BUZZMONITOR: A TOOL FOR MEASURING WORD OF MOUTH LEVEL IN ON-LINE COMMUNITIES

Alessandro Barbosa Lima
Jairson Vitorino
Henrique Rebêlo

ABSTRACT

This paper describes an application to monitor on-line Word of Mouth across different Internet services and how the collected data can be used to feed a company's on-line marketing strategy. Our motivation lies on the theory of social maps and scale-free network laws. We describe briefly buzzMonitor, which is our first prototype for the task of measuring and monitoring conversations among consumers in on-line communities.

KEYWORDS

Social Networks, Viral Communication, Internet, Word of Mouth Communication, Hubs, Buzz Marketing.

1 Introduction

Since the last decade, the number of Internet users has been growing steadily as well as the use of on-line services that allows interaction among them. Consumers are now much more empowered to freely express their opinions about products and services through the so-called on-line communities, such as blogs, e-commerce sites, forums, discussion lists and others.

Before the advent of the Internet, Word of Mouth (WOM) among consumers were impossible or far too expensive a task to perform accurately. Hence the majority of companies could seldom use WOM data on their marketing strategy. The emerging of Internet applications such as blogs, chats, bulletim boards systems and other on-line communities allowed for the first time to observe the mapping of social networks and in doing so provided new perspectives to study interpersonal online communication for marketing purposes.

WOM phenomena once restricted to the real world, has now taken over the web as a result of the interaction among specific individuals, which feel compelled to talk about their experiences with products, services and brands. Such individuals act as opinion leaders in these communities and can influence literally hundreds of people through their postings on web sites.

As conversations about products and services of a brand spread through the web, companies must be aware of the power of influentials both as a source of information that help to improve their products and services and as a way to prevent emerging crisis by anticipating possible negative WOM detected earlier by hearing these very influentials.

This paper describes a proposal of an application to monitor on-line WOM across different Internet services and how the collected data can be used to feed a company's marketing strategy. Section 2 describes the so called social maps and Barabási scale-free network laws, section 3 explains our approach for word of mouth monitoring

on the web, section 4 describes our proposed application BuzzMonitor, section 5 gives a brief for future work and section 6 concludes the paper.

2 Social maps on the Internet

Milgram demonstrated that our social networks are not random but has a scale-free topology. In 1967 he conducted an experiment that consisted in sending a package to each of 160 people that lived in Boston and Omaha (Nebraska), in the United States. These people were instructed to send each package to a person, living in another city. The participants were said not to send the package directly, but through someone who they trusted and could help using his or her network to reach the target. To monitor the pathway taken by the package Milgram instructed the participants to write down the full name on it, helping to monitor the number of people handling it before it gets to its final destination. Surprisingly, six was the average number of people that the package passed through before getting to the target. The results of this experience created the hypothesis that everyone in the world is only 6 degrees apart from anybody else. This is sometimes referred also as the six-degree of separation hypothesis [Milgram67].

To Rosen [Rosen2000], what consumers talk, regarding the past acquisitions or future purchases are protected, because *our social ties are not easily visible to the world. It also means that marketers are in the dark. If members of social networks can't see their own links to each other, these links are even more hidden to outsiders.*

In 2004 though the first applications that enabled the creation of social maps, as Orkut and LinkedIn appeared on the web. These sites use the scale-free network model to make social ties visible to a community. Barabási [Barabási2003] proposed the scale-free network model for the Internet in 1999. According to this model, the scale-free networks obey two laws:

a) **Growth**: For each given period of time a new node is added to the network.
b) **Preferential Attachment:** If each node has a choice between two existing nodes, the probability that it will choose a given node is proportional to the number of links the target node has. That is, given the choice between two nodes, one with twice as many links as the other, it is twice as likely that the new node will connect to the more connected node.

Barabási found that most networks of practical interest, from the language to the sex web, are shaped by the same universal laws and therefore share the same hub-dominated architecture. According to the author, hubs are the integral components of scale-free networks; they are the statistically rare, highly connected individuals who keep social networks together. Gladwell affirms that six degrees of separation does not simply mean that everyone is linked to everyone else in just six steps. It means that a very small number of people are linked to everyone else in a few steps, and the rest of us are linked to the world through those few [Gladwell2000], these people are hubs.

Orkut demonstrates a social scale-free network. It is possible to find people with hundreds of contacts and people with just a few dozen. However Orkut works a bit different: the nodes with more links do not necessarily attract more nodes, as a *preferential attachment*.

To deal with this issue, Barabási introduced a third law to scale-free networks. This law is called Fitness. "Fitness is your ability to make friends relative to everybody else in your neighborhood; a company's competence in luring and keeping consumers compared to other companies; an actor's aptitude for being liked and remembered relative to other aspiring actors; a Webpage's ability to bring us back on a daily basis relative to the billions of other pages competing for our attention. It's a quantitative measure of a node's ability to stay in front of the competition. Fitness may have genetics roots in people; it may be related to product and management quality for companies, to talent for actors, or to content for Websites" [Rosen2000].

If our social ties are visible so are our nodes and what we talk. Blogs and communities such as Orkut make visible not only our social networks but also what individuals talk about with each other. What was before considered invisible and impossible to measure now is available, although scattered through thousands of nodes

on the web. A company can measure these conversations among consumers, monitoring positive and negative comments that a certain product or service are generating.

Godes and Mayzlin [Godes2003] demonstrates that *WOM can and should be measured just as all of the other key metrics of a company's success are typically measured. Just because it is a difficult phenomenon to get one's hands around doesn't mean that it should be thought of as purely "qualitative."*

In the following sections we will describe how one get hands around the problem of gathering WOM data from the web by describing a search-based application that will help companies to monitor, categorize and analyze WOM among their on-line consumers.

3 The Monitoring of on-line Word-of-mouth

According to Kotler [urlAnet2005], “Buzz marketing, the effort to generate business by WOM, will increase substantially. Marketers have improved their ability to identify opinion leaders and reach them early so that they can do the work of spreading the word about a distinct product or service”. Although it is possible to use several search engines to manually seek and analyze data from conversations on the web, it is not a practical approach for delivering accurate WOM reports on a systematic basis. As it turns out, a software application built specifically for this purpose may have a significant impact on the monitoring activities, freeing the marketers from the time-consuming raw search of terms and allowing them to focus on the analysis of relevant postings filtered automatically by incoming links, terms, URLs and other specific criteria.

The first issue to be considered is where the WOM happens; we should thus define what WOM stands for. WOM happens as a result of Internet users interpersonal communication, it is not an isolate comment, but the sum of all comments about a certain product that are exchanged among people at any given time [Rosen2000]. This communication is scattered on different on-line communities. According to Godes and Mayzlin the more dispersed information is today, the more likely it will be to inform new people tomorrow [Godes2003]. So, it is important not to focus on a particular community or web site but try to measure WOM across communities. Thus our proposal is to monitor WOM across different on-line communities where spontaneous communications occur.

Another key issue is to qualify the individuals responsible for the buzz, this can be accomplished based on Barabási hubs theory. Barabási call hubs nodes with an extraordinary large number of links. In a social network we can observe people that act like hubs. They accumulate a great number of linked contacts, they are people whose blogs have the larger number of incoming links from other blogs, they got a higher number of contacts in their Instant Messaging applications and got the maximum number of friends in Orkut (1,000).

Studying a scale-free network of people connected to each other through their e-mail, Ebel [Ebel2003] found that e-mail hubs on a network are more likely to spread computers worms once infected. Their experiment also showed that *making use of the high clustering, commercial e-mail providers can identify communities of users more easily and focus marketing more efficiently* [Ebel2003].

Pastor-Satorras and Vespignani [Watts2003] found out that even if a vaccine is available to a specific kind of virus, the virus continues to infect other users keeping a stable infection rate during a long period. This study blames the existence of *hubs* (Outlook users who have more contacts on their Contact List than the usual for example) to the keeping infection rate of computer virus.

Hub-users are more visible, because the preferential attachment law states that the probability a node will choose another specific node is proportional to the number of links the chosen node has, therefore these hub-users play a fundamental role in the word of mouth dynamics. It is also important to know who is responsible for the conversations and also the influence these people – also called by different names as Mavens, Hubs, and Opinion Leaders – have to the spreading of buzz.

3.1 Finding hubs through Incoming Links

The blog phenomenon is one of the new on-line categories applications helping to create social maps. People usually link to each other from their blogs, creating a web of connected blogs that belong to the same cluster. Clusters are (…) *sets of people who share similarities in some dimension of their lives and, as a result, who frequently communicate with one another. Millions of clusters are formed according to dimensions such as age, sex, education, occupation, social class, area of interest, geography, and ethnic background.* [Rosen2000].

In word of mouth monitoring, it is essential to have systems that help the marketer to measure the hubs of a certain network and also the existence of a cluster surrounding a specific hub because this information will help the marketer not only to find out who are the opinion leaders about certain product or service, but to whom these people can directly lead.

A way of doing this is to measure the incoming links that a blog gets because people usually link to the blogs or sites they like. A recent research revealed that from 60 political web sites randomly chosen, only 15% linked to opposite ideology web sites, while 60% linked to same political ideas web sites [Barabási2003]. For marketers, knowing if a blog or a web site belongs to a certain cluster could be useful to imply social-demographic information once omitted. For example: if unidentified person belongs to a cluster of web sites which users are all from Rio de Janeiro, it will be more likely that this person is from Rio as well.

Determining the incoming links will also help us to create social maps, these maps in turn will help to understand how the word of mouth spreads across clusters and who the opinion leaders are. Although a direct relation among hubs and opinion leaders are not yet established we could use the social network topology as an indicator of ways a message could flow through a network. Localizing clusters will also help to find out if a buzz is still restricted to a community (cluster) or if it spread across the Internet. This will be very useful to marketers to detect if a negative word of mouth is about to become an emerging crisis.

4 BuzzMonitor: a tool for measuring the o-line WOM level

BuzzMonitor is a web application designed to search, store and classify information on any number of web sites. It is targeted to on-line market researchers, on-line marketers and other professionals interested in analyzing on-line word of mouth.

BuzzMonitor lets users to define WOM monitoring projects for a specific URL and then generate search results reports. A BuzzMonitor project definition must include a set of parameters that ranges from search terms to auxiliary data that will help to classify them. A typical set of parameters are (numbers correspond to the fields in figure 1):

1 – Client: every project has a client (a firm or a person).
2 – Terms: the terms of the search.
3 – Tags: a project may include any number of tags (used to label each search result).
4 – Number of search results brought.
5 – The size of the text snippet that the user wants to retrieve from matched pages.
6 – Specific URL to be searched.

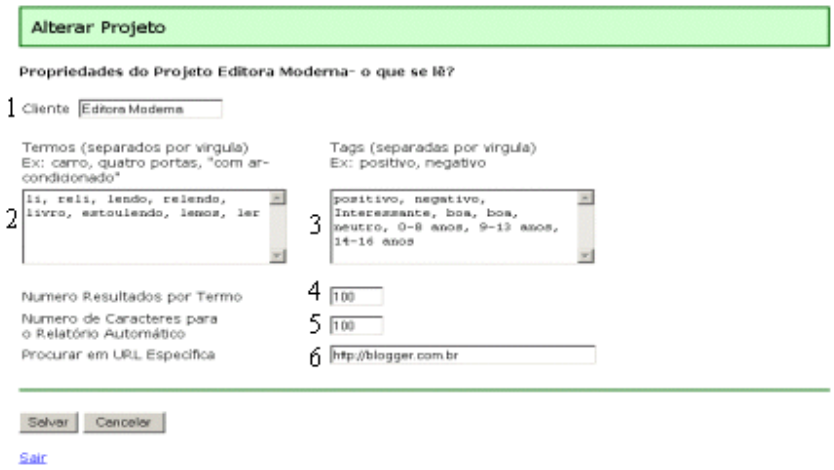

Figure 1: Project definition screen

Tags are a key feature of BuzzMonitor as they are the primary elements used to categorize search results. When visualizing the results, the user can apply any number of tags to one or more URLs brought by the search. Market researchers to categorize their results, as they like, may use tagging. For example a result could be either positive or negative or relevant or irrelevant. BuzzMonitor allows that these same results may be further classified as say male or female (meaning that they came from a posting written by a man or a woman) and so on. By providing tags to each relevant result the user will refine their search and may generate different reports from the tagged data, such as only positive results or only results posted by women. These reports are the very deliverables of the researcher, which can use the tool not only to gain productivity on their on going projects but also to keep track of every single search that she has ever made using BuzzMonitor as all results include the search date. Figure 2 shows the tagging process in action.

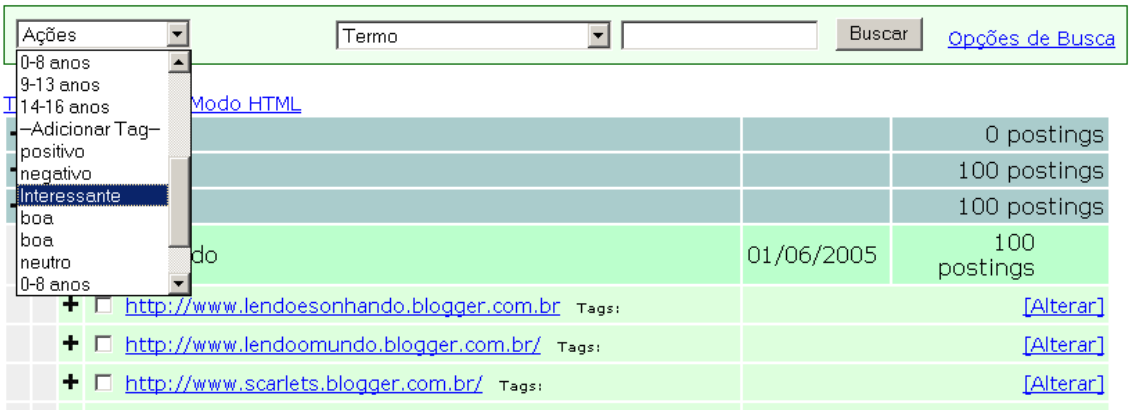

Figure 2 - Tagging the results

Once a search is finished it is possible to visualize all the results brought by the engine. These results can be furthered filtered by term, by tag, search date or a refined word search within the snippets of each result. There are two kinds of snippets, the manual snippet, which lets the user to copy and paste any relevant piece that she finds on the matched page; and the automatic snippet, which gathers a piece of the text around the term and stores in the application database, the size of this snippet, is determined in the project creation. In the visualization

mode the user can also archive the irrelevant results, this do not erase the results but hides them from the main view, meaning that the user can still retrieve the archived results and even to reactivate them. Figure 3 shows the refined search screen. The refined search lets the user chooses a time period, a keyword (to be searched within the snippets), a term and to indicate if she wants archived pages or pages that no longer are available on the web (using google cached pages feature). The refined search results encompasses all the relevant data the researcher needs, now all she has to do is to generate a report based on this data.

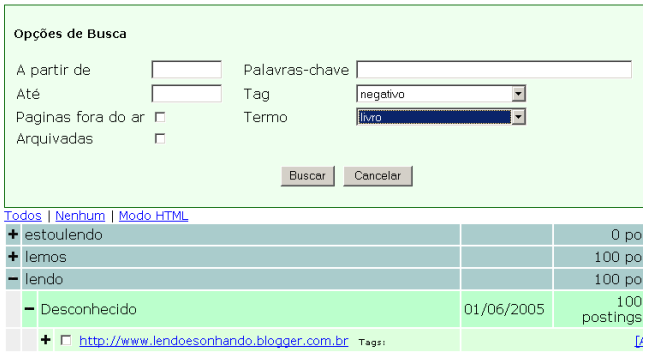

Figure 3 – Searching for categorized results

Finally after tagging properly all the results and archiving the irrelevant ones, the user goes to the last phase of the process, which is to generate the reports. There are two kinds of reports: the default report gather all the refined search results and display the manual snippet for each result on the report, together with the name of the service (e.g. www.blogger.com or www.blig.com.br), the name of the site, the author and the date of publication which may be manually entered by the user for each result. The BuzzMonitor application generates for each result the total number of incoming links and also displays this item at the report. The second report is the automated report, the only difference from the default report is that instead of inserting the manual snippets in the report it will retrieve the automatic snippet. The automated report is useful if the researcher is looking for a brief glimpse of what results she will get from the terms she picked. This may help the researcher to refine her search, tunning her terms or to rapidly generate a preliminary buzz report.

4.1 BuzzMonitor Technology

Our first version of BuzzMonitor was developed in Java using IBM open-source IDE Eclipse. BuzzMonitor is organized as three-tier architecture: the presentation or user interface tier, the business tier and the access data tier. The interface relies on JSP (Java Server Pages) for making it possible to implement several advanced features such as expanding and contracting the tree of urls results, tagging such results and generating html reports dynamically. The business tier uses the Standard Java package and the database tier runs on the open-source MySQL database. Figure 4 shows the basic class diagram for the application.

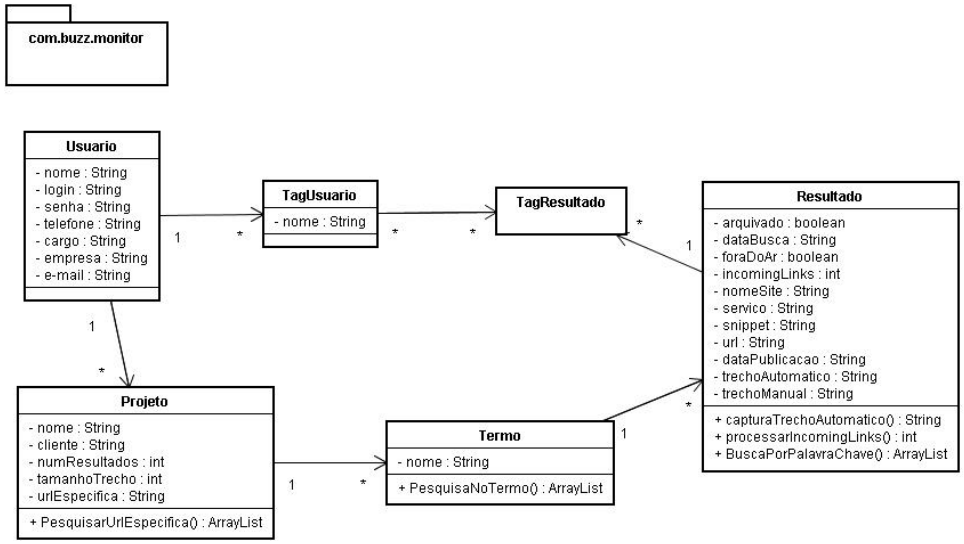

Figure 4 – Basic BuzzMonitor Class Diagram

The system works basically processing requests from the user and then launches html pages dynamically with the results. Whenever a user sends a request through her browser, this request is sent to the business tier where it is processed most of the time by writing or reading information on the MySQL database.

A significant feature of BuzzMonitor is the use of Google APIs service. Google is currently the most popular search engine that by the time of this writing indexed roughly 8 billion web pages. BuzzMonitor relies on the Java Google API freely distributed to perform searches for terms and to retrieve other important raw data such as: cached pages, number of incoming links and matched pages within specific URLs. After retrieval, this data is further processed by our system (i.e. htmls tags are removed, snippets of text with user-defined size are extracted and so on) and by the user (e.g tagging and extracting manually relevant texts from cached pages) to generate the buzz monitoring reports.

5 Future Work

There are many ideas we would like to include in the system based on the theory of social maps and scale-free networks and our own findings. In future developments of BuzzMonitor we are interested in including at least two important new features:

- A pattern-matching component based on user-defined regular expressions for gathering specific data from page results. This feature will allow the user to apply filters within the results and could, for example, provide more specialized search results such as users demographic data: such as gender, age or city based on their on-line postings.
- A social map graphical representation of hubs, i.e. users that have much more incoming links than average on the market research subject. Theses maps will help to identify such users and to better understand the dynamics of word of mouth spreading through the net. For example: if a blogger mentions he is fond of President Lula, the user might monitor other blogs that are connected to this one and check if his linked acquaintances have some opinion about President Lula as well. If so, we will

classify this opinion. In the long run we will be able not only to detect who are the influentials in many topics but whom they are influencing directly.

We also started investigating the impact of the use of a tool like BuzzMonitor on the knowledge management field. As the system lets users to store their Google searches and to further refine them, categorizing them and generating reports on the search results, we believe that BuzzMonitor could also morph into something like *Kmonitor* (K for knowledge) for helping large companies to manage the overwhelming amount of information generated everyday by their employees´ Google searches.

6 Conclusions

Our work aimed to demonstrate that it is possible to monitor word of mouth and help marketers to identify opinion leaders through their scale-free networks social maps. We believe BuzzMonitor will help companies to make systematic use of the theories related to the study of Buzz Marketing and Word of Mouth on social networks.

This work also demonstrates that it is possible to analyze on-line conversations among Internet users, not only in a qualitative way, but using quantitative data such as incoming links, presence of hubs and clusters. In the long run, systematic use of the tool will allow our group to study the implications between a social network topology and the *Fitness* of each node, as described by Barabási.

Finally we shall be concerned with user privacy when deploying these techniques and BuzzMonitor on a real world case. Knowing a person social network and making systematic use of monitored conversations may allow companies to try to influence what people talk about. These opinion leaders (hubs) will be responsible for spreading the buzz about a certain product or service. This is being called as Consumer Generated Media (CGM). Making a social network more visible implies on more lack of privacy among Internet users. Market researchers and on-line marketers should be able to make systematic use of this information without disrespecting personal privacy.

7 References

[Barabási2003] BARABÁSI, Albert-Laszló. 2003. *Linked: how everything is connected to everything else and what it means for business, science and everyday life*. New York: Plume, New York.

[Ebel2003] EBEL, H.; MIELSCH, L.; BORNHOLDT, S. 2002. *Scale-free topology of e-mail networks*. Institut fur Theoretische Physik. Available at:<http://www.theo-physik.uni-kiel.de/~bornhol/pre035103.pdf.> on March, 2004.

[Gladwell2000] Gladwell, M. 2000. *The Tipping Point: How Little Things Can Make a Big Difference*. Little, Brown, New York.

[Godes2003] GODES, David; MAYZLIN, Dina. 2003. *Using Online Conversations to Study WOM Communication.* Boston. Available at: <http://www.som.yale.edu/faculty/dm324/wom2.pdf> on March, 2004.

[Milgram67] Milgram, S. 1967. *The small world problem.* Psychology Today 2, 60–67.

[Rosen2000] ROSEN, Emanuel. 2000. *The Anatomy of Buzz: How to Create WOM Marketing.* New York: Doubleday, New York.

[Watts2003] WATTS, DUNCAN J. 2003. *Six Degrees*: The Science of Connected Age. New York: W.W. Norton, New York.

[urlAnet2005] http://www.amanet.org/books/catalog/0814472958_Conversation.htm