

Sysomos Inc.

Business Intelligence for
Social Media

contact@sysomos.com

www.sysomos.com

(866) 483 3338

120 Baldwin St, Suite 3

Toronto, ON

M5T 1L6

Whitepaper

Nov 2008

Identification, engagement and measurement:

Putting It All Together for WOM Marketers

Nilesh Bansal and Dr. Nick Koudas

Word of Mouth Research Symposium

WOMMA Summit 2008





ABSTRACT

Word of mouth (WOM) marketers face two main challenges:

- (1) Identification of seed influencers who belong in target geographic and demographic profile
- (2) Post-engagement continuous monitoring of impact of the campaign.

Human analysts are commonly employed for these tasks, making WOM campaigns expensive and time consuming. In this paper we tackle both problems utilizing information technology and algorithmic techniques demonstrating how a software platform is engineered to assist marketers. We recognize influence quantification, geographic analysis, demographic segmentation, text analytics, linguistics technology, and measurable metrics as key components required to build a unified platform to serve WOM marketers. In this paper we discuss each of these components and share our learnings from building the Sysomos' software platform.

Accurate quantification of influence is central in selection of seed influencers for a campaign. Several attributes such as inlink count, number of comments on a blog post, friend count in social networks, and number of diggs are commonly used for this task. The problem is further complicated by inclusion of restrictions on geography and demographics of matching author profiles. We explore all these alternatives in detail and discuss ways to go beyond simple counts, e.g., while analyzing inlink count taking in account the network structure, authority of participating nodes, and topic of the content.

Measuring of the impact of a campaign is important to establish the ROI. The community has proposed several measurable metrics for these tasks which we build upon. For a software platform to be able to measure these metrics, it must meet several requirements:

- comprehensive coverage from multiple sources
- scalable data mining
- multilingual text analytics
- natural language processing
- customizable reporting capability

We discuss each of these attributes in detail.



INTRODUCTION



Blogs, social networks, message forums, collaborative wikis, micro-blogging sites, and various other Web 2.0 platforms offer unique opportunities for information exchange and online collaboration.

Word of mouth (WOM) is increasingly becoming predominant in marketing and customer outreach. Although the fundamental principal of WOM is naturally rooted in human social activities, the advent of the internet and Web 2.0 provides novel opportunities for WOM marketing, product placement, and opinion solicitation. Increasingly individuals utilize such means of communication and online expression. The numbers are indicative of the adoption and penetration of such services: fifty million Americans actively read blogs (Pew Internet & American Life Project), almost half a billion individuals are registered on one or more social networking site, and millions of people are utilizing microblogging services. Given such vast numbers of user engagement, these services provide a unique opportunity for WOM marketing by enabling a wide global reach than ever before. As individuals increasingly spend more time online, communicating and expressing themselves, identifying suitable individuals and engaging with them in their online world is becoming a reality and a pressing concern.

In this paper we will present a study of issues associated with user identification, engagement, and measurement. We share our experiences in building the software platform at Sysomos that continuously scans user generated content (UGC) sources and collects, cleans and aggregates them. Intelligence gathered is made then available for subsequent tasks of influencer identification and impact measurement, delivered via the Software as a Service (SaaS) model. Working closely with professionals in the domain, we have identified a typical workflow that WOM marketers have for each online project.

Such a workflow usually consists of (a) identifying a set of individuals with influential online presence to engage in order to seed a marketable entity (e.g., product) (b) engaging with these individuals to identify their interests in acting as entity (product, brand etc) evangelists and (c) monitoring and measuring the results of such engagement in terms of impacting or fostering discussions around the entity online to identify how the perception (adoption) of the entity changes (as a function of time) at various geographical locations. The workflow is commonly cyclic in practice, whereas the first step follows the last step again.

Figure 1 Typical WOM workflow for online campaigns



In the remainder of the paper we will discuss each of these steps in isolation and highlight how a software platform can aid each of them, highlighting challenges and issues that one needs to address in order to provide an effective and usable solution.



IDENTIFYING INFLUENCERS

Identification of individuals with influential online presence (possibly around the entity of interest) is an important first step for online WOM marketing. What constitutes online influence is an issue of ongoing discussion. Although most probably there is no single 'best' definition, there are a multitude of parameters that need to be considered in assessing online influence. Such parameters also depend on the particular instance of online forum considered. The table below lists commonly used measures of influence across multiple Web 2.0 platforms.

In the case of blogs, the number of individuals reading the posts of a candidate influencer on a daily basis is the most informative factor. This number however is not always publicly disclosed, and other metrics are therefore used instead. The number of inlinks to a blog by other blogs is the most commonly used parameter since this information is easily available. While count of inlinks provides some information about influence, there is a lot more that can be inferred from the structure of the inlink network to a blog. For example, ranking a blog which is linked by many low authority blogs and one which is linked by a few high authority blogs requires sophisticated network analysis similar in spirit to what Google does with PageRank. This network analysis process however is much more complicated than that on traditional web where links are the only criterion for authority and influence. In the blog world, nodes in the network (each representing a single blog), are also associated with additional information regarding number of comments, frequency of posting by the blogger, bookmark counts, digg counts, and writing style of the blogger. The influence propagation mechanism in this network must take in account all these attributes to rank a blog surrounded by few "good" blogs higher than one surrounded by many "average" blogs.

Figure 2 Different criterions of influence.

Blogs	Social Networks	Microblogging	Video Sharing
<ul style="list-style-type: none"> • Number of readers • Number of inlinks • Structure of inlink network • Bookmark count on Delicious • Total Digg count • Post frequency • Average number of comments per post 	<ul style="list-style-type: none"> • Number of friends • Friend network structure • Profile page view count 	<ul style="list-style-type: none"> • Follower count • Update frequency 	<ul style="list-style-type: none"> • View count • Rating • Number of comments • Favorite count

The Figure below shows a simplified example network whereas each circle represents a blog and each edge represents presence of a link between two blogs. The color of circle represents its quality based on all additional attributes, darker



nodes representing higher quality. The blog A is linked by 5 *average* blogs while blog B is linked by three *good* blogs. Simple inlink count will wrongfully pick blog A as an influencer even though blog B is a better choice.

• • •

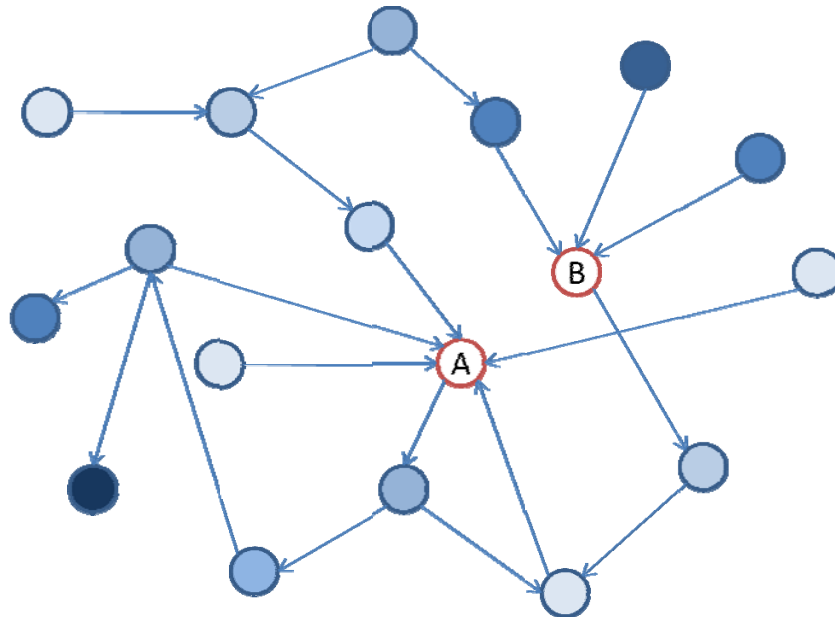
Assessing blog influence for WOM purposes should allow for the flexibility to weight the model attributes

• • •

Most campaigns are focused towards select domains. An influencer in one domain may not have the same amount of influence in other domains. For example, a popular automobile blogger is not influential for consumer electronics domains. For WOM purposes however it is imperative to associate links between blogs to the specific topics and domains. In other words, links must be isolated so that can be precisely attributed to the domain of interest of the campaign. We refer to such links as WOM-domain inlinks.

We propose a probabilistic approach that can combine all attributes available for a blog and assess influence propagation in such a network. Our approach is recursive and at each step of the recursion for each blog, several attributes are combined to derive the overall blog influence. We believe that any approach for assessing blog influence for WOM purposes should allow for the flexibility to weight in the model attributes (e.g., comments, WOM-domain inlink count, number of bookmarks, etc) in different ways. This flexibility assures that the model can be tuned differently for each campaign depending on its specific needs.

Figure 3 Example blog inlink network



In social networks friend count is frequently used as a measure of influence. Again, this simple count based measure fails to take in account the graph structure. An individual connected with other well-connected friends and popular communities is more valuable than one connected to equal number of closed community of friends. This points to the



need for sophisticated topological examination of the neighborhood of each individual in order to assess influence. Influence prorogation based friend network analysis can unearth more informative influence information than plain friend counts.



Advanced text analytics
and language processing
technology enables
effective understanding of
content



Based on project requirements, marketers start by constructing a target profile which usually consists of attributes such as age, gender, geographic location, profession, lifestyle, interests and hobbies. Then the search for individuals matching these criteria is conducted and top influencers are selected for engagement. These searches are commonly executed by employing an army of analysts who browse through thousands of different blogs, social network profile pages, and websites. Given the huge amounts of data on the web, performing this task manually is a painstaking and expensive proposition. Specialized software platforms are therefore needed to automate these tasks. We believe it is possible to offload a significant portion of the influencer identification task to computers. For a software system to qualify however, it must meet some basic requirements as discussed below.

Commonly WOM projects have a well defined geographic scope. For example a WOM campaign might commence in a large metropolitan city, specific cities in a country, or certain countries. It is imperative to be able to restrict the search for influential individuals at the desired geographical scope. Commonly individuals disclose their geographical locations at some level (city, state, country). This is true in the majority of social or blogging platforms. Such locations can be extracted, standardized and disambiguated and made available for querying and further restricting candidate influential individuals. In cases where the geographical location is not explicitly disclosed, natural language processing tools can be utilized to make a best effort guess. Same is

true for restrictions on age, gender, and profession attributes.

An additional requirement for target individuals could be their interests, hobbies, and expertise in particular domains, in addition to online influential activity. The specific ways that such expertise is expressed is usually flexible. As an example WOM projects might request the identification of individuals that frequently talk about kitchen appliances or individuals who are usually interested in automobile domains. Being able to aid in a precise definition of these attributes and then conducting searches based on them is a basic requirement for any software platform. From a technology point of view this means that the software must encompass advanced text analytics and language processing technology that enables effective understanding of the content of postings.

Social media consists of different forms of online expression by individuals. Blogs are just one part of the equation, and to get a 360 view of the universe, it is important to cover different data sources. Social networks have tremendous amounts of useful information in terms of author profiles and friend networks. There are several networks with specialized vertical communities, e.g., Cafe Mom for mothers, or Nexopia for Canadian teens. Microblogging services Twitter and Jaiku are re-defining the concept of blogging by restricting length of each post to 140 chars. Video sharing



sites, most notably YouTube, have provided a new platform to reach millions of consumers by means of a single viral video.

Given the global product market and availability of social media content in multitude of different languages, the system must provide its searching and analysis capabilities in all languages. Automated machine translation should be available to convert all content to a single language.

ENGAGEMENT



Once target individuals have been identified the next phase is to engage with them in order to identify their level of interest in participating in the campaign. The software platform can aid by extracting contact information for these individuals from their profiles if it is disclosed. In cases where the contact information is not publicly available, comment on a blog post or social network page can be used for first contact. The first contact, however, can be tricky.

It is crucial to make the message personal, otherwise the individual may discard it as spam. Automated messages usually don't work with today's sophisticated internet users. The system can help again by providing a brief summary of topics that an individual writes about to assist in the construction of a personal message for the first contact. Once the contact has been made and the individual agrees to participate in the campaign, WOM marketers move on to the third step of measurement.

MEASUREMENT

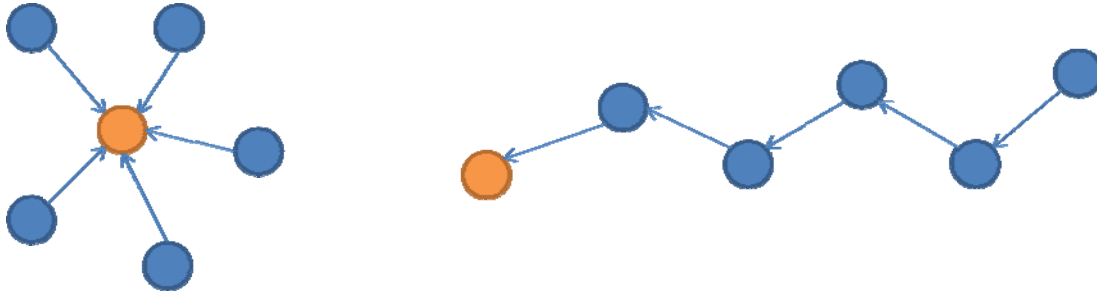
The next phase following the successful engagement of a set of influential individuals is to quantify the effectiveness of the WOM marketing campaign. Several measures can be adopted to quantify effectiveness. Based on the premise that influential individuals have been engaged to comment on or promote entities of interest, it is natural to expect that at the very least, the volume of discussions related to the entity of interest should increase in the days following the launch of the WOM campaign. Thus one measure of interest is the successful monitoring of the volume of mentions/discussions for an entity of interest in specific temporal periods.

In addition to mentions/discussions, identifying and monitoring for deeper engagement with the entity of interest is a necessary requirement. The precise definition of engagement in an online world is also open to multiple interpretations. In the blog world, number of trackbacks and comments are commonly used to quantify the engagement. Linking activity



or trackbacks usually reveal a deeper form of engagement between bloggers when compared to a comment. The reason is that a post P linking to or citing another post Q expands the contents of the original post Q by referencing it. Other bloggers may then link to the post P , generating threads of discussions in the form of post cascades for varying depths and sizes. The shape of these cascades is equally important as the size. The Figure below shows two post cascades each with 6 participating blog posts originating from a single node. In the first case, the discussion is limited with 5 direct links to the originating post, while in the second case there is a long chain of discussion. Identifying and measuring the size of large post cascades as a result of a post from an influential individual, reveals a certain form of engagement.

Figure 4 Two different types of post cascades of same size



In order to combine different measures of engagement (e.g., inlink count, number of comments, etc..) to compute the “influence” of a single post P , we propose use of an additive formula as shown in the equation below. Weights w_1 , w_2 , w_3 .. can be adjusted according to need. The “influence” of all posts in a time period can then be aggregated to quantify the total buzz.

$$\text{Influence}(P) = (w_1 * \text{numberOfInlinks}) + (w_2 * \text{bookmarkCount}) + (w_3 * \text{numOfComments}) + \dots$$

Certain spatio-temporal aspects of engagement are of interest as well. The ability to map mentions or cascades to specific geographic regions is highly important since it provides insight as to which cities, states, countries the WOM campaign is most successful in. The ability to segment engagement further by several demographic factors is also highly useful. Such segmentation provides insight into which demographic groups the WOM campaign yields results. For example one might identify that males between ages 24-35 living in Toronto, Canada are engaging successfully.

The sentiment (tonality of text) for the entity of interest is the crucial measurable quantity. Being able to characterize discussions or mentions of the entity of interest as having positive or negative tone provides an instant measure of effectiveness for the WOM campaign. Large campaigns can generate high volumes of chatter resulting in hundreds to thousands of reactions. Manually reading all this content is not always practically feasible. Software platforms must therefore provide automated tone analysis and ability to summarize these results in concise visual forms utilizing entity extraction and text summarization tools.

WOM marketing efforts are geographically global in reach and scope. Thus, support for search and analysis in multiple languages is an absolute requirement. Moreover, fully functional and integrated language translation capabilities should be available to aid comprehension and understanding of multilingual content.



Measurement and monitoring are continuous processes that must be carried throughout the duration of the campaign. Sentiment quantification coupled with advanced text analytics that provide the means to capture automatically the main themes of discussions around the entity of interest, provide a powerful driving tool for campaigns. In effect they provide powerful feedback to the WOM process helping to identify the elements of the process that are successful but also the ones that are not and aid to correct them. This imposes interactivity and real-time requirements on the software systems. In a fast moving world, analysts cannot wait for results to be produced, and sub-second querying and to-the-minute monitoring are absolute musts.

CONCLUSIONS



We provided an outline for the workflow of a WOM marketing campaign online. Identification of influential individuals online in order to engage in the campaign as well as measurement of the impact of the campaign are important phases of this workflow. A large chunk of both these phases can be automated. There are several social media product suites available in the marketplace, which come with varying levels of sophistication. These tools provide services ranging from basic keyword based monitoring functionality to full blown dashboards and cater to industry verticals of marketing and PR.

We have identified key requirements for a system suitable for WOM marketers, summarized below as a checklist. Based on items on the list above we have engineered the Sysomos' platform over the last three years to provide maximum assistance to WOM marketing professionals. This list is validated by our interactions with experts in the domain and their usage of the Sysomos' platform.



Figure 5 Checklist of key requirements for a software platform to serve WOM marketing professionals.

