Understanding Group Interaction in Blogosphere: A Case Study

Nitin Agarwal[†], Huan Liu[†], John J. Salerno[‡], Sanjay Sundarajan[†]

[†]Arizona State University, [‡]Air Force Research Lab/IFEA
[†]699 South Mill Avenue, Tempe AZ 85287
[‡]Air Force Research Lab/IFEA, Rome NY 13441
[†]{Nitin.Agarwal.2, Huan.Liu, Sanjay.Sundararajan}@asu.edu, [‡]John.Salerno@rl.af.mil

Abstract

Social interactions are an essential ingredient of our lives. People convene groups and share views, opinions, thoughts, and perspectives. Similar tendencies for social behavior are observed in the World Wide Web. This inspires us to study and understand social interactions evolving in online communities especially in the blogosphere. In this paper, we study and analyze various interaction patterns in community blogs. This would lead to better understanding of the socio-cultural ties between these communities to foster collaboration, better personalization, predictive modeling, and enable tracking and monitoring. Tapping community interactions via link analysis has its limitations due to exponentially large search space. We propose a model, circumventing the challenges with link analysis based approach, to observe interaction within community blogs via an observed event and community reaction to that by studying the opinion and sentiments of the members towards that event. We present a case study on ethnic community blogs exploiting the proposed model and report our findings and observations. During our study we encountered several challenges with the proposed model. We discuss these issues and present future directions to make the model more robust.

Introduction

In the early 1970's, Stanley Milgram performed a small study to investigate a phenomenon he called 'Familiar Strangers'. Milgram defined a familiar stranger as someone who is observed, repeatedly for a certain time period and without any interaction (Milgram, 1977). Familiar strangers are common throughout our urban existence: for example, commuters that we recognize at the bus stop or the old man who walks his dog in the morning. Analogous to the physical world, in virtual world social interaction has always been an important part of our life. In general, people are inquisitive to know about similar kind of people whom we do not know already. This has led to the development of several social community websites where people interact with their own kind to share thoughts and feelings. These people or familiar strangers do not know each other but still share common interests (Agarwal, Liu, Salerno, & Yu, 2007). Such a concept also exists in the World Wide Web and across online communities. Identifying such familiar stranger communities would be of a great interest in many ways.

Identifying similar yet unknown communities has several potential applications. The identified communities could be considered as recommendations to the original community. This would foster collaboration among these communities. It is difficult to provide personalization to niche Long Tail communities due to the lack of data for catering to each and every community. Aggregating such niche communities would help in accurate personalization. Such aggregation certainly promotes predictive modeling by employing inductive algorithms to learn predictive models from the data and predict trends. Another significant application of identifying similar yet unknown communities involves the study of interaction between these communities. Many researchers are interested in knowing how these groups interact within groups or between groups. Such studies could lead to very interesting details on ethnic behavior to social events.

The emergence of Web 2.0 paradigm along with the social web (Mikroyannidis, 2007) has revolutionized the way people use the Internet. The former consumer of information is now the producer. This has promoted services like blogs, wikis, social friendship networks, media sharing, collaborative tagging or folksonomies, witnessing an overwhelming amount of participation resulting in what is known as participatory journalism. A blog or a blog site is a website that is either maintained by an individual or a group of people to discuss about various topics, displayed in chronological order. A typical blog can combine text, images links to other blogs and web pages. The entries could either be blog post itself or comments related to that post.

According to a recent report published by Technorati[1], there are around 112.8 million blogs in existence today, with 100% increase in the number of blogs every six months. There are approximately 1.6 million posting every

[1] http://www.technorati.com/

day, roughly 18.6 posts per second[2]. Blogging is thus currently a very popular way in which mass Web users express, communicate, collaborate, share and debate about ideas, thoughts, and others. Blogs act as conduits of information that helps the information to flow at amazingly fast pace. This could be a very good medium to convene communities of like-minded people. The main highlight of the blogs is that, they are uniquely characterized compared to normal websites mainly because of their freshness in data. They are updated constantly and use a very common language that might at times be very colloquial, due to the casual environment.

Blogs could be broadly classified into two types: Individual Blogs or Community (group) Blogs. Individual or single-authored blogs could be personal blogs where people write to share their thoughts or few interesting incidents in life for which many people respond with their views and comments on it. An individual blog could also exist for a purpose to disseminate something that is highly important, like announcements, news etc. Individual blogs are more like personal accounts or daily experiences. Examples of individual blogs are: Sifry's Alerts: David Sifry's Musings[3] (Founder & CEO, Technorati), Ratcliffe Blog-Mitch's Open Notebook[4], etc. Community blogs or multi-authored blogs are like a community in the real world where each and every person in the group could write their own views and share their thoughts. This is highly useful in developing the community and to bring in a good medium for interaction among people who are separated by geographical barriers. Examples of community blogs are: Google's Official Blog site[5], Boing Boing: A Directory of Wonderful Things[6], etc. In our work, we focus on community blogs instead of individual blogs because there is a lot of social interaction in the community blogs as compared to the individual blogs.

In the recent years, blogs have played a vital role and caught the attention of researchers due to the availability of information even before it might be published in top news websites. Blogs serve a new dimension to know people where they form groups and interact to discuss similar interests. Due to the humongous increase in the number of blogs and their unique characteristics like freshness of data, searchability has become a challenge. It is always a challenge to measure the similarity of the results with the query to provide the most similar answer to the user. Moreover the user might not be able to express the query exactly. There are several popular blog search engines such as Blogpulse[7], Technorati, and Google Blog Search[8]. Most of these are based on text and link analysis, but the requirements of the people are much more. People are interested in finding similar communities or groups to theirs that are not known to them. Current blog search technologies do not provide such capabilities to the users. In this paper, we propose a group interaction based approach to identify similar yet unknown group/ community to a given group/community. Main contributions of this paper can be summarized as follows:

- Defined and formulated the problem of identifying similar yet disconnected community blogs in the blogosphere.
- Identified the challenges with the problem.
- Proposed an interaction based approach to solve the problem.
- Presented a case study with real world community blogs and reported interesting observations and encouraging results.

The rest of the paper is organized as follows: We define the problem and formulate it mathematically in the next section. We also discuss the challenges with the problem. We outline our approach next to circumvent the challenges with the problem. Next we present a case study using the proposed approach and studying real world blog communities. We report the findings and observations and challenges with the proposed model. We review the current ongoing research in the related work section and conclude with some future directions in the conclusion section.

Problem Definition

There are an overwhelming number of community blogs on the blogosphere. People in groups interact within themselves and also across various groups. These interactions could be tapped using the link analysis. The assumption is if a blog post in a community blog cites another blog post published in a different community then these communities are considered to be known to each other. Two communities are disconnected if there is no prior interaction between them. Moreover, if two communities interact with each other they are more likely to be similar, or talking on similar themes. A naïve link analysis based approach would enable us to identify groups that could be similar. However, such groups might already know each other. Thus using link analysis it is not very difficult to identify groups that are similar and connected, but it is not of great help as these groups already know each other and this is not new information for them. So the dilemma of finding similar yet disconnected groups presents novel challenges. Next we mathematically formulate the problem.

[2] http://www.sifry.com/alerts/archives/000436.html
[3] http://www.sifry.com/alerts/
[4] http://www.ratcliffeblog.com/
[5] http://googleblog.blogspot.com/
[6] http://boingboing.net/
[7] http://www.blogpulse.com/
[8] http://blogsearch.google.com/

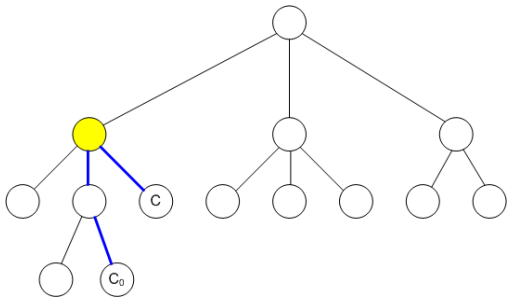

Figure 2: Community blogs C_0 and C_j in community taxonomy

Let C_0 be the original community blog. Our goal is to find an ordered set of community blogs C: $\{C_1, C_2, \ldots, C_T\}$, such that $C_0 \sim C_j \geq C_0 \sim C_k$ where j<k. The binary operator "~" defines the desired similarity metrics between the two community blogs, discussed later in this section. In terms of information retrieval, C_0 is our search query and C_j represents an element of the set C of search results.

Figure 2 shows the placement of community blogs C_0 and C_j in the community taxonomy, where the leaf nodes are the community blogs and the non-leaf nodes correspond to the topic tree. Such blog taxonomies are available through BlogCatalog[9]. We have to maximize the similarity between C_0 and C_j which is given by:

$$\frac{1}{d_{LCA}(C_0, C_j)/D}$$

where, D is the normalizing factor and is computed as 2×max(depth), and d_{LCA} is the path length between the community blogs C_0 and C_j to their least common ancestor in the taxonomy.

In our consideration we have to maximize the disconnectedness between C_0 and C_j. In other words, we prefer acquaintances over friends. To minimize reachability from C_0 to C_j we penalize short path length from C_0 to C_j by associating a cost with the path, defined as, $1/e^{\lambda k}{}_{0,j}$, where, $\lambda>0$ and $k_{0,j}$ (>0)=number of hops from C_0 to C_j. If there is no path from C_0 to C_j then $k \rightarrow \infty$. In case there exists multiple paths from C_0 to C_j then the shortest path is chosen.

Also, we need to avoid community blogs that are very famous or reside in the Short Head, since everyone will know of them already. The underlying assumption for a blog community to be in the Short Head is that it is highly cited by other community blogs, i.e. it has a high degree centrality. In order to avoid communities in the Short Head we penalize the community blogs for large group degree

[9] http://www.blogcatalog.com/

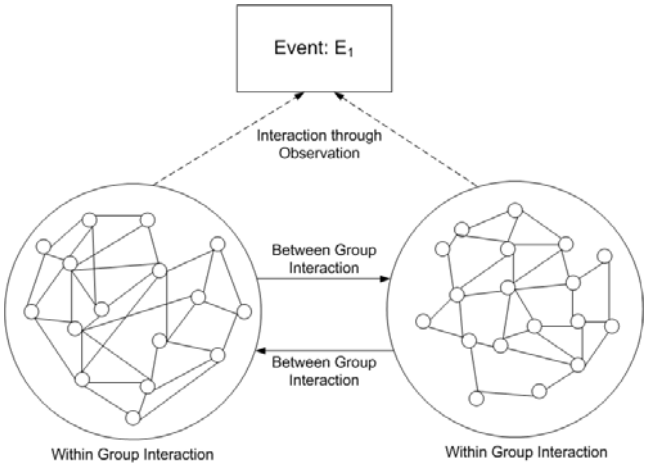

Figure 1: Different types of interactions in community blogs.

centrality. The normalized group degree centrality is given by:

$$\frac{|N_{in}(C_j)| + |N_{out}(C_j)|}{2 \times (|V| - |C_j|)}$$

where, $|N_{in}(C_j)|$ refers to the total number of inlinks for blog community C_j. Similarly, $|N_{out}(C_j)|$ refers to the total number of outlinks for blog community C_j. |V| refers to the total number of vertices or blog posts (nodes) in the blogosphere and $|C_j|$ refers to the total number of blog posts (nodes) in the community C_j.

Putting it all together we get the following expression,

$$\max_{\forall j}\left\{\alpha \frac{1}{d_{LCA}(C_0, C_j)/D} - \beta \frac{1}{e^{\lambda k_{0,j}}} - \gamma \frac{|N_{in}(C_j)| + |N_{out}(C_j)|}{2 \times (|V| - |C_j|)}\right\}$$

The above expression will yield the search results for original blog community C_0, evaluated using the desired similarity metrics.

Further analysis of the above expression results in the following observations: the first two components of the expression that correspond to maximizing the similarity and maximizing disconnectedness involve link analysis. For 'h' hops and average degree 'd', the link traversal is exponential of the order $O(d^h)$, and hence very complex. The third expression that indicates to avoid the Short Head communities is also tough to evaluate as the denominator corresponds to the number of community blogs in the blogosphere which is hard to track.

Since the above formulation is practically infeasible to solve, we look into various heuristics based solutions to our problem. We explore the possibility of interactions based solution. There could be interaction within community or between communities. For the sake of our

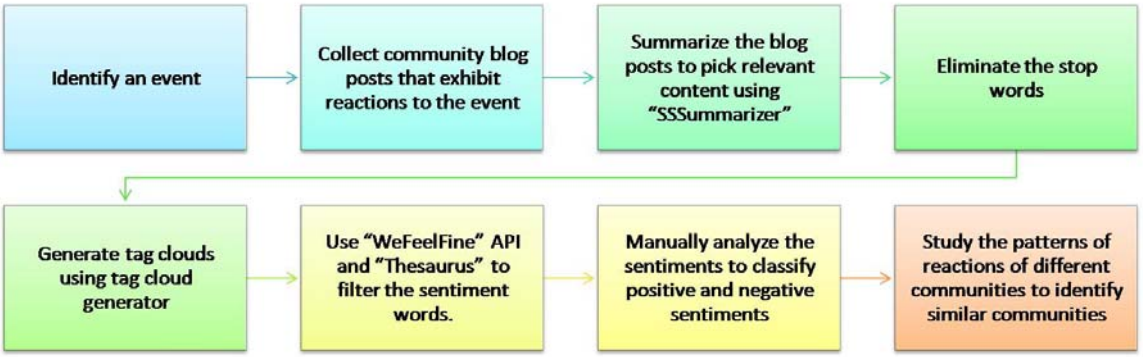

Figure 4: Flowchart of the various components of the proposed approach.

problem of identifying similar yet disconnected communities we ignore the within community interactions and focus on between community interaction. Identifying such interaction is again a very complex problem, since tracking the interactions between communities involve link analysis. Moreover, identifying similar communities based on link analysis defeats the disconnectedness constraint. In the next section we present the proposed approach to identify such community blogs.

Approach

Previously we pointed out within community and between community interaction and the problems associated with it due to link analysis. In this section we study a new type of interaction, known as "Interaction through Observation". It refers to the reaction that communities express on an event. This kind of interaction is different than the interaction that communities have within them or between them. This is also not tapped through link analysis. First major benefit of interaction through observation is the non-utilization of link analysis which is exponential in computation. Second, since blogs have very sparse link structure as compared to the websites so exploiting link analysis to identify similar community blogs gives poor results (Kritikopoulos, Sideri, & Varlamis, 2006). These three types of interactions are illustrated in Figure 1. Based on the reactions that communities have on an event/issue one can identify whether two communities are similar or not. Intuitively, if communities consistently express similar feelings on an issue or an event then they tend be similar. This concept forms the bottom-line of our proposed approach for identifying similar yet disconnected community blogs via interaction through observation. We explain the approach in more detail after a small necessary background information on types of sentiments that is important to understand the approach.

Figure 3: Types of reactions of community blogs to an event.

Background Knowledge

Reactions of a community blog to an event could be viewed as either like, dislike or indifferent. Few examples of these reactions in real world community blogs could be visualized at the following blogs:

- **Indifferent:** (http://www.tuaw.com/2007/12/30/iphone-firmware-1-1-3-breaks-unlocks/)
- **Like:** (http://alexking.org/blog/2006/05/18/macbook-vs-macbook-pro, http://blog.barslund.org/2006/10/02/my-new-macbook-is-great/)
- **Dislike:** (http://apple.slashdot.org/article.pl?sid=06/07/17/2046205,http://www.tuaw.com/2006/07/13/cook-breakfast-with-your-macbook/)

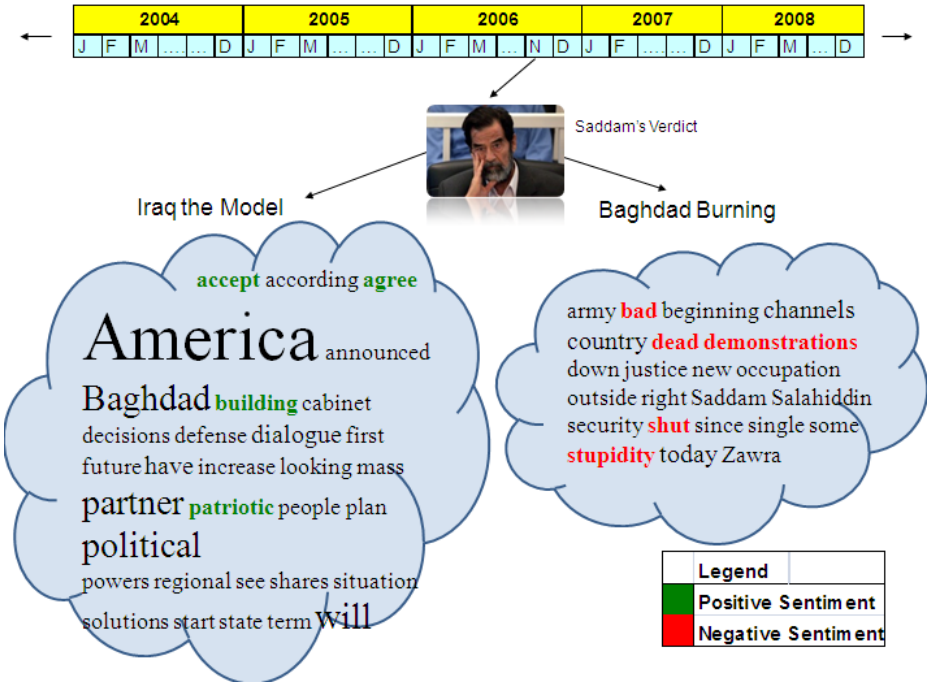

Figure 5: Blog reactions to Saddam Hussein's Verdict.

For the work in this paper we focus on Like and dislike reactions of the community blogs to an event. These reactions are illustrated in the Figure 3. The event here is 'the new Macbook' and the blogs on the left show 'dislike' reaction (because of the excessive heat generated by these laptops) while the blogs on the right show 'like' reaction (because of its great usability and cool features).

Proposed Model

Our approach initially requires identifying an event that incites reactions of sufficient posts to analyze. From the blog posts we summarize the text using the tool Subject Search Summarizer. SSSummarizer[10] (Coombs, Ulicny, Jaenisch, Handley, & Faucheux, 2008) generates and displays summaries as a list of key sentences the product extracts from documents. By presenting and translating sentences it reflect the subject of a given document thus providing the key information. The SSSummarizer allows you to choose the number of sentences to display in the summary. We tried 10, 20 and 30 sentences for different posts and compared the performance of the tool by manual analysis. We were convinced that when the number of sentences is set to 20, the tool provides relevant information that is good enough for further analysis.

The stop words are eliminated from the summarized text and it is fed into a tag cloud generator that spits out the representative words from the summarized text. The tag cloud generator[11] is an online tool from artviper that generates tag clouds for the given text. The tag cloud generator works well when it is used after the SSSummarizer. If used without the tool and raw posts are given directly as input, the identified tags are extremely noisy and irrelevant and appear in larger font increasing its significance. Moreover the number of tags generated is huge and if in case we restrict the number then the quality of results is compromised.

We then compare the identified tags with the list of sentiment key words given by WeFeelFine[12], augmented by Thesaurus[13], and tag the matching words as sentimental words. WeFeelFine provides an API that contains a list of sentiments that have been identified from the blogosphere. Each word has a number associated with it that indicates the number of times it has been identified as sentiment in the blogs. We consider a word to be a "Sentiment word" only if this number is greater than 10. Thesaurus.com has an online searchable collection of words and is grouped together with antonyms. These groups of words are used to augment the sentiment words obtained from WeFeelFine.

[10] http://www.kryltech.com/summarizer.htm

[11] http://www.artviper.net/texttagcloud/

[12] http://www.wefeelfine.org/

[13] http://www.thesaurus.com/

From this point by manually analyzing each of these words we tag them as either a positive sentiment or a negative sentiment. From this collection of positive and negative sentiment words we will be able to decide the type of reaction of the communities for that event. If the reactions of these communities remain consistently similar then these communities are more likely to be similar. Thus the approach based on interaction through observation, enables us to easily break the barriers due to link analysis and thus find two similar communities that are disconnected. A summarized flow chart of the whole approach is illustrated in Figure 4. Next we present a case study on a real world community blog and report the interesting observations. We also present the challenges that we encountered with the proposed model.

A Case Study

As an example consider an event such as Saddam Hussein's Verdict. The Sunnis opposed the event stating it to be ridiculous. At the same time the Shiites felt it was a good decision and they were supporting the event. Such interactions could be found in the blogs but there is no direct way to identify how each group reacted to such events without reading the full post. Though there are several tools available to summarize, identify concepts, themes there is no such tool to find this directly. By identifying sentiments from these blog posts, we can observe their feelings and reactions.

Identifying similar groups that are disconnected through observing an event involves the following steps.

1) We obtain the posts from the three sites – *Iraq The Model*[14], *Baghdad Burning*[15] and *East Kurd*[16] in the month of the event i.e., Saddam Hussein's Verdict.
2) We use the SSSummarizer to obtain the summary of the posts for each site.
3) The Stop words are eliminated from the summarized text of the blog posts obtained from these blogs.
4) The summarized text after stop word elimination is given as input to the Tag Cloud Generator to identify the tags.
5) These tags are then checked with the API provided by WeFeelFine (augmented by the Thesaurus) and the matching words are tagged as Sentiment words.
6) The words identified as sentiment words are then tagged manually as either positive sentiments or negative sentiments. From these words that have been identified for the three blogs we are able to observe that they have different feelings towards the event.

[14] http://iraqthemodel.blogspot.com/
[15] http:// riverbendblog.blogspot.com/
[16] http://eastkurd.blog.co.uk/

The sentiment words clearly revealed that one website i.e. *Baghdad Burning* opposed the event while the other two i.e., *Iraq The Model* and *East Kurd*, were in favor of the event. We also aligned our findings with the ground truth obtained from the news site and came to a conclusion that *Iraq The Model* aligns well with the Shia, *Baghdad Burning* aligns well with the Sunnis and *East Kurd* aligns well with the Kurds. Figure 5 shows the reactions of *Iraq The Model* and *Baghdad Burning* to Saddam Hussein's verdict. *Iraq The Model* has a very positive reaction to the event as evident by the tags like 'accept', 'agree', 'building', and 'patriotic' highlighted in green color whereas *Baghdad Burning* has a very negative reaction to the event as evident by the tags like 'bad', 'dead', 'demonstration', 'shut', and 'stupidity' highlighted in red color.

We also considered another event that was not very famous as the Saddam's verdict; the series of suicide bombings in Iraq during the month of April 2006. This event did not have as many posts compared to that of the Saddam Hussein's verdict. We considered posts from the same three blog sites to identify their reaction to this event. We identified that all the three sites had posts that indicated a negative sentiment. This clearly indicates that the three blog sites strongly oppose the event. This was also manually verified by reading the blog posts. The blogs revealed their grief in the events and people expressed how much they were affected by these bombings in their place. Based on these findings we can observe that *East Kurd* and *Iraq The Model* are very similar in terms of their reaction to these events. This analysis could be highly useful in case of a future event. The results for both the events are summarized in Table 1.

Table 1: Summary of the results of the reaction of three different blogs to the events.

Event	Iraq The Model	Baghdad Burning	East Kurd
Nov 2006 - Saddam Verdict -death sentence[17]	Accept and support the verdict.	Oppose the verdict. Feels its lynching.	Accept and support the verdict
August 2006 - Series of Suicide bomb explosions[18]	Feel bad for it and oppose mildly	Feel bad and oppose the event strongly	Feels bad and opposes

The above table represents a summary of the results obtained. Green color indicates a positive sentiment and red indicates a negative sentiment towards the event.

[17] http://www.guardian.co.uk/world/2006/nov/05/iraq.michaelhoward1
[18] http://www.historycentral.com/freeIraq/Iraqinfo/august06.html

Challenges with the Model

The current tools available provide a lot of blog posts for a given search query. It is still difficult to affiliate a community into one of the three categories Shia, Sunni or Kurd. This is primarily due to lack of ground truth that hinders in classifying these posts. Manually establishing this truth is a challenge which will enable the categorization of other posts based on this. Moreover detecting the positive or negative (bi-polar feelings) is also a challenge.

The API provided by WeFeelFine restricts the sentimental words and does not contain few colloquial words that people use to express their feelings. For instance the word "for" or "pro" that is often used in the place of "support" was not found in the list provided by WeFeelFine. Moreover it would also be eliminated as a stop word. To start with a solution to this we have refined the word list with the help of Thesaurus that contains a list of sentiment words that have been categorized into several groups. This list can be used to identify the sentiment words and also the group to which they belong to. Though we have identified the sentiments from the WeFeelFine, it is pretty tough to tag them as either positive or negative without human intervention. The sentiment list provided by Thesaurus has groups of sentiment words along with the antonyms. This list can be used as a base to color the word as either a positive or negative sentiment.

Related Work

There has been a very fast growth in the use of blogs since its inception. This has led to several research activities focusing to maximize the utilization of blogs to extract useful information. The emergence of communities in a network can frequently be found in a microscopic level (Kumar, Novak, Raghavan, & Tomkins, 2003). This section contains a detailed description of the ongoing research in community extraction in the blogosphere, group interaction, sentiment analysis and opinion mining.

Community Extraction

There has been a lot of research activities in community extraction (Coffman & Marcus, 2004), (Flake, Lawrence, & Giles, 2000), (Lin, Sundaram, Chi, Tatemura, & Tseng, 2006), (Zhou & Davis, 2006). One of the approaches for community extraction is by mutual awareness (Lin, Sundaram, Chi, Tatemura, & Tseng, 2006). The basic idea is that a blog community structure emerges through individual blogger's behavior, precisely how they read and communicate with other members in the blogosphere. The most important thing in a community blog is that should be aware of each other's presence through interaction. This is termed as mutual awareness of bloggers. Authors in (Lin, Sundaram, Chi, Tatemura, & Tseng, 2006) use a method based on this mutual awareness by analyzing it and then extract ranking-based communities from mutual awareness.

Another approach is to view the extraction of communities as a graph problem (Dourisboure, Geraci, & Pellegrini, 2007), (Brinkmeier, Werner, & Recknagel, 2007). Individuals are represented as nodes and the relationships among them are represented as edges. Now the problem is transformed into grouping the cohesive nodes in terms of relationship defined as a function on the edges.

Group Interaction

Face to Face interaction in small groups has been studied for decades. With the advancement in the computer mediated communication in all fields, it has facilitated the interaction between people (Fahy, 2005). Group interaction on the web could either be within the group or between groups. In either case it is tough to identify the interactions due to the very large number of blog posts in the web and they are all not well connected. The current methods that use link analysis will eventually identify only groups that are similar and connected. This information that is identified is already known. In case we have to identify similar and disconnected groups, novel ways as presented in the paper have to be adopted.

Sentiment Analysis

Sentiment analysis involves classifying text based on its sentiment. Most of them focus on approaches to extract sentiments associated with polarities of positive or negative for a given document whereas few of them focus on specific subjects from a given document. For example, the classification of a movie review (Pang, Lee, & Vaithyanathan, 2002), (Turney, 2002) assumes that all sentiment expressions in the review represent sentiments directly toward that movie, and expressions that violate this assumption (such as a negative comment about an actor even though the movie as a whole is considered to be excellent) confuse the judgment of the classification (Nasukawa & Yi, 2003).

Most of the approaches to sentiment analysis involve high level of Natural Language Processing (NLP) that is very complicated and incurs a very high cost. Moreover as mentioned earlier, blogs use a very common language that is colloquial in nature and hence NLP becomes even tougher. This has raised several issues and complexities in the use of NLP in blogs.

Opinion Mining

There are several works on opinion mining. One can get an opinion on a movie review or on a new product that they are about to purchase. Opinion mining basically classifies a blog post into a positive or a negative response that enables a user to have a one word answer from the entire post. Extracting opinions from customer reviews has been a great research focus. With the increase in the number of

people writing reviews, it becomes very hard for a potential customer to read and come to a decision. Feature based opinion summarization (Hu & Liu, 2004), a technique that identifies features of the products that the customers have expressed their opinions and ranks the features according to the frequency. From this the number of positive and negative features is identified.

Conclusions and Future Work

In this paper, we present a novel problem of identifying similar yet disconnected communities especially in the blogosphere. We formulate the problem and identify the challenges that are commonly associated with tapping the community interaction via link analysis. We propose a model that observes this interaction through an event and the reaction of the communities towards this event. Similarity between the communities is evaluated using the identified sentiments and opinions. We present a case study on real world community blogs and report our findings and issues with the model.

As part of the future work, we plan to identify and investigate a few more community blog sites that are highly active with regard to events in Iraq. Using websites like Sphere[19] we can connect news events with blogs and other media. We will be able to classify a new blog post into one of the three categories by identifying how well they align with the ground truth established through these events. Moreover, we consider augmenting the API provided by WeFeelFine with more words that are colloquially used to express feelings and with respect to a specific domain. Apart from identifying these sentiments provided by the WeFeelFine's API we are also looking on ways to categorize words into other groups such as religion, Law/Order, etc. We also plan to automate the process of identifying positive and negative sentiments. The data thus collected could be utilized to predict the behavior of a group to a forthcoming event based on their reaction to similar events in the past, by comparing the forthcoming event's correlation to the past event.

Acknowledgments

This work is supported in part by AFOSR and AFRL grants.

References

Agarwal, N., Liu, H., Salerno, J., & Yu, P. S. (2007). Searching for "Familiar Strangers" on Blogosphere: Problems and Challenges. *NSF Symposium on Next-Generation Data Mining and Cyber-enabled Discovery and Innovation.* Baltimore.

Brinkmeier, M., Werner, J., & Recknagel, S. (2007). Communities in graphs and hypergraphs. *In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management (CIKM)* (pp. 869-872). New York: ACM Press.

Coffman, T., & Marcus, S. (2004). Dynamic classification of groups through social network analysis and HMMs. *In Proceedings of the IEEE Aerospace Conference.*

Coombs, M., Ulicny, B., Jaenisch, H., Handley, J., & Faucheux, J. (2008). Formal Analytic Modeling of Bridge Blogs as Personal Narrative: A Case Study in Grounding Interpretation. *Social Computing, Behavioral Modeling, and Prediction (SBP).* Phoenix: Springer.

Dourisboure, Y., Geraci, F., & Pellegrini, M. (2007). Extraction and classification of dense communities in the web. *In Proceedings of the 16th international conference on World Wide Web* (pp. 461-470). Banff, Alberta, Canada: ACM Press.

Fahy, P. J. (2005). Online and Face-to-Face Group Interaction Processes Compared Using Bales' Interaction Process Analysis (IPA). *The European Journal of Open and Distance Learning (EURODL)* .

Flake, G. W., Lawrence, S., & Giles, C. L. (2000). Efficient identification of web communities. *In Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining (KDD).*

Hu, M., & Liu, B. (2004). Mining Opinion Features in Customer Reviews. *In Proceedings of Nineteeth National Conference on Artificial Intellgience (AAAI).* San Jose.

Kritikopoulos, A., Sideri, M., & Varlamis, I. (2006). Blogrank: ranking weblogs based on connectivity and similarity features. *In AAA-IDEA '06: Proceedings of the 2nd international workshop on Advanced architectures and algorithms for internet delivery and applications.*, (p. 8).

Kumar, R., Novak, J., Raghavan, P., & Tomkins, A. (2003). On the Bursty Evolution of Blogspace. *In Proceedings of the 12th International Conference on World Wide Web* (pp. 568-576). New York, USA: ACM Press.

Lin, Y.-R., Sundaram, H., Chi, Y., Tatemura, J., & Tseng, B. (2006). Discovery of blog communities based on mutual awareness. *In Proceedings of the 3rd annual workshop on webloging ecosystem: aggreation, analysis and dynamics.*

Mikroyannidis, A. (2007). Toward a Social Semantic Web. *Computer , 40* (11), 113-115.

[19] http://www.sphere.com/

Milgram, S. (1977). *The Familiar Stranger: An Aspect of Urban Anonymity*. Addison-Wesley.

Nasukawa, T., & Yi, J. (2003). Sentiment Analysis:Capturing Favorability Using Natural Language Processing. *In Proceedings of the K-CAP*. Florida.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. *In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (pp. 79-86).

Turney, P. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, (pp. 417-424).

Zhou, Y., & Davis, J. (2006). Community discovery and analysis in blogspace. *In Proceedings of the 15th international conference on World Wide Web* (pp. 1017-1018). New York: ACM Press.