

# Mapping Socio-Cultural Dynamics in Indonesian Blogosphere

**Shamanth Kumar**

Computer Science and Engineering,  
Arizona State University  
shamanth.kumar@asu.edu

**Nitin Agarwal**

Department of Information Science,  
University of Arkansas  
at Little Rock  
nxagarwal@ualr.edu

**Merlyna Lim**

School of Justice and Social Enquiry  
and Consortium of Science, Policy  
and Outcomes  
Arizona State University  
merlyna.lim@asu.edu

**Huan Liu**

Computer Science and Engineering,  
Arizona State University  
huan.liu@asu.edu

## Abstract

Understanding socio-cultural dynamics of a society requires an exhaustive study of the people who form the society. This requires collection of large amount of demographic data, which is a difficult task. Blogosphere provides a convenient solution to this problem. With the pervasion of Internet, and blogging in particular, it has become easier to collect data from the virtual world, which can be used to study the social and cultural behavior of a community. The Indonesian blogosphere offers rich examples for interactions between multi-cultural society and the Internet. With the help of a social scientist, we motivate the need to perform a study of the Indonesian blogosphere. We identify relevant theories that explain our observations. We also point out interesting problems that can be studied in the Indonesian blogosphere and present a preliminary study of the blogosphere based on the Indonesian blog data that we have collected.

## 1. Introduction

Blogosphere, the network of blogs, is growing at a phenomenal rate. Technorati<sup>1</sup>, one of the largest blog indexing service, states that it's current index size is around 133 million blog records. In addition there is an influx of roughly 900,000 blogs every day. Blogosphere has no central controlling authority and provides freedom to those who might be interested in voicing their opinions anonymously. This attribute makes blogosphere an attractive medium of communication and an interesting medium to study.

In Indonesia too, like elsewhere in the world, over the past few years blogs have become increasingly popular. In

fact, Indonesian blogosphere, is arguably one of the most dynamic and culturally rich blogosphere. Being a new democratic nation and the most populous Muslim country in the world, Indonesia is inherently unique. Its 216 million people have survived colonialism under the Dutch, a slide towards communism under Sukarno, more than 30 years of authoritarianism under Suharto, religious, regional, and ethnic violence, and an eruption of Al Qaeda-style radicalism. By all measures, Indonesia should be one of the most difficult terrains for democracy. Yet, with all its past turmoil, it is moving steadfastly through a complicated democratic election process in 2009.

Indonesia is an immensely diverse and complex country comprised of a rich history, and 922 permanently inhabited islands (or more than 17,000), each multi-ethnic, multi-linguistic, and multi-religious. Internet penetration is not ubiquitous, its use is not widespread across demographics, and Internet Service Providers are centralized predominantly in hub cities. However, Jakarta, Indonesia's capital city, is one the top 30 cities for blog postings and comments according to BusinessWeek (2007) and according to WordPress statistics (December 2008). Bahasa Indonesia is one of the top languages used by WordPress.

The Indonesian blogosphere is chosen as a case study due to the significant contribution social media provides in expanding the spheres of social, cultural, and political expressions. The goal of this paper is to produce a baseline assessment of the networked public sphere in Indonesia and its relationship to a range of emergent issues, including politics, media, religion, culture, and global affairs. This paper reveals the preliminary assessment of a broader study.

## 2. Indonesian Blogosphere

Most of the studies on blogosphere are conducted either in a democratic (the United States) or an authoritarian system (such as Middle Eastern countries). There is a lack of stud-

Copyright © 2009, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

<sup>1</sup><http://technorati.com>

ies in countries such as Indonesia where political system has been undergoing “transition” to democracy. Transition in this context is used to refer to the process of Indonesia’s conversion from an undemocratic authoritarian system under President Suharto (1965-1998) to a more democratic system marked by the first and the second presidential elections.

Media plays an important role in transition to democracy. Access to information is essential for the health of democracy for at least two reasons. Firstly, it ensures that citizens make responsible and informed choices rather than acting out of ignorance or misinformation. Secondly, information serves as a *checking function* by ensuring that elected representatives uphold their oaths of office and carry out the wishes of those who elected them. In some societies, an antagonistic relationship between media and government represents a vital and healthy element of fully functioning democracies.

Within the context of supporting democratic transitions, media in Indonesia is in transition from a government directed and controlled domain to one that is more open and has a degree of editorial independence that serves the public interest. If the media is to have any meaningful role in democracy, then the ultimate goal of media assistance should be to develop a range of diverse mediums that are credible and to create and strengthen a sector that promotes such outlets of information. This will enable the citizens to have access to information that they need to make informed decisions and to participate in society.

Even during the Suharto era, Internet development in Indonesia provided a more conducive democratic media environment when compared to other media spheres such as television, radio and print media. In addition, political experiences of the 1990s, show how the substantially unregulated Internet contributed to the political reforms that led to the downfall of Order Baru (Lim 2002). The blogosphere thus emerged in an unregulated environment of the Internet in Indonesia, which naturally was also an open, free, unregulated landscape for blogosphere to exist.

While some Indonesian bloggers had started their first blogs from 2001, blogging activity only became popular in 2004, mostly due to the progressive campaigns of some early bloggers such as Enda Nasution, the father of Indonesian blogging, Priyadi and A Fatih Syuhud as well as the availability of (currently) free blogging platforms such as Blogger, LiveJournal, Blogdrive, Wordpress, Multiply and Friendster’s blog.

The number of Indonesian bloggers is not very well documented. At the Pesta Blogger 2007, in Jakarta, Enda Nasution said that the number of bloggers in Indonesia had reached around 150,000. It is unclear how he came to that figure. However compiling some statistical data from various sources like Alexa, Multiply, Compete and Wordpress, we can come up with these figures for Table 1:

- Number of global Internet users  $U = 360,985,492^2$ .
- Total users for a social networking site  $T = p\%$  of global users for the site  $x U$ .

<sup>2</sup>Internet Stats, August 2008

Table 1: Indonesian blogger statistics for popular social media websites.

	rank in IND	IND rank globally	% IND users q	% of global users p	Total users T	Unique visitors/month (1000)	IND users I	IND bloggers (est)
Multiply	15	1	21.60%	0.42%	1,516,139	828,000	327,486	<b>261,989</b>
Wordpress	7	3	6.40%	2.56%	9,241,229	19,360,000	591,439	<b>473,151</b>
Blogger	5	5	3.70%	7.20%	25,990,955	24,660,000	961,665	<b>769,332</b>
Livejournal	41	10	2.20%	0.74%	2,685,732	3,936,436	59,086	<b>47,269</b>
Friendster	3	1	37%	1.25%	4,512,319	1,403,000	1,669,558	N/A
Facebook	8	19	1.10%	10.72%	38,697,645	41,181,000	425,674	N/A
MySpace	39	24	0.30%	6.19%	22,345,002	58,389,000	67,035	N/A
YouTube	6	18	1.10%	16.51%	59,598,705	62,311,000	655,586	N/A

- Total number of Indonesian users  $I = T \times q\%$  of Indonesian users of the social networking site
- Estimated number of Indonesian bloggers =  $80\% \times I$  (observably less than 20% of the users are not bloggers/subscribers)

As shown in Table 1, without counting blogs of Friendster (no data on that), we can roughly calculate that there are approximately 1.5 million active blogs in Indonesia (only those who reside in Indonesia), with around 500,000 to 1.2 million active bloggers.

More interesting than the numbers is the fact that Indonesia tops the charts in major social networking sites and even occupies number one position in both Multiply and Friendster. While compared to the Internet population, 25 million out of 240 million people in the country, the number of active users of these social networking websites is still small, the fact that it comprises a significant proportion of global users certainly shows that there is a certain fit between Indonesian society and social networking.

### 3. Blogging and Politics

There are two dominant threads of speculation about the role of the Internet, including social networking platforms such as blogging, in Muslim world such as middle eastern countries and Indonesia with respect to politics. The first is a view of the Internet as primarily a vehicle for radicalization. The second thread is the hope that the Internet will empower political movements that can move Muslim societies towards democratic values and governance. Previous studies on Indonesian Internet suggest that in Indonesia, the Internet plays the role of a double edged sword (Sen and Hill 2005; Lim 2002; 2003; 2009). While the fundamentalism and radicalism do appear (Lim 2002; 2003; 2005), previous studies found little support for terrorism or violent jihad in the blogosphere (Lim 2009). The studies support some aspects of the view that the Internet can empower political movements in the country since it provides infrastructure for expressing alternative views (to the mainstream) and mobilizing voices for bottom-up contributions to the public agenda.

However, while the Internet lays a good foundation for a battle of ideas, it does not necessarily favor a winner, and is not proven to have capacities to fundamentally transform society into a more democratic state.

With the rapid growth of blogosphere in Indonesia, many questions arise about its significance in politics. What is the specific role of blogosphere in political mobilization, public discourse and political outcomes? How do the features of blogs influence the discourse on political issues in Indonesia? How do these features accommodate communities to cut across traditional lines of affiliation? To what extent will the Indonesian online public organize into meaningful 'global ummah' discourse communities versus reinforcing nationalism? Do they offer promising possibilities for new type of Muslim communities to emerge, such as a more globalized community? Do they open opportunities for minority groups, such as Christians or Chinese Indonesians, to reinforce their identity? What are the implications of the blogosphere for Islamic movements, including the internal politics of Islamic groups, their capacities to mobilize, and their places in the wider public sphere? Is there tension between bloggers who want to bridge internationally and the domestic bloggers?

These are very important questions that are critical to our understanding of not just socio-political dynamics in Indonesia but also of broader social dynamics of our contemporary world. It is our ambition to be able to have a tool that will enable us to map trends and dynamics reflected from these questions. These questions, however, cannot be answered without doing a thorough longitudinal study with a rich set of data that covers the depth and the breadth of Indonesian blogosphere. Before moving further to such a study, we need at least a preliminary study to map the Indonesian blogosphere, which is exactly what this paper attempts to do.

#### 4. Mapping the Blogosphere

Considering the novelty of blogosphere, there are only few studies that have attempted to map the blogosphere. The first and, possibly, the most well-known mapping work is the map of the American political blogosphere with respect to the 2004 US elections (Adamic and Glance 2005). The authors studied linking patterns and discussion topics of political bloggers to measure the degree of interaction between liberal and conservative blogs and to uncover differences in the structure of the two communities. They concluded that the American blogosphere is politically divided, with liberals and conservatives linking primarily within their separate communities, with far fewer cross-links exchanged between them. They also found out that this division extended into their discussions, with liberal and conservative blogs focusing on different news articles, topics, and political figures. In other words, American blogosphere reflected the polarization theory (Wilhelm 2000). Polarization theory is fundamentally rooted in the principle of homophily (McPherson, Lovin, and Cook 2001) which states that people who share certain characteristics will interact more often and more closely than those who are dissimilar (Sandell 1999; Granovetter 1973). In general, homophily is found to have a

strong influence on the organization of social networks. The categories which lead most directly to strong homophily are race, ethnicity, and recently, religion.

Similar, yet broader studies were conducted on the Iranian (Kelly and Etling 2008) and Arabic (Etling et al. 2009) blogosphere. Using computational social network mapping in combination with human and content analysis, the Iranian study indicates that Iranian blogosphere is dominated by four major network formations, or poles, with identifiable sub-clusters of bloggers within these poles, which are labeled as secular/reformist, conservative/religious, Persian poetry and literature, and mixed network. The Arabic study found a similar conclusion except that the network formation is mostly based on countries as well as dominant ideological lines such as the Muslim Brotherhood.

While social networking mapping tool seems to be more established, pinpointing on these studies, we found out that there is no sufficient tool to code the content and to generate patterns of blogger behaviors. We believe it is highly important to establish a right tool that enables researchers to automatically read the patterns of bloggers behavior through data analysis.

#### 5. Challenges in Indonesian blogosphere

Some observers see blogs as a catalyst for change, and as people's media and an empowering tool. They see the rise of blogs as the beginning of an era of citizen journalism in which the marginalized can play a greater role in creating rather than merely consuming news. Others argue that public debate would be dramatically revitalized if politicians would all start blogging (Gillmor 2006). However studying political blogosphere is not very convenient because of some of the inherent problems it suffers from, which are generic to the global blogosphere.

The blogosphere suffers from an unequal distribution of readers. The distribution of links and traffic is skewed so that only a handful of bloggers get most of the readers. Generally speaking, these are either those who got established in the blogosphere early when there was little content, or were already well-known as either journalists or politicians. This tendency shows that the blogosphere is not an exemplary public sphere in which everybody's voice is heard. This could be also seen as a selection process, weeding out the bad blogs, it also favors players who got in early or who make outlandish statements to attract readers (Shirky 2003).

Blogosphere is a free media and there is no central organization to the blogosphere and little consensus among bloggers with regard to many key issues. This creates a virtual Tower of Babel in which voices tend to become so particular and so exclusionary of other views so as to be unable to communicate with each other or to a broader audience. On the other hand, the proliferation of unique point of view in the blogosphere may encourage genuinely individual voices to emerge and perhaps even foster real dialogues (as opposed to the watered-down positions distributed in mass media) (Wilhelm 2000).

The amateur status of bloggers also raises many questions. Most bloggers are part-timers for whom blogging is a voluntary endeavor. Amateur bloggers do not, in general,

have the resources and capacity to investigate material prior to publishing it. Thus, from a journalistic point of view, the credibility of blog entries generally cannot meet those of articles in mainstream media. On the other hand, the voluntary nature of blogging implies that it is also a positive way for regular people to voice their opinions without going through the filtering effects of traditional journalism.

## 6. BlogTrackers

BlogTrackers (Agarwal et al. 2009) is a tool that can be used to track and analyze the blogosphere. Analyzing blogosphere to look for patterns in user behavior involves, collecting of the blog posts and analyzing them to study the behavioral patterns of bloggers and other users in blogosphere. A social scientist would do these tasks manually. Consider a scenario where a social scientist might be interested in studying the blogs from a particular ethnic region say Iran, to identify the influential bloggers who wrote about the recent presidential elections. Once the relevant blogs have been identified, the most important task is to obtain the blog posts relevant to the topic. The analysis of these blogposts has to be performed manually identify influential blogposts. Various factors could define the influence of the blogpost such as the number of comments a blogpost receives or the number of blogposts linking to it. These tasks require considerable amount of time and effort. BlogTrackers combines these two tasks under a single platform to reduce the time required to perform such an analysis over blogs. The tool is straightforward and convenient to use and has a short learning curve. BlogTrackers currently consists of the following features which can assist the social scientists in their work:

**Crawlers** BlogTrackers consists of two types of crawlers, the batch crawler and the RSS crawler. The batch crawler can be used to crawl the archived blogposts from a blog. We use regular expressions to parse the data from the HTML pages. This is helpful since most blogs do not provide API's to access their archives. Other crawlers can also be easily integrated into BlogTrackers. The second crawler in the BlogTrackers system is the RSS crawler. Unlike the batch crawler, the RSS crawler extracts blogposts from the RSS feed of the blogsites. This crawler can also be scheduled to run at a specific time of the day to automatically crawl blogposts on a regular basis.

**Event Detection** Events of interest can be identified using the Traffic Pattern Graphs feature in BlogTrackers. A trend that we have observed over a period of time is that a real world event causes a surge in blog traffic. In this feature we try to take advantage of this particular trend to identify such events. The graph displays the aggregated count of blogposts for a specific time period at various scales ranging from daily to yearly. Both the time period and the scale can be varied to generate graphs that depict the blog traffic at different granularity levels for fine grained analysis. The graphs also allow the user to analyze the keywords and the blog posts associated with the time period, which can be

used to pin point an event.

**Influential Blogger** The influential blogger component can be used to identify influential bloggers using a generic model of influence derived from the number of inlinks, number of outlinks, the number of comments and the length of the blogposts. We identify the maximum influence among all the blogposts of a blogger to summarize the blogger's influence. A more detailed explanation is mentioned in Section 7.

**Blogger influence and time matrix** One of the features of BlogTrackers is the ability to correlate influence with a blogger's activity levels. Activity level of a blogger can be measured by counting the number of blog posts posted by a blogger. The correlation is presented in a convenient confusion matrix that displays the bloggers in terms of their activity levels and influence. A blogger with a large number of blog posts can be easily mistaken to be influential. This feature helps segregate active bloggers from the influential bloggers.

**Watchlist and Alerts** BlogTrackers system can identify relevant keywords from the indexed blogposts and suggest them to users for setting up offline and online alerts. Every user can have a watchlist consisting of keywords that he/she would like to keep track of. The system automatically sends an e-mail notification to the user when a blogpost containing one of the keywords is crawled and indexed. These future notifications are important and help in keeping track of specific topic related keywords over time with little human intervention.

## 7. Influential Bloggers

An influential blogger is one who has a significant proportion of the users following his opinions on relevant issues. To identify influential bloggers for our experiments as described in the later sections we use a generic model of computing influence scores of individuals (Agarwal et al. 2008) based on both link and content information with tunable weights. Some of the desirable properties of an influential individual are summarized as follows:

**Recognition:** An influential individual is recognized by many. His writings,  $p$  are referenced by other individuals. The more influential the referring individuals are, the more influential the referred individual becomes. Recognition is measured through the inlinks ( $\iota$ ). Here  $\iota$  denotes the set of inlinks to an individual's writings  $p$ .

**Activity Generation:** An individual's capability of generating activity can be indirectly measured by how many comments he receives, the amount of discussion he initiates. A large number of comments ( $\gamma$ ) indicates that the individual *affects* many such that they care to write comments, and therefore, the individual can be influential. Some of these comments could be spam which could be eliminated using the existing work in (Kolari, Finin, and Joshi 2006; Lin et al. 2007).

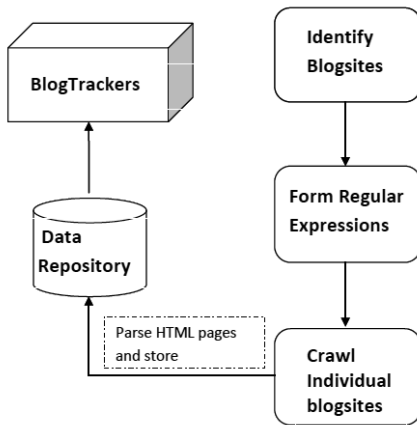
**Novelty:** Novel ideas exert more influence as suggested in (Keller and Berry 2003). Hence, the outlinks ( $\theta$ ) is an indicator of novelty. If an individual’s writing refers to many other articles it indicates that it is less likely to be novel. An individual’s writing  $p$  is less novel if it refers to more influential articles than if it referred to less influential articles.

**Eloquence:** An influential individual is often eloquent (Keller and Berry 2003). Given the informal nature of the social media, there is no incentive for an individual to write a lengthy piece. Hence, a long writing often suggests some necessity of doing so. Therefore, we use the length ( $\lambda$ ) as a heuristic measure for computing eloquence.<sup>3</sup>

## 8. Data Collection

Our data source consists of selected Indonesian blogs which were chosen because of their ranking in the Indonesian blogosphere and because of their political orientation. These blogs were selected from the Top 100 Indonesian blogs published by Indonesia Matters<sup>4</sup>. The ranking here is based on a number of different factors such as Google Page Rank, Yahoo Backlinks, BlogLine RSS Subscribers, Alexa Traffic Rank, Technorati Blog Rank and Del.icio.us Social Bookmarks. More details about the ranking can be found at the FAQ page in Indonesia Matters. From the selected 100 blogs we then manually short-listed 19 blogsites which had politically oriented blogposts.

Figure 1: Data Collection Methodology



A visual representation of our methodology for data collection is shown in Figure 1. One of the major challenges in collecting data from the Indonesian blog sites was the lack of API support. The blog sites, for this reason, had to be crawled by scraping the HTML pages. We formed regular expressions for each one of the 19 blogsites to parse the

<sup>3</sup>This property is most difficult to approximate using statistics. Eloquence of an article could be gauged using more sophisticated linguistic based measures.

<sup>4</sup>Indonesia Matters <http://www.indonesiamatters.com>

Table 2: Dataset Characteristics

Number of Blog sites	19
Total number of blogposts	13046
Average number of blogposts per blogsite	686
Average number of comments per blogpost	20
Average number of inlinks	4
Average number of outlinks	1
Average length of blogposts	2469 characters

HTML pages and extract information from it. This task was further complicated because of the lack of fixed formatting in the HTML pages among the blog sites. The information extracted from the blog sites for every blog post included: tags, categories, number of comments, number of outlinks, number of inlinks, blog post length, blog post text, post date and the author. For determining the number of inlinks of a blog post, we used the link search feature of Technorati<sup>5</sup>. The calculation of inlinks is therefore limited to only those blogs which were indexed by Technorati. Collected data is stored in a relational database and is indexed by our tool (Agarwal et al. 2009) to enable fast querying and search.

Our dataset consists of 19 Indonesian blog sites. All the blog posts in the blog sites are written in Indonesian using the English alphabet. The blog sites are mostly individual blog sites which record the opinions and thoughts of the bloggers. Some characteristics of our dataset are mentioned in Table 2

## 9. Research Problems

### The relationship between popularity and influence

The importance of a blogger in a community can be measured using two different measures. Popularity can be defined as the ability of a blogger to attract readers to his content. Popularity is generally measured by the traffic statistics of a weblog. Influence, on the other hand, can be defined as the amount of trust readers place on the content generated by the blogger on a specific topic. This specificity distinguishes influence from popularity. In this section, we attempt to determine whether popularity is a measure by which we can judge the influence of a blog.

In our experiment, we ranked the bloggers on the basis of both, popularity and influence, and compared the two ranking. The popularity based ranking in this experiment was collected from another Indonesian blog site Indonesia Matters. The blogsite provides ranking of Indonesian blogs based on a combination of factors. Some of the factors included Google Page Rank, Yahoo link count which measures the number of links to a blogsite, number of RSS subscribers, number of social bookmarks on the social bookmarking site Del.icio.us, Technorati blog rank and Alexa traffic ranking which is based on the amount of traffic that is attracted by a website. A score is generated based on these parameters and this score is used to generate the ranking. We

<sup>5</sup><http://www.technorati.com/search/>

Table 3: Blogger Rankings comparison based on Influence and popularity

Blogger	Popularity based rank	Influence based rank	Rank distance $\delta$
Harry Sufehmi	10	1	9
Rendy Maulana	19	2	17
Agam Rosyidi	18	3	15
Ndoro Kakung	4	4	0
Fatih Syuhud	5	5	0
theunspunblog	13	6	7
anotherfool	9	7	16
Iman Brotseno	14	8	6
Adinoto Kadir	11	9	2
Wimar	3	10	-7
Uncle Tyo	2	11	-9
Asruldin Azis	17	12	5
Anick	12	13	-1
Maseko	1	14	-13
Enda Nasution	6	15	-11
Ndobos	15	16	-1
Rovicky	7	17	-10
Jakartass	8	18	-10
Anang Surabaya	9	19	-10

obtained this ranking for each of the 19 blogsites and normalized the ranks based on their position among our list of 19 blogs. The influence based ranking was generated based on the influence model described in Section 7. The results of the two ranking methods are presented in Table 3.

We define a parameter  $\delta$  that tells the distance between the popularity rank and the influence rank of a blogger. We calculate the distance  $\delta$  as follows; let the popularity rank of blogger  $i$  be  $P_i$  and the influence rank of blogger  $i$  be  $I_i$ . Then  $\delta$  can be defined as  $\delta = P_i - I_i$ . The value of  $\delta$  tells us the change in the number of positions from popularity rank to influence rank for a particular blogger. Positive distance indicates that the blogger was ranked higher in the popularity based ranking but was ranked that many ranks lower in the influence based rank. On the other hand a negative distance indicates that the blogger was ranked higher in the influence based ranking when compared to the popularity ranking.

From this experiment we observed that the bloggers who had a high popularity rank were displaced down by an average of 8 positions in their influence rank. This is worse in the case of the topmost ranked blogger “maseko” who was displaced by 13 positions in the influence ranking. On

the other hand, the bloggers with comparatively lower ranks in the popularity ranking, on an average 9 positions in the influence ranking. The only exceptions were bloggers “Ndoro Kakung” and “Endo Nasution” whose position remained consistent in both the ranking. The results show that popularity does not directly translate to influence. Popularity based rank shows the reach of the blogger because of the high traffic and large user following. Whereas, influence based rank shows the expertise of the blogger and the number of people who follow the opinions of that particular blogger. The results also generate further interesting questions on the reasons for such a phenomenon. Some of the questions that need to be answered include the reasons behind the consistent rank of the two bloggers in the two ranking. One possible reason for this could be that Enda Nasution is also known as the father of Indonesian blogging (Lim 2009) and has been blogging for a long period of time and has therefore become both popular and influential. These results indicate that further analysis is required to understand such observations.

### Classification of Bloggers based on their Political Orientation

Bloggers can be classified into different political categories based on their views on political issues in their blogposts. Some of these categories include Islamic, nationalist, extremist, fundamentalist, etc. Identifying the bloggers political orientations can help us gauge the overall sentiment among people on any particular issue. To understand the uses of this kind of analysis, let us consider an election scenario where a candidate would like to gather support from groups that do not share his/her viewpoint. If we can identify influential bloggers for whom we know the political orientation, then those individuals whose interests do not match the candidate’s own can be targeted to understand their opinion and to appease them. This would likely help in gathering the support of other individuals who follow the blogger too.

Initial part of this task comprises of manual identification of the categories by the analysis of blogposts to look for keywords that can hint us towards the political orientation of a blogger. These relevant keywords can be identified and associated with the categories and may also need to be updated over time. Manually this task can be formidable because of the loquaciousness of blogosphere and the number of bloggers and their blogposts. However, with the help of BlogTrackers tool it can be simplified. Initial portion of the blogposts could be annotated manually and could be treated as the training set for a classifier to train a model. Rest of the blogposts could be classified using the trained model and a bloggers orientation could be identified by looking at the proportion of the blogposts that fall into each of the above mentioned categories. The tool could be enriched with such a technique to automate the task and greatly reduce the time required for such an analysis.

### Temporal Study of the Indonesian Blogosphere to Track Sentiment

Blogosphere is still in its nascent stages and is constantly evolving. This is also due to the fact that the bloggers opin-

ions change with time and this is reflected in blogosphere. One of the recent and most interesting uses blogosphere has been utilized for campaigning in the field of politics. Politicians have realized the potential of blogosphere and taken advantage of the ease with which one can reach a vast majority of people and understand their needs and sentiments. This was clearly evident in the 2004 and the recent U.S. presidential elections (Adamic and Glance 2005) where the current President Barack Obama utilized social media services like blogs and Twitter, a microblogging service, to communicate with people profusely and very effectively. This gave people an opportunity as well, to understand the presidential candidate and decide their vote.

Figure 2: Tag Cloud for the month of June 2009



Figure 3: Tag Cloud for the month of July 2009



Indonesian blogosphere has been active in terms of politics for a few years now. To the best of our knowledge, no study has yet been performed on the temporal dynamics of this particular subset of blogosphere, which is unique in its cultural and social aspects. We propose to perform a study of the current state of Indonesian blogosphere with a political viewpoint. As discussed above, the state of blogosphere is

constantly changing and so is the opinion and sentiment of bloggers towards politics. Associating temporal aspects to the analysis of such sentiments can give meaningful insight into the development of the thought process of the society. This can be performed with the help of relevant keywords. Blogposts that relate to the keywords can be analyzed to gather temporal changes in sentiments of a particular ethnic group towards political issues. BlogTrackers can contribute in this analysis by automating this process. The data collection capabilities of BlogTrackers can be combined with a keyword tracking mechanism to monitor the evolution of opinions. This can also help us to predict changes in the opinions of bloggers in the future.

We have performed a preliminary analysis of the Indonesian blogosphere through our current dataset described in Section 8 for the period Jan 2009 to July 2009 on a monthly scale. We generated important keywords for the above periods using our tool BlogTrackers and visualized the tag clouds using Wordle (Feinberg 2008). Our aim here was to identify current topics in the blogosphere, in view of the presidential elections that were held in Indonesia in July 2009. The tag clouds for the month of June and July are shown in Figure 2 and Figure 3. We observed many keywords that pointed to the democratic elections in Indonesia for example “pemilu”, “negara”, “politik”, “indonesia”, and also the names of candidates like “sby”, “prabowo”, etc.

The font of the keywords in the word cloud represents the frequency with which they appeared in the blog posts. From the figures we can observe that keywords which described elections were far more frequent than others. This indicates the proliferation with which the elections were being discussed in the blogosphere during the occurrence of the event. Another interesting thing to observe in Figure 2 is the prominence of keywords “antasari” and “corruption”. These keywords described the arrest (Barrowclough 2009) of the head of the Corruption Eradication Commission (KPK) of Indonesia in relation to a murder case which was a big blow for the current President who was running for a re-election. Such events indicate that it is indeed possible for us to track the events and its associated occurrences in blogosphere by analyzing keywords extracted from blogposts.

### Automatic Coding of Indonesian Blogosphere

Some of the recent studies on different ethnic blogosphere have been done by manual coding of the blogosphere. Studies like (Etling et al. 2009; Kelly and Etling 2008) are intended to perform social network mapping of the blogosphere with human and automated content analysis, to analyze different ethnic blogosphere. The human coding aspect of the studies requires a lot of human effort and is very expensive compared to an automated approach. The studies intended to map the network of bloggers in the ethnic blogosphere by identifying the network links manually. These studies find different political factions in the blogosphere like Reformists and Conservatives. They also identify demographic characteristics of the bloggers in the two blogosphere. Human Coders were used to identify the themes of the blogposts and classify them into categories like domestic policy, economy and politics. This is a fairly time consum-

ing task and is also dependent on external factors like the availability of human coders for the who can understand the language and idiosyncracies associated with them. Our aim is to automate this process with the help of some human coding in the initial stages through our tool BlogTrackers which can help us realize such a study in considerably less amount of time and with less effort. We intend to start this process from Indonesian blogosphere and later apply it to other ethnic blogosphere as well. This task will aid social scientists to analyze bloggers and blogposts and help compute demographic and network statistics quickly and efficiently.

## 10. Conclusion

Cultural differences between societies brings major differences in reactions to social issues, which can be identified relatively easily by the study of blogosphere. In this paper we described the utility of blogosphere in studying social responses to global and local issues with the focus on Indonesian blogosphere. We identified relevant problems, in performing such a study, which we think must be resolved. In Section 6 we introduced a tool that can help automate some of the tasks proposed in Section 9. The main intention of the tool was to implement an automated framework for monitoring socio-political blogospheres in a way that reflects their characteristics and structure. We also proposed to extend the tool in future, to perform other tasks outlined in the later sections which currently have to be performed manually, like the coding of the blogosphere and link analysis of the blogosphere. We also presented some initial results we have obtained on the relation between popularity and influence of a blogger in Indonesian blogosphere. These results highlight the need for a more extensive study of the Indonesian blogosphere, which can give us better insight into the behavior of an ethnic group.

## 11. Acknowledgements

This work is supported, in part, by ONR N000140810477 (2008), ONR N00014-09-1-0165 (2009).

## References

- Adamic, L. A., and Glance, N. 2005. The political blogosphere and the 2004 US election: Divided they blog.
- Agarwal, N.; Liu, H.; Tang, L.; and Yu, P. S. 2008. Identifying the influential bloggers in a community. In *Proceedings of the First ACM International Conference on Web Search and Data Mining (Video available at: [http://videlectures.net/wsdm08\\_agarwal\\_iib/](http://videlectures.net/wsdm08_agarwal_iib/))*.
- Agarwal, N.; Kumar, S.; Liu, H.; and Woodward, M. 2009. Blogtrackers: A tool for sociologists to track and analyze blogosphere. In *Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media (ICWSM)*.
- Barrowclough, A. 2009. Indonesia graft-buster Antasari Azhar arrested for murder. <http://www.timesonline.co.uk/tol/news/world/asia/article6224117.ece>.
- Etling, B.; Kelly, J.; Faris, R.; and Palfrey, J. 2009. Mapping the Arabic blogosphere: Politics, culture, and dissent. Technical report, Berkman Research Center, Harvard Law School.
- Feinberg, J. 2008. Wordle, beautiful word clouds[online]. <http://www.wordle.net>.
- Gillmor, D. 2006. *We the media: Grassroots journalism by the people, for the people*. O'Reilly.
- Granovetter, M. S. 1973. The strength of weak ties. *American Journal of Sociology*.
- Keller, E., and Berry, J. 2003. *One American in ten tells the other nine how to vote, where to eat and, what to buy. They are The Influentials*. The Free Press.
- Kelly, J., and Etling, B. 2008. Mapping Iran's online public: Politics and culture in the Persian blogosphere. Technical report, Berkman Research Center, Harvard Law School.
- Kolari, P.; Finin, T.; and Joshi, A. 2006. SVMs for the blogosphere: Blog identification and splog detection. In *AAAI Spring Symposium on Computational Approaches to Analysing Weblogs*.
- Lim, M. 2002. Cyber-civic space in Indonesia: From panopticon to pandemonium? In *International Development and Planning Review*, volume 4, 383–400.
- Lim, M. 2003. From warnet to netwar: The Internet and resistance identities in Indonesia.
- Lim, M. 2005. The Internet, Islamic radicalism and narratives of conspiracy.
- Lim, M. 2009. *Internationalizing Internet studies: Beyond Anglophone paradigms*. Routledge. chapter Global Muslim Blogosphere: Mosaics of Global-Local Discourses, 178–195.
- Lin, Y.-R.; Sundaram, H.; Chi, Y.; Tatemura, J.; and Tseng, B. L. 2007. Splog detection using self-similarity analysis on blog temporal dynamics. In *AIRWeb '07: Proceedings of the 3rd international workshop on Adversarial information retrieval on the web*, 1–8. New York, NY, USA: ACM Press.
- Mcpherson, M.; Lovin, L. S.; and Cook, J. M. 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27:415–444.
- Sandell, R. 1999. Organizational life aboard the moving bandwagons: A network analysis of dropouts from a swedish temperance organization 1896-1937. *Acta Sociologica* 42(1):3–15.
- Sen, K., and Hill, D. 2005. *Media, Culture and Politics in Indonesia*. Routledge.
- Shirky, C. 2003. Power laws, weblogs and inequality.
- Wilhelm, A. G. 2000. *Democracy in the digital age: Challenges to political life in cyberspace*. Routledge.