**A MODEL OF INFLUENTIAL BLOGGER IDENTIFICATION BASED ON SOCIAL PROOF, COMMENTS AND EMERGE TOPIC**

**BY**

**AWOTUNDE JOSEPH BAMIDELE**

**12/52HE003**

A Ph.D RESULT SEMINAR ONE (1)

**SUPERVISOR:**

**DR. R.G. JIMOH**

**DEPARTMENT OF COMPUTER SCIENCE, FACULTY OF COMMUNICATION AND INFORMATION SCIENCES, UNIVERSITY OF ILORIN, ILORIN, NIGERIA.**

**(MAY, 2018)**

**CERTIFICATION**

This is to certify that this Ph.D. thesis was written by **AWOTUNDE, Joseph Bamidele (12/52HE003)** in the Department of Computer Science, Faculty of Information and Communication Sciences, University of Ilorin, Ilorin, Nigeria as part of the requirement for the Award of Doctor of Philosophy degree in Computer Science.

Dr. R.G. Jimoh Date

(Supervisor)

Dr. D. R. Aremu Date

(Head of Department)

Dr.(Mrs.) Tinuke. O. Oladele Date

(Postgraduate Coordinator)

**TABLE OF CONTENTS**

Title page i

Table of Contents iii

**CHAPTER ONE: INTRODUCTION**

1.1 Background to the study 4

1.2 Statement of the problem 8

1.3 Aim and Objectives of the Study 8

1.4 Justification of the Research 8

1.5 Scope of the Study 9

1.6 Overview of the Method 9

1.7 Definition of Terms 10

1.8 Summary of the Chapter 11

**CHAPTER TWO: REVIEW OF LITERATURE**

2.1 Introduction 14

2.2 Social Media and its Impact 13

2.3 Blog as an Instance of Social Media 16

2.4 Types of Blog 21

2.5 Modeling the Blogosphere 22

2.6 Review of Related Work 30

2.6.1 Influence Flow Method 31 2.6.2 Productivity and Temporal Approach 35 2.6.3 Productive and Influential Bloggers 38

**CHAPTER THREE: RESEARCH METHODOLOGY**

3.1 Proposed Influence Model 42

3.2 Proposed Influence Factors 44

3.3 Data Collection 53

**CHAPTER FOUR: RESULTS AND DISCUSSION**

4.1 Introduction 54

4.2 Evaluation of the Proposed Model 54

4.3 Experimental Results 54

**CHAPTER FIVE: CONCLUSION AND EXPECTED CONTRIBUTIONS**

5.1 Conclusion 80

5.2 Expected Contributions 82

Reference 83

**CHAPTER ONE**

**INTRODUCTION**

This chapter introduces the concepts of the study, the critical background to the study, statement of the problem, research aim and objectives, scope of the study, significance of the study, definition of operational terms and chapter summary.

**1.1 Background to the Study**

Web 2.0 has contributed greatly to the way people think and interact with each other (Andzulis, Panagopoulos, & Rapp, 2012).The introduction of the social computing especially blog has contributed immeasurably to the ways of lives of an individual or group of people (Baruah, 2012).These are used not only to read a topic of interest but also contributed to the issue discussed (Agarwal & Liu, 2008). The use of blog is becoming very useful in almost all areas of human endeavour because people do read blog post by a blogger and gather information on things that are important to them before taking major decision on such(Agarwal, Liu, Tang, & Yu, 2008).

The bloggers share their knowledge and understanding on a topic for others to comments on them, other blogger or reader share their own contribution or knowledge about the issue been discussed (Veletsianos & Navarrete, 2012). Web 2.0 (blogs) have been the major communication channel for many users over the years (Collin, Rahilly, Richardson, & Third, 2011). Bloggers always share their opinion, thought, ideas, and experience with others bloggers or commenters through blog posts (Groups, 2009).

A useful blog post normally comprises of text, image, video, or even links of other blogs or web pages (Sheedy, 2011). Commenters and other users normally visit and read posts in weblog sites and comment on them based on their understanding of the topic been discussed. After reading a post by a blogger, people who get motivated, concerned or convinced by the post idea will eventually and finally choose their decisions in favor or against the ideas borrowed from the blog posts (Akritidis, Katsaros, & Bozanis, 2011). It has been established in America that one in ten people suggest to others where to eat, what to buy, who to vote etc and these happen on a daily basis (Keller & Berry, 2003).

The experts in social computing call the former as influential, as their influence will eventually affects the decision of individuals or group of people that read posts in making their final decisions. Similarly bloggers who write influential blogs can influence readers and other people, who visit, read or comments on their posts. Experts called these bloggers an influential bloggers (Agarwal, Liu, Tang, & Yu, 2008).

The arrival and use of Web 2.0 has generated a flow of expected intelligence by blogs. Many people are now creating interest in blog writing and involving in Web 2.0 activities and happenings. These have also created enormous extents of mutual wisdom or mechanism of brainpower among many people. The web 2.0 has not only allow people to contribute and edit posts through blogs, twitter, and wikis, but it expand and deepen the existing content by creating and providing tags or labels that make the former information users to become the new makers (Agarwal, Liu, Tang, & Yu, 2012).

Giving the users prospect to contribute their thoughts and ideas or even edit the posts has increased relationship among users of Web 2.0 unlike in Web 1.0 where access to the content was circumscribed and limited to a selected few users. The experts in social media claimed that the system has transformed the way we communicate, share and interact with each other and this has also created platforms and it publicizing networks by bloggers to achieve global reach and influence (Couldry & Dijck, 2015). Gaining access to them can be a powerful form of marketing move and gain, since influential bloggers have a favourite, reliable and large audience (Keller & Berry, 2003). These gives one a complete access to their posts and network. It has become research interests to find influential blog sites in the blogosphere, to understand how some blog sites affect and influence the outside world and within the blogosphere.

Blogosphere continues to grow speedily since people’s special interest in the world of blogosphere (Agarwal et al., 2012). This special interests have brought current research awareness on Blogosphere, which include finding the influential bloggers on the web. Regardless whether a blog is being significant of not, this can still generate and create its influential bloggers in a community of many bloggers. It is of great importance to find the influential blogger on the web, which will help to know whether a blogger have a significant influence on his/her followers or a fellow blogger that always follow or read their posts (Agarwal et al., 2008).

The existing models used for finding and identify influential bloggers lack some important parameter that make up the aspects of the blogosphere (Shola, 2012). In this research, the research attempts to study the existing models for possible areas of improvement by combining and expanding the existing methodologies and ideas of the previous models.

The areas of improvement to existing models are as follows:

To get a list of effective blogs in the blogosphere, implement search using robust search engine such as Google or Technorati. This is done not to design and implement a search engine but to use the existing one to get a list of blogs effective in blogosphere. By going through the list one can understand which blogs have influence with their followers, this will be achieve by creating mathematical model in classification of comments by using Positive, Negative and Neutral to find out commenters consent towards the post. Then go and check out their blog and see what they publish. Is it interesting? Does it have community engagement with many comments? And does it well written?

Researchers can also measure influential bloggers by determine the bloggers social proof. This is obvious by the number of retweets, Facebook and Google+ shares etc. A good rule of browse through is that the more retweets, facebook likes and shares herein, the more influential the blog is.

In identifying influential blogger the emerging topics can be used, given the constant talk on the blogosphere, much insight can be gathered by examining arrangements of what people are blogging about to find emerging areas of conversation. One commonly used approach to capture this perception of burning topics is to identify frequently occurring key-phrases. Such an approach may identify that some phrase is been mentioned in more than one blog site e.g. “Boko Haram” and “Chibok Girl” are words that frequently appear together, and as a phrase they are mentioned many times in bringing our girl blogs.

Blogs have come to be a possible way by which new ideas, knowledge and information spreads speedily on the World Wide Web. Observing and following both influential bloggers on the web and their ideas, thoughts, feelings, sentiments, views and opinions on the blogosphere, can therefore have a significant number of applications in the realm of politics, academic, technology, public relations and product marketing etc.

The social media called blogs can be used by profitable companies to serve as their unofficial spokesman to promote their products and services. The important aspects of Web 2.0 can still be used to affect the voting behavior of an individual, which make the political parties to use it to influence and benefit their vote bank and forge their political agendas (Shola, 2012). This can be replicated in other domains such as academic, technology, arts, public relations etc.

**1.2. Statement of the Problem**

The increased flow of interest in Web 2.0 has attracted great research interests to explore how communities and individuals spread influence among themselves (Akritidis, Katsaros, & Bozanis, 2009). The global growing interest in social network has brought about many models to determine influential bloggers in a community (Agarwal et al., 2012).

The significance of realizing competitive advantage in a blog community by identifying influential blogger has created several research gaps in this context ( Akritidis, Katsaros, & Bozanis, 2011). The present model shows that (Akritidis et al., 2011):

Influential bloggers are not necessarily active,

Active bloggers are not necessarily influential,

The present models still lack some important parameters in measuring and identifying influence flows of bloggers on the web (Akritidis et al., 2011).This is due to modification of the weights associated with the parameters of the present models, one can examine how different parameters impact the influence ranking for different needs, and hybridized the models for better performance.

Present models can still serve as starting point for identifying influential bloggers which can be extended by integrating extra parameters to discover different patterns in a blogosphere. This has become a significant problem as the bloggers can take special roles in the society (Akritidis et al., 2011; Agarwal et al., 2012).

This study aims at proposing enhanced model for identifying influential bloggers.

**1.3 Aim and Objectives of the Study**

The aim of this research is to identify influential bloggers at a blog site regardless of its influence. In achieving this aim, the following specific objectives are formulated to:

1. determine the metrics for evaluating influential bloggers.
2. develop a mathematical model for determining influential bloggers on the web.
3. stimulate the proposed model
4. compare the existing methodologies with the developed model.
5. evaluate the developed model with lasion study and pairwise correlation

**1.4 Justification of the Study**

This study proposes a more inclusive, hybrid and dynamic algorithm to enhance the present models in measuring influential blogger in the blogosphere. The newly proposed algorithm assists to acquire detailed parameters and information from blogosphere which play a key role in identify influential bloggers on the web, thereby giving clearer understanding of required statistics and data for measure this important aspect of the blogosphere.

**1.5 Scope of the Study**

This study covers influential blogger identification based on new introduced parameters. The study uses the Endadget community blog site as a basis for the work, the blog site contains all the parameters needs for purpose of this research. The work scoring each blog post, posted by bloggers based on influential factors and raking them accordingly.

**1.6 Overview of the Method**

It has now become of importance to know and measure the influence pattern of a blog post since bloggers have a direct impact on their readers to influence them in taking important decisions. A common techniques for identifying influential bloggers were studied, inspired from the interesting understandings in the literature; existing algorithms were analyzed to find out required areas of improvement. A thorough investigation of blogosphere were carried out to discover additional influential factors which further improved the influential mining system.

The researcher also measured the novelty of each blog post and utilized the parameter in the new algorithm so that the score of each blog post reflect quality and goodness of blog post. An experiment was carried out which exposed the new findings of the study. The study then compares the new results against the previous model, to further justify the proposed model.

**1.7 Definition of Terms**

**Web:** This is sometimes called World Wide Web (W3) is a system of internet servers that support formatted documents using markup language called Hypertext Markup Language (HTML) that links to other documents, likes texts, graphics, audio, and video.

**Web 2.0:** This are the second generation of the W3 that focused on the ability for people to collaborate and share information online within themselves. It is basically a transition from static HTML website to a more dynamic web based on serving Web applications to users, which is more organized.

**Social Computing:** This is an area of computer science that studied the intersection of social behaviour and computational systems. It is based on producing or restructuring social conventions and social frameworks through the use of software and technology.

**Social network**: This is the process of sharing personal information and profiles online through text messaging, chatting, blogging, or viewing online communities (such as Facebook, MySpace, or Webkinz).

**Blog:** This is a regularly updated online personal journal or diary about an individual. It is a platform to express yourself, share your thoughts and your passions to the world

**Blogger:** A person who writes content for a blog

**Blogging:** The act of writing a post for a blog

**Blogosphere:** The collective community of all blogs is known as the *blogosphere*. Since all blogs are on the internet by definition, they may be seen as interconnected and socially networked. The online community of blogs and bloggers.

**Influential post:** This is a post that have an impact or have the power of change the thinking or behaviour of his/her readers or follower.

**An influential blogger:** Is recognized as such if he/she has written influential posts recently or if the posts have had an impact recently.

**Social proof:** This is the impression that people will imitate or follow the actions of others under the assumption/impression that those actions are reflective of the correct behavior. It is a psychological occurrence where people adopt the action of others in an effort to imitate correct behaviour for a given situation.

**Comments:** This is an essential method for interaction between bloggers and blog readers in the blogs. It is an action where blog readers leave a comment to a blog post or article, reply to readers’ comments by other visitors or the blogger, which is the author of article by answers or replies to visitors’ questions and comments.

**TF-IDF:** It stands for term frequency-inverse document frequency, and the tf-idf weight is a weight often used in information retrieval and text mining. This weight is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the corpus. Variations of the tf-idf weighting scheme are often used by search engines as a central tool in scoring and ranking a document's relevance given a user query.

**1.8 Chapter Summary**

The introduction covers the background to the study, motivation of the study, problem statement, research aim, research objectives, scope of the study, significance of the study, limitation of the study, definition of operational terms and chapter summary. The next chapter gives the literature review of the study. Chapter three presents the methodology adopted for the research while chapter four discusses the results obtained for the research, analysis and evaluation of the model with the existing models. Finally chapter five gives the concluding remark, makes some recommendation and future prospects of the research.

**CHAPTER TWO**

**LITERATURE REVIEW**

**2.1 INTRODUCTION**

Social computing is an aspect of computing through social media. It involves the study and understanding of human complex interactions/relations with social media (Shin, Singh, Cho, & Everett, 2015). These social media like blogs, Facebook, twitter, wikis, etc have created research issues in social computing and online media sharing through computational means (Mohan, Agarwal, & Al-Doski, 2013). Research in social computing builds on participatory Web characterized by rich Web applications, user generated contents, user enriched contents, user developed widgets, and collaborative environment of participatory web and citizen journalism (Agarwal & Liu, 2009). Social media have witnessed a remarkable development in past few years (Andzulis et al., 2012). Social computing is a lively and new field with many research interests including remarkable growth, dynamism, long tail singularity, sparse link structure, lack of ground truth, information quality, and data collection (Agarwal & Liu, 2009).

Social computing as a field is a multi-disciplinary research platform with emphasis on human, cultural, and behavioral characteristics. It brings together experts from various disciplines like: anthropology, cognitive science, computer science, economics, linguistics, mathematics, neuroscience, political science, psychology, sociology, statistics, and theology. Social computing refers to the intersection of social behavior and computational systems (Agarwal & Liu, 2009). It is often defined as modeling complex human interactions that are expressed on a variety of social media (Mohan et al., 2013). Social media, or commonly known as the Social Web, consists of an ant-colony of services including blogs, media sharing, micro blogging, social bookmarking, social news, social friendship networking websites, and wikis (Sungkur & Rungen, 2014).

Social media have transformed the use of original web and has changed World Wide Web into the social web by which users can create their own content (Baruah, 2012). This change remained promising due to social media platforms like forums, wikis, and blogs. Blog is known for is widely used as a form of effective communication to express an opinion about an event, taught, behaviour, product or experience and can reach a large audience (Baruah, 2012), (Mohan et al., 2013), (Alejandro, 2010). Users can influence others to buy a product, have certain political or social views, etc. Therefore, identifying the most influential bloggers has become very significant as this can help us in the fields of commerce, advertisement and product knowledge searching (Shin et al., 2015).

**2.2 Social Media and its Impact**

Social media is the combination of online communications channels dedicated to community-based input, interaction, content-sharing and collaboration (Wise & Shorter, 2014). These are websites and applications that are committed to forums, microblogging, social networking, social bookmarking, social curation, and wikis, which are among the different types of social media (Rogers, Liddle, Chan, Doxey, & Isom, 2007).

Social Media Marketing (SMM) takes advantage of social networking to increase brand exposure and broaden customer reach of a company. The main goal is to usually create content convincing enough for users to share on their social networks.

With the passage of time, social media have gathered large number of users through the use of internet or mobile devices (Valk & Rashid, 2010). Due to decreasing cost of the mobile set, it enable users to use social media apps to share their contents with the social networks. Now Facebook has 1,590,000,000 users, WhatsApp has 1,000,000,000 users, Instagram has 400,000,000 users, Twitter has 320,000,000 users, Skype has 300,000,000 users, Viber has 249,000,000 users worldwide(Chaffey, 2016;Valk & Rashid, 2010). There are millions of contents sharing on daily basis(Valk & Rashid, 2010).

Social media sites have changed the lifestyle of the major population (Sheedy, 2011). Now people who were waiting for television or newspaper for sports or weather forecasts use the social media sites or internet sites at any time to get updated news about sports, weather forecasts, entertainment etc (Sensis, 2015). Now more than 80% of American are using the internet and 60% from them are using the social media sites (Sensis, 2015). An Internet research company, Pew Research Center, claims that "more than half of internet users (52%) use two or more of the social media sites measured (Facebook, Twitter, Instagram, Pinterest) to communicate with their family or friends (Conole, Galley, & Culver, 2011).

Social media have improved the creativity between different age groups especially in school going children (O’Keeffe et al., 2011). Now students can interact with their class mates, peers and experts using social media sites and develop their assignments (O’Keeffe et al., 2011). And along with possessing features for increase communication, creativity, social media sites have greatly distracted the attention of the different age groups from physical activities, face to face communication (Collin et al., 2011). Most of the children invest their time in hovering around different sites of their interest especially social media sites (Couldry & van Dijck, 2015); (Shullich, 2011). Advertising on social media sites strongly influenced the teens and effect on their future buying habits. A lack of social integration with people from different backgrounds is damaging social skills in younger people (Shullich, 2011).

**2.3 Blogs as an Instance of Social Media**

Blogs have been the major communication channel for many internet users for many years. Bloggers are sharing their ideas, opinions, and experience with others through blog posts. A blog post might comprise of text, image, video, and links of other blogs or web pages. Many users visit blog sites to read the posts and comment on them. People who get attracted or convinced by a post idea will eventually choose their decisions in favor of the ideas borrowed from blog posts (Shola, 2012).

Many users maintained a blog and write posts to express their opinion, experience and knowledge about a product, an event and every subject of general or specific interest. More users visit blogs to read these posts and comment on them. This “participatory journalism” of blogs has such an impact upon the masses that Keller and Berry argued that through blogging “one American in tens tells the other nine how to vote, where to eat and what to buy” (Akritidis et al., 2009).

Blogging, in particular, distinguishes itself through both popularity and impact (Kayes, Qian, Skvoretz, & Iamnitchi, 2012). For example, WordPress alone, a free and open source blogging tool, is used by over 14.7% of Alexa Internet’s “top 1 million” websites and as of August 2011 manages 22% of all new websites. Citizen journalism had high impact in major events such as South Asia tsunami, London terrorist bombings, and New Orleans Hurricane Katrina. The blogosphere, the virtual universe of the blogs on the web, provides thus a conducive platform for different aspects of virtual and real life, such as viral marketing, sales prediction, business models, and counter terrorism efforts (Richards, 2016).

A blog (also referred to as a “web log”) is a personal journal published on the World Wide Web consisting of discrete entries (“posts”) typically displayed in reverse chronological order (Lee, Hwang, & Lee, 2006). Blogs are usually the work of a single individual, occasionally of a small group, and are often themed on a focused topic (Nardi, Schiano, Gumbrecht, & Swartz, 2004). Blogging platforms allow the creation of online profiles in which links to other bloggers are specified (Richards, 2016). This blogger to blogger ties specify the blogger’s interest and endorsement of other bloggers, creating a social network through which blog updates are automatically disseminated (Kayes, Zuo, Wang, & Chakareski, 2014).

The influence bloggers have on forming public opinions is significant. First, bloggers influence other bloggers’ opinions: in 2011, 68% of bloggers claimed to be influenced by the blogs they read. Second, they can influence the opinions of the masses: 38% of bloggers talk about brands positively and negatively on their blogs (Gruhl, Guha, Kumar, Novak, & Tomkins, 2005). Studies show that 83% of people prefer consulting family, friends or an expert over traditional advertising before trying a new restaurant, 71% of people do the same before buying a prescription drug or visiting a place (Gruhl et al., 2005).

For the most part, blogs were an obscure part of cyberspace until the 2001 terrorist attacks on the U.S. initiated a spike in the number of blogs and blog readers who discovered these sites as a way to share their grief and get firsthand accounts of the events of the day (Kaye, 2007; Oravec, 2003). Since then weblogs have become powerful voices for and against the media and the government (Kaye, 2007).

In 1999 the 50 or so existing weblogs were just a blip in cyberspace, but by the end of 2004 an estimated 8 million individuals had created a blog or Web-based diary (PEW/Internet, 2005). By early 2005, about 27 percent (32 million) of all Internet users accessed blogs and 12 percent had posted comments or links on these sites ((PEW/Internet, 2005); The State of the News Media, 2005). Additionally, several studies report that blog users are young, well-educated, high-income males, who are Internet veterans (Daeid, 2008). More recent data, however, indicates that usage by women, minorities, and individuals between 30 - 49 years of age is experiencing "greater than average, "growth" (PEW/Internet, 2005).

The advantages of identifying influential bloggers are already evident: influential bloggers are often market-movers (Kayes et al., 2012). Identifying these bloggers can help companies better understand key concerns, identify new trends, and smartly affect the market by targeting influential bloggers with additional information to turn them into unofficial spokespersons (Pal & Kawale, 2011). About 64% of the companies are shifting their focus to blogging (Agarwal et al., 2008); (Pal & Kawale, 2011).

In a physical world, according to (Keller & Berry, 2003), 83% of people prefer consulting family, friends or an expert over traditional advertising before trying a new restaurant, 71% of people do the same before buying a prescription drug or visiting a place, and 61% of people talk to family, friends or an expert before watching a movie (Keller & Berry, 2003). In short, before people buy or make decisions, they talk, and they listen to other’s experience, opinions, and suggestions. The latter affect the former in their decision making, and are aptly termed as the influential (Keller & Berry, 2003).

Influence has always been unabated interest in business and society. As the pervasive presence and ease of use of the Web, an increasing number of people with different backgrounds flock to the Web a virtual world to conduct many previously inconceivable activities from shopping, to making friends, and to publishing (Agarwal et al., 2008). By drawing parallels between physical and virtual communities, among citizens of the blogosphere, experts are intrigued by the questions like whether there exist the influentials in a virtual community (a blog), who they are, and how to find them (Agarwal & Liu, 2009).

Since the bloggers can be connected in a virtual community anywhere anytime, the identification of the influential bloggers can benefit all in developing innovative business opportunities, forging political agendas, discussing social and societal issues, and lead to many interesting applications. For example, the influentials are often market*-*movers. Since they can influence buying decisions of the their fellow readers and other bloggers, identifying them can help companies better understand the key concerns and new trends about their products (Agarwal et al., 2008).

The term web-log, or *blog*, was coined by Jorn Barger in 1997 and refers to a simple webpage consisting of brief paragraphs of opinion, information, personal diary entries, or links, called *posts*, arranged chronologically with the most recent first, in the style of an online journal (Majhi, 2010); (Doctorow, 2002). Most blogs also allow visitors to add a *comment* below a blog entry (Doctorow, 2002).

This posting and commenting process contributes to the nature of blogging (as an exchange of views) in what Yale University law professor, YochaiBenkler, calls a ‘weighted conversation’ between a primary author and a group of secondary comment contributors, who communicate to an unlimited number of readers (Benkler, 2006). It also contributes to blogging's sense of immediacy, since ‘blogs enable individuals to write to their Web pages in journalism time that is hourly, daily, weekly whereas the Web page culture that preceded it tended to be slower moving: less an equivalent of reportage than of the essay’ (Benkler, 2006).

Each post is usually ‘tagged’ with a keyword or two, allowing the subject of the post to be categorized within the system so that when the post becomes old it can be filed into a standard, theme-based menu system. Clicking on a post’s description, or tag (which is displayed below the post), will take you to a list of other posts by the same author on the blogging software’s system that use the same tag.

Linking is also an important aspect of blogging as it deepens the conversational nature of the blogosphere and its sense of immediacy. It also helps to facilitate retrieval and referencing of information on different blogs but some of these are not without inherent problems:

The *permalink* is a permanent URI which is generated by the blogging system and is applied to a particular post. If the item is moved within the database, e.g. for archiving, the permalink stays the same. Crucially, if the post is renamed, or if the content is changed in any way, the permalink will still remain unchanged: i.e. there is no version control, and using a permalink does not guarantee the content of a post.

*Trackback* (or *pingback*) allows a blogger (A) to notify another blogger (B) that they have referenced or commented on one of blogger B’s posts. When blog B receives notification from blog A that a trackback has been created, blog B’s system automatically creates a record of the permalink of the referring post. Trackback only works when it is enabled on both the referring and the referred blogs. Some bloggers deliberately disable trackback as it can be a route in for spammers.

The *blogroll*is a list of links to other blogs that a particular blogger likes or finds useful. It is similar to a blog ‘bookmark’ or ‘favourites’ list.

Blog software also facilitates *syndication,* in which information about the blog entries, for example, the headline, is made available to other software via RSS and, increasingly, Atom. This content is then aggregated into feeds, and a variety of blog aggregators and specialist blog reading tools can make use of these feeds.

The large number of people engaged in blogging has given rise to its own term *blogosphere* to express the sense of a whole ‘world’ of bloggers operating in their own environment (Nardi et al., 2004). As technology has become more sophisticated, bloggers have begun to incorporate multimedia into their blogs and there are now photo-blogs, video blogs (vlogs), and, increasingly, bloggers can upload material directly from their mobile phones (mob-blogging). For more on the reasons why people blog, the style and manner of their blogging and the subject areas that are covered (Nardi et al., 2004).

**2.4 Types of Blog**

The virtual universe that contains all blogs is known as the Blogosphere and accommodates two types of blogs ( a) individual blogs, maintained and updated by one blogger (the blog owner), and (b) community blogs, or multi-authored blogs, where several bloggers may start discussions about a product or event. Since in the former type of blogs, only the owner can start a new line of posts, the present article focuses only on community blogs.

Blogs often serve as on-line diaries, evangelical platforms, or as an informal medium for reporting on events. Most are maintained by an individual (individual blogs or single-author blogs). However, blogging by groups of individuals (community blogs or multi-authored blogs) with similar viewpoints or common purpose is increasing (Agarwal & Liu, 2009).

**Individual vs. community**

Single authored (Individual blog sites)

|  |  |
| --- | --- |
| **Individual Blog Sites** | **Community Blog Sites** |
| Owned and maintained by individual users. | Owned and maintained by a group of like-minded users. |
| More like personal accounts, journals or diaries. | More like discussion forums and discussion boards. |
| No or almost negligible group interaction. | High degree of group discussion and collaboration. |
| No or almost negligible collective wisdom. | Enormous collective wisdom and open source intelligence. |

**2.5 Modeling the Blogosphere**

The most important aspect of blogosphere is to develop an appropriate model. The foremost questions asked by researchers is which model will be best describes the characteristics and structure of the blogosphere (Agarwal & Liu, 2008). Since such a model will help have greater impart in gaining deeper insights into the relationships between bloggers and commenters, the blog posts and comments, viewers/readers, and different blog sites in the blogosphere (Agarwal & Liu, 2008). This will surely help in gaining deeper understanding of the blogosphere and allow us to define various concepts of the blogosphere at the beginning of the research (Akritidis et al., 2009). These can also be of significant help in finding and tackling several other challenges of the blogosphere (Agarwal & Liu, 2008).

A well design model for the blogosphere will be very useful in generating an artificial dataset to modify the parameters to develop and simulate a distinct situation and compare different algorithms and studies. The designed models can be of great important in studying the peculiarities of blogosphere on the web and Such a model will also help in studying peculiarities in the blogosphere and deduce hidden patterns and structures, which could be used to explain certain occurrences like blogs influence, spam blogs, information discussion, topical in the blogosphere, and community discovery, etc.. Modeling the web and blogosphere is often related. Web is always represent by researchers as a web-graph, the webpage will forms a node and hyperlinks between each other as edges. The illustration show direct cyclic graph, which weights can be connected with these edges. This can be used to converts the web into a graphic model is widely exploited. One important used in the domain is the search engine based model of the web to rank web pages which relies on this graph (Kleinberg & M., 1999); (Brin, Page, Brin, & Page, 1998).

Researchers debate the used of web models in modeling the blogosphere, which also seem to be appropriate method for modeling the blogosphere, but there key differences that prevent it reusing in the blogosphere domain. Firstly, the models adopts a dense graph organization due to huge number of interrelating hyperlinks within web pages, which does not hold in the blogosphere, since the interconnection hyperlink structure in the blogosphere is very scarce, as shown in (Kritikopoulos, Sideri, & Varlamis, 2006). Secondly, the important aspect of this web 2.0 which is the level of interaction in terms of comments and replies to a blog post makes the blogosphere different from the web. Lastly, the extremely forceful and “short-lived" nature of the blog posts will not allow it to be simulated using the web models. The web models do not consider this important aspect of dynamicity in the web pages at is always considered in a blogosphere. The assumption is that web pages gather and accumulate links over time. It is impossible to construct a static graph in blogosphere like the one in the web, since in a blog network blog posts are the nodes. These important differences in both web and blogosphere require the need for new model that will be more useful and measure the characteristics of the blogosphere. There are many types of models developed for the web such as random graph(Watts & Strogatz, 1998), preferential attachment graph (Barabási & Albert, 1999), hybrid graph (Pennock, Flake, Lawrence, Glover, & Giles, 2002), and random walk on graph (Hubert, Hubert, Hubert, Mugizi, & Rwebangira, 2006). The proposed random graph fails to show the power law of degree distribution or scale-free graph since it constructs edges between each pair of nodes with some probability (Agarwal & Liu, 2008). For this purpose random graph models will not be suitable for used as a model in the blogosphere. The preferential attachment graph models technique follow the characteristic and phenomenon of the “rich gets richer", in which the likelihood of a new edge to a node to be added is based on its degree (Agarwal & Liu, 2008). The chances of a node to be connected with other node is the increase in the degree of a node (Hubert et al., 2006).

The used of these models show the power of law distribution. The hybrid graph models are combination of both random graphs and preferential attachment models, this is done to give better performance and prediction to the graph model and so as to give a lucky poor" a chance to get rich" (Agarwal & Liu, 2008). The hybrid model that be used in blogosphere with some modifications (Agarwal et al., 2012). To find a solve to the problem of isolated subgraph with strong connectedness, which is irreducibility in nature, a better model random walk on a graph model called a random jump used 0:8 and 9:0 as fixed probability in hybrid the preferential attachment model (Agarwal et al., 2008). Also the random jump models have been used to model the blogosphere with some modifications. But the these models still lack some important aspect that be used to explain the blogosphere better and precisely (Agarwal & Liu, 2008). The above mentions shortcoming of these models has motivated researchers to come up with models are better and specific for the blogosphere (Huang, Shen, Lin, & Chang, 2007). (Leskovec, Mcglohon, Faloutsos, Glance, & Hurst, 2006) looked at how blog posts are linked, and what is the link density by studied how often people create blog post using the temporal patterns of the blogosphere, burstiness and popularity. The results show that these occurrences follow the power law distributions. They developed a cascade model from the epidemiology similar to SIS (susceptible-infected-susceptible) model based on their findings (Leskovec et al., 2006).

The randomly picked of blog can affect its unaffected immediate neighbors blog significantly, but repeating the same process until no node remains uninfected. This will gives a blog network at the end. With the assumption that blogrolls have links that related and similar to blog post (Kumar et al., 1999); (Kumar, Novak, Raghavan, & Tomkins, 2003)use the blogrolls given on a blog post to create a network of connected posts (Agarwal & Liu, 2008). Many research has been carryout that uses and posits a known network structure of the blogosphere to model the problem domain. The mathematical models are very problem specific domains and influence in Blogs and Propagation (Agarwal & Liu, 2008).

The communities as been known for change and evolve over time, this also made the bellwethers or leaders of the communities who possess the power to influence the topical. The word-of-month that is the new marketing style has been found to be more effective than that of the traditional advertising in physical communities. These have been proved from the studies of Keller and Berry, which says that before people buy, talk or watch, they listen to others people or experts (Keller &Berry, 2003). Experts have been see to have an influence, which can affect decisions of people, these are called the influentials. A blog posts by bloggers that affect the decision and actions of other members’ and opinions are termed Influential bloggers. These bloggers have enlarge respect in the community in recent and over time. People have to listen to what the influentials say before making vital decisions. The models for the identification of these influential bloggers (Agarwal et al., 2008) could lead to several motivating applications.

This can be a potential market-movers since influentials influence people choices. They affect buying decisions of majority, the companies can make used of them and promote them as hidden brand ambassadors for their products. Blogs tend to cloud numerous vibrant argument on different issues which include services, new products, marketing strategies and their comparative studies, being such a highly interactive medium. Sometimes this discussion acts as word-of-month by advertising several products and services. Many advertising companies, which are almost 64% have acknowledged this fact and are now moving towards blog advertising and identifying these influentials (Efimova & Hendrick, 2005).

These influentials could control opinions and decisions of people in political campaigns, elections and reactions to government programmes or policies(Farrell & Drezner, 2008). Since they know many people and gathers large amount of information, this makes the influentials to stand out as a sources of advice, insight and knowledgeable(Farrell & Drezner, 2008). From a research by Farrell and Drezner, About, 84% of the influentials in a physical communities are attracted in politics and are consult and sought out by others for their perceptions on politics and government, 55% on a regular basis, which shows that some influentials sought out in both politics and on regular issues (Farrell & Drezner, 2007). They also help in customer services, support and troubleshooting. These contributed to the interest of many companies these days to host their own customer blogs, where people could laid complaint or discuss issues related to a product (Agarwal & Liu, 2008). The influentials normally discussed on these blogs to analyzed and troubleshoot the problems consumers are having and sometimes prospered solutions, which could be trusted because of the sense of authority and expertized these influentials possess. These influentials also often offer suggestions to improved their products (Agarwal & Liu, 2008).

These priceless comments could be indeed supportive for companies and customers. Companies can focus on the influentials' blog posts instead of going through each member's blog posts or comments (Agarwal & Liu, 2008). For example, the blogs post of 500 people who write about Macromedia's technology were categorizes, searches and aggregates by Macromedia (Agarwal & Liu, 2008). A 100% increase in the size of the blogosphere every six months were the recent numbers from Technorati (Agarwal & Liu, 2008). These have shown growth of over 60 times during the past three years, since it is observed that approximately two new blog posts appear every second on the web. New blog posts that generated these blazing fast rate on the web, has make it to be a difficult and challenging to keep track of what is going on in the blogosphere (Agarwal & Liu, 2009). Many blog readers/subscribers just want to know the most perceptive and convincing stories before probing into the topics of discussions.

The purpose of blog posts from influential bloggers is to exactly serve the drive by standing out as representative articles of a blog site. The influentials can be the platforms to showcases of a group on the blogosphere. These stimulating claims have attracted a flow of research in identifying influential bloggers as well as influential blog sites. Many researchers have try to create models that can be used to find influential blog sites in the blogosphere and study their influence on the external world and within the blogosphere (Gill, 2004). The problem of ranking blog sites or identify influential bloggers differs from that of finding authoritative web pages, therefore models used in finding authoritative web pages cannot be used in the former. Kritikopoulos et al., pointed out that blog sites are very sparsely linked in the blogosphere and it is not suitable to rank blog sites (Kritikopoulos et al., 2006) using Web ranking algorithms like PageRank (Page, Brin, Motwani, & Winograd, 1998) and HITS (Kritikopoulos et al., 2006). The Random Surfer model of webpage ranking algorithms (Page et al., 1998)cannot be used because it will not work well for sparsely linked structures. The chronological aspect of blog is most significant in the blogosphere (Agarwal & Liu, 2008). While a webpage may attain specialist over time “its adjacency matrix gets denser”, a blog post or a blogger's influence weakens and diminishes over time. Consequently, the graph which considered as “adjacency matrix of blogs” will get scarcer as thousands of new sparsely-linked blog posts shown up and appear every day (Agarwal & Liu, 2008).

In a research conducted by (Kritikopoulos et al., 2006), they add inherent links to increase the concentration of link information based on topics. It is possible for two blogs to be talking about the same topic, by this concentration an edge can be added between these two blogs based on the topic similarity or information increases. The building links based on the topic model still remains an area of research and interest in the blogosphere (Agarwal & Liu, 2008). (Adar, Zhang, Adamic, & Lukose, 2004) adopted and consider the implicit link structure of blog posts, in their model called iRank algorithm predict whether or not two blogs should be linked by building a classifier (Agarwal & Liu, 2008). The main purpose of this model is to find out how one piece of information is propagated, which is the path of infection. The model *iRank* attempts to find the blogs which recruits the waves (Agarwal & Liu, 2008). It is worth noted that an originat or of the topic might not be an influential as they might affect only limited blogs (Agarwal & Liu, 2008). Those people that play a key role in the information waves are called influentials. Gruhl et al., model information diffusion of different topics in the blogosphere between different blog sites, the algorithm was created from theory of infectious diseases. (D. Gruhl, Liben-Nowell, Guha, & Tomkins, 2004)study.

Another model called general cascade model was adopted by (Goldenberg, Libai, & Muller, 2001). The model was originating from independent cascade model and was called the general cascade model by relaxing the individuality theory (Agarwal & Liu, 2008). Their model tried to associate `read' probability and `copy' probability with each edge of the blog graph representing the trend of a blog to be read and copied, respectively (Agarwal & Liu, 2008).

Furthermore, the model parameterize the stickiness of a topic which is similar to the virulence of a disease. It motivated difficult aspects in relation to viral marketing (Richardson & Domingos, 2002; Kempe, Kleinberg, & Tardos, 2003) used the maximize total influence among the nodes (blog sites) by selecting a fixed number of nodes in the network. In selecting the most influential node in each iteration after removing the selected nodes, a greedy approach can be adopted in exploit this. This method overtakes PageRank, HITS and ranking by number of records, and is robust in clarifying splogs (spam blogs) (Java, Kolari, Finin, & Oates, 2006).

Finding influential blog sites is different from identifying influential bloggers, model used in one cannot be applied in the other. Given the nature of the blogosphere, influential blog sites are few (Agarwal & Liu, 2008). Where abundant new marketing, new business, and development opportunities can be explored, a large number of non-influential sites belong to the long tail (Agarwal, Liu, Salerno, & Yu, 2007).

(Agarwal & Liu, 2008) created a model that studied and identify the influence of a blogger on a community blog site irrespective of the site being influential or not. They used inherent link structure, which include inlinks and outlinks as edges and treating different blogger as nodes using blog site as a graph. The influence flow of different bloggers is observed recursively using the link structure. The models also implored other blog post parameters statistics such as blog post quality and comments' information to achieve better results. The model used different algorithms and weights to normalize the impact of different parameters. These models, weights and algorithms can be modified to obtain different types of influential bloggers. Since influential bloggers are not automatically active bloggers at a blog site (Agarwal & Liu, 2008).

Some blog websites ranked and list top bloggers or top blog posts using time frame (e.g., daily, weekly or monthly). Choosing these influentials blogger and those top lists are usually based on some trace information and parameters like how many blog posts a blogger posted or number of comments a blog post attracted (Kathy 2004). It is becoming difficult, if at all possible, to physically track the growth and happening in the blogosphere, especially with the speedy growth of the blogosphere (Akritidis et al., 2009).Now that many bloggers actively participate in discussion, at many blog sites, receiving information, examining and looking for answers, and voicing their complaints and needs (Agarwal & Liu, 2009).

**2.6 Review of Related Work**

In this section the research will discuss three different related solutions proposed to solve the problem of identifying influential bloggers.

First, Influence Flow Method in which authors Agarwal, Liu, Tang & Yu have utilized Inlinks, Comments, Outlinks, and the length of the blog post as influential factors to identify influential bloggers in a community. The parameters used in the method has set up ground truth for the problem. Second, Akritidis, Katsaros, & Bozanis, have introduced two metrics called MEIBI (Metric for Evaluating and Identifying Blogger’s Influence) and MEIBIX an extended version of MEIBI indexes to evaluate influence by considering temporal aspects and productivity along with inlinks and comments. They have ignored parameters such as the length of the post and outlinks of posts in the second model.

Third, the same authors Akritidis, Katsaros, & Bozanis, have modified the metrics MEIBI & MEIBIX indexes to BP & BI indexs to differentiate between productive and influential bloggers. The following subsections elaborate various aspects of the three models in detail.

**2.6.1 Influence Flow Method**

Agarwal, Liu, Tang & Yu (2008) have first introduced the problem of identifying influential bloggers to the literature. The authors proposed Influence flow method to find out the influential bloggers in a web community. According to the authors in the real world an influential person is one s/he is well recognized by others, often does good work that is novel, well received and creates some activity in society. The authors considered these social parameters such as recognition, activity generation, novelty, and eloquence as influential factors for scoring each blog post of a blogger. But how can we relate these social gestures to blogosphere? Each blog post in blogosphere is associated with parameters like inlinks, outlinks, comments, length of the post etc. the thesis can use these parameters to quantify the influence. The blog post parameters can be related to social gestures as following:

**Recognition -** An influential blog post is recognized by many. This can be equated to the case that an influential post *p* is referenced in many other posts, or its number of inlinks (*ι*) is large. The influence of those posts that refer to *p* can have different impact: the more influential the referring posts are, the more influential the referred post becomes.

**Activity Generation -** A blog post’s capability of generating activity can be indirectly measured by how many comments it receives the amount of discussion it initiates. In other words, few or no comment suggests little interest of fellow bloggers, thus non-influential. Hence, a large number of comments (*γ*) indicate that the post *affects* many such that they care to write comments, and therefore, the post can be influential.

**Novelty -** Novel ideas exert more influence as suggested by (Keller & Berry, 2003). Hence, the number of outlinks is an indicator of a post’s novelty. A large number of outlinks (*θ*) may suggest that a post refers to many other blog posts or articles, indicating that it is less likely to be novel. The number of outlinks is negatively correlated with the number of comments which means more outlinks reduces people’s attention.

**Eloquence -** An influential is often eloquent (Keller & Berry, 2003). This property is most difficult to approximate using some statistics. Given the informal nature of the blogosphere, there is no incentive for a blogger to write a lengthy piece that bores the readers. Hence, a long post often suggests some necessity of doing so. Therefore, it use the length of a post (*λ*) as a heuristic measure for checking if a post is influential or not. The blog post length is positively correlated with number of comments which means longer blog posts attract people’s attention.

The parameters used in influence flow method can be summary as:

**Recognition:** How well people have received the idea? This can be measured by the inlinks received by the blog post.

**Activity Generation:** The participatory activity generated i.e. the comments received by the blog posts. More the comments more the activity generated.

**Novelty:** Conveys a new idea shared in the blog post. Can be measured through the outlinks used in the post. The more the outlinks less is the novelty.

**Eloquence:** The goodness of the blog post can be measured by the length of the post.

Finally the Influence Score for a blog is defined as a function of social gestures. These parameters help the model to trace the intuitive influence flow among the blogs. The influence flow of blog post *p* is calculated as

(2.1)

Where *Win*, and *Wout*are the weights that can be used to adjust the contribution of incoming and outgoing influence respectively. And *I(pm), I(pn)* are the influence flow based on inlinks and outlinks, respectively.

For a blog post *p*, that has iinlinks and *θ* outlinks, the influence score can be calculated by the following equation:

(2.2)

Where *w(λ)* is the weight function chose based on the length of a blog post, Pγ denotes the number of comments received by *p, Wcom* is weight parameter used to normalize the contribution of the number of comments.

The influence score is calculated for each blog post P that belongs to blogger B. The maximum of all influence scores of Blogger B is represented as the influenceIndex of blogger B.

InfluenceIndex (B) = MAX(I(Pi)) (2.3)

Where 1<= i<= N

Where N is the total number of posts written by blogger B. The bloggers who have achieved top influenceIndex values are the most influential ones of total bloggers.

**2.6.2 Productivity and Temporal Aspects**

The influential flow method was able to set some ground truth for influential bloggers problem but model has couple of drawbacks with it. Isolating a single post to identify influential bloggers made this approach over simplistic and it overlooks productivity of the bloggers. The output of the model highly dependent on the user defined weights.

Most importantly, this model ignores temporal aspects of blogs which is very crucial for Blogosphere:

Time: A blog that has influenced most today might not be influential after two months. i.e. proposed model should be able to decay the influence score with time.

Considering the above issues Akritidis, Katsaros, &Bozanis, have proposed a new model that investigates the problem by two easily computable blogger scoring methods. These methods define two metrics MEIBI & MEIBIX that not only includes temporal aspects but also takes care of productivity of the blogger. The influential factors that have been considered for this model are inlinks, comments, productivity, and time i.e. age of the blog post and inlinks to the blog post. The outlinks factor was ignored in this model. Bloggers use outlinks to support or define their ideas. But it doesn’t mean it works against the novelty of the blog post.

Beyond any doubt, the number of incoming links to a blog post is strong evidence of its influence. Similarly, the number of comments made to a post is another strong indication that this blog post has received significant attention by the community. The case of outlinks is more subtle. In Web ranking algorithms like PageRank and hyperlink induced topic search (HITS), the links are used only as a recognition of (or to convey) authority. The influence-flow method of (Agarwal et al., 2008) assigns two semantics to a link: it is the means of conveying authority, and it is also a means of reducing the novelty. This mechanism results in two significant problems: (1) It misinterprets the intention of the link creators; and (2) it causes stability and convergence problems to the algorithm for the influence score calculation due to the existence of two sums (one for inlinks and one for outlinks). It is characteristic that the authors admit( Agarwal et al., 2008) that the presence of outlinks in novel posts is quite common and it is used “to support the post’s explanations.” Therefore, they argue that outlinks are not relevant to the post’s novelty, and all links should have a single semantic, that of implying endorsement (influence).

It is generally acceptable that the longer documents are possibly of higher informational value than the shorter ones. This intuition is also present in some of the most successful Web ranking functions, such as BM25 (Robertson, Zaragoza, & Taylor, 2004), where the length of a document is a factor that determines its score during query processing. Regarding blog communities, although the length of a post is not a safe indication of its influence, they accept that longer posts are likely to cause stronger reactions from other bloggers or readers than the shorter ones.

The temporal dimension is of crucial importance for identifying the influentials in a rapidly changing environment such as blogosphere. Time is related to the age of a blog post and also to the age of the incoming links to that post. Moreover, the age of the comments made to a post is also of significant importance. In the former case, the time involves the age of the post (e.g., in days since the current day) and in the latter case, the time involves the age (e.g., in days since the current day) of the incoming links and of the comments.

An influential blogger is recognized as such if he/she has written influential posts recently or if the posts have had an impact recently. Specifically, *Proximal impact:* Denotes the influence that a blogger has on the regular members/readers of the community. It is mainly visible by the comments made to a post. Table 2.1 show the details and explains the symbol used in MEIBI and MEIBIX model

**Table 2.1: Notations used in MEIBI & MEIBIX model**

|  |  |
| --- | --- |
| **SYMBOL** | **MEANING** |
| **BP(j)** | Blogger j’s blog posts. |
| **Bpj (i)** | Blogger j’s post i. |
| **Cj (i)** | Number of Comments on post i of blogger j. |
| **Rj (i)** | Inlinks referring post i of blogger j. |
| **ΔTPj(i)** | Difference between current date and posted date of post i of blogger j. |
| **ΔTP(i)** | Difference between current date and date referral post i submitted. |

**MEIBI: (Metric for Evaluating and Identifying Blogger’s Influence)**

MEIBI metric takes into the account the number of comments, inlinks received by blog post, along with the date of publication. A blog post posted today will become outdated after few months. The score for each blog post *i* that belong to blogger *j* is calculated as follows:wps_clip_image-32720 (2.4)

Where γ and δ are constants assigned values 4 and 1 respectively. After calculating (i) score for all blog posts the model assigns Blogger *j* a MEIBI index value equal to *m*, if and only if *m* blog posts of blogger *j*, has got a score >= *m* and remaining BP(j) –*m* posts got a score <*m*.

This MEIBI takes care of both productivity and influence of a blogger. The top k MEIBI index bloggers are the top k-influential bloggers.

**MEIBIX: (MEIBIX extended)**

Instead of considering the age of the post the authors decay the score for each blog, based on the age of the inlinks received. For each incoming link authors assign a weight that depends on the age of the inlink. The idea is formulated and a new score is calculated as follows:

(2.5)

After calculating score for all blog posts the model assigns Blogger *j* a MEIBIX index value equal to *x*, if and only if *x* blog posts of blogger *j*, has got a score >= *x* and remaining BP(j) –*x* posts got a score <*x*.

The model does not depend on any user defined weights, and evaluates both blog posts and bloggers effectively. A blogger will be influential if and only if he/she blogs influential posts very often.

**2.6.3 Productive and Influential Bloggers**

The MEIBI & MEIBIX indexes have considered both temporal aspects and productivity of the blogs, and it was able to show significant improvement over the preliminary model “Influence Flow Method”. Though these metrics were able to define unique index values to each blogger, they fail to differentiate between productivity and influence exactly. Considering this drawback the same authors Akritidis, Katsaros, & Bozanis, have refined their previous model in Feb 2011 to differentiate between productive bloggers and influential bloggers. In their defense productive bloggers are those who posts very often i.e. active bloggers, and influence bloggers are those whose writing’s impact others decisions. Many influential bloggers are also active bloggers. Although productivity and influence do not coincide, the influence of a blogger decays with time.

An influential blogger is recognized as such if he/she has written influential posts recently or if the posts have had an impact recently (Akritidis et al., 2011). The impact can occur as summary and described below(Akritidis et al., 2011).

**Proximal Impact:** Indicates the blogger’s influence on regular users/members of the community. This is measured through the comments received on a post.

**Wide Impact:** Indicates the blogger’s influence outside the community. This is measured through inlinks used by other bloggers outside the community. Table 2.2 shows and explains the symbol used in productive and influential bloggers model.

**Table 2.2: Notations used in “Productive and Influential bloggers” model**

|  |  |
| --- | --- |
| **SYMBOL** | **MEANING** |
| **Pjt** | Blogger j’s Productivity at instance t |
| **Ijt** | Blogger j’s Influence at instance t |
| **Nj** | Blogger j’s blog posts |
| **nji** | Blogger j’s post i. |
| **Cji** | Comments on blog post i of blogger j. |
| **Rji** | Inlinks received by blog post i of blogger j. |
| **Lji** | Length of blog post i of blogger j |
| **L** | Average length of blog post of blogger. |
| **T** | Time variable |
| **Tji,p** | Time stamp of post i of blogger j |
| **Tjx,l** | Time stamp of post x referring blogger j |
| **Tjx,c** | Comment x’s time stamp on post of blogger j. |

**Productivity Calculation:**

For each blog post *i* of blogger *j*, a score Uji,p (t) is measured as

(2.6)

where γ,, are predefined constants, whose values are 100, 86400, and 1, respectively. The score *Uji,p* connects the post with its age and its score decays with time and the blogger’s productivity index is defined next.

After calculating score *Uji,p (t)* for all blog posts the model assigns Blogger *j* a BP- index value at a given instance *t*, equal to *Pjt*, if and only if *Pjt* blog posts of blogger *j*, has got a score >= *Pjt* and remaining BP(j) – *Pjt* posts got a score <*Pjt*.

BP-index defines productive bloggers in terms of recent lengthy blog posts submitted by a blogger.

**Influence Calculation:**

For each blog post *i* of blogger *j*, the dual nature of influence i.e. proximal impact and wide impact can be quantified as follows.

(2.7)

Where *w1*= 100 and *wc*= 10. The score *Vji,p(t)* connects the comments and inlinks received by post with corresponding age and its score decays with time. And the blogger’s influence index is defined as following.

After calculating score *Uji,p (t)* for all blog posts the model assigns Blogger *j* a BI- index value, at a given instance *t*, equal to *Ijt*, if and only if *Ijt* blog posts of blogger *j*, has got a score >= *Ij t*and remaining BP(j) – *Ijt* posts got a score <*Ijt*.

The BI-Index defines influential bloggers in terms of recent influence. Also it appreciates blog posts that receive many inlinks and comments.

**CHAPTER THREE**

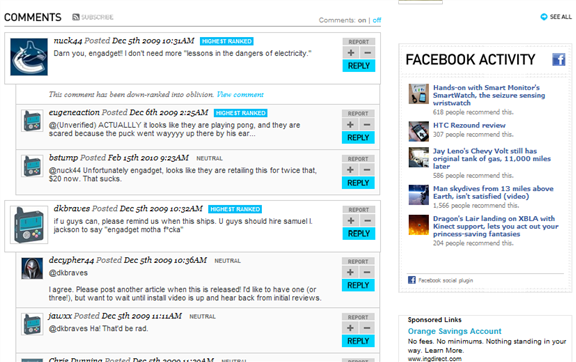
**RESEARCH METHODOLOGY**

**3.1 Proposed Influence Model**

In this section, the research will discuss different influential factors considered in the new model and the procedure followed to find influential bloggers in a blog community. Each blogger in a web blog community maintains a blog space, where bloggers blogs multiple posts time to time. For each blogger, the research calculate the influential score for all the posts and define Influence index (I-index) based on the scores obtained. In the new model take ideas and methodologies in the existing work and modified inefficient areas of the algorithm for better model. Figure 3.1 and figure 3.2 shows one of the blog posts from Engadget web blog community. As highlighted in the figure, each blog consists of title, post body, author name date of post, comments, social proof (number of f-share, twitter shares, Buffer, in share, pin it, etc).

****

**Figure 3.1: Cropped blog post with Author’s Name, Social Proof, Date of post, Number of comments**



****

**Figure 3.2: Cropped blog post with comments and Facebook activity**

**Figure 3.3: Cropped social proof from blog post**

Previous models did not used social proof as one of the parameters in finding influential bloggers.

Social proof is a psychological phenomenon where people adopt the actions of others in an effort to imitate correct behaviour for a given situation. This is also known as informational social influence, smart marketers used the social proof as a marketing strategies or approaches for easing the minds of worried customers. This system of arouse people awareness will always be productive because customers are becoming more informed all the time. With the power of the internet at their fingertips, customers can know an immense amount of information about your business before ever speaking with a salesperson.

Almost all the exiting model used comments as of the parameter in measuring influence of a blogger. But the exits model failed to weight the comments based on their nature, the former model biased in the use of comment, it make used of the number of comments received by a blogger, which make the model to be weak. Comments can be positive, negative and some are neutral. Also, not all comments of a blog is always relevant to the topic of blog discussion.

Some commenters agreed with the bloggers idea, some were neutral and some commenter disagreed with the blogger’s idea, so in other to get the truth nature of the influence of a blogger based on comments the research measure the comments of blogger by using positive, negative and neutral in measure the commenters of a blog post.

Therefore the comments on blog should be treated differently so as to know the mind of each commenters by this it will be very easy to see an experts comment that can enhanced the influence of post.

**3.2 Proposed Influential Factors:**

Beyond any doubt, the number of incoming links to a blog post is strong evidence of its influence. Similarly, the number of comments made to a post is another strong indication that this blog post has received significant attention by the community. Many users express their opinion on the post through comments. Some people agree with the author’s idea and some people disagree.

Outlinks are one other factor work against the influence of a blogger. This is shown in Influence Flow Method, outlinks in a post indicate usage of other’s idea to support blogger’s own idea, which work against the novelty of the post (Agarwal et al., 2008). Temporal (time) aspects of blogosphere also play a major role in influence, as the blog post that exerted impact on public today may not have any influence after two months or so (Akritidis et al., 2009). The research hasn’t considered the length of the post since length of a post doesn’t matter in as much the post’s content is of good quality.

Since bloggers are regularly keeping up-to-date of the latest news and often talk about new trends before they peak, recent research has focused on extracting opinions and identifying buzz from blogs (Glance et al., 2005). (Gruhl et al., 2005) have found strong correlation between spikes in blog mentions to Amazon sales ranks of certain books. More recently, (Lloyd et al., 2005) found similar trends for named entities in blog mentions and RSS news feeds.

Blogs are topical in nature and their link structures constantly evolve as new topics emerge. (Kumar et al., 2003) study the word burst models (Kempe et al., 2003) and community structure on the Blogosphere (Kumar et al., 1999). They find a sustained and rapid increase in the size of the strongly connected component on the Blogosphere and explain that the community structure is due to the tendency of the bloggers to topically interlink with posts on other blogs.

Finally, the thesis have considered the latest parameters that is of importance associated with blog posts Social Proof (i.e. facebook likes, G+ 1’s, sharing twitter) and length time in network on social networking websites as new influential factors. These latest parameters are also the indication of direct influence exerted by blogs, same as inlinks and comments. Table 3.1 show and explains the symbol used in the proposed new model for identifying influential bloggers

**Table 3.1: Notations used in the new approach**

|  |  |
| --- | --- |
| **Symbol** | **Meaning** |
| **Ijt** | Blogger j’s Influence at instance t |
| **BP(j)** | Blogger j’s blog posts |
| **nji** | Blogger j’s post i. |
| **Spji** | Social proof count of ith blog of blogger j. |
| **Cji** | Comments on blog post i of blogger j. |
| **Rji** | Inlinks received by blog post i of blogger j. |
| **Inf(bc)** | Influence of the blogger that issued comment c. |
| **SF(bc,x)** | Sentiment factor of blogger bc on comment x. |
| **TC(bc)** | Total number of comments of the blogger bc |
| **T** | Time variable |
| **tji,p** | Time stamp of post i of blogger j |
| **tjx,l** | Time stamp of post x referring blogger j |
| **tjx,c** | Comment x’s time stamp on post of blogger j. |
| **tb** | Living time length score of tb of blog site b |
| **ETt(i,j)** | Topical relevance of bloggers to a topic t |
|  | The number of messages of topic t by blogger j and inlinked by blogger i |
|  | The number of messages from blogger j by blogger i |

Social proof is the impression that users will adjust, adapt and change their behaviour according to what other people are doing. This always happen for examples when some see a queue of people waiting to eat at a restaurant or a photo of a personality drinking a brand of certain brand of soft drink, it helps in promoting the gravitas and quality to the product.

Social media has different ways to provide this kind of social proof. Facebook widgets that show other Facebook friends that “like” a brand, Twitter’s displays of people you follow that also follow another person, and the various ways that company offer rewards for referring others to the brand are all examples of this.

Social sharing buttons showing number of shares are also a great form of social proof. If a person sees that a landing page was shared a lot on social, it gives it more credibility.

Before reading random articles, news, web blog or a personal profile, many of us tend to look at the number of Social Proof hits (visitors), likes (fb-likes, G+ 1s etc.) or sharing (facebook, twitter etc.) on them, so that we can get an intuitive idea of goodness of current reading.

Nevertheless, these social proof hits, likes and sharing can be used as influential factors for the current model. The act of clicking on fb-like, G +1 or sharing on social web sites most likely indicates the interest or approval of the content read. The number of views, f-likes, G+1’s on a web blog post can be treated as the direct influence received by post same as inlinks. People by sharing on facebook and twitter are supporting and voting for the web content, so capturing those shares and votes can help us in evaluating support or acceptance that the web content has received. But keeping in mind over fitting error problem the thesis have considered the combination of social proof likes and shares as “social proof count” parameter and introduced it as a new influential factor into the model. So given social proof count a time varying Sp-score for a blog i of blogger j is given by

(3.1)

, are predefined constants whose values are 1, 86400 respectively. represents the time difference between the current date and publication date of the post in seconds. Wsp the weight parameter of social proof count factor whose value is 5.

**Algorithm I** Calculating the Social Proof Influence of a Blogger

1: Let t be a likes from the set of Social Proof S.

2: Let s be a shares from the set of Social Proof S.

3: Let Sp be a set of Social Proof of the form , where i is the

social proof likes post p produced by blogger j.

4: Let be the time difference between the current data

and publication data of the post per second

5: Let Wsp be the weight parameter where value is 5

6: Let, be constant where value are 1 and 86400

7: **for all** Sp do

8: **if** blogger i has social proof t then

9:

10: **else**

11: = 0

12: **end ifif** the Let be the number of likes and shares related to

post t by blogger i

13: **end for**

Research in the area of information propagation was motivated by a large body of work in the areas of model topic propagation. As described in (Gruhl et al., 2004) this model applies well in the Blogosphere where a blogger may have a certain level of interest in a topic and is thus prone to talking about it. By discussing the topic he/she may infect others and over time might recover. The research use this approach in characterizing individuals into various phases of a topic in which they are more likely to become infected. The research model individual propagation and use an expectation maximization algorithm to predict the likelihood of a blogger linking to another blogger. They also study the different types of topics present in the dataset and describe an approach to categorize topics into subtopics. Certain topics are more infectious than others and spread through the social network of bloggers. Automatically predicting such topics and developing models to accurately identify the propagation patterns on the Blogosphere is another focus of this research. Emerge topic can also be used as a parameter in identify influential of a blogger.

First identifying the set of blogs revolved about some topic, and then characterizing the different patterns into which the collection of blogs about the topic may fall. Then assumed that topics are mostly composed of a union of conversation (ongoing discussion whose subtopic flow is largely determined by decisions of the blogger.

The perception is that bloggers’ writing “interest" can provide a better insight of the influence of a blogger based on topical relevance. The topical influence of a blog comment depends on the topical diversity contained on his posts and the number of times other bloggers inlink his posts. The more inlink a post generates the larger the topical audience it reaches. The same applies for the case of an entity.

Let IL be a set of triples of the form (i; k; j), representing the inlink relationship between blogger i and blogger j through post k. The thesis define the topical relevance of the inlink relation (i; j) to a given topic t as follows:

Given the inlink relationship (i; j), the topical relevance of this pair of bloggers to a topic t is defined as:

(3-2)

Where, represents the number of messages of topic t written by blogger j and inlinked by blogger i; and represents the number of messages from blogger j inlinked by blogger i.

This definition captures the notion that the more a blogger iinlink posts from blogger j about topic t, the higher the relevance of blogger j to this topic. Generally this leads to a higher influence on i, corresponding to a higher rank conferred from blogger i to blogger j.

**Algorithm II** Calculating the Topical Influence Checking Vector

1: Let t be a topic from the set of topics T.

2: Let d be the total number of bloggers posting subtopic related

to topic t.

3: Let IL be a set of inlink triples of the form (i; k; j), where i

is the blogger inlinking post k produced by blogger j.

4: Let be the checking vector for topic t

5: **for all** triples in IL do

6: **if** blogger i has Topic t then

7:

8: **else**

9: = 0

10: **end ifIf** the Let jKij be the number of messages related to

topic t inlinked by blogger i

11: **end for**

When considering the activity generated “Post-Reply relationship” (Cai & Chen, 2010) by a blog post there are important and valuable information embedded with it. First, the influence of each comment may have different impact power, depending on who issues it. The comment might be from an expert on the topic or a general blogger or a normal user.

Their comments on blog post should be treated differently, and it is easy to see an expert’s comment would enhance the influence of post more. Secondly, the comments from other bloggers could be positive, negative or neutral, and these sentimental factors also affect the post’s influence among commenters. So given set of comments to blog post *Cji*, for each comment *x* the thesis ought to find the person who issued it and his consent towards the blog post.

The work use this information to find the comment score achieved by the blog post. Also each comment is associated with the time stamp which can be used to keep up with the dynamic nature of the blog post. For a received set of comments Cji,p (t) the comment score for ith post of blogger *j* is calculated as follows:

(3.3)

Wc is the weight parameter of comment factor whose value is 8.

**Algorithm III** Calculating the received set of comments

Influence of a Blogger

1: Let *x* be a comment from the set of Comment C.

2: Let *c* be total comments from the set of Comments TC.

3: Let C be a set of Comment of the form , where i is the

comment post p produced by blogger j.

4: Let be the time difference between the current date

and publication date of the comment per second

5: Let Wc be the weight parameter of comment factor whose value is 8

6: **for all** C do

7: **if** blogger i has comment *x* then

8:

9: **else**

10: = 0

11: **end ifIf** then Let be the number of comments related to

post t by blogger i

12: **end for**

Also, when scoring each blog post, the model should be able to mine the post content. The score of blog post should also reflect the goodness and quality of the content. Novelty is one such parameter that is directly related to blog post and can convey quality of content present. So the thesis include novelty in scoring the algorithm. Novelty is inversely proportional to outlinks present in the post (Agarwal et al., 2008). i.e. the more the number of outlinks present in blog post, the more the number of ideas borrowed from others to support the idea present in the post. Also to measure the novelty the researcher should know how novel the content is for which need to build a word space that represent a general space of ideas. In the new approach the thesis considered each blog post as a document. All the stop words were removed from the documents and then using TFIDF weighing scheme to extract all the important keywords from the documents to create word space. After creating word space the thesis project each latest blog post into the word space to estimate the degree of novelty. And the novelty is calculated as follows.

Nji = Wn (3.4)

Where *W*n is novelty weight parameter and *Wo* is outlinks weight parameter whose values are 30 and 1/10, respectively.

A blog site with longer living time may have more opportunities to interact with more other bloggers, thus the impact on the entire blog social network will be deeper and wider than the blog site which with shorter living time. Living time length score of (τ*b*) of blog site *b* is calculated as:

Tji,p = T(b)/max βεB {T(B)} (3.5)

Where *T* (*b*) is the time interval between the post time stamps of first and the latest posted blogs in Uji.

So finally the influence score for each blog post is calculated as follows.

The score for post *i* of blogger *j* is:

(3.6)

Where INLKji,p is the inlink score received by the blog post i of blogger j given by (Leonidas, Dimitrios, & Panayiotis, 2009),

(3.7)

Wl is the inlink weight parameter whose value is 75.

After calculating the post score *PSji,p (t)* for all blog posts using equation 3.7 the model assigns Blogger *j* a Influence- index value at a given instance *t*, equal to Ijt, if and only if Ijt blog posts of blogger *j*, has got a score >= Ijt and remaining BP(j) – Ijt posts got a score <Ijt.

**3.3 Data Collection**

Data collection is the first and one of the serious work in this work. The main aim of this research is to find inﬂuential bloggers on the web. There are many blog data that are available from different blog website that can be used to find inﬂuential bloggers, but most of this site lack some of the parameters used for this research e.g. Buzz Metric dataset research experiments but it lack some key statistics required for this research. Hence, for the purpose of this research endadget blog site will be used for this work. The site is a multilingual technology blog site with daily coverage of gadgets and consumer electronics. It currently operates a total number of ten (10) blogs. The site ranked among the top five in the Technorati top 100 in the past and one of the best blogs of 2010 by Time. It is used for news and reviews founded in 2004 and has been operated by AOL since October, 2005. The site available in five different languages which includes English, Chinese (traditional and simplified), Japanese, Spanish, and German.

**CHAPTER FOUR**

**RESULTS AND DISCUSSION**

**4.1 Introduction**

In this chapter the research discuss the dataset for experimental data and select a blog site that have all the parameters that will be used for this research, design various experiments with the developed model and used the mathematical model to find influential blogger based on the experimental results. It discuss how the experiments are carried out with the results obtained and compare the results against the previous model. First, the research need blog post dataset with necessary fields and parameters and make used of the dataset that contain blog posts from engadget web blog community. The research develop and elaborate an evaluation procedure for effective comparison with previous model used in find influential blogger.

**4.2 Evaluation of the proposed model**

The research perform comparison against other models used to find influential blogger in a community and on the web. To check if all the three additional parameters are necessary, the research design and perform a lesion study and correlation study. One or more parameters used may be correlated with each other or one of them maybe redundant. Pairwise correlation analysis was conducted among all the three parameters add to the models and lesion study is conducted taking one parameters out each time and comparing the results.

**4.3 Experimental results**

The research make used of the dataset from authors of “Productive and Influential Bloggers (Akritidis, Katsaros&Bozanis, 2011) that have a small dataset containing blog posts from engadget web blog community. Table 4.1 gives the sample of the dataset used from engadget web blog community.

**Table 4.1: Sample blog post from Endagdet web blog community**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Post ID** | **Post Title** | **Post Author** | **Post Date** | **Post Comment** | **Post Inlinks** |
| 11224 | Engadget give away: win some Beatles MusicSkins official Apple now offering | Laura June | 3/22/2010 0:00 | 550 | 5 |
| 11113 | ipones contact free (upedat: not unlocked) Official: Palm Pre Plus and Pixi Plus | Joshua Topolsky | 3/22/2010 0:00 | 311 | 29 |
| 13557 | For AT &amp, T in the coming months’ | Joshua Topolsky | 3/22/2010 0:00 | 220 | 23 |
| 14890 | Kindle for iPad and tablets makes the scene. Google.cn now rerouting to Hong | Thomas Ricker | 3/22/2010 0:00 | 129 | 32 |
| 11557 | Kong domain, an ‘entirely legal’ workaround to censorship woes sprint to announce. | Ross Miller | 3/22/2010 0:00 | 112 | 26 |
| 10445 | ‘groundbreaking new device’ (HTC Supersonic?) tomorrow Samsung to announce Galaxy S | Ross Miller | 3/22/2010 0:00 | 161 | 18 |
| 11000 | Smartphone, content initiatives this week | Chris Ziegler | 3/22/2010 0:00 | 20 | 31 |
| 13890 | Early reports show IE not faring well in the post-ballot screen days | Tim Stevens | 3/22/2010 0:00 | 182 | 15 |
| 15112 | Nokia Nuron for T-Mobile review | Chris Ziegler | 3/22/2010 0:00 | 112 | 15 |
| 12779 | Tata Motors’ Nano lights up the streets of Mumbai | Joseph L. Flatley | 3/22/2010 0:00 | 114 | 13 |
| 11335 | George Takei can’t show you Sharp’s fourth pixel can still blow your mind | Richard Lawler | 3/22/2010 0:00 | 114 | 12 |
| 12890 | Spring Design Alex review | Joanna Stem | 3/22/2010 0:00 | 52 | 15 |
| 12112 | Impossible Project’s Polaroid film goes on sale this week | Donald Melanson | 3/22/2010 0:00 | 45 | 15 |
| 12557 | Motorola’s Android powered il launching at CTLA | Sean Coper | 3/22/2010 0:00 | 56 | 14 |
| 11779 | Motorola makes il official, melds Android and push-to-talk this summer on Sprint | Ross Miller | 3/22/2010 0:00 | 61 | 13 |

The dataset used has all the needed parameters postdate, post title, post content, author name, comments, comment dates, inlinks and inlinks dates etc. The research crawled through each blog post for the additional parameters that are missing for the experiment, the social proof which comprises Facebook likes, twitter shares, G +1s etc and add them together to give us total number used as social proof. The research used the previous model Productive and Influential Bloggers by other authors as new sample and introduced the new parameters one by one to see how it work. To avoid over fitting error by adding too many parameters into the model that can jeopardize the overall performance of the model, the experiment was carried out in different stages.

The research introduced the first parameters (new influential factor) Social Proof into the model and study the changes in the results.

**Table 4.2: Individual blog post scores when Social proof is introduced into model**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Post ID** | **Post Author** | **Score\_Y** | **Score\_Z** | **PS\_SP** |
| 11224 | Laura June | 18 | 781 | 799 |
| 11113 | Joshua Topolsky | 16 | 768 | 1186 |
| 13557 | Joshua Topolsky | 93 | 694 | 1268 |
| 14890 | Thomas Ricker | 19 | 569 | 655 |
| 11557 | Ross Miller | 31 | 519 | 907 |
| 10445 | Ross Miller | 12 | 447 | 548 |
| 11000 | Chris Ziegler | 21 | 437 | 447 |
| 13890 | Tim Stevens | 9 | 418 | 475 |

In Table 4.2 columns Score\_Y&Score\_Z represent the productive & influence scores of posts according to “productive and influential bloggers” model, and PS\_SP represents the influence score of posts according to resultant model when social proof is introduced.

To monitor the blog score when social proof is introduced, the research used random blog posts posted on 03/22/2010 to monitor the changes when social proof is introduced. The influence score of blog score entirely depends on other parameters like comments and inlinks since the date for all the blog posts is the same. Study the influence scores of score\_Z, and PS\_PS for blog posts with post\_id 14890 and 11557, there is a subtle difference in number of comments and inlinks but blog post 11557 has greater social proof than the post 11890, which has favored the influence score of 11557 to be more in resulting model unlike “productive and influential blogger” model.

The result model scores each blog post and defines an influence index from the score obtained for each blogger. The results of the influence index has been used to rank the bloggers among other bloggers in the community. When social proof is introduced into the model, the results from table 4.3 shows how the final ranking of the bloggers has been changed. Productive blogger is represents by BP-index and influential bloggers is represents by BI-index of the blogger according to “productive and influential bloggers (Akritidis, Katsaros & Bozanis, 2011)” and I-SP represents Influence-index of blogger according to resulting model. BP-index measures how often blogger produces lengthy posts is defined out of all score\_Y, which is not related to influence, hence, BI-Index was used to compare the result model. BI-index is calculated using all score\_Ys which measures the influence of blog post and I-PS is calculated using all PS\_PS of blogger.

**Table 4.3: BI vs I-PS**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Author ID** | **Author Name** | **BP** | **BI** | **I-PS** |
| 20 | Darren Murph | 23 | 144 | 124 |
| 13 | Chris Ziegler | 18 | 94 | 101 |
| 92 | VladislavSavov | 17 | 91 | 111 |
| 53 | Laura June | 13 | 88 | 98 |
| 64 | Nilay Patel | 13 | 74 | 87 |
| 87 | Thomas Ricker | 13 | 72 | 83 |
| 49 | Joshua Topolsky | 16 | 72 | 76 |
| 67 | Paul Miller | 20 | 67 | 78 |
| 23 | Donald Melanson | 12 | 64 | 74 |
| 88 | Tim Stevens | 14 | 63 | 74 |
| 45 | Joseph L. Flatley | 16 | 56 | 66 |
| 73 | Ross Miller | 15 | 54 | 63 |
| 42 | Joanna Stern | 19 | 42 | 49 |
| 75 | Ryan Block | 3 | 39 | 39 |
| 71 | Richard Lai | 9 | 36 | 40 |

**Figure 4.1: Influential Bloggers using Social Proof**

The resultant model ranked V Savov second and C Ziegler third as shown in Table 4.3. R Block and R Lai ranks are interchanged and J Topolsky and P Miller ranks are also interchanged when introduced the new parameter (social proof).

In the second stage mining comments was introduced into the sample model to see if there is any change in the ranking of the scores of blog posts. Last column represent the influence score of blog post when mining comment is introduced into the new model.

**Table 4.4: Individual blog post scores with mining comments technique**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Post ID** | **Post Author** | **Post Comment** | **Score\_Y** | **Score\_Z** | **PS\_Comment** |
| 11224 | Laura June | 550 | 18 | 781 | 908 |
| 15112 | Chris Ziegler | 112 | 119 | 322 | 371 |
| 12779 | Joseph L. Flatley | 114 | 12 | 308 | 271 |
| 11335 | Richard Lawler | 114 | 9 | 299 | 357 |
| 12890 | Joanna Stem | 52 | 127 | 261 | 252 |
| 12112 | Donald Melanson | 45 | 8 | 256 | 242 |
| 12557 | Sean Coper | 56 | 7 | 247 | 236 |
| 11779 | Ross Miller | 61 | 16 | 246 | 229 |

The final influence score for mining comment favored 11335 unlike productive and influential bloggers model (Score\_Z) though blog posts with post\_id 12779 and 11335 have almost same number of comments.

The comment body was used to find the tone of each comments and was given weight accordingly. Content of the comment was used as the tone for each responds to the blog post, the author that issued the comment and the comment tone was evaluated by the model respectively. To calculate the comment tone, the resultant model considers influence of comment\_author along with his/her tone to calculate the comment score. The sentiment factor (SF in eq. 8) for each positive comment is +2, -1 for each negative comment and +1 for a neutral comment. The technique capture the true acceptance of topic unlike prototype model by mining comments and finding out commenters consent towards the post. Also the comments are scored based on the users issued the comments and their influence (eq. 8).

**Table 4.5: Statistics of mined comments**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Post ID** | **Positive** | **Negative** | **Neutral** | **Total**  **Comments** | **Avg Influence**  **of Authors** |
| 12779 | 15 | 17 | 82 | 114 | 83.19 |
| 11335 | 26 | 19 | 69 | 114 | 85.65 |

Post ID 12779 and 11335 was used as sample because both comments has the same number of comments. The blog post 11335 has received more number of positive comments from the classification of comments as shown in Table 4.5. Post ID 11335 has 85.65 average influence score of authors against 83.29 for post ID 12779, with highest number of positive comments in post Id 11335 the above example explains the reason why blog post 11335 is being scored more than 12779.

Table 4.6 shown the result of the mining comments (Influence-indexes)of the bloggers when applied on the dataset. The Influence-index result of the model is represented by I-Mining comment column when comment score part (second half of eq. 6) in “productive and influential bloggers” has been replaced by “mining comments” technique (eq. 8).

**Table 4.6: BI vs. I-Mining comments**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Author ID** | **Author Name** | **BP** | **BI** | **I-MC** |
| 20 | Darren Murph | 23 | 114 | 104 |
| 13 | Chris Ziegler | 18 | 94 | 92 |
| 92 | VladislavSavov | 17 | 91 | 91 |
| 53 | Laura June | 13 | 88 | 74 |
| 64 | Nilay Patel | 13 | 74 | 74 |
| 87 | Thomas Ricker | 13 | 72 | 73 |
| 49 | Joshua Topolsky | 16 | 72 | 66 |
| 67 | Paul Miller | 20 | 67 | 65 |
| 23 | Donald Melanson | 12 | 64 | 66 |
| 88 | Tim Stevens | 14 | 63 | 65 |
| 45 | Joseph L. Flatley | 16 | 56 | 56 |
| 73 | Ross Miller | 15 | 54 | 52 |
| 42 | Jonna Stern | 19 | 42 | 45 |
| 75 | Ryan Block | 3 | 39 | 25 |
| 71 | Richard Lai | 9 | 36 | 35 |

**Figure 4.2: Influential Bloggers using Mining Comments**

The introduction of mining comments equation have changed the ranking among among J Topolsky, P Miller, D Melanson and T Stevens have been interchanged as shown in table 4.8. Also Ryan Block and Richard Lai’s ranks were interchanged. The methods followed the true influence of blog post by capturing the tone of the commenter even though most of the ranking order doesn’t change.

In the third stage of experiment the research applied “emerge topic method into the prototype model. The parameter changed the scores of blog posts as shown in Table 4.7 the last column PS\_ET represent influence score of blog post when mining of the emerge topic is introduced into the sample model.

**Table 4.7: Individual blog post scores with emerge topic technique**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Post ID** | **Post Title** | **Post Author** | **Post Inlinks** | **Score\_Y** | **Score\_Z** | **PS\_ET** |
| 11224 | Engadget give away: win some Beatles MusicSkins official Apple now offering | Laura June | 5 | 18 | 781 | 48 |
| 11113 | iphones contact free (update: not unlocked) Official: Palm Pre Plus and Pixi Plus | Joshua Topolsky | 29 | 16 | 768 | 139 |
| 14890 | Kindle for iPad and tablets makes the scene. Google.cn now rerouting to Hong | Thomas Ricker | 32 | 19 | 569 | 146 |
| 11557 | Kong domain, an ‘entirely legal’ workaround to censorship woes sprint to announce. | Ross Miller | 26 | 31 | 519 | 129 |
| 11000 | Smartphone, content initiatives this week | Chris Ziegler | 31 | 21 | 437 | 138 |
| 13890 | Early reports show IE not faring well in the post-ballot screen days | Tim Stevens | 15 | 9 | 418 | 77 |
| 12779 | Tata Motors’ Nano lights up the streets of Mumbai | Joseph L. Flatley | 13 | 12 | 308 | 73 |
| 12112 | Impossible Project’s Polaroid film goes on sale this week | Donald Melanson | 15 | 8 | 256 | 73 |
| 12890 | Spring Design Alex review | Joanna Stem | 15 | 127 | 261 | 77 |
| 11335 | George Takei can’t show you Sharp’s fourth pixel can still blow your mind | Richard Lawler | 12 | 9 | 299 | 70 |
| 11346 | Gmail for Android update brings cleaner conversation view and improved multi-select | Darren Murph | 38 | 13 | 216 | 170 |
| 21134 | Google's next Android Music app leaks out, gives hints of a 3.0 UI overhaul | Paul Miller | 22 | 22 | 229 | 115 |
| **11347** | RIM's BlackBerry Remote Stereo Gateway on sale for $80 | **Evan Blass** | **3** | **12** | **215** | **33** |
| 11245 | WSJ: Apple going to CES 2010. Reality: Nope | Ryan Block | 14 | 19 | 232 | 85 |
| 11359 | No guy in a coma, no missed iPhone launch, no kidding | Peter Rojas | 11 | 6 | 113 | 79 |

The blog posts with post\_id 11346, 11000 and 11557 have almost same number of mining emerge topic since they have highest number of inlinks which is the most determinant for mining emerge topic and the final influence score has been favored 11346 to be more in the result model(PS\_ET) unlike “productive and influential bloggers” model (Score\_Z).

**Table 4.8: BI vs. I-ET**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Author ID** | **Author Name** | **BP** | **BI** | **I-ET** |
| 20 | Darren Murph | 23 | 114 | 117 |
| 13 | Chris Ziegler | 18 | 94 | 101 |
| 92 | VladislavSavov | 17 | 91 | 99 |
| 53 | Laura June | 13 | 88 | 96 |
| 64 | Nilay Patel | 13 | 74 | 89 |
| 87 | Thomas Ricker | 13 | 72 | 104 |
| 49 | Joshua Topolsky | 16 | 72 | 48 |
| 67 | Paul Miller | 20 | 67 | 71 |
| 23 | Donald Melanson | 12 | 64 | 68 |
| 88 | Tim Stevens | 14 | 63 | 70 |
| 45 | Joseph L. Flatley | 16 | 56 | 68 |
| 73 | Ross Miller | 15 | 54 | 55 |
| 42 | Jonna Stern | 19 | 42 | 70 |
| 75 | Ryan Block | 3 | 39 | 73 |
| 71 | Richard Lai | 9 | 36 | 67 |

**Figure 4.3: Influential Bloggers using Emerge Topic**

The introduction of emerge topic equation have changed the ranking among Chris Ziegler, Thomas Ricker, Joshua Toplsky and Ross miller, the result have been interchanged as shown in table 4.8. Also Tim stevens and Jonna Stern ranks were the same. The methods show the true influence of blog post by capturing those bloggers that people reference their post always compare to other bloggers, this make most of the ranking order to change.

In stage 4 the research combined social proof and mined comments into productive and influential blogger model to see how the result will rank the bloggers. The I-(SP + MC) represents influential-index of the result model when both the two parameters were combined. It was observed that there are many changes in the ranking order of the result model.

**Table 4.9: BI vs. I-(SP+MC)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Author ID** | **Author Name** | **BI** | **I-(SP+MC)** |
| 20 | Darren Murph | 114 | 112 |
| 13 | Chris Ziegler | 94 | 97 |
| 92 | VladislavSavov | 91 | 105 |
| 53 | Laura June | 88 | 77 |
| 64 | Nilay Patel | 74 | 82 |
| 87 | Thomas Ricker | 72 | 76 |
| 49 | Joshua Topolsky | 72 | 67 |
| 67 | Paul Miller | 67 | 71 |
| 23 | Donald Melanson | 64 | 69 |
| 88 | Tim Stevens | 63 | 71 |
| 45 | Joseph L. Flatley | 56 | 59 |
| 73 | Ross Miller | 54 | 56 |
| 42 | Jonna Stern | 42 | 46 |
| 75 | Ryan Block | 39 | 25 |
| 71 | Richard Lai | 36 | 38 |

Table 4.9 shows the merged results of social proof and mining comments with better judgment between bloggers. The raking of bloggers that is highlighted has been interchanged.

In stage 5 the research combined social proof and emerge topic into productive and influential blogger model to see how the result will rank the bloggers. The I-(SP + ET) represents influential-index of the result model when both social proof and emerge topic parameters were combined. It was observed that there are many changes in the ranking order of the result model.

**Figure 4.4: Influential Bloggers using Social Proof and Mining Comments**

**Table 4.10: BI vs. I-(SP+ET)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Author ID** | **Author Name** | **BI** | **I-(SP+ET)** |
| 20 | Darren Murph | 114 | 118 |
| 13 | Chris Ziegler | 94 | 99 |
| 92 | VladislavSavov | 91 | 88 |
| 53 | Laura June | 88 | 83 |
| 64 | Nilay Patel | 74 | 84 |
| 87 | Thomas Ricker | 72 | 95 |
| 49 | Joshua Topolsky | 72 | 81 |
| 67 | Paul Miller | 67 | 81 |
| 23 | Donald Melanson | 64 | 73 |
| 88 | Tim Stevens | 63 | 69 |
| 45 | Joseph L. Flatley | 56 | 65 |
| 73 | Ross Miller | 54 | 78 |
| 42 | Jonna Stern | 42 | 61 |
| 75 | Ryan Block | 39 | 56 |
| 71 | Richard Lai | 36 | 52 |

Table 4.10 shows the merged results of social proof and emerge topic with better judgment between bloggers. There is reordering among the blogger especially among Laura June, Nilay Patel, Thomas Ricker, Paul Miller and Joshua Topolsky.

In stage 6 the research combined mining comments and emerge topic into productive and influential blogger model to see how the result will rank the bloggers. The I-(MC + ET) represents influential-index of the result model when both mining comments and emerge topic parameters were combined. It was observed that there are many changes in the ranking order of the result model.

**Figure 4.5: Influential Bloggers using Social Proof and Emerge Topic**

**Table 4.11: BI vs. I-(MC +ET)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Author ID** | **Author Name** | **BI** | **I-(MC+ ET)** |
| 20 | Darren Murph | 114 | 103 |
| 13 | Chris Ziegler | 94 | 113 |
| 92 | VladislavSavov | 91 | 83 |
| 53 | Laura June | 88 | 82 |
| 64 | Nilay Patel | 74 | 88 |
| 87 | Thomas Ricker | 72 | 98 |
| 49 | Joshua Topolsky | 72 | 81 |
| 67 | Paul Miller | 67 | 75 |
| 23 | Donald Melanson | 64 | 74 |
| 88 | Tim Stevens | 63 | 75 |
| 45 | Joseph L. Flatley | 56 | 71 |
| 73 | Ross Miller | 54 | 80 |
| 42 | Jonna Stern | 42 | 61 |
| 75 | Ryan Block | 39 | 58 |
| 71 | Richard Lai | 36 | 56 |

Table 4.11 shows the merged results of mining comments and emerge topic with better judgment between bloggers. There is reordering among the blogger especially among Nilay Patel, Thomas Ricker, Joshua Topolsky, Paul Miller, Donald Melanson, Tim Stevens, Joseph .L. Flatley and Ross Miller.

**Figure 4.6: Influential Bloggers using combined Emerge Topic and Mining Comments**

In stage 7 the researcher applied novelty along with Social proof, emerge topic and mining comments into model which is the final model. To find novelty of a blog the research need to build word space which represents a general space of ideas. In the new approach the work considered each blog post as a document. The work first removed the entire stop words from the documents and then using TFIDF weighing scheme to extract all the important keywords from the documents to create word space. After that the researcher projected each latest blog post into the word space to estimate the degree of novelty. The outlinks used in blog post also plays major role in measuring degree of novelty. And novelty is calculated using eq. 9.

**Table 4.12: Bi vs. I-New Approach**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Author ID** | **Author Name** | **BP** | **BI** | **I-New Approach** |
| 20 | Darren Murph | 23 | 114 | 103 |
| 13 | Chris Ziegler | 18 | 94 | 99 |
| 92 | VladislavSavov | 17 | 91 | 94 |
| 53 | Laura June | 13 | 88 | 88 |
| 64 | Nilay Patel | 13 | 74 | 96 |
| 87 | Thomas Ricker | 13 | 72 | 115 |
| 49 | Joshua Topolsky | 16 | 72 | 80 |
| 67 | Paul Miller | 20 | 67 | 77 |
| 23 | Donald Melanson | 12 | 64 | 74 |
| 88 | Tim Stevens | 14 | 63 | 74 |
| 45 | Joseph L. Flatley | 16 | 56 | 71 |
| 73 | Ross Miller | 15 | 54 | 68 |
| 42 | Jonna Stern | 19 | 42 | 55 |
| 75 | Ryan Block | 3 | 39 | 48 |
| 71 | Richard Lai | 9 | 36 | 47 |

When the researcher applied the final model on dataset the ranking has been definite and more accurate than the other models. The column I-New Approach indicates the Influence Indexes of final model. As shown in Table 4.12 the ranking has been reordered, the bloggers were ranked without any overlapping among them. Darren Murph has been ranked 2nd and C Ziegler ranked 3rd. Comparing with I-(SP + MC) result, Vladislav Savov has taken 5th rank and Laura June took 6th. And the ranking among J Topolsky, P Miller, D Melanso, and T Stevens is interchanged without any overlapping. And when compared the new approach with I-(MC + ET) result, Vladislav Savov has taken 5th and Laura June took 6th, the ranking among Thomas Ricker, Chris Ziegler, Thomas Ricker and Nilay Patel without overlapping.

Finally all the results of the seven (7) experiments were put together side by side in Table 4.13 for better comparison and understanding.

**Table 4.13: BI vs. I-indexes**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Author ID** | **Author Name** | **BI** | **I-SP** | **I-MC** | **I-ET** | **I-**(PS+**MC**) | **I-(SP+ET)** | **I-(MC + ET)** | **I-New Approach** |
| 20 | Darren Murph | 144 | 124 | 104 | 117 | 112 | 118 | 103 | 103 |
| 13 | Chris Ziegler | 94 | 101 | 92 | 101 | 97 | 99 | 113 | 99 |
| 92 | VladislavSavov | 91 | 111 | 91 | 99 | 105 | 88 | 83 | 94 |
| 53 | Laura June | 88 | 98 | 74 | 96 | 77 | 83 | 82 | 88 |
| 64 | Nilay Patel | 74 | 87 | 74 | 89 | 82 | 84 | 88 | 96 |
| 87 | Thomas Ricker | 72 | 83 | 73 | 104 | 76 | 95 | 98 | 115 |
| 49 | Joshua Topolsky | 72 | 76 | 66 | 48 | 67 | 81 | 81 | 80 |
| 67 | Paul Miller | 67 | 78 | 65 | 71 | 71 | 81 | 75 | 77 |
| 23 | Donald Melanson | 64 | 74 | 66 | 68 | 69 | 73 | 74 | 74 |
| 88 | Tim Stevens | 63 | 74 | 65 | 70 | 71 | 69 | 75 | 74 |
| 45 | Joseph L. Flatley | 56 | 66 | 56 | 68 | 59 | 65 | 71 | 71 |
| 73 | Ross Miller | 54 | 63 | 52 | 55 | 56 | 78 | 80 | 68 |
| 42 | Joanna Stern | 42 | 49 | 45 | 70 | 46 | 61 | 61 | 55 |
| 75 | Ryan Block | 39 | 39 | 25 | 73 | 25 | 56 | 58 | 48 |
| 71 | Richard Lai | 36 | 40 | 35 | 67 | 38 | 52 | 56 | 47 |

**Figure 4.7: Influential Bloggers for the New Approach**

**4.4 Lesion study**

Lesion study was performed on the model by removing one parameter in turn. That is, the research compute the influence scores using only the remaining two parameters. Ranking the most six influential bloggers by leaving one parameter out one after the other and thus obtain the ranking results and compared the results obtained with the result of All-in-All (all the three parameters). The performance was conducted to check if there been a parameter that did not contribute to the influence score of a blogger, if removing it would not result in any difference in the ranking of these bloggers. Figure 8 presented the results obtained from the lesion study, x-axis denotes different ranking schemes to find the influential. For example, “No Emerge Topic” means the ranking of influential bloggers was computed using social proof and mining comments without the emerge topic. The results excitingly shown that all the top six influentials remain unchanged, but their relative ranks differ. The results shown that no blogger maintains the same rank in all the five variations and no two ranked lists are the same. Therefore, it is evident that the three parameters contribute to the proposed model in determining influential blogger.

**Figure 4.13: Lesion study for evaluating significance of each of the parameters.**

**5.1 CONCLUSION**

Blogs have made it possible for people to express their thoughts, voice their opinions, and share their experiences and ideas. Individuals experience a sense of community, a feeling of belonging, a bonding that members matter to one another and their recess needs will be met through online interactions.

Blogs are becoming one of the most popular media of communication and interaction among blog community. Blogs have demonstrated an ability to influence, to shape events, news, educations, politics etc. This makes the technology and those who use it key players in day to day activities of human endeavor. It have been a major communication channel for many internet users where they sharing ideas, opinions, and experience with others through comments.

Identify influentials blogger will not only allow us to better track and understand the interesting things happening in a virtual world, but also present unique opportunities for academic scholars, political scientist, industry, sales, and advertisements etc.

There exist model that proposed various solutions to identify influential bloggers in web community. Attempts to mathematically determine which blogs are the most influential are currently characterized by some shortcoming and failed to consider important aspects of the blog posts. The existing methodology mainly depended on numbers associated with the blog post i.e. comments, inlinks, outlinks and post lengths etc. They failed to considered the user tone toward the blog post (i.e. agree or disagree) which is very important while calculating (evaluate) the influence neither the process measures influence by considered the novelty of post which indicates the goodness and purity of the content.

Encouraged from the literature review and classy solutions that the current literature lacking, the research proposed a method that evaluate each post based on their quality and user consent toward the content present. The competing methods have not taken into account the quality and user consent towards the content present, which we argue are also important ones when evaluate influence of bloggers in a virtual community.

We introduced new parameters which can be used as influential factor (social proof and emerge topic) into the literature that further evaluated blog posts proficiently. The main motivation for the introduction of social proof and emerge topic is that they are very important and helped in identify the influence of a blog in web community.

The final model on the dataset used, the ranking has been definite and more accurate than the previous model (Identifying the Influential Bloggers in a Community and Identifying the Productive and Influential Bloggers in a Community). The new approach indicates that the ranking has been reordered and the bloggers were ranked without any overlapping. Thomas Ricker was ranked 1st, Darren Murph has been ranked 2nd and C Ziegler ranked 3rd. Comparing with I-(SP + MC), I-(MC + SP) and I-(ET + MC) results, some of the ranking among the bloggers were interchanged without any overlapping.

When the results of the seven (7) experiments were put side by side, it can be seen that the New Approach could better identify influential of bloggers on the web and more robust than the existing models ( i.e. MEIBI and MEIBIX), which show that the new approach has improved over the previous models.

The experimental results shows that current approach could significantly identify influential of bloggers on the web and the proposed model have better performance than other approaches such as Influence Flow Method, MEIBI (Metric for Evaluating and Identifying Blogger’s Influence) and MEIBIX (MEIBI extended). This show the current approach has improved over the previous models.

**5.2 EXPECTED CONTRIBUTION**

There is no doubt influential blogger have contribute greatly to many life in term of changing their way of living, thinking and behaviour by having direct impact on their personal life. The major contribution of the research is that it explore the blogosphere to see how it can improved on the existing model on how to find influential bloggers on the web.

Precisely, the thesis make the following contributions:

* Identified new parameters in finding influential bloggers on the web. Three more parameter were introduced to the existing model. They are (i) comments on a blog, (ii) determine the bloggers social proof by the number of retweets, Facebook, and Google+ shares etc, and (iii) emerging topics on the blogosphere. These collectable statistics in the Blogosphere were used to quantify a blogger’s influence on the web.
* Defined and formulated a mathematical model for the influence of a blogger in terms of the presenting parameters.
* Proposed an algorithm to formulate a program in calculates the influence score of each blogger in a community on the blogosphere.

**REFERENCE**

Adar, E., Zhang, L., Adamic, L., & Lukose, R. (2004). Implicit structure and the dynamics of blogspace. *HP Information Dynamics Lab*, *13*(1), 1–8. https://doi.org/10.1021/jp961373i

Agarwal, N., & Liu, H. (2008). Blogosphere. *ACM SIGKDD Explorations Newsletter*, *10*(1), 18. https://doi.org/10.1145/1412734.1412737

Agarwal, N., & Liu, H. (2009). Modeling and Data Mining in Blogosphere. *Synthesis Lectures on Data Mining and Knowledge Discovery*, *1*(1), 1–109. https://doi.org/10.2200/S00213ED1V01Y200907DMK001

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 (NGDM)*, 0–4.

Agarwal, N., Liu, H., Tang, L., & Yu, P. S. (2008a). Identifying the influential bloggers in a community, *41*(5), 207–218. https://doi.org/10.1145/1341531.1341559

Agarwal, N., Liu, H., Tang, L., & Yu, P. S. (2008b). Identifying the influential bloggers in a community. In *Proceedings of the international conference on Web search and web data mining - WSDM ’08* (p. 207). New York, New York, USA: ACM Press. https://doi.org/10.1145/1341531.1341559

Agarwal, N., Liu, H., Tang, L., & Yu, P. S. (2012). Modeling blogger influence in a community. *Social Network Analysis and Mining*, *2*(2), 139–162. https://doi.org/10.1007/s13278-011-0039-3

Akritidis, L., Katsaros, D., & Bozanis, P. (2009). Identifying influential bloggers: Time does matter. *Proceedings - 2009 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2009*, *1*, 76–83. https://doi.org/10.1109/WI-IAT.2009.18

Akritidis, L., Katsaros, D., & Bozanis, P. (2011). Identifying the Productive and Influential Bloggers in a Community. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *41*(5), 759–764. https://doi.org/10.1109/TSMCC.2010.2099216

Alejandro, J. (2010). Journalism in the Age of Social Media. *Reuters Institute for the Study of Journalism*, 1–47. Retrieved from http://reutersinstitute. politics.ox.ac.uk/publication /journalism-age-social-media

Andzulis, J. “Mick,” Panagopoulos, N. G., & Rapp, A. (2012). A Review of Social Media and Implications for the Sales Process. *Journal of Personal Selling and Sales Management*, *32*(3), 305–316. https://doi.org/10.2753/PSS0885-3134320302

Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, *286*(5439).

Baruah, T. D. (2012). Effectiveness of Social Media as a tool of communication and its potential for technology enabled connections: A micro-level study. *International Journal of Scientific and Research Publications*, *2*(1), 1–10. https://doi.org/ISSN 2250-3153

Benkler, Y. (2006). *Political freedom part 2: Emergence of the networked public sphere*. *The Wealth of Networks: How Social Profduction Transforms Markets and Freedom*. https://doi.org/10.1016/j.infoecopol.2007.03.001

Brin, S., Page, L., Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, *30*(1–7), 107–117. https:// doi.org/10.1016/S0169-7552(98)00110-X

Cai, Y., & Chen, Y. (2010). MASS: a multi-facet domain-specific influential blogger mining system. In *2010 IEEE 26th International Conference on Data Engineering (ICDE 2010)* (pp. 1109–1112). IEEE. https://doi.org/10.1109/ICDE.2010.5447798

Chaffey, D. (2016). Global social media research summary 2016 Our compilation of the latest social media statistics of consumer adoption and usage, 1–13.

Collin, P., Rahilly, K., Richardson, I., & Third, A. (2011). Literature Review: The Benefits of Social Networking Services, (April), 29. Retrieved from http://www.fya.org.au/wp-content/uploads/2010/07/The-Benefits-of-Social-Networking-Services.pdf

Conole, G., Galley, R., & Culver, J. (2011). Frameworks for understanding the nature of interactions, networking, and community in a social networking site for academic practice. *International Review of Research in Open and Distance Learning*, *12*(3), 119–138. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Couldry, N., & van Dijck, J. (2015). Researching Social Media as if the Social Mattered. *Social Media + Society*, *1*(2), 1–7. https://doi.org/10.1177/2056305115604174

Daeid, N. N. (2008). *Educating the next generation*. *Science and Justice* (Vol. 48). https://doi.org/10.1016/j.scijus.2008.03.007

Doctorow, C. (2002). *Essential blogging*. O’Reilly.

Efimova, L., & Hendrick, S. (2005). In search for a virtual settlement : An exploration of weblog community boundaries. *Communities&Technologies’05*. Retrieved from https://doc.novay.nl/dsweb/Get/Document-46041/weblog-community-boundaries.pdf

Farrell, H., & Drezner, D. W. (2007). The power and politics of blogs. *Public Choice*, *134*(1–2), 15–30. https://doi.org/10.1007/s11127-007-9198-1

Farrell, H., & Drezner, D. W. (2008). The power and politics of blogs. *Public Choice*, *134*(1–2), 15–30. https://doi.org/10.1007/s11127-007-9198-1

Gill, K. E. (2004). How can we measure the influence of the blogosphere ? *WWW2004, New York*, 1–8.

Glance, N., Hurst, M., Nigam, K., Siegler, M., Stockton, R., & Tomokiyo, T. (2005). Deriving marketing intelligence from online discussion. *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, 419–428. https://doi.org/10.1145/1081870.1081919

Goldenberg, J., Libai, B., & Muller, E. (2001). Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth. *Marketing Letters*, *12*(3), 211–223. https://doi.org/10.1023/A:1011122126881

Groups, H. (2009). How To Create a Blog Website With WordPress.

Gruhl, D., Guha, R., Kumar, R., Novak, J., & Tomkins, A. (2005). The Predictive Power of Online Chatter. *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, 78–87. https://doi.org/10.1145/1081870.1081883

Gruhl, D., Liben-Nowell, D., Guha, R., & Tomkins, A. (2004). Information diffusion through blogspace. *ACM SIGKDD Explorations Newsletter*, *6*(2), 43–52. https://doi.org/10.1145/1046456.1046462

Huang, C.-Y., Shen, Y.-Z., Lin, H.-X., & Chang, S.-S. (2007). Bloggers’ Motivations and Behaviors: A Model. *Journal of Advertising Research*, *47*(4).

Hubert, A. B., Hubert, A. B., Hubert, T., Mugizi, C., & Rwebangira, R. (2006). A Random-Surfer Web-Graph Model. *IN ANALCO ’06: PROCEEDINGS OF THE EIGTH WORKSHOP ON ALGORITHM ENGINEERING AND EXPERIMENTS AND THE THIRD WORKSHOP ON ANALYTIC ALGORITHMICS AND COMBINATORICS*, 238246. Retrieved from http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.61.2018

Java, A., Kolari, P., Finin, T., & Oates, T. (2006). Modeling the Spread of Influence on the Blogosphere. *Work*, (August 2015), 1–7. Retrieved from http://ebiquity.umbc.edu/\_file\_directory\_/papers/262.pdf

Kaye, B. K. (2007). Web Site Story: An Exploratory Study of Why Weblog Users Say They Use Weblogs. *Blogging, Citizenship and the Future of Media*, (August), 127–148.

Kayes, I., Qian, X., Skvoretz, J., & Iamnitchi, A. (2012). How Influential Are You: Detecting Influential Bloggers in a Blogging Community. *Social Informatics*, 29–42. https://doi.org/10.1007/978-3-642-35386-4\_3

Kayes, I., Zuo, X., Wang, D., & Chakareski, J. (2014). To Blog or Not to Blog: Characterizing and Predicting Retention in Community Blogs. *Proceedings of the 2014 International Conference on Social Computing*, 7:1--7:8. https://doi.org/10.1145/2639968.2640061

Keller, E., & 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, New York*, *25*(5), 1–8. Retrieved from http://www.johnbesaw.com/The\_Influentials.pdf

Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’03* (p. 137). New York, New York, USA: ACM Press. https://doi.org/10.1145/956750.956769

Kleinberg, J. M., & M., J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, *46*(5), 604–632. https://doi.org/10.1145/324133.324140

Kritikopoulos, A., Sideri, M., & Varlamis, I. (2006). BlogRank. In *Proceedings of the 2nd international workshop on Advanced architectures and algorithms for internet delivery and applications - AAA-IDEA ’06* (p. 8). New York, New York, USA: ACM Press. https://doi.org/10.1145/1190183.1190193

Kumar, R., Novak, J., Raghavan, P., & Tomkins, A. (2003). On the bursty evolution of blogspace. In *Proceedings of the twelfth international conference on World Wide Web - WWW ’03* (p. 568). New York, New York, USA: ACM Press. https://doi.org/10.1145/775152.775233

Kumar, R., Raghavan, P., Rajagopalan, S., Tomkins, A., Kumar, R., Raghavan, P., … Tomkins, A. (1999). Trawling the Web for emerging cyber-communities. *Computer Networks*, *31*(11–16), 1481–1493. https://doi.org/10.1016/S1389-1286(99)00040-7

Lee, S., Hwang, T., & Lee, H. (2006). Corporate blogging strategies of the *Fortune* 500 companies. *Management Decision*, *44*(3), 316–334. https://doi.org/10.1108/00251740610656232

Leskovec, J., Mcglohon, M., Faloutsos, C., Glance, N., & Hurst, M. (2006). Cascading Behavior in Large Blog Graphs Patterns and a model Cascading Behavior in Large Blog Graphs Patterns and a model. *Science*, (October).

Lloyd, L., Lloyd, L., Kaulgud, P., Kaulgud, P., Skiena, S., & Skiena, S. (2005). Newspapers vs. blogs: Who gets the scoop. *Proceedings of the AAAI Symp. Computational Approaches to Analysing Weblogs (AAAI-CAAW 2006)*, 8. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Newspapers+vs.+Blogs:+Who+Gets+the+Scoop?#0

Majhi, S. (2010). Innovative Web 2.0 technologies for integrating the learning process. *11th MNLIBNET National Conference*, *Secuderaba*, 324–329. Retrieved from http://hdl.handle.net/10760/15432

Mohan, S., Agarwal, N., & Al-Doski, L. (2013). Mobile network-aware social computing applications: A framework, architecture, and analysis. *Journal of Ambient Intelligence and Humanized Computing*, *4*(1), 43–56. https://doi.org/10.1007/s12652-011-0066-y

Nardi, B. A., Schiano, D. J., Gumbrecht, M., & Swartz, L. (2004). Why we blog. *Communications of the ACM*, *47*(12), 41. https://doi.org/10.1145/1035134.1035163

O’Keeffe, G. S., Clarke-Pearson, K., Council on Communications and Media, C. on C. and, Ito, M., Horst, H., Bittani, M., … Pediatrics, A. A. of. (2011). The impact of social media on children, adolescents, and families. *Pediatrics*, *127*(4), 800–4. https://doi.org/10.1542/peds.2011-0054

Oravec, J. A. (2003). Oravec 2003 Weblogs as an emerging genre in higher education.pdf, *14*(2), 21–44.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The PageRank Citation Ranking: Bringing Order to the Web. *World Wide Web Internet And Web Information Systems*, *54*(1999–66), 1–17. https://doi.org/10.1.1.31.1768

Pal, A., & Kawale, J. (2011). Connecting Mutually Influencing Bloggers. *Artificial Intelligence*, 574–577.

Pennock, D. M., Flake, G. W., Lawrence, S., Glover, E. J., & Giles, C. L. (2002). Winners don’t take all: Characterizing the competition for links on the web. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(8), 5207–11. https://doi.org/10.1073/pnas.032085699

PEW/Internet. (2005). T 14: The State of Blogging. *Director*, 1–4.

Richards, J. (2016). To Blog or not to Blog? *The Open University H800 Module Blog, 23 April*, 7:1--7:8. https://doi.org/10.1145/2639968.2640061

Richardson, M., & Domingos, P. (2002). Mining knowledge-sharing sites for viral marketing. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’02* (p. 61). New York, New York, USA: ACM Press. https://doi.org/10.1145/775047.775057

Robertson, S., Zaragoza, H., & Taylor, M. (2004). Simple BM25 extension to multiple weighted fields. In *Proceedings of the Thirteenth ACM conference on Information and knowledge management - CIKM ’04* (p. 42). New York, New York, USA: ACM Press. https://doi.org/10.1145/1031171.1031181

Rogers, P. C., Liddle, S. W., Chan, P., Doxey, A., & Isom, B. (2007). WEB 2.0 learning platform: Harnessing collective intelligence. *Turkish Online Journal of Distance Education*. https://doi.org/10.17718/tojde.89177

Sensis. (2015). TWEET Sensis Social Media Report May 2015. Retrieved from www.sensis.com.au/socialmediareport

Sheedy, C. S. (2011). Social media for social change: A case study of social media use in the 2011 Egyptian revolution. *Capstone Project*, *28*(4), 1–58.

Shin, W.-Y., Singh, B. C., Cho, J., & Everett, A. M. (2015). A New Understanding of Friendships in Space: Complex Networks Meet Twitter, 1–17. Retrieved from http://arxiv.org/abs/1507.02206

Shola, S. P. (2012). Identifying Influential Bloggers. Retrieved from http://scholarworks.sjsu.edu/etd\_projects

Shullich, R. (2011). Risk Assessment of Social Media. *Global Information Assurance Certification Paper*, 0–45.

Sungkur, R. K., & Rungen, D. (2014). Harnessing Collective Intelligence for the Learning Communities ; a Review of Web 2 . 0 and Social Computing Technologies, *2*(2), 111–129.

Valk, J., & Rashid, A. (2010). Using mobile phones to improve educational outcomes: An analysis of evidence from Asia. *The International Review of Research in*, *11*(1), 13–14. https://doi.org/1492-3831

Veletsianos, G., & Navarrete, C. (2012). Online social networks as formal learning environments: Learner experiences and activities. *The International Review of Research in Open and Distributed Learning*, *13*(1), 144–166.

Watts, D. J. J., & Strogatz, S. H. H. (1998). Collective dynamics of “small-world” networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Wise, E. K., & Shorter, J. D. (2014). Social networking and the exchange of information. *Issues in Information Systems*, *15*(2), 103–109. Retrieved from http://www. tamuk. edu/cba/research/documents/95\_iis\_2014\_103-109.pdf\nhttp:// iacis.org /iis /2014 /95\_ iis\_ 2014 \_103-109.pdf