

## Language Networks on LiveJournal

Susan C. Herring, John C. Paolillo, Irene Ramos-Vielba, Inna Kouper, Elijah Wright,  
Sharon Stoerger, Lois Ann Scheidt, and Benjamin Clark

*Indiana University, Bloomington*

{herring, paolillo, iramosvi, inkouper, ellwright, sstoerger, lscheidt, bjc2}@indiana.edu

### Abstract

*Language use in 1,000 randomly-selected and 5,025 crawled LiveJournals was analyzed in order to determine the overall language demographics, the robustness of four non-English language networks (Russian, Portuguese, Finnish, and Japanese), and the characteristics of individuals who bridge between different languages on LiveJournal.com. The findings reveal that English dominates globally but not locally, network robustness is determined mostly by population size, and journals that bridge between languages are written by multicultural, multilingual individuals, or else they have broadly accessible content. Implications of these findings for cross-cultural conversation via blogs are considered.*

### 1. Introduction

In recent years, patterns of sociability among weblogs (blogs) have been analyzed using link analysis [13], including within large blog hosting communities such as LiveJournal [20]. Social networks have been identified based on blog topic (e.g., politics), common interests (e.g., fandom), and online and offline friendship connections; bloggers converse within these networks by linking to and commenting on one another's content [21]. More basic than these forms of social glue, however, is language. Indeed, a shared language would seem to be the *sine qua non* for meaningful interconnection and conversation; thus, we also expect to find blogs that use the same language(s) linking to one another. Yet although the number of languages in which people blog continues to grow, no research has yet investigated language networks in the blogosphere.

This study analyzes language-based networks on LiveJournal.com, one of the first and largest blog hosting services in the world. LiveJournal is international in scope: More than two-thirds of its 11 million users self-report being outside the United States, and the site's interface allows users to set their

journal pages to appear in 32 different language varieties. At the same time, LiveJournal is based in the U.S., and English is the predominant language on the site as well as on the Internet in general [19]. This research asks to what extent other languages are actually used in LiveJournals, and for what intended audiences, where 'audience' is operationalized as other LiveJournalers designated as 'friends' (also known as 'friending') by individual journalers. Our assumption is that journalers 'friend' journals whose language they can read, and—since 'friending' is often reciprocal—that the friends can also read their journals. A question of particular interest is: At what point do non-English 'friend' networks cease to be linguistically self-contained and become integrated with other language networks, and what kind of language use occurs at those transitional boundaries?

To address these questions, we randomly sampled and identified the languages used in 1,000 LiveJournals, selecting four of the most commonly-found non-English languages—Russian, Portuguese, Finnish, and Japanese—as case studies for further analysis. Through crawling the 'friends' links of monolingual source journals from each language, and coding the languages used in the resultant sample, patterns were revealed that we represent using network visualization techniques.

Our findings show that, with the exception of Russian, non-English journals comprise a small minority on LiveJournal.com. The four non-English languages analyzed display differing degrees of network density and linguistic homogeneity, with the largest network, Russian, being the most self-contained, followed by Portuguese, Finnish, and then Japanese. LiveJournals that bridge between languages and language networks tend to have broadly accessible, non-linguistic content, or else they are written by multicultural, multilingual individuals such as expatriates and foreign language learners. We conclude by considering the implications of these findings for cross-cultural conversation and linguistic diversity on LiveJournal.com and on the Internet in general.

## 2. Background

### 2.1. Connectivity in the blogosphere

To our knowledge, this is the first study of online language networks that compares different languages. Previous studies of networks on the Web have not focused on language. Because of their tendency to start sampling from English-language sites, the default language of the networked entities has typically been English [2, 14, 20]. For example, the sources used in one study [14] included *blogger.com*, *www.meme-pool.com*, *metafilter.com*, and *blogs.salon.com*, which consisted at the time mostly of English-language blogs. In an essentially monolingual blogspace, it may safely be assumed that authors of interlinking blogs understand (even if they do not necessarily agree with) one another.

There is evidence that bloggers link to similar-minded bloggers. Adamic and Glance [1] analyzed political blogs during the 2004 U.S. presidential election and found linking among, but rarely between, blogs with different ideological positions. Conservative bloggers linked to each other more densely than did liberal bloggers. Herring et al. [13] crawled links from English-language blogs randomly selected through the blog tracking site *blo.gs*, and identified topical cliques of interlinked blogs, including Catholic bloggers, homeschooling bloggers, and (conservative) political bloggers; of these, the Catholic bloggers linked to each other most densely. In contrast, in a study of *LiveJournal.com*, Paolillo and Wright [20] found that networks of 'interests' listed in user profiles were independent of their 'friend' (link-based) networks.

Two recent studies examined blog conversations in terms of blog entries and comments, in addition to links. Herring et al. [13] found evidence of conversational interaction through blog posts and comments in interlinked blog pairs within the three cliques they identified. Efimova and de Moor [10] analyzed an extended conversation carried out through entries and comments in the first author's knowledge management weblog community. Their analysis revealed some sub-conversations in German, in addition to English. The multiple language usage in this case can be attributed to the international nature of this weblog community, the members of which were located in both Europe and the U.S.

### 2.2. Language demographics in the blogosphere

English has predominated historically in the blogosphere, and by most estimates, continues to do

so. In August 2004, the National Institute for Technology & Liberal Education (NITLE) estimated that 61.9% of the 2.1 million weblogs visited by their crawlers were written in English [17]. By November 2005, NITLE's estimate of English weblogs had risen to 68.7% of the 2.9 million weblogs visited [18].

At the same time, as blogging increases in global popularity, more bloggers are utilizing their native languages on their sites. As early as August 2002, Brazilians were reported to be the second-largest users of Blogger software, comprising 13% of Bloggers' 750,000 users [11]. By 2004, the growth of Portuguese-speaking Brazilian users on *orkut*, a Google-owned social networking site that allows for diary weblogs, had become a point of contention with the site's English-speaking community. The Portuguese-speaking population of the site was estimated at 41.2% of the 769,000 users, while English-speakers made up only 23.5% [11]. As of September 11, 2006, the number of Brazilian *orkut* users had reached 65% of the total users, followed by the U.S. with 13.5% [26].

Portuguese speakers are not the only group increasing its presence online. In 2002, Iranians were the third largest nationality on *orkut* [11]; some diasporic Iranians are also active political bloggers [8]. Diary blogging has been reported to be popular in Poland, especially among young females [24]. By January 2005, it was claimed that 25% of all South Koreans had weblogs [4]. Other developed countries, such as Germany, have shown a low blogging adoption rate [15]. While these estimates are by country rather than language used, it is likely that many of the sites counted were non-English language weblogs.

Reports of language demographics in the blogosphere vary considerably depending on their source. In NITLE's [18] estimates, Catalan ranks a distant second after English, followed by French, Spanish, and Portuguese. In contrast, Technorati reports that Asian languages are the current top blogging languages, with Japanese-language weblogs making up the largest language tracked with 37%, compared to 31% for English [22]. Chinese-language weblogs are the third largest category with 15%, and Spanish comes in a distant fourth with 3%, followed by Italian, Russian, French, Portuguese, and Dutch at 2% each [22].

Part of the discrepancy between these reports is due to the fact that large Asian blog hosting services are not systematically tracked by U.S.-based blog tracking sites such as *blo.gs* and NITLE. Neither, for that matter, is *LiveJournal.com*, which is home to a large Russian community [12]. The automated language identification tools used by some tracking sites are also unreliable; for example, the tool used by NITLE, *TextCat*, has a tendency to misidentify Spanish,

(2007, January). *Proceedings of the Fortieth Hawai'i International Conference on System Sciences (HICSS-40)*. Los Alamitos: IEEE Press.

French, and Russian blogs as Catalan.<sup>1</sup> We attempted to use automated language classification tools in this study, and also found them lacking, as described in section 3.3.2.

As this section has shown, weblogging has become popular not only in English-speaking countries but in many places around the world. From the limited data available, it appears that the enthusiasm with which blogging is taken up does not necessarily correspond to the population size or degree of technological advancement of a country, but may depend on political, historical, and cultural factors; these have yet to be systematically elucidated. In the meantime, there is a need for reliable descriptive, comparative data about language use in the blogosphere.

### 3. Methodology

#### 3.1. LiveJournal as a data source for language analysis

We chose to study language use on LiveJournal.com for several reasons. First, it is a large, self-contained blog hosting service, reporting over 11 million blogs created since 1999, of which a significant portion are active. Second, it explicitly targets an international population, and two-thirds of LJ users self-reported blogging from outside the U.S. at the time of our analysis in March 2006. Moreover, as a hosting environment it is welcoming to non-English bloggers, in that it offers blog templates in 32 language varieties and supports Unicode. The constraints on use of non-ASCII fonts identified in previous discussions of multilingualism on the Internet [7] do not seem to apply to LiveJournal; every language we were able to identify displayed in its native font in our browsers, except when bloggers chose to represent it in another font.

LiveJournal.com also has features that facilitate network analysis. These include a 'friends' page for each journal on which recent entries by other LiveJournalers designated as one's 'friend' can be read together, and FOAF (friend-of-a-friend) data in user profiles that allow links to 'friends' journals to be tracked automatically and reliably.

Along with these advantages there are limitations. As a relatively self-contained 'community' (as it refers to itself), LiveJournal has its own culture and practices that do not necessarily represent those of other blog hosting services or the blogosphere as a whole [16]. Most relevant to the present study, it is based in the U.S., and for that reason may be expected to over-

represent North American culture and the English language. However, many Internet services based in North America are used extensively by people in other countries, due to the U.S.'s historical dominance of networking technology [19]. At least one, orkut, has been thoroughly embraced by non-English users, despite being owned by Google, a corporation headquartered in the U.S. LiveJournal.com is thus an example of a not uncommon phenomenon, an 'international' social networking site based in the U.S.,<sup>2</sup> with the power dynamics and potentials that this situation entails.

#### 3.2. Research questions

In this study we aimed to answer three questions. First, what languages are actually used on LiveJournal? The self-reported user statistics on geographical location available on the LiveJournal site cannot be taken as a reliable indicator of language use: People do not necessarily blog in the language of the country they reside in, and some users report inaccurate, even obviously fraudulent, locations (such as 'Antarctica')<sup>3</sup>. We aim to address this question empirically.

Second, how robust are the networks formed by LiveJournalers in languages other than English? This question stems from an interest in how well other languages can sustain themselves in an English-dominant online environment, and what factors favor their maintenance and spread [7]. We consider a network to be 'robust' if its nodes link to other nodes in the same language, and if it is large and dense with a well-defined core, as represented in social network visualizations [cf. 25]. In this study, robustness is a relative, rather than an absolute, measure.

Third, where are the transition points in the networks between one language and another, and what are the characteristics of bloggers who occupy a bridging position in the language network structure? Sociolinguistic network theory tells us that individuals who belong to multiple speech communities may bridge linguistically between them, like couriers carrying news from one locale to the next, in the process uniting them in a larger community and promoting diffusion of linguistic practices [6]. We aim to shed light on the blog 'conversations' taking place at the transitions between language networks, and consider their implications for the larger question of

<sup>1</sup> <http://www.languagehat.com/archives/000942.php>

<sup>2</sup> LiveJournal.com has been owned since January 2005 by Six Apart, the creators of MovableType and TypePad weblog software, which is based in San Francisco, California.

<sup>3</sup> Whereas 4,467 LiveJournalers report blogging from Antarctica, the total population of Antarctica is less than 4,000 in the summer and 1,000 in the winter (<http://www.indexmundi.com/antarctica/population.html>, retrieved September 7, 2006).

(2007, January). *Proceedings of the Fortieth Hawai'i International Conference on System Sciences (HICSS-40)*. Los Alamitos: IEEE Press.

cross-language interaction, i.e., to what extent do the different language communities on LiveJournal understand and communicate with one another?

### 3.3. Analytical methods

**3.3.1. Random journal analysis.** To address the first research question, we randomly sampled 1,000 LiveJournals using the 'view a random journal' feature of the LiveJournal web site, and coded them for the languages used. Language identification was done manually by the authors, the first two of whom have advanced degrees in linguistics and have studied 18 languages between them, including Japanese, and the third and fourth of whom have native expertise in Romance languages and Slavic languages, respectively. The criteria for coding a journal as a particular language were 1) that a majority of its entries be in that language, and 2) that the coded content have been created by the blogger (rather than quoted from an external source). The language of the template, and, to a lesser extent, the name of the journal were disregarded in making the language assignment, as these often differed from the language used in the entries. A journal was coded as mixed (e.g., Japanese/English) if it made roughly equal use of two or more languages in its entries.

**3.3.2. Language network analysis.** In order to investigate the structure of language networks on LiveJournal, we used the findings from the coding of the random sample to select four languages: Russian, Portuguese, Finnish, and Japanese. These languages were selected because they were found to be among the most common non-English languages on LJ, and because they contrast genetically and in their writing systems. Russian is a Slavic language written in the Cyrillic script. Portuguese is a Romance language written in the Roman alphabet modified by diacritic marks, i.e., ã. Finnish is a Finno-Ugric language written in the Roman alphabet modified by different diacritics, e.g., ä. Japanese has not definitively been established to be related to any other known language (its likely closest relation is Korean); it has three writing systems: Chinese-style characters; a syllabary for native Japanese words; and a syllabary for foreign words. By way of contrast, English is a Germanic language written in Roman script with no diacritics. We hypothesized that writers of languages more similar to English might be more attracted to LiveJournal and find it easier to use than writers of completely unrelated languages.

For each of the four languages, we selected six monolingual LiveJournals (for a total of 24) as sources or 'seeds' from which to crawl the language networks.

In the case of Russian, the seeds were drawn from the initial random sample of 1,000 LJs. For the other three languages, the random samples were supplemented by searching within LJ for the words 'music,' 'dance,' and 'computer' (common interests among LJ users) in the respective languages, and selecting the first monolingual journals with friends links that came up in the search results. In the case of Japanese, this method did not produce a sufficient number of monolingual journals; the list was completed by identifying monolingual Japanese members of two Japan-focused LJ community journals.<sup>4</sup>

From the 24 seed journals, we then constructed a snowball sample by crawling two degrees out along the network. The network crawls were carried out by retrieving the RDF/XML FOAF profiles for the users, extracting the friend lists for each, and using that information to seed the next level of profile retrievals. This information was stored and imported into a PostgreSQL database for later analysis.

For journals identified by the crawl, we initially attempted to use TextCat (<http://www.let.rug.nl/~van Noord/TextCat/>) for automatic language identification, but abandoned it when it failed to recognize Unicode (e.g., Russian UTF-8). We then tried a different tool, languid (<http://languid.cantbedone.org/>), which had been trained on 74 languages including some using UTF-8, but it also returned an unacceptably high error rate (about 40%). We finally resorted to hand coding the journals for language as we had done for the original random sample, using the same criteria. Hand coding, while time-consuming, produced more reliable language identifications. All journals were coded at degrees D0 and D1; for journals located in degree D2, we sampled a 1:20 selection and determined the languages used. Degree 0 contained 24 journals; degree 1 contained 1,871 journals, and degree 2 contained 63,900 journals, from which we sampled a set of 3,130, for a total of 5,025 journals coded.

By combining the database of friend relationships with the languages from our hand coding, we were able to extract user and network edge lists with attached language attributes. We used the R statistical computing environment (<http://www.r-project.org/>) to load these lists, generate appropriate matrices for social network analysis, and produce network visualizations for each of the four languages. The visualizations use the Fruchterman-Rheingold layout algorithm of the SNA package in R [5].

---

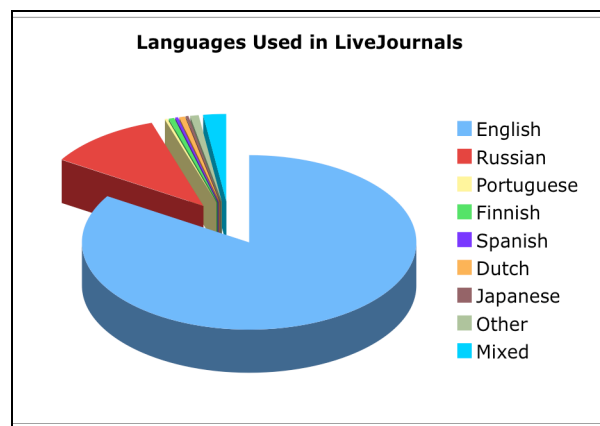
<sup>4</sup> No community journals were included in the seed sample or the crawls, however.

**3.3.3. Bridge analysis.** After the journals in the crawl were coded and the network visualizations generated, it was possible to identify journals that, through their position in the network and language use, bridge between different languages. For this stage of the analysis, three types of bridges were selected from the PostgreSQL database: 1) journals coded as using two or more languages, i.e., 'mixed' types (N=59); 2) journals occupying the middle position in a directional chain of 'friending' (i.e.,  $A \rightarrow B \rightarrow \text{not } B$ , where A and B are different language codes) (N=839); and 3) journals that connect different language networks, regardless of the directionality of linking (N=68).<sup>5</sup>

All of the bridge journals of types (1) and (3), and a random sample of 300 of type (2), were examined qualitatively in an attempt to explain their position between different languages. Factors considered were language competence (does the bridge blogger appear to know the languages of the neighboring journals?), blogger demographics (e.g., gender, age, occupation, place of residence), blog features (e.g., presence of non-verbal content), and number of friends.

## 4. Results

### 4.1. Language used on LiveJournal.com



**Figure 1. Languages of 1,000 randomly-selected LiveJournals, sampled March 2006**

English is overwhelmingly the language used by LiveJournal bloggers. English was the language of 84% of the journals in the random sample, including many written by people living outside the U.S. and by people, both in English speaking countries and elsewhere, for whom English was not their native

<sup>5</sup> The three categories overlap to some extent. Directional chains (Type 2) include 29 mixed-language journals (Type 1). Inter-network bridges (Type 3) include two chains (Type 2) and one mixed-language journal (Type 1).

language. The next most commonly-used language was Russian with 11%, with Portuguese, Finnish, Spanish, and Dutch tying for third place with .4%, followed by Japanese with .3%. Mixed language journals made up 2.3%, with all other languages combined adding up to .8%. This distribution is shown in Figure 1.

From this broad perspective, with the exception of Russian, LiveJournal does not appear to be very multilingual. In the next sections, we look more closely at non-English LiveJournalers, setting aside the broader English context to focus on bloggers in minority languages and their friends.

### 4.2. Networks of four non-English languages

Our second research question concerns the robustness of non-English language networks on LiveJournal.com. Several measures bear on this question. First, we observed that journals in the different seed languages have different numbers of friends, as shown in Table 1. The language from these four that is most represented on LiveJournal, namely Russian, is also the one whose users have the greatest number of friends. Similarly, Portuguese and Finnish users are less represented on LJ and have fewer friends, and Japanese has the smallest presence and the fewest friends per journal.<sup>6</sup> It appears that in general, the greater the presence of a language on LJ, the larger the universe of potential friends its speakers have to choose from, and the more friends they are likely to have.

**Table 1. Number of friends per journal by language**

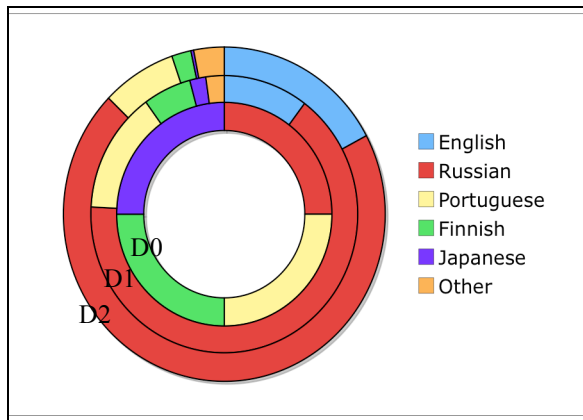
Language	Mean	Range
Russian	231.5	92-380
Portuguese	169.5	38-480
Finnish	60.0	17-180
Japanese	20.5	5-48

Number of friends has evident consequences for the size of the networks produced by crawling out from friend links. Figure 2 shows the proportion of journals in each language at each crawl degree. At degree D0 in our crawled sample, the four languages are represented equally (25% each); at degree D1, Russian constitutes 65% of the sample, growing to 70% at degree D2, while the proportions of journals in languages with fewer friends decrease accordingly. In Figure 2 and other figures, the same colors are used to indicate each

<sup>6</sup> The difference in number of friends between Portuguese and Finnish journals, which appeared with roughly equal frequency in our random sample, may reflect cultural differences in sociability; see the discussion of Portuguese use of the orkut social networking site in section 2.2 and section 5.



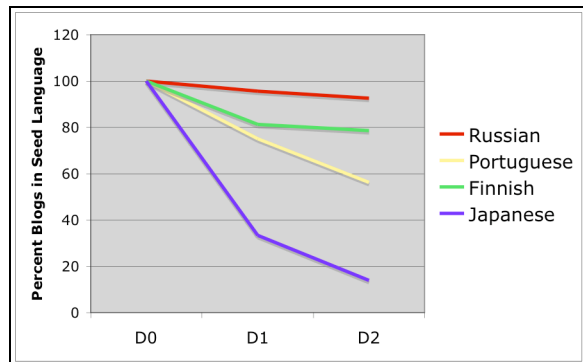
language in our sample: English (blue), Russian (red), Portuguese (yellow), Finnish (green), Japanese (violet), and other (orange).



**Figure 2. Distribution of seed languages at three degrees**

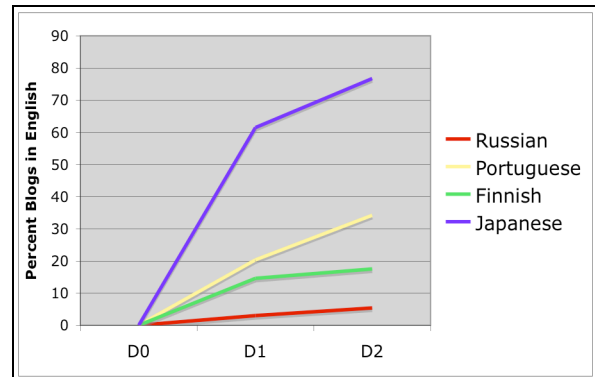
Figure 2 also shows the rate and extent to which English permeates the sample. English journals account for 0% at D0, 10.3% at D1, and 17.3% at D2. The rapid introduction of English is to be expected, given that 84% of LiveJournals are in English. What is striking is that at this level of analysis, it is not English, but Russian, that dominates.

The size of the presence of a language in LiveJournal also correlates roughly to the persistence of that language in the network, where 'persistence' is defined as proportion of journals in the seed language at each successive degree of crawl. As Figure 3 shows, Russian is highly persistent, followed by Finnish and Portuguese. Japanese, in contrast, drops rapidly from 100% to 33% at D1 and 14% at D2.



**Figure 3. Persistence of seed languages in their networks**

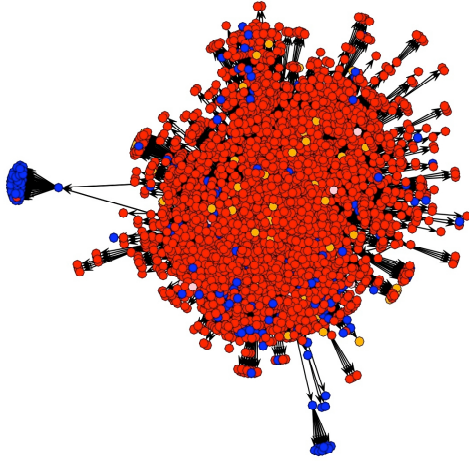
Figure 4 shows the proportion of journals friended at each degree that are in English. The pattern is almost the mirror image of that in Figure 3: When non-English journals friend a journal in another language, that language is almost always English. Figure 4 shows that all four languages converge towards English, albeit at different rates. Japanese converges especially rapidly, while Russian shows minimal English incursion even at D2.



**Figure 4. Convergence with English in seed language networks**

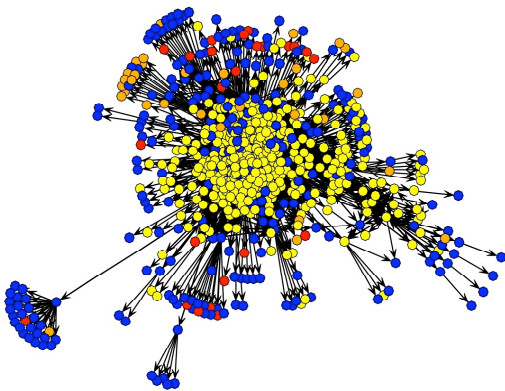
The results for number of friends/network size and network persistence suggest a continuum of language network robustness—corresponding roughly to the number of journals present on LiveJournal.com in a given language—with Russian being most robust, followed by Portuguese or Finnish (depending on the measure), and Japanese being the least robust.

The network visualizations add further support to this observation in terms of network density and centrality. The Fruchterman-Rheingold layout used in the plots pulls together well-connected nodes, and pushes to the edges nodes that are less well-connected, enabling a clear visual comparison of density and centrality (presence or absence of a core) across language networks. Figures 5 through 8 were created by selecting the seed journals from one language and displaying them as nodes, along with all of the coded journals that were reachable in two degrees. This means that all of the journals at degrees D0 and D1 are displayed, as well as those from the random sample at D2 that connect to the start set. In all four figures, mixed-language blogs (e.g., Japanese/ English) are represented in a shade related to, but lighter than, the color of the non-English language. For optimal effect, these figures should be viewed in color.



**Figure 5. The Russian language network**

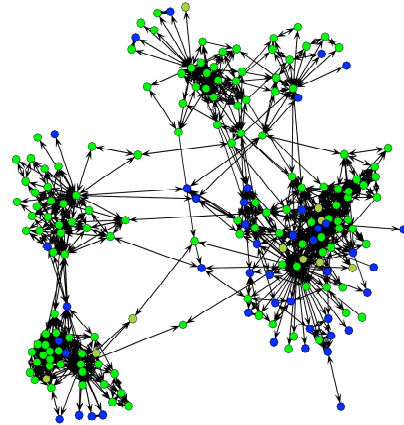
The Russian language network in Figure 5 has a large number of well-connected nodes that are almost entirely Russian. A few English and other language journals (e.g., Ukrainian and Romanian) are found in the dense central core, while the occasional concentrated clump of English is pushed to the edges. For Russian users of LJ, it appears that it would not be too difficult to connect through friends to English journals, although most of one's LJ experience would be in Russian. Monolingual Russians would find an abundance of journals to read and interact with on LJ.



**Figure 6. The Portuguese language network**

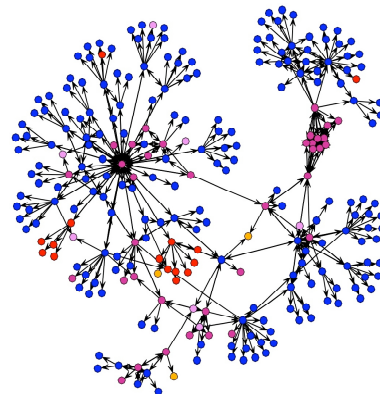
The Portuguese network in Figure 6 is similar to the Russian network of Figure 5, but looser in composition. There is a fairly large, tight cluster of Portuguese journals at the center, with a periphery characterized by a preponderance of English. A few Russian journals are also scattered throughout the periphery of the network, in addition to other Romance languages (e.g., Spanish and Italian). The smaller population of Portuguese LJ users appears to be reflected in the somewhat smaller network, the slightly

greater variety of languages, and the larger proportion of English found in the periphery. A Portuguese LJ user could converse in Portuguese with numerous journals, but would be reminded often of the presence of English LiveJournalers.



**Figure 7. The Finnish language network**

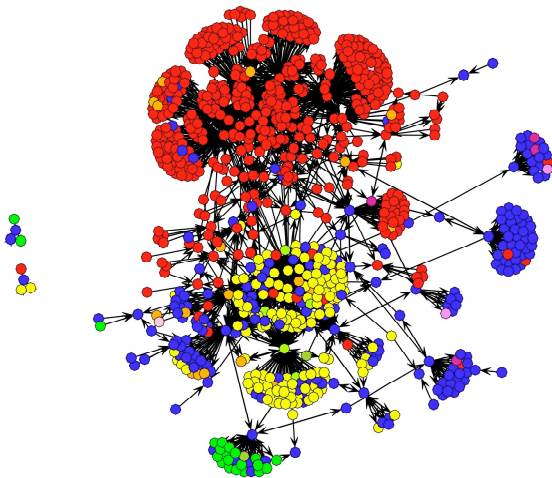
The Finnish network in Figure 7 is qualitatively different. It continues the trend observed in the Portuguese network in that it is smaller and looser, but unlike the previous networks, it lacks an apparent center. Rather, each seed journal and its friends constitutes an independent cluster. Moreover, rather than appearing mainly on the margins of the network, English appears in the centers of the clusters, and even in positions bi-directionally bridging between Finnish journals. This pattern is suggestive of a high degree of Finnish-English bilingualism among Finnish LiveJournalers; indeed, most of the English journals in this network appear to be written by Finns. Thus Finns have conversations on LJ in both Finnish and English, but mostly among themselves. The relation of this network to other language networks appears to take place through connection to English journals.



**Figure 8. The Japanese language network**

The Japanese network in Figure 8 is even sparser than the Finnish network. With the exception of a small tight clump of Japanese that appears towards the upper right, the source journals neither have dense clusters of links nor do they connect closely to one another. English or mixed English/Japanese surrounds most of the Japanese journals. At the same time, some presence of Russian is apparent, so the Japanese network is more multilingual than the Finnish network. It appears that it would be difficult to have extensive conversations in Japanese on LJ; Japanese users would also have to know English.

Some of the journals in Figures 5-8 connect to seed journals in more than one language; hence, they are actually plotted in more than one language network. By designing the necessary SELECT operations in the relational database we were able to identify the journals that bridge between different language networks, along with their immediate neighborhoods (journals linking to or linked from them in their friends lists). The resulting network is visualized in Figure 9, using the same layout method and node coloring as in the previous figures.



**Figure 9. Network surrounding the bridges connecting the four language networks**

Figure 9 provides a view of the relations among the different language networks. The four language networks are clearly individuated in this diagram, but their interconnections are also revealed. The densely-connected Russian network is visible at the top of the diagram, linked to some extent to Portuguese (middle of the figure) and Japanese (right of the figure, mixed with English). Portuguese occupies the most central position in this network, with many links to clumps of English, while Japanese and Finnish are peripheral, and linked through English. Thus English plays a dual role in the overall network, as both periphery (the larger

English-dominant context) and connective glue (as a lingua franca between groups that otherwise do not share a common language). The characteristics of journals that bridge between languages are examined further below.

### 4.3. Bridging journals

Our qualitative analysis led to the identification of three main patterns among the bloggers who bridge between different languages on LiveJournal. The first pattern corresponds to students of a foreign language who wish to practice their skills in that language and to reach a broader audience through LJ. They may post in their native language to communicate with their friends, but also use the language they are learning to contact people with different degrees of linguistic competence, independent of their interlocutors' geographical location. For example, a German woman in her third year of Asian studies keeps an LJ in German for her German friends, and another in Japanese that links to Japanese LJs as well as LJs of German and English Japanophiles. Similarly, a young Malaysian man living in Australia is studying Japanese; his LJ alternates among Malay, English, and Japanese. His friends and those who friend him include many bilingual or partially bilingual individuals, among them Anglo learners of Japanese, and Japanese learners of English.

The second pattern is expatriates bridging between their native language and the language of the place where they currently live. Expatriates use LJs to maintain their former social networks and also to participate in new networks. An example where the emphasis is on maintaining contact with former networks is the friends-only LJ of a Russian living in Egypt, whose description is in Arabic, but which lists Russian friends, and is linked to by many Russians. Despite the fact that the LJ presents itself in Arabic, its neighborhood of friends situate it in the core monolingual Russian LJ network. An example of a fully-integrated expatriate is a young Japanese man studying in the U.S., who blogs in English and links mostly to English-language blogs in the U.S. and Japan. He is fully bicultural, but orients more towards his place of current residence than where he is from.

Extensive use of non-verbal content is characteristic of the third pattern of LiveJournal bridges. Bridge journals—especially those that bridge between two different language networks—tend to favor universal modes of expression such as photographs, graphics, and popular song lyrics that allow users to interact with little or no understanding of the language of the journal. A number of inter-network bridges are professional or semi-professional photographers;



(2007, January). *Proceedings of the Fortieth Hawai'i International Conference on System Sciences (HICSS-40)*. Los Alamitos: IEEE Press.

others include celebrity fan site maintainers, graphic designers, and fashion designers. Such bridges tend to link to and be linked from several different languages, and have many friends. Some are explicitly multilingual; for example, a photo-journaler based in Spain invites comments in Spanish, Galician, Portuguese, or English. Others blog in one language only, but link to many others, such as the Portuguese photo-journaler whose hundreds of friends include English, German, Russian, and Portuguese journals.

Bridge journalers thus tend to be multilingual and multicultural, although some simply blog about content with international appeal. As a population, they probably reside outside the place where their first language is natively spoken more often than do non-bridge LJ authors; they also appear to be slightly older and more educated. Whereas the average LiveJournaler is about 18 years old, according to self-report statistics on the LJ site, bridge journalers are typically college student-aged, with some older professionals. They appear equally likely to be male as female, despite the female majority (67.6%) self-reported on the LJ site. This may reflect the preponderance of Russians in our sample—it is our impression that more male than female Russians have LJs. Gorny [12] also claims that Russians on LJ are older than LJ users on average.

Just as not all bridge journalers know foreign languages, not all use of foreign languages is found in bridges. LiveJournals frequently use languages other than the main language of the journal in limited ways—in templates, titles, photo captions, lyrics, poems, and quotations. English is especially commonly found in these positions, but French, German, Spanish, Japanese, and other languages also appear in formulaic or emblematic uses. In this way, journalers can connect symbolically with speakers of other languages, without possessing real competence in foreign languages. The frequency of this practice suggests that many LiveJournalers, even those who lack the linguistic or non-verbal means to interact directly with speakers of other languages, are aware that LJ is a cosmopolitan environment, and orient to that fact, albeit superficially. We take this as further evidence that the language networks on LJ are not entirely discrete, but that transitional places exist at the boundaries of the networks where mutual other-language and other-culture awareness enables varying degrees of cross-network interaction.

## 5. Discussion

In contrast to what might be concluded from the site's user location statistics, the results of this study show that English is the dominant language on

LiveJournal.com. They also support Gorny's [12] claim that Russian dominates the non-English portion of LJ, in numbers and also in linking activity. Given that Russians are not among the most frequent participants on other social networking sites, the question arises as to why they flock to LJ. Gorny [12] suggests that the reasons for this are historical and cultural. The first Russian LJ users heard about the site from American friends while studying in the U.S. in 1999, and subsequently promoted it among their friends in Russia. Gorny [12] attributes their community-oriented behavior online to the Russian 'national character' which includes a need for intense affiliation, which is supported by the 'friending' and 'community journal' features of LJ. Russian participation on LJ lends further support to the observation that even within a U.S.-based service, other language groups can build a significant presence.

In explanations of their presence on orkut, Brazilian Portuguese speakers have also been claimed to be highly community oriented as well as geographically mobile, resulting in a large Portuguese diaspora in which online social networking is valued as a way to maintain contacts [3]. These cultural tendencies may help to explain the attraction of LJ to Portuguese speakers as well. However, social networking sites are also very popular in Japan [23], so culture-based explanations of variability in LJ language networks must be balanced with other explanations.

Regarding the robustness of non-English language networks, our main finding is that population size matters. The number of LJs in the languages in our study corresponds roughly to degree of network persistence, density, and centrality. Russian and Portuguese LJs have attained a critical mass sufficient for a (relatively) persistent, dense, centralized network to emerge for each language, while Finnish and Japanese have fewer LJ users and less persistent, sparse, decentralized networks. Our hypothesis is also partially supported that languages similar to English will be found more on LJ than languages that are genetically and typographically different. Russian and Portuguese, both Indo-European languages, are well-represented on LJ, whereas Finnish and Japanese, which are genetically unrelated to English or each other, have a more marginal presence. According to the criterion of typographical similarity, however, Portuguese should be used more than Russian, which is not the case.

Our qualitative examination of the transitions between networks revealed that bridge bloggers play an important, albeit probably unintentional, role in connecting language networks on LJ, much as bridge speakers connect face-to-face social networks [6]. Young, multilingual, geographically mobile bloggers

(2007, January). *Proceedings of the Fortieth Hawai'i International Conference on System Sciences (HICSS-40)*. Los Alamitos: IEEE Press.

link to, and are linked by, journals in different language groups, creating de facto bridges across cultures. Visual cultures and English-language popular culture also provide widely-accessible content that connects bloggers who otherwise may not share a common language, calling into question our initial assumption that meaningful interconnection among blogs can only take place through verbal language. The rise in popularity of videoblogs and amateur video dissemination on the Internet may lead to further cross-cultural contact, although it is a matter for future debate whether such contact should be considered 'conversational.' Finally, consistent with previous studies of Internet multilingualism [9], we found that English serves as a lingua franca for many LJers from different language backgrounds who choose to blog in English rather than their native language.

## 6. Conclusions

LiveJournal.com, although international in scope, is not as multilingual as it might be, despite the friendliness of the site design to other languages. This is due to historical, social, and economic factors that favor English use in cross-cultural online contexts [9, 19], especially when the context is based in the U.S. At the same time, the Russian network on LJ is healthy and growing, and other languages could increase in number and build networks in similar fashion, creating language-specific blogospheres within LJ. The 'friending' and 'community journal' features of LJ, along with the 'friends only' option and the general autonomy that the service affords users, support both monolingual interaction (for those who wish to blog in their native languages) and cross-language interaction.

Given the growth and importance of social networking sites such as LiveJournal.com, these findings have implications for cross-cultural conversation and linguistic diversity on the Internet more generally. Although English retains the founder's historical advantage in many online contexts, the Internet is globalizing and becoming increasingly multilingual [7]. While use of English as a lingua franca is also increasing [9], evidence such as that documented here of robust and largely monolingual non-English weblog networks may serve as a corrective to the tendency of many North American observers to view language use online as a zero-sum game. Rather, the present study suggests that trends towards English use and other language use co-exist on the Internet, along with the tendency for bridging individuals to blur the boundaries between language groups.

Inevitably, this study has certain limitations. Our findings are based only on LJ; they may not apply to other blog hosting services in the U.S., or to similar services in other countries. Moreover, only four languages on LJ were examined; further research is needed to determine the robustness of languages such as Spanish and Dutch that appeared in our random sample roughly as often as Portuguese and Finnish, to determine what factors besides population size influence network patterns. Further, for reasons of manageability, our network analysis was based on only two degrees of crawl, and only a partial analysis of the second degree. Subsequent degrees should be crawled and visualized in order to generate a comprehensive picture of the language networks and how they fit together. Finally, more research is needed on the phenomenon of bridge bloggers, to identify the factors that lead bilingual individuals to blog in one language rather than another, and to investigate more fully the linking and communication patterns of bloggers who favor non-verbal over verbal content. These practices appear to play a crucial role in integrating multiple language networks in the blogosphere.

## 7. References

- [1] Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 U.S. election: Divided they blog. *Workshop on the Weblogging Ecosystem, WWW2005*. Retrieved June 9, 2006 from <http://www-idl.hpl.hp.com/blogworkshop2005/adamic.pdf>
- [2] Adar, E., Zhang, L., Adamic, L. A., & Lukose, R. M. (2004). Implicit structure and the dynamics of Blogspace. *Workshop on the Weblogging Ecosystem, WWW2006*. Retrieved June 9, 2006 from <http://www.hpl.hp.com/research/idl/papers/blogs/blogs-pace-draft.pdf>
- [3] Baker, L. (2006, March 9). Why Brazil loves orkut! *Search Engine Journal*. Retrieved September 12, 2006 from <http://www.searchenginejournal.com/?p=3082>
- [4] Blog statistics and demographics II (2006, April). *Caslon Analytics*. Retrieved June 5, 2006 from <http://www.caslon.com.au/weblogprofile1.htm>
- [5] Butts, C. T. (2004). *sna: Tools for Social Network Analysis. R package version 0.50-0*. Retrieved June 8, 2006 from <http://erzuli.ss.uci.edu/R.stuff>
- [6] Chambers, J. (2002). *Sociolinguistic Theory: Linguistic Variation and Its Social Significance*, 2nd Edition. Oxford: Blackwell.

(2007, January). *Proceedings of the Fortieth Hawai'i International Conference on System Sciences (HICSS-40)*. Los Alamitos: IEEE Press.

[7] Danet, B., & Herring, S. C., Eds. (In press). *The Multilingual Internet: Language, Culture and Communication Online*. NY: Oxford University Press.

[8] Doostdar, A. (2004). 'The vulgar spirit of blogging:' On language, culture, and power in Persian weblog-estan. *American Anthropologist*, 106(4), 651-662.

[9] Durham, M. (2003). Language choice on a Swiss mailing list. *Journal of Computer-Mediated Communication*, 9(1). Retrieved June 14, 2006 from <http://jcmc.indiana.edu/vol9/issue1/durham.html>

[10] Efimova, L., & De Moor, A. (2005). Beyond personal webpublishing: An exploratory study of conversational blogging practices. *Proceedings of the Thirty-Eighth Hawai'i International Conference on System Sciences (HICSS-38)*. Los Alamitos: IEEE.

[11] Gallagher, D. (2002, August 6). Free weblog service and a vampire, too. *New York Times*. Retrieved June 1, 2006 from <http://www.nytimes.com/2002/08/26/technology/26BLOG.html?ex=1149566400&en=0b19cfdc999ccafe&ei=5070>

[12] Gorny, E. (2004). *Russian LiveJournal: National specifics in the development of a virtual community*. Retrieved June 14, 2006 from [http://www.ruhr-uni-bochum.de/russ-cyb/library/texts/en/gorny\\_rl\\_j\\_2.htm](http://www.ruhr-uni-bochum.de/russ-cyb/library/texts/en/gorny_rl_j_2.htm)

[13] Herring, S. C., Kouper, I., Paolillo, J. C., Scheidt, L. A., Tyworth, M., Welsch, P., Wright, E., & Yu, N. (2005). Conversations in the blogosphere: An analysis 'from the bottom up.' *Proceedings of the Thirty-Eighth Hawai'i International Conference on System Sciences (HICSS-38)*. Los Alamitos: IEEE.

[14] Kumar, R., Novak, P., Raghavan, S., & Tomkins, A. (2003). On the bursty evolution of Blogspace. *Proceedings of the Twelfth International World Wide Web Conference*, Budapest, Hungary.

[15] Lumma, N. (2004). *The German blogosphere: Some facts and figures*. Presented at the meeting of the *BlogTalk 2.0*, Vienna, Austria.

[16] Marlow, C. (2006). Investment and attention in the weblog community. *Proceedings of the AAAI Spring Symposium on Computational Approaches to Analysing Weblogs*. Palo Alto, CA: AAAI Press.

[17] NITLE Weblog Census (2004, August 16). *National Institute for Technology & Liberal Education (NITLE)*. Retrieved August 16, 2004 from <http://www.blogcensus.net/>

[18] NITLE Weblog Census (2005, November). *National Institute for Technology & Liberal Education (NITLE)*. Retrieved January 15, 2005 from <http://www.blogcensus.net/>.

[19] Paolillo, J. C. (2005). Language diversity on the Internet. In S. Ellis (Ed.), *Measuring Linguistic Diversity on the Internet* (pp. 43-89). Paris: UNESCO.

[20] Paolillo, J. C., & Wright, E. (2005). Social network analysis on the Semantic Web: Techniques and challenges for visualizing FOAF. In V. Geroimenko & C. Chen (Eds.), *Visualizing the Semantic Web*, 2<sup>nd</sup> ed. Berlin: Springer.

[21] Scoble, R., & Israel, S. (2006). *Naked Conversations: How Blogs are Changing the Way Businesses Talk with Customers*. Hoboken, NJ: John Wiley & Sons.

[22] Sifry, D. (2006, May 1). State of the blogosphere, April 2006 Part 2: On language and tagging. *Sifry's Alerts*. Retrieved May 18, 2006 from <http://www.sifry.com/alerts/archives/000433.html>

[23] Terada, S. (2006, June 21). Web connects users to new reality. *The Japan Times*. Retrieved September 12, 2006 from <http://classified.japantimes.com/supp/articles/JT20060621.html>

[24] Trammell, K. D., Tarkowski, A., Hofmohl, J., & Sapp, A. M. (2006). Rzeczpospolita blogów [Republic of Blog]: Examining the motivations of Polish bloggers through content analysis. *Journal of Computer-Mediated Communication*, 11(3), article 2. Retrieved June 6, 2006 from <http://jcmc.indiana.edu/vol11/issue3/trammell.html>

[25] Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

[26] Wikipedia (2006). *Orkut*. Retrieved September 12, 2006 from <http://en.wikipedia.org/wiki/Orkut>