

8 Conclusion

Blogviz is not the most accurate or the most reliable method for visualizing topics' diffusion in the *blogosphere*. However, it's a well thought attempt. It's also a tryout in the process of leveraging the understanding of memetics, as it portraits to the comprehension of the enveloping information dynamics in our surroundings. Blogviz is a visualization model, and has implicit in its deliberation the notions of trial and prototype. I like to consider this project as an experiment in the field of Information Diffusion.

Blogviz has still a long way to go, but the first step was made. I believe the MFA program was a crucial context for its embryonic development, but there are still so many other features I want to implement in the model, that it's hard to enunciate them individually. I also think there's an increased motivation when developing a one year thesis project and expect to continue expanding it beyond its immediate delivery deadline.

Information Diffusion Models (IDM) are not abundant in the research arena, compared for example with Diffusion of Innovations Models (DIM), which strongly captivate commercial interest. Due to DIM's applicability in management and economics theory, they've become an important measurement in new product releases by most industries. Nonetheless, even if IDM's commercial significance is still not so obvious, Information Dynamics is becoming an area of great interest among several researches. In a world increasingly overflowed with petabytes of information, it's critical to understand how this flow of information behaves. I believe the fields of Information Architecture and Information Visualization, among others, have not only the responsibility of making information more useful and understandable, but also to inquisitively investigate how the information itself propagates. This is an area of study where other sciences such as sociology, semiotics, linguistics, human-computer interaction, cognitive psychology, memetics, epidemiology, and diffusion of innovations, play a major role in deciphering the intricacies of Information dynamics. And this is categorically the main reason why Blogviz came to life.

I have this higher goal, or perhaps delusion, of building this perfect visualization system that doesn't need any extra panels or descriptive features to amplify its understanding. I'm conscious, though, that it's quite utopian to idealize such a system. From the moment one decides to let other people interact with the application, there are vast arrays of learning curves to be considered. Different people have different levels of expectation, knowledge and patience. One cannot expect for a system, as simple and intuitive it may seem for the creator, to be as straightforward to other users.

There was no user testing on Blogviz, mostly due to lack of time. The model, even if it's not intended to be used and understood by every single Internet user (since it's where it resides), it has to be comprehensible and insightful to its immediate audience, exposed in this paper. Therefore, Blogviz will evolve from the feedback and judgments of its users.

I'm still not sure if the right visualization choices were made and that's something I will probably never feel confident enough. I had extremely high ambitions for this project, that relate closely to my own nature, which sometimes might prove to be counterproductive. I tried hard to come up with a visual schema that would be somehow revolutionary and could even be adapted to different models in other fields. I'm sure I wasn't able to do that. But one thing that comforts me is knowing that I put all my effort and dedication in this project, and even if the result doesn't prove to be as groundbreaking as initially planned, its development was an extremely enriching experience.

The thinking process, as it may be illustrated in the *Methodology* chapter, was a long and laborious course with several important iterations. I have to admit, at this stage, that I'm not certain that the last iteration was the best resolved one in terms of visual depiction of my subject of analysis. I guess this state of uncertainty is softened by the thought that under the natural progression of the project, this resolution was the final result based on improvements of the previous.

Sometimes I also feel I'm overflowing Blogviz with too much data, and by doing it, the visualization features are becoming diminished in this assortment of info-bars and text. But at the same time there's so much information I want to include, so much interesting

comparisons to be made, that I feel it would be nonsense to discard them. It's quite a challenge to find a balance within these two forces. On one side I'm trying to keep the visualization simple and intuitive, and on the other, wanting to lengthen the data analysis for an even richer knowledgeable experience. I guess it comes down to the eternal balance between form and function.

This project was indubitably a technical achievement for me. This was the first time I deeply explored technologies such as: ColdFusion, MySQL, XML, Flash Remoting, Flash Communication Server, among others. It was my first solid attempt to deal with dynamic data exchange, middleware and database communication technologies. In the end I was able to overcome a series of obstacles that made me very self-doubting from the beginning. Throughout the thesis development process I thought to myself, many times, that I might not be able to overcome many of the technological and programming hurdles needed to make this project a reality. Fortunately enough, I did. The outcome, within this framework, was exactly as I expected and I must admit that the achievements made in this process were the ones that made me jump the most. On this context I think this project was very successful and rewarding to me. I cannot wait to apply the technical knowledge I acquired in other endeavors.

It's rather curious that my main fixation, from the beginning of this thesis process, it's not currently represented in Blogviz. This obsession is associated with the qualitative dimension of the analyzed pattern. I believe that in terms of pattern depiction, as it relates to the core concept of Blogviz, both its quantitative and temporal dimensions were satisfactory resolved. The missing link, that leverages an even more critical insight on the pattern evolution, concerns the existing relationships between *weblogs*. From my first prototypes, my main obsession has always been to visualize the inbound and outbound links among *blogs*. It was naïve on my part to believe that by understanding the linkage I would be able to explain how the information spreads. The linkage practice, as said before in this paper, is not elucidating enough to explain the adopting process of information in *blogspace*. This was clearly exposed in the HP Information Dynamics Lab study, also mentioned before; where it was said that roughly 75% of *blogs* for each topic don't have any direct link between them. However there are still levels of visualization I plan to implement in order to portrait the relationships between topics' adopting units (*weblogs*). Some of these levels will be constrained to specific timeframes and generator groups.

8.1 Initial Results

However scarce a small sample of 330 blogs might be, I believe a few patterns are starting to emerge from Blogviz, some of which I was not expecting in the beginning. I'm aware that these initial results are still too immature to fundament any reliable conclusion, but I though I should mention them nonetheless.

- The first blogs to start a topic are usually not highly popular. It seems the most popular blogs only start appearing within a certain number of days from the initial date of spreading. It will be interesting to analyze if this behavior follows some sort of ratio or consistency through the analysis of a larger portion of topics. However, this is an outcome I was somehow expecting, based on previous studies made in this area.
- There are some topics clearly dominated by a particular type of generator. Interestingly enough these appear to be more common within LiveJournal users, who count themselves among the less popular *weblogs* in *blogspace*.
- 8 topics, from the 9 analyzed so far, have the peak day within 1 or 2 days from the first day of spreading, even when they extend through long periods of time.

8.2 Next Steps

- **Create an automatic script to input the data in the existing database.** This method will allow a much larger number of *blogs* to be seamlessly incorporated into Blogviz. The more data Blogviz gathers, the more plausible will be its subsequent results on sustaining the key intent of the project: to improve our understanding on the dynamics of information propagation in *blogspace*.
- **Improve the code.** This is probably a never-ending task, but nonetheless, it's important to acknowledge that there's always space for improvement in the application coding.
- **Implement nodes visualization (*blogs*) and subsequent layers of analysis.** So far, Blogviz is employing most of its visual depiction in portraying patterns' evolution, or in other words, topics' inner structure and transmission through

time. The next step will be to include a series of visual techniques to represent single adopting units' (blogs) behaviors, relationships, and levels of popularity, adoption and innovation.

- **Extent the current time span** of the application, to the possible point of including real-time data. Presently Blogviz is mapping the evolution of topics within the first 64 days of 2005, respectively, from January 1st to March 5th, 2005. This decision was made under the time limitations for this thesis development; nonetheless, the model was built in order to accommodate other timeframes.
- **Feedback.** As mentioned before, there was no time to conduct any kind of user testing, so I plan to build a bug reporting and comment form to help solving existing problems and hopefully receive interesting suggestions.