

1 Introduction

Blogging presents one of the most interesting **social phenomenons** of our time. This change in the **flow** of online **information** might radically change the way we look at **news** providers and large media conglomerates. It also provides an extraordinary **online laboratory** to analyze how **trends, ideas** and **information** travel through **social communities**.

1.1 Concept

Blogviz is a non-commercial research project developed with the intent of disentangling this highly **complex network** for further study, research and analysis. The main goal of Blogviz is to improve our understanding of the dynamics of information propagation among *weblogs*.

An underlying question to Blogviz is: "How can we measure meme as a unit of cultural evolution?". The answer is not easy. Memes, due to their widespread trait and frequent untraceable evolutionary track, become extremely hard to measure accurately. In opposition to this commonly undetectable meme pool, the *blogosphere* offers a discernible and documented map of thousands of memes, with clear trails of progression, structured by date and time.

There are many possible ways of looking at information diffusion in *blogspace*. It can be based on conversation threads, comment threads, key sentences, themes, tags, or top links. Blogviz analyzes top links, occasionally called topics, which represent the most cited URLs appearing in *blog* entries in any given day. These popular links represent particular memes that provide an idea of sources, stories and themes that have occupied the attention of *bloggers* over a certain period of time.

By exploring the evolution of these topics through time, Blogviz will not only be able to track its popular dispatchers and key innovators, but also, follow its dissemination pattern from the beginning to an eventual tipping point, where it might leap the *blog* community and reach the mainstream.

Blogviz embodies a flash driven interactive visualization model with extensive use of **information visualization** and **information architecture**. Why is Information Visualization central to Blogviz? Information Visualization can be defined as "the use of computer-supported, interactive, visual representations of abstract data to amplify cognition" (Card, Mackinlay & Shneiderman, 1999). Information Visualization does not only makes data easier for human interpretation but it also discovers and highlights relationships in data elements, usually reducing the processes of searching by gathering information in a small rich space.

Therefore, Blogviz employs Information Visualization with the key intent of uncovering hidden patterns in the data and deriving plausible conclusions, which promote an advanced knowledge of information dynamics in *blogspace*. By unraveling the modus operandi behind the *blogosphere* we might be able to improve our knowledge on the mechanics of **online social communities** and, to some extent, the mechanics of **complex social networks**.

Blogviz is currently a portrait of blogosphere's topic activity during the months of January and February 2005. The selection of a time period was purely arbitrary. In order to make this project a reality within the thesis development time limitations, a decision was made in order to constrain the project to a more specific time span. Nevertheless, the model was developed to easily incorporate different timeframes. Blogviz will continue to expand in the future, to the possible point of including real-time data.

Blogviz uses existing data from three different *blog* search engines organized in a database that will soon be available for public access. (see *Technical Sources* for additional information)

1.2 Memetics

From a conversation with my Thesis Writing instructor, Mark Stafford, I was able to understand how my thesis had become closely related to the concepts of memetics or meme behavior. We came to the conclusion that I was developing a “topological model of meme activity”, even if until then I was somehow oblivious to it. That title actually remained for a while when characterizing Blogviz. But later on I decided to change it, since the word *meme* was slightly audience limiting and the expression *topological* could result in inadequate interpretations. I still question why the notion of Memetics didn’t come up in my research earlier, but what is particularly interesting is that it was there from the beginning, immersed in every iteration of my work. I think I was too much concentrated in the idea of a word-of-mouth behavior, an expression used by Malcolm Gladwell in “The Tipping Point” and by Duncan Watts in “Six Degrees: The Science of a Connected Age”.

The vital point is that Memetics is the principle theory when contextualizing Blogviz, and because of that, understanding the theory of Memetics is a crucial measure to comprehend the underlying concept of Blogviz.

1.2.1 What’s a Meme?

The term was first coined by Richard Dawkins’s, in 1976, on his notorious book “The Selfish Gene”. In the words of Dawkins the word “meme” refers to “a unit of cultural transmission, or a unit of imitation”. More specifically, a meme can be defined as a self-propagating unit of cultural evolution, a unit of information, held in an individual’s memory or in an outside artifact (e.g. book, record or tool), which is likely to be communicated or copied to another individual’s memory or retention system. Examples of memes are ideas, catch-phrases, melodies, technologies, icons, theories, inventions, languages, designs, fashions, and traditions. This covers all forms of beliefs, values and behaviors that are normally taken over from others rather than discovered independently.

A meme is basically a pattern of information that induces people to repeat it. People try to “infect” each other with memes they find most appealing, despite of the memes’ objective value or truth.

1.2.2 What is Memetics?

Memetics is the study of evolutionary models of information transmission based on the concept of the meme. In spite of its roots in evolutionary biology and computer simulation, memetics has become more of a social science, focusing primarily on the spread of information within human society. Rather than debate the inherent "truth" or lack of "truth" of an idea, memetics is largely concerned with how that idea itself gets replicated.

Another definition of Memetics declares it is the theoretical and empirical science that studies the replication, spread and evolution of memes. As portrayed in the *Journal of Memetics**: "It's core idea is that memes differ in their degree of 'fitness', i.e. adaptation to the socio-cultural environment in which they propagate. Because of natural selection, fitter memes will be more successful in being communicated, 'infecting' a larger number of individuals and/or surviving for a longer time within the population. Memetics tries to understand what characterizes fit memes, and how they affect individuals, organizations, cultures and society at large".

Since the premise of Memetics is to investigate the evolutionary mechanisms that determine the propagation of information within a population of human, animal or artificial agents, we can easily perceive why this science is vital to the understanding of cults, ideologies, or marketing campaigns of all kinds.

A meme is acknowledged as a self-propagating unit of cultural evolution, analogous to the gene (the unit of genetics). And because of memes' similar behavior to life forms, Memetics embraces the analytical techniques of diverse sciences, such as, epidemiology, evolutionary science, immunology, diffusion of innovations, linguistics, and semiotics.

* Journal of Memetics (<http://jom-emit.cfpm.org>)

1.3 Diffusion of Innovations

I believe any type of Information Diffusion Model (IDM) in Social Networks must derive extensive practical knowledge from the sciences of epidemiology and diffusion of innovations. These two domains help us understand many of the factors that characterize the spreading of information and adoption process in social communities. Epidemiology and Diffusion of innovations also share many similarities and are surprisingly linked together. For these reasons I decided to include in this thesis a short description of these areas, since in addition to the concept of Memetics, they create an extraordinary context to the understanding of Blogviz.

I don't make wide explanations of each domain but rather comparisons between them on how they relate to this thesis's assertion. In order to delineate a common ground for the following definitions, this paper assumes that an innovation can be characterized as a new meme, given that it is also described as a new idea. In the context of information diffusion in the blogosphere, it assumes the process of adoption to be the process by which a blogger, aware of the existence of a new meme (or innovation), decides to mention it on his/her own personal blog, in the form of a post or part of a post. This action can be understood as an "adoption" by the *blogger* of this particular unit of information, therefore contributing to its replication.

The study of innovation adoption and diffusion has its origins in the Midwestern United States. In an Iowa State University study, Ryan and Gross (1943) showed that the pattern of adoption and diffusion of a maize hybrid was systematic, hence opening the door for further research.

Diffusion is the process by which an **innovation** is communicated through certain **channels** over **time** among the members of a **social system** (Everett M. Rogers, 1995). The *innovation* includes "any thought, behavior, or thing that is new because it is qualitatively different from existing forms" (Jones, 1967). The characteristics of an innovation, as perceived by members of a social system, determine its rate of adoption.

Just by analyzing these last statements one can easily grasp a series of similarities with the notion of Memetics. Even to the point that the theory of Diffusion of Innovations also considers the unit of adoption not exclusive to an individual person, but extending to other types of retention systems.

The four main elements in the diffusion of new ideas are:

- (1) The innovation
- (2) Communication channels
- (3) Time
- (4) The social system (context)

1.3.1 The Innovation

These are the characteristics that determine an innovation's rate of adoption:

- Relative advantage
- Compatibility
- Complexity
- Trialability
- Observability to those people within the social system.

1.3.2 Communication Channels

A communication channel is the means by which messages get from one individual to another. Mass media channels are more effective in creating knowledge of innovations, whereas interpersonal channels are more effective in forming and changing attitudes toward a new idea, and thus in influencing the decision to adopt or reject a new idea. Most individuals evaluate an innovation, not on the basis of scientific research by experts, but through the subjective evaluations of near-peers who have adopted the innovation. (Everett M. Rogers)

In a broad sense, the communication channel in the context of Blogvizi is indubitably the Internet. Without it there wouldn't even be any kind of communication between bloggers. However, without blogrolls and posting citations within each blog, the restrict channels among them would be very difficult to perceive. Blogrolls are the backbone of blog communities, the edges that keep all the nodes interconnected, and therefore, are the key factors in understanding how information develops across the blogosphere. In fact, a major characteristic of online social communities is that they are based on communication channels, not on physical co-location. A blogroll is a listing of websites that often appear as links on weblogs, usually on a left or right frame of the page. This list of links is used to relate the site owner's interest or affiliation with other webloggers.

1.3.3 Time

The Diffusion of Innovations theory divides the element of Time in three main dimensions, in which only two can be fully applied to the context of information diffusion in the blogosphere.

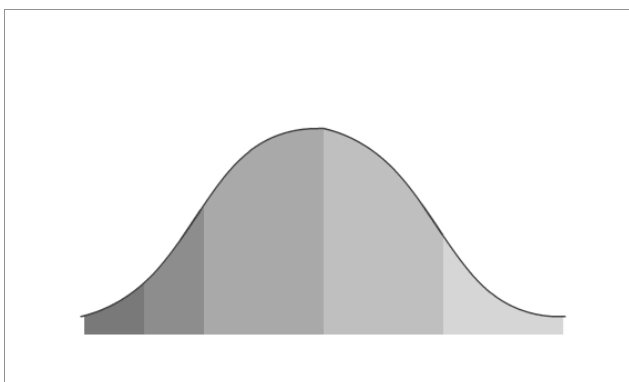
> **Innovation-decision** – The innovation-decision process is the mental course of action in which an individual passes from first knowledge of an innovation to forming an attitude toward the innovation, to a decision to adopt or reject it, and if adopting it, to implement this new idea and confirm the decision.

In the case of a *blogger* deciding to post or not a specific meme in his/her weblog, this decision process is so fast that it's almost impossible to measure. It applies to other memes, and definitely to other innovations, but it's not relevant as a measurement in top links replication.

> **Innovativeness** – Innovativeness is the degree to which an individual is fairly faster in adopting new ideas in relation to other members of a social system. Innovativeness, in opposition to the innovation-decision process, is an extremely significant measurement in top links replication, as in most information diffusion models.

There are five adopter categories, or member classifications of a social system, based on their level of innovativeness:

- **Innovators**
- **Early adopters**
- **Early majority**
- **Late majority**
- **Laggards**



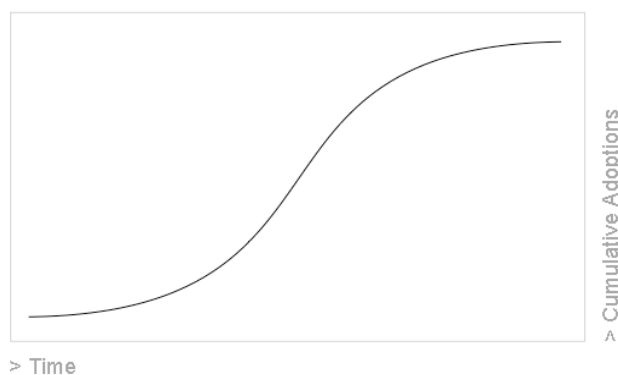
- innovators (2.5%)
- early adopters (13.5)
- early majority (34%)
- late majority (34%)
- laggards (16%)

Bell-shaped curve showing categories of individual innovativeness and percentages within each category

Innovativeness among social systems is characterized by a bell-shaped curve where time and incidence of adoption are the two main vectors. This concept, in the context of Blogviz, is further explored in the *Methodology* chapter of this thesis.

Many search engines and community tools analyzing the blogosphere, assume a direct correlation between blogs popularity and innovativeness. I believe this assumption is incorrect. Their thinking is very simple. If a specific blog has a high number of inbound links and therefore a sizeable readership, it must imply that it's in the frontline in finding and publishing original information. The **HP Information Dynamics Lab** study on the "Implicit Structure and the Dynamics of Blogspace" (Eytan Adar *et al*) showed exactly the opposite. The study demonstrated that popular blogs are rarely among the first ones to start a specific trend. Many popular blogs claim most of their "discoveries" by not citing their original source, which are usually smaller unfamiliar blogs. The level of popularity of each blog might be directly related to its scale of influence, but not necessarily to its level of innovativeness. So who are these unknown bloggers that bring fresh ideas to the blogspace? Who are these innovators or trendsetters? Blogviz will allow an exposure of these anonymous sources, crucial in the dynamics of topics diffusion.

> **Rate of adoption** – The rate of adoption describes how fast an innovation is adopted by members of a social system in a given time period. When mapping the cumulative adoption time path or temporal pattern of a diffusion process, the resulting distribution can generally be described as taking the form of an S-shaped (sigmoid) curve. Time and cumulative adoption (or infected population) are the plot main vectors.



1.3.4 The Social System

The fourth main element in the diffusion of new ideas is the social system, which basically creates a boundary for the diffusion and adoption of an innovation to occur. A social system is defined as a set of interrelated units that are engaged in joint problem-solving to accomplish a common goal (Everett M. Rogers). The members or units of a social system may be individuals, informal groups, organizations, and/or subsystems.

In regards to the replication of top links among weblogs, the social system is undoubtedly the *blogosphere*, depicted as a fertile network of endless social communities. This vast communication network consists of interconnected individuals (bloggers) who are linked by shared interests and patterned flows of information.

At a first glance, considering the highly interconnected web of links, connections and shared interests among bloggers, it might seem easy to understand the adoption process of a particular unit of information or innovation. However, another crucial conclusion exposed by the HP Information Dynamics Lab study, mentioned before, declared that “for URLs appearing on at least 2 blogs, 77% of blogs do not have a direct link to another blog mentioning the URL earlier. For those URL’s present on at least 10 blogs, 70% are not attributable to direct links”.

There have been several studies on how the system’s social structure, and norms or established behavior patterns, affect the diffusion of innovations within a particular social system. But another area of research that is closely linked to Blogviz relates to opinion leadership. It can be described as the degree to which an individual is able to influence informally other individuals' attitudes or explicit behavior in a desired way with relative frequency. Blogviz allows a broad understanding of opinion leadership in blogspace by tracking and exposing the most influential and innovative topic proliferators.

1.4 Epidemiology

Throughout this thesis I use several times the terms *contamination* and *infection* when describing the adoption process of memes. Even though this practice might lead to unwanted interpretations, its use is not arbitrary, and it actually facilitates the comprehension of information diffusion dynamics.

Epidemiology in its broadest sense is the study of disease patterns in human populations (Wikipedia). Epidemiology can also be described as the study of the determinants, occurrence, and distribution of health and disease in a defined population. Infection is the replication of organisms in host tissue, which may cause disease. A carrier is an individual with no overt disease who harbors infectious organisms. And the notion of dissemination is understood as the spread of the organism in the environment.

In the above description, regardless of the different terms, we start noticing several similarities with the domain of diffusion of innovations. This analogy is even more explicit when characterizing the three major elements in disease occurrence, the so-called chain of infection:

- (1) The etiologic agent (parallel to the innovation)
- (2) The method of transmission (parallel to the communication channel)
- (3) The host (parallel to a unit of a social system)

Further along in characterizing the disease evolution, the *epidemiologic descriptive study* organizes data by time, place and person. It is unquestionably the closest approach to the concept of Information Diffusion. It divides the element of Time into four main trends; respectively, *secular trends*, *periodic trends*, *seasonal trends* and *epidemics*. What's interesting in this typology of Time is that it applies equally well to the evolution of top links across the *blogosphere*. Because of that I assume a series of parallelisms between them.

The **secular trend** describes the occurrence of disease over a prolonged period. This continual development is less usual than the *seasonal trend* in the context of *blogspace*. This trend usually describes commercial or very popular websites that never lose entirely the bloggers' interest and as a result have a continuous existence among them.

The **periodic trend** basically expresses a temporary modification in the overall secular trend. It conveys a sudden new interest in a specific meme that is part of a continual trend.

The **seasonal trend** reflects seasonal changes in disease occurrence following changes in environmental conditions that enhance the ability of the agent to replicate or be transmitted. This short transitory trend is the most common in *blogspace*. A new meme that spreads quickly and rapidly loses interest, dying in a short period of time.

The **epidemic** incidence of a disease happens generally when it surpasses a threshold of 7% of the target population. An epidemic is a sudden and boost in occurrence due to prevalent factors that support transmission. An information epidemic in blogspace might originate a **tipping point**, where a specific meme escalates and leaps the *blogspace*, reaching the mainstream.