# Information Network Analysis to Understand the Evolution of Online Social Networking Sites in the Context of India, Pakistan, and Bangladesh

Shameem Ahmed and Jana Diesner
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign, IL, USA
ahmed9@illinois.edu, jdiesner@illinois.edu

## ABSTRACT

Online Social Networking Sites (SNS) are becoming very popular in India, Pakistan, and Bangladesh; impacting the personal, professional, and social life of millions of people. We report on how SNS are evolving in this region over time and with respect to important events. We also show how different stakeholders involved in governing SNS in these countries are influencing each other. For our research, we use relational text analysis of open source documents. Our results show that both direct and indirect influences among stakeholders are gradually increasing; religious issue and social caste system had a great impact even couple of years back, however their impact is gradually decreasing; governments are imposing new rules and regulations to SNS to have more control over content; and more people is becoming attracted to SNS due to its availability in local languages and accessibility through cell phones.

## Categories and Subject Descriptors

I.2.7 [**Natural Language Processing**]: *Text analysis.*

## General Terms

Measurement.

## Keywords

Online Social Networking Sites (SNS), Co-occurrence analysis

## 1. INTRODUCTION

After becoming successful in western countries, Online Social Networking Sites (SNS) such as Facebook, MySpace, Orkut, Twitter, LinkedIn, etc. are currently spreading their reach to the rest of the world, e.g. Asia, Africa and Latin America. Among others, India, Pakistan, and Bangladesh are countries where SNS are experiencing a tremendous growth in terms of number of users. Of the world population, 22.2% come from these three countries [17]. Despite tensions in several contexts, e.g. politics, among these neighboring countries, their citizens share similar cultural backgrounds, education systems, and social norms. Though a large majority of people in these countries lives below the poverty level and there are many places where basic needs such as sufficient supplies of potable water are not met yet, SNS have been broadly adopted even in those remote and underdeveloped regions. The following quote illustrated this trend

(Business Week, August 2011): *"In a two-room shanty with no running water in northern Mumbai, Darshana Verma makes tea on a small stove. On a bench nearby, her 18-year-old son, Vishal, messages Facebook friends on the keypad of his Nokia smartphone".*

**Table 1: Facebook statistics in three countries**

|  | **India** | **Pakistan** | **Bangladesh** |
|---|---|---|---|
| Facebook users (million) | 45.8 | 6.4 | 2.5 |
| Position in the world in terms of number of Facebook users | 3 | 28 | 55 |
| % of population using Facebook | 3.8% | 3.4% | 1.6% |
| Growth of Facebook use in last six months | 20.5% | 16.7% | 14.2% |

**Table 2: Population and technology use, in millions (% of population in parenthesis)**

|  | **India** | **Pakistan** | **Bangladesh** |
|---|---|---|---|
| Population | 1,189 | 187 | 159 |
| Facebook users | 45.8 (3.8%) | 6.4 (3.4%) | 2.5 (1.6%) |
| Internet users | 121 (10.2%) | 29.1 15.5%) | 5.5 (3.5%) |
| Cell phone users | 919 (77.3%) | 114.6 (61.2%) | 86.6 (54.6%) |

Table 1 shows Facebook adoption statistics for India, Pakistan, and Bangladesh [1]. India holds one of the top three positions worldwide in terms of number of Facebook users even though Facebook penetrates only 3.8% of India's total population. Pakistan and Bangladesh have similar statistics (3.4% and 1.6% respectively). Strong growth rates (20.5%, 16.7%, and 14.2% in the last six months) and the possibility of reaching to the enormous "remaining" population makes these countries lucrative potential market for Facebook and other SNS. Table 2 shows that India, Pakistan, and Bangladesh have a readily available infrastructure [1, 6] for SNS to reach the remaining population. A large share of the population of these countries has access to cell phones (77.3%, 61.2%, and 54.6%). This is relevant as half of all Internet users in India access the Internet via cell phones and 346 million Indian cell phone users are currently subscribed to data package [2]. Similar settings apply to Pakistan [3] and Bangladesh. Since a large portion of Internet traffic comes from cell phones which are available even in the remotest areas of these countries, it is possible and likely that more people will be connected to the Internet including SNS via cell phones in the near future.

The younger generation (under age of 25) is the main users of Facebook in India, Pakistan, and Bangladesh (60%, 65%, and 61%) [1]. More than 50% of the people are under the age of 25 in those countries [2]. This fact further strengthens the potential for SNS adoption in this region.

Our idea is to analyze, over a number of years, progression of SNS and factors that influence its growth and shape in this region. By examining the patterns in the adoption, development, and impact of SNS, we have identified several stakeholders, including SNS, ICT providers, advertisers, government, users, and community, who are playing major roles in this growth. To the best of our knowledge, no prior research is available on the role and influence of these stakeholders. Related to that, there is a lack of research on how SNS are evolving over time and in terms of major SNS-related events in this particular region. We address these gaps by identifying the landscape of SNS stakeholders in these three countries from textual data. With this work, we answer the following questions:

- How do SNS evolve over time in India, Pakistan, and Bangladesh?
- How do SNS evolve with respect to important and critical SNS-related events in these three countries?
- How do different stakeholders influence each other (directly or indirectly)?

Our data source is a collection of news articles treating SNS in these countries, which cover the period. From the text of these articles we first extract both central SNS-related knowledge and central SNS-related agents. Then we use word co-occurrence analysis to build a set of semantic networks that represent the relationships above, over time. With this temporal segmentation, we can analyze how agents, knowledge, knowledge-networks, and their inter-relationships evolve. These semantic maps will inform us about a number of social and conceptual developments coupled to the emergence of SNS in these areas of the world.

## 2. BACKGROUND

Although a large amount of research on multiple aspects of SNS is available, only a few publications discuss SNS in the context of India, Pakistan, and Bangladesh. Prior work on this topic includes empirical studies based on interviews and online surveys - to understand usage pattern, including the amount of time spend on SNS, popularity of SNS, socio-demographic differences in use, etc. [7, 8, 10, 12, 13, 15, 16]. A few empirical studies tried to identify positive impacts of SNS on users, such as women empowerment, educational improvement, and freedom of speech [8, 11, 14]. Gupta et al. performed content and activity analysis on Twitter messages on the Mumbai bomb blast to understand the pattern in micro-blogging during an emergency [9]. However, there is no prior work on the role and influence of stakeholders related to SNS. Also, to the best of our knowledge, no article reported the SNS evolution in these three countries in terms of time and major events. In our research, we are trying to bridge these gaps.

## 3. DATA

We collected reports from major business periodicals[1] which published SNS-related articles in the context of India, Pakistan,

---

and Bangladesh. We were able to retrieve only 62 news articles which we categorized in two ways: according to (a) publication year and (b) important SNS-related events. Since only few articles were found for the first four years, we combined these years (2006-2009). Table 3 summarizes publication statistics.

**Table 3: Statistics of SNS-related articles collected**

| Year | Number of Articles |
|------|--------------------|
| 2006-2009 | 11 |
| 2010 | 17 |
| 2011 | 26 |
| 2012 (up to April) | 8 |
| **Total** | **62** |

## 4. METHOD

### 4.1 Data Preprocessing and Node Identification

We converted collected articles into TXT format using an online tool (www.w3.org/services/html2txt). Next, we identified important words from these articles by using the $tf * idf$ algorithm where

$$tf * idf(t, d, D) = \frac{\# \ of \ occurenece \ of \ term \ t \ in \ document \ d}{\log\left(\frac{Total \ \# \ of \ documents \ \ D}{\# \ of \ documents \ containing \ term \ t}\right)}$$

We have used AutoMap, a text mining tool, to generate $tf * idf$ per word and selected the first 70 highest ranking terms. These terms serve as nodes for the semantic network. To construct such a network, we classified these 70 terms into more general categories (average number of categories per year is 35). For instance, Facebook, MySpace, and Twitter are classified as Social (SNS); India, Pakistan, and Bangladesh are classified as Government, and so on. These general terms are then categorized as either "knowledge" or "agent." We have six agents, namely SNS, ICT providers, advertisers, government, users, and community. All other terms were coded as knowledge.

### 4.2 Network Construction

Co-occurrence analysis is a proximity-based node linkage technique [4]. The link weight typically represents the cumulative frequency of observing a word pair. Co-occurrence approaches differ in the semantic units, e.g. sentence level, paragraph level, and document level. Within that level, the window size $n$ specifies how many words may at most appear between two words. In our experiment, we linked the identified words or nodes via co-occurrence analysis to generate the following networks: Agent by Agent Network, Agent by Knowledge Network, and Knowledge by Knowledge Network. We have used AutoMap for this purpose.

We considered six network metrics to identify the top-ranked agent and knowledge nodes namely Total degree centrality, In-degree centrality, Out-degree centrality, Eigenvector centrality, Closeness centrality, and Betweenness centrality [5]. We used ORA to visualize and analyze the resulting networks. We further verified our findings by analyzing relevant texts reported in our dataset. In the remaining of the paper, we report observations collectively from our network analysis and from textual analysis.

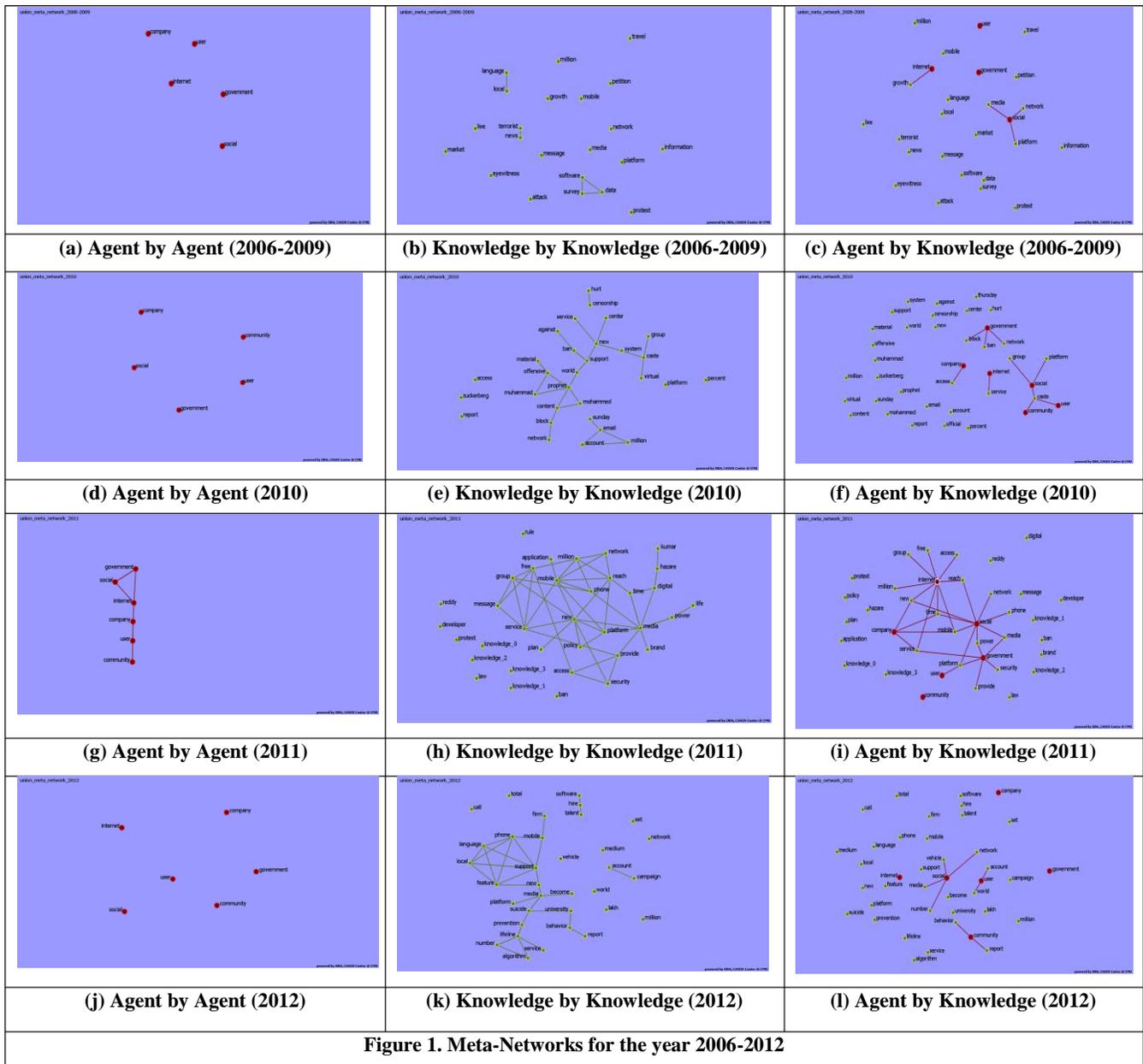| | | |
|---|---|---|
| **(a) Agent by Agent (2006-2009)** | **(b) Knowledge by Knowledge (2006-2009)** | **(c) Agent by Knowledge (2006-2009)** |
| **(d) Agent by Agent (2010)** | **(e) Knowledge by Knowledge (2010)** | **(f) Agent by Knowledge (2010)** |
| **(g) Agent by Agent (2011)** | **(h) Knowledge by Knowledge (2011)** | **(i) Agent by Knowledge (2011)** |
| **(j) Agent by Agent (2012)** | **(k) Knowledge by Knowledge (2012)** | **(l) Agent by Knowledge (2012)** |

**Figure 1. Meta-Networks for the year 2006-2012**

# 5. EXPERIMENTAL RESULTS
## 5.1 SNS Evolve Over Time

We generated one meta-network per time slice (four meta-networks in total) for 2006-2012, where each meta-network has either five or six agents. The number of unique knowledge nodes varied from 23 to 36.

### 5.1.1 Meta-Network Analysis for 2006-2009

Figures 1a, 1b, and 1c show the meta-networks for 2006-2009, respectively. For the social network (Agent by Agent), there is no direct link among agents (Figure 1a). This indicates that the stakeholders did not have any direct impact on each other during this period. Since SNS just started out during this time, agents might still be uncertain about their roles and relationships with others. For the knowledge networks (Knowledge by Knowledge)

where agent and knowledge nodes are both present. This network allows for analyzing whether agents are linked indirectly through knowledge nodes. Although there are edges between agent and knowledge nodes, there are no such indirect social links. Additional observations include:

- Top-ranked knowledge nodes are: survey, software, data, and local.
- Internet growth was enormous and many users started using the Internet through cell phones.
- SNS started working as a platform for social networking and spreading news.
- SNS started to consider offering services in local languages.
- Many surveys were conducted to uncover SNS usage statistics which generated huge amounts of data.
- SNS became popular among the software developers.

### 5.1.2 Meta-Network Analysis for 2010

Figures 1d, 1e, and 1f show the three meta-networks for 2010. Still, there are no direct links among agents (Figure 1d). Figure 1e shows a significantly different Knowledge by Knowledge network where many nodes are directly and indirectly linked to each other. Figure 1f shows how agents are influencing each other through knowledge nodes (indirect links), e.g.: "Internet" connects "Company" via "access", "SNS" connects "Community" via "caste", and "Government" connects "SNS" via "ban". Additional findings include:

- Top-ranked knowledge nodes are: Prophet, Mohammad, caste, and new.
- Tensions arise between SNS and Government with respect to religion-related topic.
- Many SNS groups were formed based on social caste.
- SNS started opening new support centers in India.
- SNS started new products (e.g. Message system) which were graciously accepted by the users of these three countries.

### 5.1.3 Meta-Network Analysis for 2011

Figures 1g, 1h, and 1i show the meta-networks for 2011. Here are some findings about these networks:

- Agents started being directly linked to each other; indicating direct influence. For instance, SNS is connected to Government. This might be explained by the fact that SNS employed lobbyist to influence the Government.
- Top-ranked agent nodes are: SNS, Government, and Internet.
- Many indirect links exist among agents and knowledge nodes (e.g. "Internet" with "Company" via "access", "SNS" with "Community" via "caste").
- Top-ranked knowledge nodes are: media, mobile, and new.
- More people started using cell phones for accessing SNS.
- SNS started encouraging software developers for designing new applications.
- SNS started providing new services to the users.
- Government issued new regulations to control SNS content to ensure countries' security.
- SNS started playing an important role in social movement. For instance, people started supporting civil activist Ann Hazare's Anti-Corruption Movement via SNS.
- Users started using SNS as a new platform for announcements and as a new communication medium for informing others about their personal and professional life.

### 5.1.4 Meta-Network Analysis for 2012

Figures 1j, 1k, and 1l show the meta-networks for 2012 (up to April). Findings for this period are:

- There is no direct link among agents (Figure 1j). This might be due to two reasons: First, agents do not directly influence other agents anymore. However, the pattern from previous years makes this unlikely. Second, we only have four months of data, which might be insufficient to show such links. By the end of this year, when we will have more data available, we might see more of these links.
- Several indirect links exist among agents such as: "Internet" connects to "SNS" via "mobile" and "SNS" connects to "User" via "world".
- Top-ranked knowledge nodes are: local, support, language, feature, and phone.
- SNS finally launched their services in local languages.
- Rise of cell phone network increased number of SNS users.

### 5.1.5 Comparative Summary of Results

Table 4 shows a comparative summary of how SNS evolved during the study period in India, Pakistan, and Bangladesh. The results show that the influence (both direct and indirect) among agents gradually increased over time. Religion-related topics started creating tension among user, SNS, and government in 2010; later SNS and government became more cautious about this issue, which might explain why we did not observe such tension later on. Social caste started playing an important role, especially for group formation in SNS, mainly in 2010. However, later, users realized the importance of going beyond the social caste, at least in SNS. Hence, this trend declined over time. Government started imposing new rules and regulations to SNS, and this trend is still increasing. To attract more local people, from the very beginning, SNS started working to provide their services in local languages. From 2011 on, people started finding a way for engaging with social movements via SNS. Being inspired by the success of SNS in the middle-east, they started SNS-campaigns to show their support for social movements. Another important trend is that people are using their cell phones a lot to access SNS.

**Table 4: A Snapshot of SNS evolution over time**

|  | 2006-2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|
| Direct connection among agents | No | No | Yes | No |
| Indirect connection among agents (via knowledge) | No | Yes | Yes | Yes |
| Religious Issue | No | Yes | No | No |
| Impact of social caste on SNS | No | Yes | No | No |
| Government impose new regulations to SNS | No | Yes | More | No |
| SNS in local language | Yes | No | Yes | Yes |
| Social Movement via SNS | No | No | Yes | No |
| SNS in Cell Phones | Yes | Yes | Yes | Yes |

## 5.2 SNS Evolve Over Important Events

We have identified the following three major events since 2006:

- SNS started their services in local languages (May - June 2009)
- Government blocked SNS for publishing a blasphemous cartoon of Prophet Muhammad (May - June 2010)
- SNS started setting up local servers (April 2012)

### 5.2.1 SNS in local languages

Figures 2a, 2b, and 2c show the meta-networks generated for the event of SNS in local languages. Here are some observations:

- Top-ranked knowledge nodes are: language, transliteration, Google, official, aircel, and mobile.
- To attract more local people, SNS started planning to launch their services in major local languages.
- Use of local language keyboards is limited in these regions. Hence, SNS started working on transliteration so that the English keyboard can be used for writing in local languages.
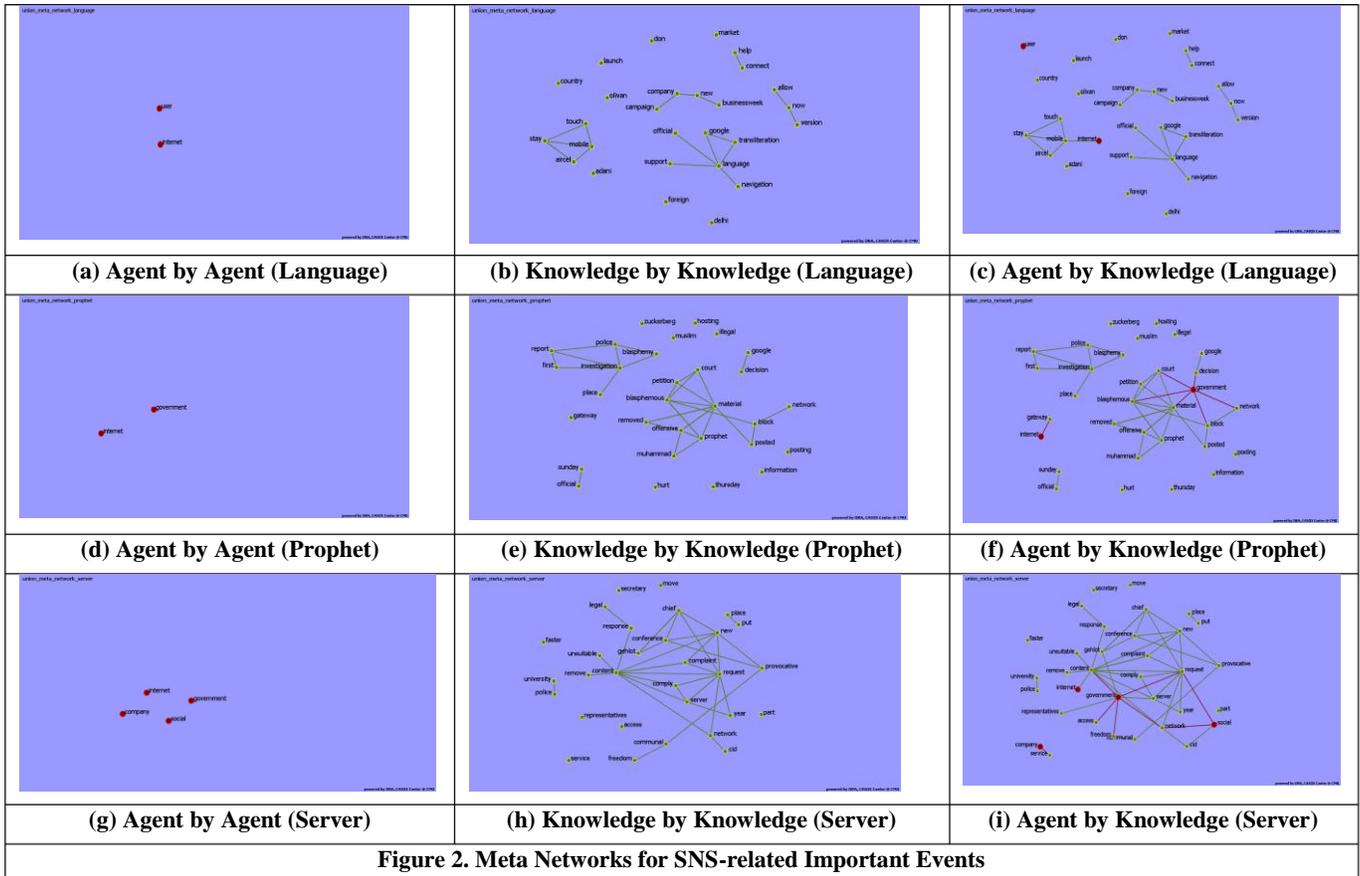
| (a) Agent by Agent (Language) | (b) Knowledge by Knowledge (Language) | (c) Agent by Knowledge (Language) |
| (d) Agent by Agent (Prophet) | (e) Knowledge by Knowledge (Prophet) | (f) Agent by Knowledge (Prophet) |
| (g) Agent by Agent (Server) | (h) Knowledge by Knowledge (Server) | (i) Agent by Knowledge (Server) |

Figure 2. Meta Networks for SNS-related Important Events



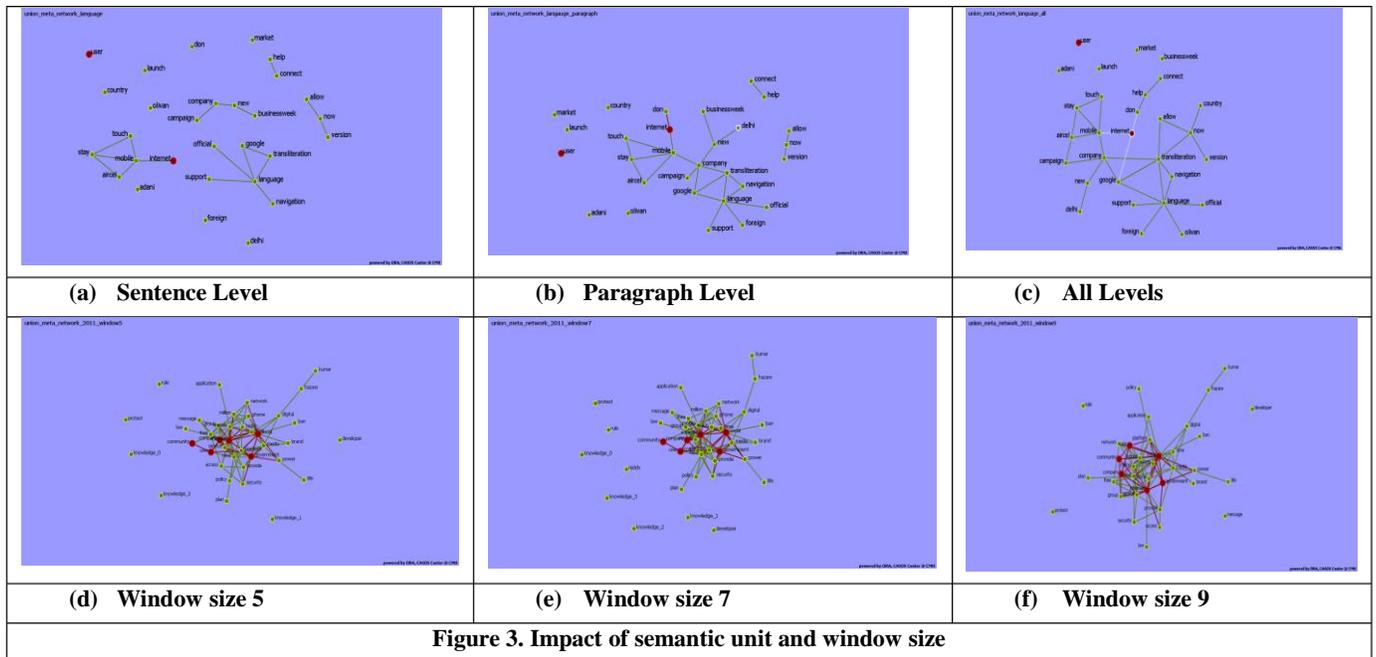| (a) Sentence Level | (b) Paragraph Level | (c) All Levels |
| (d) Window size 5 | (e) Window size 7 | (f) Window size 9 |

Figure 3. Impact of semantic unit and window size

### 5.2.2 Cartoon of Prophet Muhammad

Figures 2d, 2e, and 2f show the meta-networks generated from the event when a "blasphemous" cartoon about Prophet Muhammad was published in Facebook. These are our observations:

- Top-ranked knowledge nodes are: blasphemous, prophet, material, offensive, court, block, removed.
- In Facebook, a contest was organized to draw cartoons of Prophet Muhammad. Considering that such a controversial religious issue might upset millions of people, Pakistan's and Bangladesh's governments banned Facebook for two weeks. Mark Zuckerberg faced police investigation for blasphemy in Pakistan.

### 5.2.3 Local SNS Server

Figure 2g, 2h, and 2i show the meta-networks generated from the event related to local server establishment by SNS. Our observations are:

- Top-ranked knowledge terms are: request, content, new, chief, comply, complaint, and legal.
- Governments requested SNS to establish local server in this region so that the government can control "objectionable content" posted on SNS and if needed, those contents can be removed within a specific time frame.

## 6. DISCUSSION

We have used sentence level co-occurrence analysis for our study. To test how our results are influenced by our coding choices, we considered three different semantic units, namely sentence level, paragraph level, and document level (Figures 3a, 3b, and 3c show the Agent by Knowledge network). In all cases, the chosen window size was seven. The networks become denser when the semantic unit size is increased (from sentence to paragraph to document level). However, we did not find significant differences among top-ranked knowledge nodes. In fact, the top three recurring knowledge nodes are: language, transliteration, Google (sentence level), language, transliteration, company (paragraph level), and language, Google, Transliteration (document level). This finding shows that key entities are fairly robust towards changes in this particular coding choice.

Figures 3d, 3e, and 3f show the Agent by Knowledge network for the articles of year 2011; showing how the information network changes along with the window size. Perhaps, there are some basic differences among these networks; however, further investigation is needed to find a precise answer to this question. The top-ranked knowledge nodes are: media, mobile, new, access, group (window size: 5), media, mobile, new, access, group (window size: 7), and million, mobile, media, reach, time (windows size: 9). Again, the top-ranked nodes are fairly robust to changes in window size.

Our results suggest that network visualizations and analysis are not sufficient to correctly understand knowledge networks, but are suited for providing a first-pass understanding of the nature of the network. Extensive prior knowledge of the subject domain is crucial for interpreting findings within the applicable context.

## 7. CONCLUSION

Social Networking Sites (SNS) such as Facebook, MySpace, Orkut, Twitter, LinkedIn, etc. are being introduced and adopted across the world. Due to the large population of India, Pakistan, and Bangladesh, which together comprise 22.2% of the world's population, SNS have been very interested in branching out to these three countries. India already has the third largest Facebook user population.

Although SNS became popular only very recently in these three countries, SNS already started impacting millions of people in their personal, professional, and social lives. Besides users, some other stakeholders (government, SNS, community, ICT provider, and advertiser) are taking active part in shaping the SNS network in these regions. In this paper, we take a first step in analyzing the relationship among important stakeholders, tracking knowledge circulating in the SNS arenas, and identifying structure of knowledge networks - aiming to examine patterns in the adoption, development, and impact of SNS. Our findings indicate that both direct and indirect influences among stakeholders are gradually increasing; governments are imposing new rules and regulations to have more control over SNS content; and more people is becoming attracted to SNS due to its availability in local languages and accessibility through cell phones.

## 8. REFERENCES

[1] Social Bakers, http://www.socialbakers.com/

[2] BBC (Jan 3, 2012). Is 2012 the year for India's Internet? http://www.bbc.co.uk/news/business-16354076

[3] International Herald Tribune (Oct 28, 2011): Internet users in Pakistan cross 20 million mark http://goo.gl/6ZrFB

[4] Danowski, J. A. (1993). Network Analysis of Message Content. Progress in Communication Sciences, 12, 198-221.

[5] Wasserman, S., & Faust, K. (1994). Social Network Analysis: Methods and Applications: Cambridge University Press.

[6] Internet Usage in Asia: http://goo.gl/SlT5L

[7] Agarwal et al. An Exploratory Study of Indian University Students' Use of Social Networking Web Sites: Implications for the Workplace. *Business Communication Quarterly 72*, 1 (2009).

[8] D'Silva et al. Influence of Social Media Marketing on Brand Choice Behavior among Youth in India: An Empirical Study. *Intl Conference on Technology and Business Management*, (2011).

[9] Gupta, A. and Kumaraguru, P. *Twitter Explodes with Activity in Mumbai Blasts! A Lifeline or an Unmonitored Daemon in the Lurking?* IIITD, 2011.

[10] Haneefa, M. and E, S. Perception and Use of Social Networking Sites by the Students of Calicut University. *Journal of Library & Information Technology*, (2011).

[11] Gangadharbatla, H.S. and Khedekar, D.H. Social Media as Tools of Personal Empowerment for Indian Women. *Asian Media Info & Communication Center Annual Conference*, (2011).

[12] Marshall et al. Social Networking Websites in India and the United States: A Cross-National Comparison of Online Privacy and Communication. *Issues in Information System 9*, 2 (2008).

[13] Mishra, S. Participation of youth in social networking sites in India. *Intl. Journal of Business Innovation and Research*, (2010).

[14 Preeti, M. Use of social networking in a linguistically and culturally rich India. *The International Information & Library Review 41*, 3 (2009), 129–136.

[15 Shaheen, M.A. Use of social networks and information seeking behavior of students during political crises in Pakistan: A case study. *Intl. Information & Library Review*, (2008).

[16] Talukder, M. and Yeow, P.H. A comparative study of virtual communities in Bangladesh and the USA. *Journal of Computer Information Systems 47*, 4 (2007), 82–90.

[17] CIA – The World Factbook: http://goo.gl/JIpCx