Assignment 2

UALR

IFSC 4360/5360: Social Computing

Course Instructor: Dr. Nitin Agarwal, Ph.D.

Due Date: 10/05/2023

It's important to be able to discover who influential bloggers are in a blogging community. Throughout most of the articles I read they gave relatively the same reasons as why it's important to figure out who influential bloggers are. These influential bloggers can be thought of as market-movers because their opinions are so well received that they can influence their readers into buying certain products, watching a certain program, as well as eating at a certain place, these people in a sense can be spokesmen for many companies and products (Agarwal et al., 2008). There is also more power that influential bloggers have than selling products, they have the power to sway elections, to influence how others vote (Agarwal et al., 2008). Based on these few reasons given it is easy to see why companies and even political figures would want to know who these influential bloggers are, they can be used to help push their product or agenda.

I began to read these articles in the order I printed them and realized that some would reference studies in previous articles, so instead I sorted them chronologically and read them from oldest to newest. This helped me to compare the articles and see the growth in this research throughout the years. According to Agarwal et al. (2008), their "effort was the first attempt to find influential bloggers." They paved the way in this type of research and others picked up their efforts and added or changed certain parts they believed would be better to find the true influential bloggers.

According to Agarwal et al. (2008), they discovered that active bloggers are not always influential, and influential bloggers are not always active. Because of this they came up with four types of bloggers: "active and influential, active and non-influential, inactive and influential, and inactive and non-influential" (Agarwal et al., 2008, p. 209). They came up with a preliminary model that took statistics that could be collected via blog sites and assigned influence scores

based on their findings (Agarwal et al., 2008). Some of the information that they looked at included: recognition (how many inlinks a post had), activity generation (how many comments a post had), novelty (looking at how many outlinks a post had), and eloquence (length of post) (Agarwal et al., 2008). Through this they came up with the theory that those bloggers who have a high influence score are therefore influential (Agarwal et al., 2008). They recognized that there was room to expand on their findings, "the preliminary model can serve as a baseline in identifying influential bloggers and can be extended by incorporating additional parameters to discover different patterns" (Agarwal et al., 2008, p. 216).

Akritidis et al. (2009), felt like the existing models for finding influential bloggers were missing a key aspect. They felt that the temporal aspect needed to be considered, because "time is the most critical aspect of the blogosphere" (Akritidis et al., 2009, p.76). Akritidis et al. (2009) further explains why they feel time is an important aspect to consider when trying to identify influential bloggers:

> The temporal dimension is of crucial importance for identifying influentials. The time is related to the age of a blog post and also to the age of the incoming links to that post. An influential is recognized as such if s/he has written influential posts recently or if its posts have an impact recently. (p.78)

By looking at previous work, and with their new ideas they decided on the factors they wanted to use to measure a blogger's influence (Akritidis et al., 2009). Akritidis et al. (2009) decided that high amounts of incoming links, and number of comments made to a post both were signs of influence, just like Agarwal et al. (2008) had before. They also agree that productivity was a strong indication of influence, and this needed to be considered (Akritidis et al., 2009). Based on all the factors they came up with they developed a formula to estimate the influence of

a blog post (Akritidis et al., 2009). “An issue being discussed in a blog post at the present time and is now of major importance, may be totally outdated after two months” (Akritidis et al., 2009, p.78). So, they accounted for this, with a formula and using this they came up with MEIBI (Metric for Evaluating and Identifying a Bloggers Influence) (Akritidis et al., 2009). According to Akritidis et al. (2009):

> An old post may still be influential. How could we deduce this? Only if we examine the age of the incoming links to this post. If a post is not cited anymore, it is an indication that it negotiates outdated topics or proposes outdated solutions. On the other, if an old post continues to be linked to presently, then this is an indication that it contains influential material… Instead of assigning to a bloggers old post smaller scores depending in their age, we can assign to each incoming link of a blogger’s post a smaller weight depending on the link’s age. (p. 78-79)

“The obtained results attested that the new methods are able to better identify significant temporal patterns in the blogging behavior, and reveal some latent facts about the blogging activity” (Akritidis et al., 2009).

The article “Identifying the Productive and Influential Bloggers in a community,” is the updated work of Akritidis et al. (2009), in this article they point out the weakness of their previous proposed models, MEIBI and MEIBIX. “These metrics assign a unique value to each blogger and they cannot provide a straightforward separation between influence and productivity” (Akritidis et al., 2011, p.759). They create mathematical equations to represent the different aspects of a blogger’s influence (Akritidis et al., 2011). Like previous studies before it, Akritidis et al. (2011) agree that incoming links to a blog post, as well as number of comments

on a post each point to a higher level of influence (Akritidis et al., 2011). The two metrics that Akritidis et al. (2011) introduces are:

> BP-index, [which] is used to evaluate the productivity of a blogger with respect to recency. The second metric, BI-index reflects the influence of a blogger inside and outside a community by taking into consideration the number and the age of the incoming links and the comments. The combination of these two values is then used to characterize the bloggers. Hence, a blogger can be characterized as recently influential or recently productive, or both or none. (p.763)

"Modeling Blogger Influence in a Community", by Agarwal et al. (2011) seems to be a continuation of their work from the previous article they did, "Identifying the Influential Bloggers in a Community". They essentially pick up their work and improve and expand upon it. According to Agarwal et al. (2011),

> The primary focus of this [new] work is to identify such influential bloggers of a community blog that could be considered as representatives by first identifying influential blog posts. This characteristic property of the bloggers is evaluated by studying certain indicators that assist in quantifying community's reactions towards bloggers' postings. (p. 140)

Based on their previous work, and the work of others they propose a model, called iFinder, that utilizes similar aspects of their previous work: inlinks, outlinks, comments, and blog post length (Agarwal et al., 2011). "iFinder starts with little knowledge and with each iteration tries to improve the knowledge about the influence of blog posts until it reaches a stable state or a fixed number of iterations specified a priori" (Agarwal et al., 2011, p.145). They were able to expand

on their previous work and make a model that was more effective in identifying influential bloggers in a blogging community (Agarwal et al., 2011). This method can be utilized on other blog sites if the needed information is able to be retrieved (Agarwal et al., 2011). They admit the limitations of their work is to community blogs alone, and there are many individual blogs that their module won't work on, they would eventually like to expand their work to include individual blogs as well (Agarwal et al., 2011).

Aziz & Rafi (2010), "There are many researchers that proposed influential bloggers mining systems, but all these systems suffer from drawbacks like: domain driven, generalized shallow influential measure and validation and verification" (p. 1). Aziz & Rafi's (2010) idea was to use certain influence measuring factors, like factors "based on contents semantics of their blog-posts and quantitative analysis of the contents and fellow readerships with their comments on the post" (p.2). The algorithm they created was known as SIIB which stands for Semantically Identified Influential Bloggers (Aziz & Rafi, 2010). The algorithm works by first identifying distinct bloggers from the data set, then utilizing sentence-based similarity measure, Semantic similarity measure, sentence-wise length, and influence flow they measure the influence of each individual blogger (Aziz & Rafi, 2010). Based on their work they were able to identify influential bloggers, they also admit that that by adding comments into the measuring factors could further expand their work, making it more accurate (Aziz & Rafi, 2010).

In the article "How Influential Are You: Detecting Influential Bloggers in a Blogging Community", Kaye et al. (2012) takes a different approach than any of the other articles I have wrote about thus far. Instead of measuring factors about the blog articles themselves, they measure influence based on the bloggers position in the blogosphere (Kaye et al., 2012). "Based on previous research that correlated a node's position on the network to its influence, we

conjecture that the influence of a blogger is represented by its location in the blogging network" (Kaye et al., 2012, p.30). Kaye et al. (2012) utilizes centrality metrics to find influential bloggers. "We propose to aggregate different representative centrality metrics into an influence score" (Kaye et al., 2012 p.30). Kaye et al. (2012) selected 6 representative centrality metrics as the focus of their study: degree (how many links a node has), betweenness (measures the extent to which a node lies on the shortest path between other nodes), closeness (mean distance from a node to other nodes), eigenvector (looks at the influence of the nodes neighbors), hub (the nodes point to other nodes) and communicability (the sum of closed walks of all lengths starting and ending at a node) centrality. By measuring these centrality metrics, they found that "some bloggers span significant influence on fellow bloggers due to their strategic location in the network" (Kaye et al., 2012, p.30). They also discovered that "Influential bloggers form a densely connected core, while non-influential bloggers remain at the periphery of the network, less likely to connect to each other" (Kaye et al., 2012, p.30).

Moh & Shola (2013) use similar methods of measurement that Agarwal et al. (2008) and Akritidis et al. (2009) used. But they feel they can find a better model to identify influential bloggers. Moh and Shola (2013) introduce:

> two new factors: uniqueness and FacebookCount. The former measures the originality of a post, combining with outlink count they represent the novelty of the post. The latter reflects the influence of the emerging social network platforms, and can be extended to include twitter share, G +1's, etc. The proposed model also adopts other effectual factors including the number of inlinks, outlinks, and the comments, the timing of the posts and comments, and the influence of commenters. In addition, to capture the true influence of

> a post we mine through each comment on a post to identify the sentiment, or the tone of the comment. (p. 18)

The influential factor that Moh and Shola (2013) measure known as FB score and inlink score, is like past models in that it considers the amount of inlinks a post has, the new aspect they bring to the table is that they measure how many Facebook shares and likes a post received. Another measurement that Moh and Shola (2013) utilize is sentimental factor to look at the comments, they feel that not only do the number of comments count, but they feel the emotion of the comment is even more important, whether it is a positive, negative, or neutral comment. Not only do they feel the emotions of the comments are important, but they also measure the influence of the commenters, so if an expert comments on your post, that means it's more influential (Moh and Shola, 2013). Also, like others before them they realize how important time is in the blogosphere (Moh and Shola, 2013). The final measurement of Moh and shola (2013) is uniqueness and outlinks, they recognize like the Influence Flow Method the importance of novelty, but they use the TFIDF machine learning method to determine a blogs uniqueness. Unlike some of the other studies Moh and Shola (2013) do not feel that length of a post is a good indicator of the quality of a blog post. Moh and Shola (2013) state that the results from their new and improved model, "have shown that [their] proposed model is able to effectively rank individual post bloggers with distinct scores, and has achieved a consistent improvement over all stages" (pg. 26).

The article by Khan et al. (2015), "MIIB: A Metric to Identify Top Influential Bloggers in a community" presents a new metric, MIIB (Metric for Identification of Influential Bloggers), based on various features of bloggers' productivity and popularity. "Productivity refers to bloggers' blogging activity and popularity measures bloggers' influence in the blogging

community" (Khan et al., 2015). "The task is to find the top influential bloggers based on certain features which are related to bloggers, such as the ability to create new blogs, and blogs such as how many posts are there, how many users post their content etc" (Khan et al., 2015, p.4). The weights assigned to the features depict the significance of the features" (Khan et al., 2015, p.4). Khan et al. (2015) used the works that had come before and expanded upon it, the features they believed were important to measure are: Activity (how many blogs a blogger post), Activeness (total number of days a blogger is active), Consistency (measure that the blogger post on a regular basis), Recognition (number of comments on a blog), Authority (number of inlinks), Novelty (uniqueness of post), BlogRank (bloggers are more influential if they post on higher ranking blogs), PostLength (eloquence), and NormalizedPostLength (takes the sum of length of posts, and divides by number of posts of a blogger). According to Khan et al. (2015), modules of MIIB are Productivity, popularity and BlogRank. "The score of each module is calculated separately and each feature is given a certain weight" (Khan et al., 2015). At the time of its writing MIIB compared to the other models was able to "identify the influential bloggers in a more effective manner" (Khan et al., 2015, p.14).

Ishfaq et al. (2017) felt that all the other proposed models to find influential blogger fail "to consider the sentiment expressed in the blog posts of a blogger" (p. 506). Ishfaq et al. (2017) again looked at previous models and saw what worked, and brought those aspects into their model, such as number of blog posts of a blogger, avg length of post, number of inlinks & outlinks, number of comments. The two new parts they have incorporated are the sentiment analysis of the blog itself, some earlier methods looked at the sentiment of comments, but not the post itself. They also look at the number of distinct commenters, which I don't remember reading

any other research team incorporating this into their model. The results from utilizing the model created by Ishfaq et al. (2017)

> "prove that the modules play a significant role in identifying the top influential bloggers with higher accuracy than the baseline metrics and the novel feature, Sentiment analysis of the blog posts, effectively captures the blogger productivity by taking into account the (non-sentimental) objective information a blogger writes in the blog posts. Therefore, sentiment analysis should be considered an important feature in identifying the influential bloggers." (p. 520-521)

Article, "Modelling to identify influential bloggers in the blogosphere: A Survey", was different than any of the other papers I read. All the other articles I read were about a new approach to find influential bloggers, this article was an overview of all the models that were out at the time of its writing. Khan et al. (2016) classifies the different models into feature-based and network-based categories. "The feature-based models consider the salient factors to measure bloggers' influence. The network models, on the other hand, consider the graph-based social network structure of the bloggers to identify those who have the most impact on fellow members" (Khan et al., 2015, p.64). This article lists many of the models that I read about during this assignment such as: iIndex model, iFinder, MIIB, MEIBI and MEIBIX, and many more. There were a lot listed that I had not read of before such as: WisClus, CR algorithm, MIV, TDIR and more. This article is helpful to see all the research that was available at the time regarding all the different methods that had been implemented in trying to figure out who the most influential blogger was. This article gives a brief description of each method, and sometimes these brief descriptions are easier to understand than a complex paper is.

All the articles that I read had similar approaches to figure out who the influential bloggers were, the aspects that they measured varied slightly, with each model trying to improve upon the earlier models. Each research team had the same goal, to come up with a model to be used to figure out who the influential bloggers were. Each team seemed to come up with a more accurate model than the previous one, but they all seem to recognize their model had room to be improved upon.

References

Agarwal, N., Liu, H., Tang, L., & Yu, P. S. (2008). Identifying the influential bloggers in a community. *Proceedings of the International Conference on Web Search and Web Data Mining - WSDM '08*. https://doi.org/10.1145/1341531.1341559

Agarwal, N., Liu, H., Tang, L., & Yu, P. S. (2011). 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. *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*. 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

Aziz, M., & Rafi, M. (2010). Identifying influential bloggers using blogs semantics. *Proceedings of the 8th International Conference on Frontiers of Information Technology*. https://doi.org/10.1145/1943628.1943635

Ishfaq, U., Khan, H. U., & Iqbal, K. (2017). Identifying the Influential Bloggers: A Modular Approach Based on Sentiment Analysis. *Journal of Web Engineering*, *16*, 505–523.

Kayes, I., Qian, X., Skvoretz, J., & Iamnitchi, A. (2012). How influential are you: Detecting influential bloggers in a blogging community. *Lecture Notes in Computer Science*, 29–42. https://doi.org/10.1007/978-3-642-35386-4_3

Khan, H. U., Daud, A., & Malik, T. A. (2015). MIIB: A metric to identify top influential bloggers in a community. *PLOS ONE*, *10*(9). https://doi.org/10.1371/journal.pone.0138359

Khan, H. U., Daud, A., Ishfaq, U., Amjad, T., Aljohani, N., Abbasi, R. A., & Alowibdi, J. S. (2017). Modelling to identify influential bloggers in the blogosphere: A survey. *Computers in Human Behavior*, *68*, 64–82. https://doi.org/10.1016/j.chb.2016.11.012

Moh, T.-S., & Shola, S. P. (2013). New factors for identifying influential bloggers. *2013 IEEE International Conference on Big Data*. https://doi.org/10.1109/bigdata.2013.6691792