32143
6	201	Northern Virginia	10.79268
7	950	Santa Cruz	9.515152
8	112	Brooklyn	9.095238
9	900	Los Angeles	8.894737
10	554	Minneapolis	8.463415

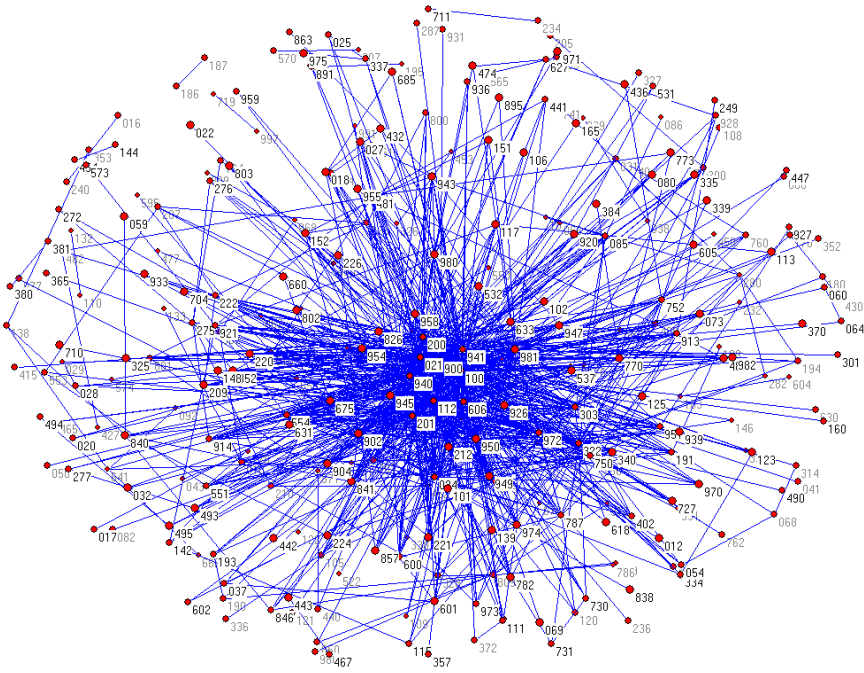


Figure 3. Blog network based on centralities

Figure 3 shows the hyperlink network based on the centrality of each city node. Network density is a measure of the relative number of connection, and it is ranged from 0 to 1. A fully connected network has a density of 1. The density for blog hyperlink network is 0.0662, which indicates that the network is far from highly interconnected.

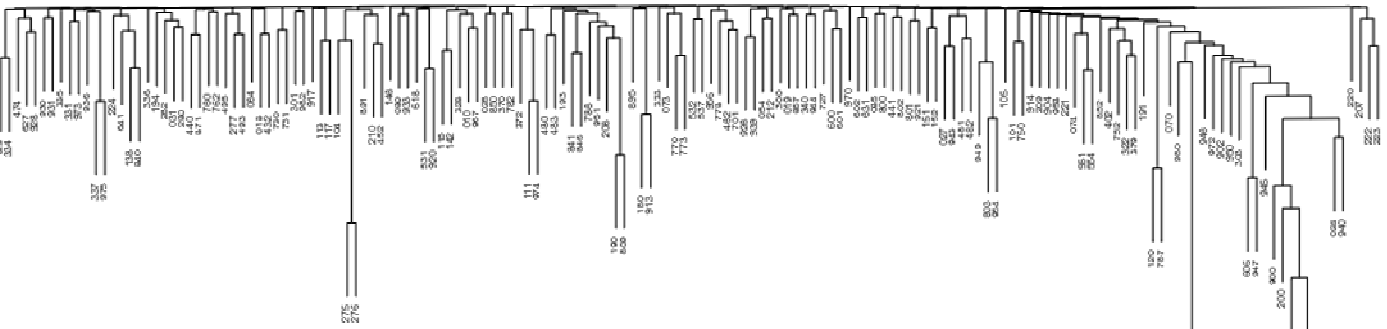
### Cluster analysis

The dendrogram from the agglomerative hierarchical clustering is shown in figure 4. There is a large cluster which consists of the major cities in the United States.

Cluster analysis finds about 30 significant clusters and dyads. The largest cluster sits in the center of the network and connects to almost all nodes and subgroup networks (Figure 3). Shown below are the sub-networks grouped by their connections, and the order of the listing is based on the clusters' distance to the central cluster (cluster 1 shown in Figure 5).

Clusters 1 (Strongest connection):  
101 (Manhattan), 112 (Brooklyn), 941 (San Francisco), 940 (south bay of San Francisco), 945 (East San Francisco Bay), 947 (Berkeley), 021 (Boston), 024 (Suburbs of Boston), 201 (DC area), 200 (DC area), 606 (Chicago), 900 (Los Angeles)

Figure 4. Dendrogram from the agglomerative hierarchical clustering



The next level of smaller subsets has much sparser connections (fewer hyperlinks), which may be the result of the fewer blogs in these smaller city units. However, these subsets show a clear geographical pattern.

- Cluster 2: 222 (Arlington), 223 (Alexandria), 220 (Fairfax), 207 (Southern Maryland)
- Cluster 3: 981 (Seattle), 980 (West Seattle), 068 (Norwalk, suburbs of NYC)
- Cluster 4 (West coast): 946 (Oakland), 950 (Santa Cruz), 902 (West Los Angeles), 972 (Portland)
- Cluster 5: 191 (Philadelphia), 120 (Albany), 787 (Austin),
- Cluster 6 (Southern League): 322 (Jacksonville), 379 (Knoxville), 752 (Dallas), 402 (Louisville)
- Cluster 7: 551 (Saint Paul, MN), 554 (Minneapolis, MN), 074 (Paterson, NJ)
- Cluster 8: 949 (North Bay), 954 (North Bay), 803 (Boulder, CA)
- Cluster 9: (Weak connections): 441 (Cleveland), 921 (San Diego), 602 (Evanston), 631 (Saint Louis), 085 (Princeton), 800 (suburbs of Denver), 802 (Denver),

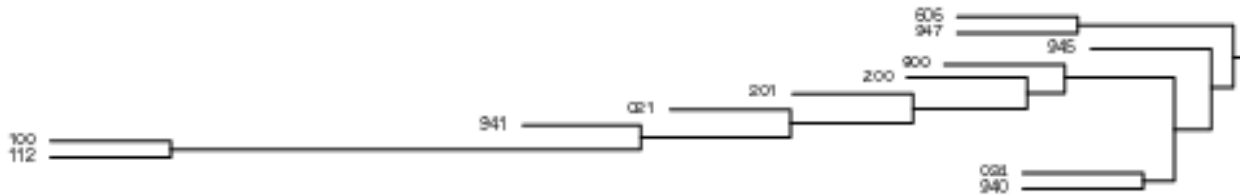


Figure 5. Clusters' distance to the central cluster

- Cluster 6 (Southern League): 322 (Jacksonville), 379 (Knoxville), 752 (Dallas), 402 (Louisville)
- Cluster 7: 551 (Saint Paul, MN), 554 (Minneapolis, MN), 074 (Paterson, NJ)
- Cluster 8: 949 (North Bay), 954 (North Bay), 803 (Boulder, CA)
- Cluster 9: (Weak connections): 441 (Cleveland), 921 (San Diego), 602 (Evanston), 631 (Saint Louis), 085 (Princeton), 800 (suburbs of Denver), 802 (Denver),
- Cluster 10: 926 (Santa Anna), 939 (Salinas), 054 (Vermont), 212 (Baltimore)
- Cluster 11 (Weak): 973 (Salem), 975 (Medford), 936 (Fresno), 337 (Miami), 331 (St. Petersburg)
- Cluster 12: 600 (suburbs of Chicago), 601 (suburbs of Chicago), 727 (suburbs of Oklahoma City)
- Cluster 13: 770 (Houston), 773 (North Houston), 073 (Newark, NJ)

The following subsets consist of dyadic notes that group with each other exclusively:

- Dyad 1: 481 (Ann Arbor), 482 (Detroit)
- Dyad 2: 532 (Milwaukee) , 537 (Madison)
- Dyad 3: 151 (Pittsburgh) , 152 (Suburbs of Pittsburgh)
- Dyad 4: 841 (Salt Lake City), 846 (Suburbs of Salt Lake city)
- Dyad 5: 480 (North suburbs of Detroit), 483 (North suburbs of Detroit)
- Dyad 6: 275 (Suburbs of Raleigh), 276 (Raleigh)
- Dyad 7: 113 (Queens), 117 (Mid Island)
- Dyad 8: 730 (Suburbs of Okalahoma City), 731 (Okalahoma City)
- Dyad 9: 760 (Fort Worth), 762 (Forth Worth)
- Dyad 10: 930 (Ventura), 931 (Santa Barbara)
- Dyad 11: 631 (San Louis), 606 (Chicago)

Subsets with underlines indicate the close geographical distance between the nodes. The list indicates that most of the clusters with strong connections also share geographical proximity.

## DISCUSSION

Blogs represent the collection of experiences and opinions of individual internet users. The networks presented in this research are based on dynamics of on-line communications among individuals. Therefore, each node in the network represents the geographical location of individual content providers instead of hosting servers. We assume that people in the same place, measured by 3-digit zip code, tend to possess some more or less collective traits. While individual difference is important, we take the view that the sum of individuals in one city reflects the general trend at the macro level.

This research finds that networks among American cities, as indicated by weblog hyperlinks, are densest among metropolitan cities on the West and the East coasts. Cities with cultural-political prominence, like Boston, San Francisco, New York, Washington and Los Angeles, traditionally seedbeds of national opinion, forge a highly connected cluster in the center of the national networks (figure 4). Meanwhile, satellite cities or suburbs around some of these cities also play a significant role in this central cluster, consistent with migration patterns in recent years. The San Francisco Bay Area has become an active cultural and technological hub, especially since the high-tech boom in 1990s. With the sharp increase of population and rising cost of housing in the area, more middle class and young people, including large number of creative workers, have moved to the suburbs, energizing and urbanizing these traditionally more conservative areas. Blogs in these cities, including surrounding areas, reciprocate hyperlinks and maintain coherent clusters. We call this group of cities the “super metropolitan cluster.” In a sense, this cluster transcends geography, though the urban/suburban connection often remains intact. Our research finds that the bloggers in these cities tend to receive more inlinks than those in other areas, beyond what would be expected given the concentration of population. The strong connection among these cities also supports Fischer’s observation that for cosmopolitan residents, close friendship are often long-distance, since urban residents move around more frequently during their lifetime and they accumulate more weak social ties (1982).

For the cities of less cultural-political significance, the connections are first and foremost with places that near them. Other than these two types of clusters, we also found less prominent clusters of connections that are based on similar city profiles. For example, the cluster made of Philadelphia, Albany and Austin, and the cluster made of Pittsburgh, Evanston, Saint

Louis, Princeton, Cleveland, Denver and San Diego seem to correspond to the prominent educational institutions and new cultural dynamics in these areas. Regional ties seem also play role in shaping the clusters of cities. The cluster made of Portland, Oakland, Santa Cruz and West Los Angeles, and the cluster made of Jacksonville, Knoxville, Louisville and Dallas seem to reflect a west-coast and a southern block sensibility, respectively.

If blogging makes possible the decentralization of publication and news resources, it might be also prompt the decentralization of large cities. However, when it comes to readership (as reflected in hyperlinked commentary on other blogs), choices made by audiences are not particularly geographically decentralized. The overwhelming number of links pouring into cities like New York, Boston, San Francisco, Washington, and Los Angeles shows that these centers of cultural and news production still attract the most attention nationwide.

This research project is limited by the technical difficulties of extracting valid link data from blogs, as well as their geographical locations. There is a lack of blog and link data from small cities, especially those in the midwest. The scale-free distribution of the network we observe based on these data suggests that a larger data set would yield similar results. Certainly a larger data set may provide more finely-grained information about local connections and structural relationships—which areas are hubs, bridges or tree nodes in the network. Nonetheless, the work here demonstrates the possibility of using links among weblogs to measure socio-geographical relationships, and suggests some interesting national patterns of discursive clustering. Future work that allows for ongoing monitoring of such relationships should provide an interesting barometer of social exchange.

## REFERENCES

- Adamic L & Glance N. 2005. The political blogosphere and the 2004 U.S. election: Divided they blog. *LinkKDD-2005*, Chicago, IL, Aug 21, 2005.
- Barnett GA, Sung EJ, Lin J, Hung H and Park H. 2003. An examination of Canada's inter-provincial networks. Paper presented in Sunbelt XXIII International Social Network Conference. Cancun, Mexico February, 2003.
- Barnett GA. 1999. The social structure of international telecommunications. In Sawhney, H and GA Barnett. (Eds.) *Progress in communication sciences, Volume XV: Advances in telecommunications*, 151-186. Stanford, CT: Ablex.
- Barnett GA & Choi Y. 1995. Physical distance and language as determinants of the international telecommunication network. *International Political Science Review*, 16 (3), 249- 265.
- Barnett GA & Park HW. 2005. The structure of international Internet hyperlinks and bilateral bandwidth. *Annales des Telecommunications*, 60(9-10), 1115-1132.
- Brunn S & Dodge M. 2001. Mapping the “worlds” of the World-Wide Web: Restructuring global commerce through hyperlinks. *American Behavioral Scientist* 44(10), 1717-1739
- Blood R. 2000. Weblogs: a history and perspective. Retrieved March 20, 2006 from: [http://www.rebeccablood.net/essays/weblog\\_history.html](http://www.rebeccablood.net/essays/weblog_history.html)
- Cairncross F. 1997. *The death of distance: How the communications revolution will change our lives*. Boston, MA: Harvard Business School Press
- Cornfield M, Carson J, Kalis A, Simon E. (2005). Buzz, blogs, and beyond: The Internet and the national discourse in the fall of 2004. Pew Internet and American Life report. Retrieved February 26, 2006 from: [http://www.pewinternet.org/ppt/BUZZ\\_BLOGS\\_BEYOND\\_Final05-16-05.pdf](http://www.pewinternet.org/ppt/BUZZ_BLOGS_BEYOND_Final05-16-05.pdf)
- DudaR, Hart P & Stork D. 2000. *Pattern Classification*, New York: John Wiley & Sons, Inc.
- Fischer C. 1982. *To Dwell among Friends : Personal Networks in Town and City*. University Of Chicago Press.
- Florida R. 2002. *The rise of the creative class*. NY, Basic books
- Halavias A. 2000. “National borders on the world wide web. *New media & society*, 1,3, 7-28
- Halavias A. 2002. Blogs and the “social weather”. Conference of Internet Research 3.0. Maastricht, Netherlands.
- Ihaka R and Gentleman R. 1996. R: A language for data analysis and graphics, *Journal of Computational and Graphical Statistics* 5, 299-314.
- King AD. 1991. Urbanism, Colonialism, and the World economy: Cultural and spatial foundations of the world urban system. London: Routledge.
- King AD. 1990. *Global Cities: Post-imperial and the Internationalisation of London*. London: Routledge.
- Kotkin J. 2001. *The new geography: how the digital revolution is reshaping the American landscape*. NY: Random House.
- Kumar R, Raghavan P, Rajagopalan S, & Tomkins A. 1999. Trawling the web for emerging cyber-communities. *WWW8 / Computer Networks*, 31, 1481-1493.
- Lin J & Halavias A. 2004. “Mapping the blogosphere in America. *Thirteenth International World Wide Web Conference*, New York City, May.
- Lin J & Halavias A. 2005. Geographical distribution of Weblogs in the US. Paper presented to ICA general conference, New York City, May
- Lin J & Halavias A. 2005. Comparison of political weblogs and mainstream media. Paper presented at the International Communication Association 2005 General Conference. NYC, May 2005
- Logan J & Molotch H. 1987. *Urban fortunes*. Berkeley: University of California Press, 1987

- Garton L, Haythornwaite C, & Wellman B. 1997. Studying online social networks. *Journal of Computer Mediated Communication*: 3(1). Retrieved February 13, 2006, from: <http://jcmc.indiana.edu/vol3/issue1/garton.html> .
- McLuhan M. 1968. *War and Peace in the Global Village*. New York: Bantam
- Mitchell W. 1999. *E-topia: "Urban life, Jim-but not as we know it"*. MA: Boston, The MIT Press
- Park R. 1984. The city: Suggestions for the investigation of human behavior in the urban environment. In Robert Park, Ernest Burgess (Eds.), *The city*. Chicago: The University of Chicago Press.
- Park HW. 2003. Hyperlink network analysis: A new method for the study of social structure on the web. *Connections*, 25(1), 49-61.
- Sassen S. 1991. *The global city*, Princeton N.J.: Princeton University Press
- Wellman B. 2001. Computer networks as social networks. *Science*, 243. 2031-2034
- Wellman B, Hasse A, Witte J & Hampton K. 2001. Does the Internet increase, decrease, or supplement social capital? *American Behavioral Scientist*, 45(3), 436-455
- Wellman B & Gulia M. 1999. Net surfers don't ride alone: Virtual communities. In Kollock P & Smith M. (Eds.) *Communities and cyberspace*.
- Wellman B, Gulia M & Potter S. 2002. The Social Network Basis of Interpersonal Resources for Coping with Stress. In *Socioeconomic Conditions, Stress and Mental Disorders: Toward a New Synthesis of Research and Public Policy*.
- Weiss M. 1989. *The clustering of America*. New York: Harper & Row Publishers.
- Winer D. 2003. *What makes a weblog a weblog?* Retrieved April 20, 2004 from: <http://blogs.law.harvard.edu/whatMakesAWeblogAWeblog>
- Zook MA. 2001. Being connected is a matter of geography. *netWorker*, 5(3), 13-16.

