

3 Context

The contextual narrowing of my thesis proposal starts on the broad area of **Complex Networks**, tightens its limits on **Social Networks** and ends at its ultimate contextual boundary, **Online Social Communities**.

Even though this Thesis proposition places itself on the center of a broad group of domains, I decided to deeply explore its closest and more direct domain – **Online Social Communities**, and the main subject of analysis – **Blogs**. Nevertheless, besides the omnipresent field of complex networks, the context of this thesis incorporates the domains of **Information Diffusion**, **Memetics**, **Information Architecture**, **Data Visualization**, **Information Theory**, **Diffusion of Innovations**, **Epidemiology** and **Small Worlds**.

3.1 Online Social Communities

Online Social Communities, although much more concise than the Science of Complex Networks, is still a wide-ranging field that can include mostly every type of online interpersonal communication medium, from e-mail listings/threads, to Usenet groups, MUDs, chat environments, instant messaging, community forums, *weblogs*, online gaming, interest groups, among others.

Online Communities offer an interesting change on the parameters that until now have defined social interaction. Several years after Milgram's notorious small-world test, Russell Bernard and Peter Killworth did what they called a "reverse small-world experiment". They interviewed hundreds of individuals, explaining Milgram's experiment and asking them what personal criteria would they use to get a specific package to someone they didn't know. Bernard and Killworth's study found that most of the subjects used only a couple of dimensions to get their message sent to the next recipient. Most predominant dimensions were geography and occupation.

Jon Kleinberg, a computer scientist who attended Cornell and MIT, was also motivated by Milgram's small-world study, and questioned how did the individuals actually find the paths within the network. Kleinberg concluded that people have generally a strong sense of distance, which they use to distinguish themselves from others. A notion of

distance can have several factors in which geographical distance is just one of them. Profession, race, religion, income, class, education, are other elements added to the equation, that describe how distant a specific person is from us.

From the beginning of human existence, communities were created for the benefits of their own members. Usually by means of expediency, either in relation to the exchange of goods or improved security against enemies, these groups of people occurred as emergent systems by means of social convenience. Geography always played an essential role and without a common shared space most of these communities wouldn't even exist. With the posterior developments of mail, and more recently, telephone, telex, and fax, human communication became highly enhanced and geography started diminishing its major influence. However, these new "technologies" only improved the way people communicated with each other, by giving them more tools and decreasing the time span and subsequently the distance; other than that, there were no major changes in the way social communities were formed. No matter how fast and easy it became for someone in Europe to talk with someone in America or China, there were never communities created on the basis of telephone calls.

If we explore the word syntax structure of most communication tools prior to the Internet, such as telegraph, telex, telegram, and telephone, we encounter the constant presence of the prefix *tele-*. *Tele* is a greek word that means "at a distance", usually implying "to be distant" or "over a distance". The first use of the prefix *tele* was in the word telescope which was actually adapted from Galileo's Italian word *telescopi*, followed by the word telegraph, meaning "writing at a distance". Therefore, Telecommunications is the field that embodies all the systems that intent to communicate "at a distant" or "over a distance". Once again we see the importance of geography as a crucial domain for human communication, where the advancement of technology, since the beginning, has been trying to diminish its constraints, by allowing people to communicate over an ever-present and disturbing distance. I find this analysis particularly interesting in such a way that the Internet, and all features associated with it, has completely abandoned the prefix *tele-*, drastically assuming the medium, and replaced it with the prefix *e-*. From e-mail, to e-commerce, and e-business, the prefix *e-* is usually associated with the latest heat of technological revolution, an abbreviation of the word electronic and an obvious association with the word *cyber*.

The advent of the Internet and the World Wide Web changed these secular communal constraints, possibly forever. The Internet became not just a medium for social gathering and communication, but it absorbed it, and the medium became truly the message. The transmission of information on the Internet is regularly measured in milliseconds, and the time it usually takes for a message to leave a computer in Tokyo and arrive at a computer in New York is more or less the same as a message sent to you, from your next-door neighbor. The difference is merely a few milliseconds, which is by itself a measurement difficult to perceive. Geography, as a crucial criterion for the birth of social communities, has been utterly disregarded by online social communities. Without the limitations of geography and physical interaction and identification, online communities had to rely on a more abstract, but equally distinguishing criteria, interests. By analyzing most current online communities, from online players to chat rooms, blogs and newsgroups, we find out that in the absence of physical recognition, social values like trust, confidence, respect and even friendship are ultimately based on a set of shared interests. And of course, this “virtual” interaction would not be possible without specific communication channels, portrayed as technological sub-systems of the larger medium, the Internet.

Personal interests are a central element of our social identity, and subsequently, a highly considered factor in relationships. Paraphrasing Duncan Watts in regards to peer-to-peer networks, “social identity is what leads networks to be searchable”. The fabulous aspect of online communities is the possibility of not only searching these clusters of shared interests, but also tracking the exchange of conversations, ideas and messages between them. By analyzing this data, it’s possible to understand, to some extent, how information travels through these virtual environments. **Weblogs**, in this conjecture, represent units of a remarkable social laboratory. It’s relatively easy to track their connectivity, but also, due to their highly clustering nature, it’s possible to examine in specific communities, how do news and trends travel through individual *bloggers*.

3.2 Weblogs

Weblogs (alternate: blogs) are not just a new fad among Internet users and they are much more than a collection of online digital diaries of spread interest groups. *Blogs* represent a change in online information flow and they are becoming a rising news source for many people. We might not even be aware of how influential blogs will be in the future but one thing is sure, there are currently *blogs* with close to half a million visitors a day, more than many large newspapers, magazines and news broadcasters.

Jorn Barger coined the term in 1997 and in 1999 Peter Merholz coined its alternative abbreviation “blog”. As Jorn Barger stated:

"Weblogs are often-updated sites that point to articles elsewhere on the web, often with comments, and to on-site articles. A weblog is kind of a continual tour, with a human guide [whom] you get to know. There are many guides to choose from and each develops an audience. There's camaraderie and politics between the people who run weblogs. They point to each other in all kinds of structures, graphs, loops, etc."

The most common definition of a *blog* is that of an online diary of thoughts, links, events, or actions posted on a web page with a dated log format. These posts are often, but not necessarily, in reverse chronological order, and are updated on a daily or very frequent basis with new information about a particular subject or range of subjects. Despite this dry classification, the usefulness of a *weblog* is incredibly rich.

Blogs are the vital elements of the personal publishing revolution. If we go back a few years, before the rise of online publishing, the only way someone could write something for general public would be through a letter to the editor, and hope for its message to be published in the magazine's next issue. For the first time in the history of human communication, any single person has the opportunity to reach millions with their message, as the cliché proclaims, with “the touch of a button”. Instead of being passive consumers of information, Internet users are becoming active participants. This power to the people is debatably a positive trend, since many people subjectively consider this measure adds to the existent “junk” flowing on the Web. Since most *blogs* don't obey to any kind of editorial process or peer review and sometimes “play” with anonymity, their public posts also raise legal concerns about intellectual property, defamation, and alike.

Controversies apart, *blogs*, as the World Wide Web, are free democratic resources that embody the concept of free speech, which is unquestionably a right for all.

Blogs also exemplify the true concept of diversity. Besides being oblivious to who might use this personal tool, blog content is as varied as the Web itself. The authors of *Essential Blogging* explain this diversity by pointing out that “creating a taxonomy of the blogiverse is a fruitless task”, since “there’s no good, central directory of blogs that puts each one in its own pigeonhole, because even the most topical blogger will stray from the subject from time to time to celebrate some personal victory or warn his readers off a terrible movie”.

One might also argue that in fact, this personal publishing revolution started with the first website, and consequently with the birth of the Internet. This is obviously true, however, until the first *blog* publishing tools became available, anyone who wanted to circulate their own ideas online, had to be fluent in HTML, web hosting, and aware of most webdesign applications available. Even after GeoCities launch in 1996, offering free web hosting to non-commercial personal pages, web pioneers had to be HTML-savvy people who would spend the evenings working on their websites. Also, these few personal webpages that start populating the Web in the mid 90’s were just a scattered collection of isolated opinions, with no regular updates and unconnected from each other. The big *blog* phenomenon started escalating in the summer of 1999, when a small web company called Pyra Labs released a product called *Blogger*. From that point on the *blog* community exploded and the more *bloggers* came into scene, more online *blog* tools became available. This was the beginning of the personal publishing revolution.

The inclination towards personalization is reaching every industry, from clothing to cars, from software to medicine. News and Information are just new elements added to the equation. In my opinion, the reasons why many *blogs* are so successful are due to two major factors: personalization and comforting lassitude. *Blogs* are usually maintained by a single person who filters the huge amount of available information according to his/her own preferences. For people who share common interests with the *blogger*, it’s not only exciting to get information from that source, since it’s going to match their inclination to some degree, but it also saves them a lot of time by avoiding the large, more abstract, and sometimes incongruent, news sources. In countries such as the US, where large media sources are becoming increasingly dry and biased, *blogs* might also represent an oasis of independent information.

3.3 Blogosphere

Blogosphere (alternate: *blogsphere*), or *blogspace*, is the collective term encircling all *weblogs* (alternate: *blogs*). It's almost impossible to determine with precision the existing number of *weblogs*, or even the ones currently active. Technorati is a leading search engine for the *blogosphere*, similar to Google or Yahoo, but exclusive to *blogs*. Technorati, as of February 2005, was tracking 7,245,866 *blogs*, and this number is far from stagnating. Out of curiosity, when reviewing this paper on April 6, 2005, I checked Technorati to see how the latest number had changed. To my not-so-surprised amazement, Technorati declared to be tracking 8,469,023 weblogs. It translates in an increase of more than 1 million blogs in less than two months.

The latest *Pew Internet* study estimates that about 27%, or about 32 million, of American Internet users are regular *blog* readers. They say a new *weblog* is created every 2.2 seconds, which means there are about 38,000 new *weblogs* a day. *Bloggers* update their *blogs* regularly; there are about 500,000 posts daily, or about 5.8 posts per second.

When we're faced with a number of blogs higher than eight million (at least), it becomes hard to consider its whole as a single community. The blogosphere, in analogy to its medium, the Internet, does not represent a single community but a vast collection of endless communities. These communities shape a complex web of more than 8 million nodes and are key factors in the outburst and further development of trends, fads and innovations. Also, due to its inherent diversity, any kind of classification regarding the blogosphere is a mere exercise of oversimplification.