Title	Influence diffusion detection using the influence style (INFUSE) model
Author(s)	Tan, Luke Kien-Weng; Na, Jin-Cheon; Ding, Ying
Citation	Tan, L. K.-W., Na, J.-C., & Ding, Y. (2015). Influence diffusion detection using the influence style (INFUSE) model. Journal of the Association for Information Science and Technology, 66(8), 1717-1733.
Date	2015
URL	http://hdl.handle.net/10220/40481
Rights	© 2015 ASIS&T. This is the author created version of a work that has been peer reviewed and accepted for publication by Journal of the Association for Information Science and Technology, ASIS&T. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [http://dx.doi.org/10.1002/asi.23287].

Influence Diffusion Detection Using the Influence Style (INFUSE) Model

Luke Kien-Weng Tan

Wee Kim Wee School of Communication and Information, Nanyang Technological University, 31 Nanyang Link, Singapore 637718. Telephone: (65) 6790-5844, Fax: (65) 6791-5214, Email: w080078@ntu.edu.sg

Jin-Cheon Na

Wee Kim Wee School of Communication and Information, Nanyang Technological University, 31 Nanyang Link, Singapore 637718. Telephone: (65) 6790-5011, Fax: (65) 6791-5214, Email: tjcna@ntu.edu.sg

Ying Ding

School of Informatics and Computing, Indiana University, 1320 E. 10th ST., LI 025, Bloomington, IN 47405, Telephone: (812) 855-5388, Fax: (812) 855-6166, Email: dingying@indiana.edu

Blogs are readily available sources of opinions and sentiments that in turn could influence the opinions of the blog readers. Previous studies had attempted to infer influence from blog features, but they ignored the possible influence styles that describe the different ways or manner in which influence is exerted. In this paper, we propose a novel approach of analyzing bloggers' influence styles, and using the influence styles as features to improve the performance of influence diffusion detection between linked bloggers. The proposed influence style (INFUSE) model describes bloggers' influence through their engagement style, persuasion style, and persona. Methods used include similarity analysis to detect the creating-sharing aspect of engagement style, subjectivity analysis to measure persuasion style, and sentiment analysis to identify persona style. We further extend the INFUSE model to detect influence diffusion between linked bloggers based on the bloggers' influence styles. The INFUSE model performed well with an average F1-score of 76% when compared to the in-degree and sentiment-value baseline approaches. While previous studies had focused on the existence of influence between linked bloggers in detecting influence diffusion, our INFUSE model is shown to provide a fine-grained description of the manner in which influence is diffused based on the bloggers' influence styles.

Introduction

The advent of Web 2.0 has led to an increase in user-generated content on the Web such as Blogs. A blog or weblog is a specialized web site that allows an individual or a group of individuals to express their thoughts, voice their opinions, and share their experiences and ideas. The entries are called blog posts, while the individuals who authored the blog posts are referred to as bloggers. A blog site could contain a single blogger and is referred as a personal blog, while a blog site containing a group of bloggers is known as a multi-author blog. Multi-author blogs have certain characteristics amongst their bloggers such as common shared goals and guidelines, and the editor would guide and organize the activities of the blog. The reasons for having multiple authors on a blog are the wider variety of opinions and ideas that could cater to more readers. Nevertheless, the greater variety of opinion expressions could lead to differences in the bloggers' influence styles. Influence style refers to the manner in which the blogger exerts influence through the blog postings and could be described based on how the blogger engages the readers, the persuasion style of the blogger, or whether the blogger's posts are received positively or negatively by the readers. For example, a linking blogger could exhibit a sharing style in engaging the readers through posting similar content from the linked blog post as seen in Figure 1. To illustrate persuasion as an influence style, the blogger of the blog post in Figure 2 exerts influence in an opinionated manner as seen from the numerous subjective words and phrases used in the post compared to the minimal use of subjective expressions by the blogger of the blog post in Figure 3, indicating a more objective persuasion style. In this paper, we propose the INFlUence StylE (INFUSE) model with the aim to analyze bloggers' influence styles and use the influence styles as features to improve the performance of influence diffusion between linked bloggers. We define the influence between bloggers as consisting of three possible styles: engagement style, persuasion style, and persona style. Engagement style indicates the frequency, scope, originality, and consistency of the bloggers in their postings. Persuasion style refers to appeals to reasons or emotions displayed in the bloggers' posts, and persona is the degree of compliance shown by linking blog posts towards the bloggers.

Phil Schiller Interview on NFC, wireless charging, and the Lightning dock connector for many years to come
by Seth Weintraub posted Sep 12th, 2012 at 1:47 PM

AllThingsD got a few words in with Apple Senior Vice President of Marketing Phil Schiller. The Apple keynoter extraordinaire defended the decision to keep Passbook NFC-free, noting it had all of the functionality needed.
"It's not clear that NFC is the solution to any current problem, Schiller said. "Passbook does the kinds of things customers need today." As for wireless charging, Schiller notes that the wireless charging systems still have to be plugged into the wall, so it's not clear how much convenience they add ……That said, Schiller said that Apple doesn't take changing the connector lightly. "This is the new connector for many years to come," he said."

Retrieved from http://9to5mac.com/2012/09/12/phil-schiller-interview-on-nfc-wireless-charging-and-the-lightning-dock-connector-for-many-years-to-come/on 1st May 2013.

Interview: Phil Schiller on why the iPhone 5 has a new connector but not NFC or Wireless Charging
by Ina Fried posted Sep 12th, 2012 at 1:36 PM

While apple managed to pack a bunch of new technologies into the iPhone 5 ……

It's not clear that NFC is the solution to any current problem, Schiller said. "Passbook does the kinds of things customers need today." As for wireless charging, Schiller notes that the wireless charging systems still have to be plugged into the wall, so it's not clear how much convenience they add…… That said, Schiller said that Apple doesn't take changing the connector lightly. "This is the new connector for many years to come," he said.

Retrieved from http://allthingsd.com/20120912/interview-phil-schiller-on-why-the-iphone-5-has-a-new-connector-but-not-nfc-or-wireless-charging/ on 1st May 2013.

FIG. 1. Blogger of linking blog post exhibiting a sharing style through posting similar content from the linked blog post.

Skype scores an iOS update with improved calling UI and some bug fixes *by Terrence O'Brien posted Mar 7th, 2013 at 5:31 PM*

Not every update from the Microsoft-owned VoIP service needs to be high-profile. Every so often a ***nice subtle*** tweak is all it takes to add a ***lovely layer of polish*** to an already ***beloved*** app. Both the iPad and iPhone versions of Skype were ***bumped*** to version 4.6 today. There are, of course, a number of bug fixes included in the update…… ***Hit up*** the iTunes app store now for your update and check out the source for a complete changelog.

Retrieved from http://www.engadget.com/2013/03/07/ skype-scores-an-ios-update on 1st May 2013.

FIG. 2. Blog post showing subjective persuasion influence style

Skype for iPhone and iPad updated with improved calling interface, general improvements
Written by: MARK GURMAN, March 7, 2013 / 9:10 am

Today, Skype has updated both its iPhone/iPod touch and iPad applications with an improved calling user-interface. Additionally, the company has added a new ***handy*** feature to mark all recent text-based conversations as read in a ***quicker*** fashion……

Retrieved from http://9to5mac.com/2013/03/07/skype-for-iphone-and-ipad-updated-with-improved-calling-interface-general-improvements/ on 1st May 2013.

FIG. 3. Blog post showing objective persuasion influence style

Blog posts are readily available sources of opinions and sentiments that in turn could influence the opinions of blog readers. The importance of understanding blog influence has led to increasing influence analysis research in the blogosphere digital space (Agarwal & Liu, 2009; Leskovec, Huttenlocher, & Kleinberg, 2010). Previous studies had proposed using blog features (Adar & Adamic, 2005; Agarwal & Liu, 2008), similarity comparison (Matsumura, Yamamoto, & Tomozawa, 2008; Song, Chi, Hino, & Tseng, 2007), and community detection (Agarwal, Liu, Tang, & Yu, 2008; Ghosh & Lerman, 2008) to detect blog influence. However, the studies had not considered the context of the blog posts and thus, would not measure influence accurately as influence is subjective in nature and is often dependent on the contextual details. Context in this case refers to the sentiments expressed on the topics in the blog posts as well as the influence styles exerted by the linked blogger. For example, a blog-feature based study would have used the in-links to blog post A to indicate the presence of influence flow from A to the other blog posts as shown in Figure 4. However, a blog post with many in-links may not necessarily indicate the blogger is influential as the linking posts could contain disagreement towards the blogger as shown in Figure 5 where negative sentiments expressed between blog post A and B could show non-influence flow despite the existence of the link between them. Recent studies had further considered the context of blog posts by analyzing the sentiments between linked blog posts (Leskovec, Huttenlocher, & Kleinberg, 2010; Cai, Bao, Yang, Tang, Ma, Zhang, & Su, 2011; Li, Bhowmick, & Sun, 2011). These studies focused on the concept that influence exists between blogs with positive sentiments, but ignored the details in the manner in which influence could be exerted. On the other hand, a detailed description of the linked bloggers' influence style would have provided a clearer understanding of how influence is propagated. Our study further identified the influence styles of bloggers to better understand the influence exerted by the bloggers in detecting influence diffusion. This can be seen in Figure 6 where the author of blog post A is described to be participating through frequent postings, focused on specific topics, creating based on originality in postings, and consistently linked throughout a given period in engaging the readers; objective in persuasion style; and having a persona that is positively received by readers. With the identified influence styles, we could further understand the manner in which influence flow occurs between the blogs. We further extend the INFUSE model based on the Independent Cascade (IC) (Goldenberg, Libai, & Muller, 2001) and Linear Threshold (LT) (Granovetter, 1978) approaches to detect influence diffusion between linked bloggers.

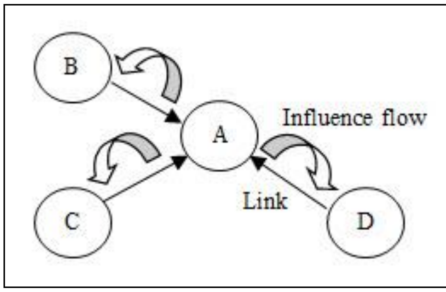

FIG. 4. Link frequency approach.

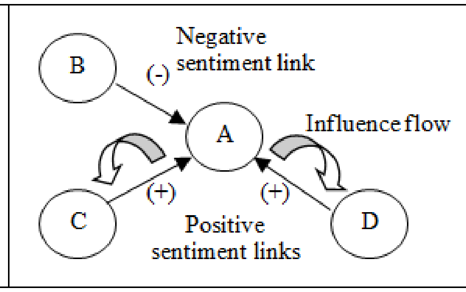

FIG. 5. Sentiment links approach.

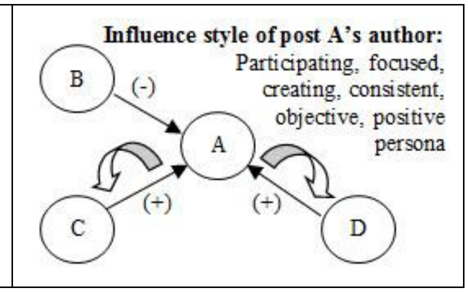

FIG. 6. Proposed Sentiment links and influence style approach.

The extended models INFUSE-IC and INFUSE-LT, based on the IC approach and LT approach respectively, were evaluated on a real-life blog posts dataset, and the results show that the additional influence style features could better fit the influence characteristics of the bloggers, and hence improve influence diffusion detection performance. Our study focuses on the influence diffusion between immediate linked bloggers due to the limitation of the coded data. This study extends the work of (Tan, Na, & Theng, 2013) where we further study the use of Linear Threshold approach in the INFUSE model to detect influence diffusion amongst the linked bloggers.

The next section describes related work followed by the research design where details of the INFUSE model are given. Next, we present the experiment procedures and results, followed by the discussion and conclusion.

Literature Review

Previous studies had attempted to detect influence using various approaches, but did not provide details on the influence types and styles of the bloggers in the influence diffusion process which would have provided a clearer understanding of the influence diffusion between the linked bloggers. Adar & Adamic (2005) related the number of blog links to influence and applied link inference techniques to find implicit graph links that could further identify influence. In the study by Agarwal & Liu (2008), an influential blogger was defined based on the number of in-links, comments for the posts, and the number of out-links. A high authority value based on a large number of in-links to the blog postings indicates high number of followers. However, having many in-links may not necessary indicate influence. For example, a linking blog post opposing the ideas and opinion of a high authority blog post could express disapproval and is not influenced while connected to the linked blog post.

Content similarity was used in (Song, Chi, Hino, & Tseng, 2007; Agarwal & Liu, 2009) to detect influence in the blogosphere, where the approaches used involved measuring document similarity between linked blog posts using techniques, such as cosine similarity. These approaches mainly checked whether the linked blog posts were discussing the same topics as part of the influence detection method. An InfluenceRank algorithm was proposed in (Song et al., 2007), which uses information novelty to rank blogs according to how influential and informative they are in the network. Using cosine similarity distance, information novelty was measured based on how dissimilar the linked blog posts were from one another. Agarwal & Liu (2009) analyzed the blog content for similarity to find influential blogs that were representative and diverse. However, content similarity is not a good measurement of direct influence as the in-link blog author could replicate similar content as reference, but express contrasting opinions on the common content. In our study, similarity comparison is used to determine the engagement style of bloggers based on the level of content creation or sharing, and not to detect influence directly.

Community identification was also used to detect influence in the blogosphere based on the assumption that blogs within the same community would influence each other as close affiliates would. Ghosh & Lerman (2008) generalized the notion of network connectivity to the number of paths that exist between two nodes and relates the network connectivity to influence in the blogosphere. Agarwal, Liu, Tang, & Yu (2008) assumed that influential bloggers could be identified within a community and proposed to detect blog communities through aggregating similar individual blogs. However, community identity alone would not be able to accurately detect influence or detect all cases of influence because blogs within a community could differ in opinions on certain topics, while blogs outside of the community could still share similar opinions.

While the above influence detection studies had ignored the details on bloggers' influence styles, our work studies bloggers' traits in relation to their influence style to determine their ability to influence bloggers. Previous studies (Guadagno, Okdie, & Eno, 2008; Yarkoni, 2010) had measured bloggers' characteristic and personality with regards to their propensity to blog, but they did not consider the bloggers' ability to influence. Guadagno et al. (2008) measured the personality of bloggers to predict blogging based on the five key personality traits: *neuroticism*, *extra-version*, *agreeableness*, *openness to experience*, and *conscientiousness* as observed by Costa & McCrae (1992). The study indicated that people who are high in openness to new experience and high tendency to experience distress (neuroticism) are likely to be bloggers. The results indicated that personality factors impacted the likelihood of being a blogger and have implications for understanding who blogs. Similarly, Yarkoni (2010) analyzed blogger personality on word use in their blog postings to study the association between personality and language of bloggers, and did not studied into the manner in which bloggers exert their influence.

Our study relates bloggers' influence style in terms of their engagement, persuasiveness, and persona. Characteristics used in our study to evaluate the engagement style include authority, reciprocity, commitment and consistency, and scarcity. Cialdini (2001) defined several human characteristics in relation to influence. *Authority* refers to subject experts in our context. *Reciprocity* is the behavior whereby people tend to return a favor. People are usually *committed* and *consistent*, and do not like to be self-contradictory. Once they commit to an idea or behavior, they are averse to changing their minds without good reason. In addition, s*carcity* of resources will generate demand.

Brehm (1966) observed that emotion and disposition may affect likelihood of conformity or anti-conformity. The arousal and affective states have an effect of discrete emotions on targets' cognitions as well as on the eventual outcome of the influence attempt (Cialdini & Goldstein, 2004). Persuasion is a process of influence through appeals to reason or appeals to emotion. Subjectivity, that is, opinionated or emotional phrases expressed in the blog contents could provide clear evidence of the blogger's persuasion style. Our study considers persuasion as a component of influence and evaluates it through analyzing the subjectivity expressed in the blog posts content.

Influence is often related to conformity, which is the act of matching attitudes, beliefs, and behaviors to group norms (Hogg & Vaughan, 2005). Kelman (1958) identified conformity as form of compliance, which is the expression of agreement towards the people or groups who are influential to the individual, while possibly keeping one's own original beliefs. In our study, we measure compliance through analyzing similar sentiments on target topics in the textual content between the linked blog posts. Milgram (1963) observed that social influence is the strongest when the group perpetrating it is consistent and committed. Even a single instance of dissent can greatly wane the strength of an influence. This means that the detection of both agreement (positive sentiments) and disagreement (negative sentiments) expressed in the content is equally important in influence analysis. Leskovec, Huttenlocher, & Kleinberg (2010) adapted a framework of trust and distrust in an attempt to infer the attitude directed towards each other using the observed positive and negative relations. Cai, Bao, Yang, Tang, Ma, Zhang, & Su (2011) defined social influence through positive and negative social relationships and Li, Bhowmick, & Sun (2011) similarly considered the positive and negative edges of the nodes, and computed the influence index and conformity index in their attempt to detect influence in the social network. However, these studies had focused on detecting the presence of influence between the linked bloggers, and had not studied in details the manner in which the bloggers could exert the influence.

Granovetter (1973) and Krackhardt (1992) observed that the effects of the social influence by users of varying personae are different. The three general types of persona identified were positive, negative, and controversial. We identify the bloggers' persona through detecting sentiments expressed on common target topics between the linked blog posts. Previous sentiment analysis studies had used a linguistic approach, in consideration of the linguistic rules found in phrases and sentences to analyze sentiments. These studies had leveraged on the semantic dependencies between words to predict sentiments (Moilanen & Pulman, 2007; Shaikh, Prendinger, & Ishizuka, 2008). A common approach to derive the semantic and syntactic dependencies between words is through typed dependency parsing (Thet, Na, & Khoo, 2010; Wilson, Wiebe, & Hoffmann, 2009; Hassan, Qazvinian, & Radev, 2010). An example of a parsed typed dependency is the adjectival relationship between the adjective "nice" and the noun "phone". Thet et al. (2010) proposed a linguistic approach of computing the sentiment of dependency structure of the clause using typed dependency polarity pattern rules. As an example, in the adjectival typed dependency polarity pattern rule "AMOD(help:[+], tremendous:[+])→[intensified+]", the adjective "tremendous" in the phrase "tremendous help" intensifies the positive noun "help". Thus, the adjectival phrase "tremendous help" becomes more positive than the noun "help" itself. A linguistic approach which considers the grammatical relations between words was used in our study to predict the sentiments expressed on topics discussed in the blog posts in the persona evaluation.

Our study adopted the Independent Cascade (IC) (Goldenberg et al., 2001) and Linear Threshold (LT) (Granovetter, 1978) approaches to detect influence diffusion between linked bloggers. In the IC model, each edge could independently activate their neighboring nodes based on an activation probability. While in the LT model, each edge has a weight and each vertex has a threshold value such that the vertex is activated if the weighted sum of its active neighbors exceeds its threshold. The two models characterize two different aspects of social interaction. The IC approach models individual and independent interaction and influence among linked nodes. On the other hand, the LT model focuses on the threshold influence probability in influence propagation, where consideration is given to all linked neighbors in influencing the node.

Both the IC and LT models leverage on the influence probability between the linked bloggers to predict influence diffusion. Earlier works (Kimura & Saito, 2006; Chen, Yuan, & Zhang, 2010) had not considered how to obtain the diffusion probability values which were often randomly assigned an identical value to all edges within a network for analysis. More recent studies (Xiang, Neville, & Rogati, 2010; Saito, Nakano, & Kimura, 2008; Goyal, Bonchi, & Lakshmanan, 2010; Lim, Kim, Kim, & Park, 2011) had attempted to derive the values of diffusion probabilities to improve the performance of influence diffusion detection. Xiang et al. (2010) observed that both strong and weak ties exist between nodes in a network, and treating all relationships equally will increase the level of noise in the learned models and likely to degrade performance.

On the other hand, performance should improve through better understanding and further differentiation of the links between the bloggers. Saito et al. (2008) presented a method based on the Expectation Maximization algorithm to obtain the set of information diffusion probabilities which maximizes the objective function that maps the probability of a node becoming active. Goyal et al. (2010) proposed a model that directly leverages available propagation traces to learn how influence flows in the network and used this to estimate expected influence spread under the General Threshold model. Lim et al. (2011) used the Independent Cascade model to explain information diffusion in blog network, where the diffusion probability assigned to each edge between a pair of bloggers in the blog network was derived from the number of posts, comments, trackbacks, and scraps. While each of the above studies had considered influence probabilities in their own aspect, none had considered the influence styles of bloggers in computing the influence probabilities, which would have provided a clearer understanding of how influence is propagated.

Research Design

Influence Style Model

In our proposed model, influence is further described in terms of *Engagement style*, *Persuasion style,* and *Persona style* in profiling the individual bloggers. An overview of the INFUSE model with descriptions of the influence styles and types, together with the corresponding models is shown in Table 1.

Engagement Style. We adapted the four general influence types from the Klout Influence Matrix[1], which include *Participating-Listening*, *Broad-Focused*, *Creating-Sharing,* and *Consistent-Casual* to describe the engagement style of the bloggers. Engagement style refers to the frequency, scope, originality, and consistency in which the influencer presents the blog post content. For example in the Klout Influence Matrix, a Broadcaster is an essential source of information with an audience that is wide and diverse. Cialdini (2001) observed there is a propensity for people to comply with authority, and related the authority of an influencer to the level of expertise on the subject domain. Blogger's authority could be derived from the depth and extent of dominance in the target topics, measured through the *Participating-Listening, Broad-Focused,* and *Consistent-Casual* type respectively in our study. The *Participating-Listening* type describes the bloggers' participation level in posting specific target topics. The listening type usually do not post much, but rather follow the blog posts with a low level of posting activities, while a participating type would actively post articles on target topics and share information readily. The *Broad-Focused* type describes the scope of the topics discussed in the blog posts. A broad type blogger would post a wide range of topics, and a focused type would concentrate on specific topics which they are domain experts. Cialdini (2001) further proposed that a perceived limitation or scarcity of resources will generate demand, which means that a more focused scope indicates higher dominance in the target topics. *Creating-Sharing* type describes the originality of the blog posts' content. A sharing blogger tends to restate the content of the linked blog posts. On the other hand, a creating blogger posts original content. As stated in Cialdini (2001) that a scarcity of resources will generate demand, this means that original blog post content would indicate higher dominance in the target topics. The objective consensus approach by (Mackie, 1987) states that individuals are more likely to systematically process a majority-endorsed message because people assume that the majority view reflects reality and people believe they share similar attitudes to the members of the majority. This means that bloggers would also tend to share blog post content of influential blogs. In our model, we consider the originality of the bloggers through their *Creating-Sharing* style. For the *Consistent-Casual* style, the *consistent* type refers to bloggers that could continually attract in-links over a timeline, whereas a *casual* blogger is irregular or intermittent in their number of in-links. Cialdini and Trost (1998) stated that individuals are driven to be consistent and do not like to be self-contradictory. Once they commit to an idea or behavior, they are averse to changing their minds without good reason. Hence, the extent to which one's commitments are made actively could be a determinant of the likelihood of compliance, which our model measures through the consistency of the target blogger's ability to garner in-links over a timeline.

Persuasion Style. Persuasion is a process of influence through appeals to reason or appeals to emotion. Brehm (1966) observed that emotion and disposition may affect likelihood of conformity or anti-conformity, while Cialdini and Goldstein (2004) observed that the arousal and affective states have an effect of discrete emotions on targets' cognitions as well as on the eventual outcome of the influence attempt. *Subjectivity*, that is, opinionated or emotional phrases expressed in the blog content could provide clear evidence of the blogger's persuasion style. Our study considers persuasion as a component of influence and evaluates it through analyzing the subjectivity expressed in the blog posts content. A *subjective* blogger is opinionated and expressive in stating his or her viewpoints. On the other hand, an *objective* blogger expresses his or her views through stating facts and reasons.

---

[1] http://klout.com

Persona Style. Kelman (1958) defined c*ompliance*, a form of conformity as the expression of agreement towards the people or groups who are influential to the individual, while possibly keeping one's own original beliefs. The degree of compliance towards the blogger is measured based on the persona of the blogs or bloggers in our model. *Persona* style shows the degree of compliance exerted by the influencer, and is assessed through analyzing the sentiments on common topics between the in-link blog post and target blog post. *Positive persona* describes bloggers with high positive influence, where their links from others often indicate approval and agreement. Milgram (1963) observed that dissent within a social network can greatly wane the strength of an influence. Our model measures dissent through evaluating the negative persona of bloggers. *Negative persona* represents bloggers with high negative influence, and their links from others usually express disagreement or distrust. The *Controversial persona* represents bloggers that are both likely to be challenged or supported by many (Cai et al., 2011), which is shown in the high number of agreement and disagreement in-link blog posts.

Influence Style Analysis Framework

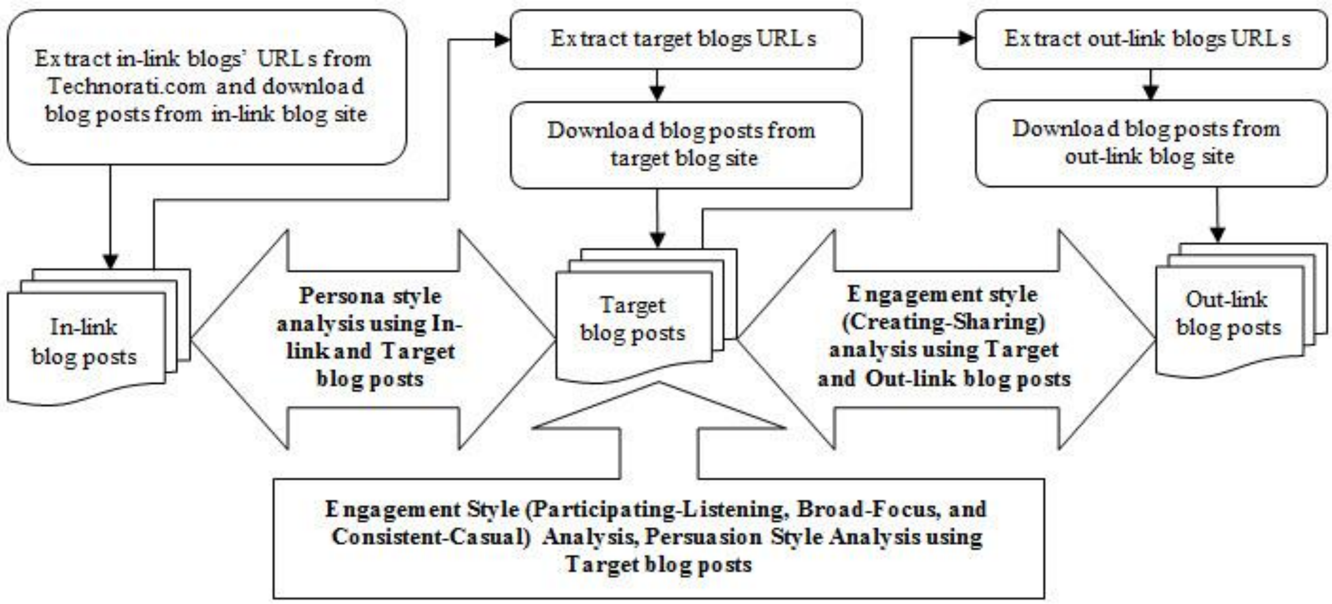

FIG. 7. Influence style analysis framework.

Figure 7 shows the influence style analysis framework. We extracted the in-link blogs URLs from Technorati.com, and downloaded the in-link blog posts, also known as back-links posts from the respective in-link blog sites. The "*search for more reactions*" feature found in Technorati.com was used to extract the in-link to the target blog site. This is followed by downloading of the target blog posts using the target blog URLs extracted from the in-link blog posts. Subsequently, the out-link blogs URLs are extracted from the target blog posts, and used to download the blog posts from the out-link blog sites. The *Participating-Listening*, *Broad-Focus*, and *Consistent-Casual* types of the Engagement style are analyzed from the target blog posts, while the *Creating-Sharing* type is analyzed between the target blog posts and the out-link blog posts. The *Persuasion* style is determined based on the subjectivity study on the target blog posts, and the *Persona* style analysis is performed between the in-link blog posts and the target blog posts.

Engagement Style Analysis. We evaluate the *Participating-Listening (PL)* type as the active participation of the bloggers in specific target topics, and measure it by the number of blog post titles that contain the target topic and related feature terms. Feature terms are words related to the aspect (attributes) of the target topics' features, for example, "*Wifi*" is a feature term in the "*Connectivity*" aspect of the target topic "*iPhone*". In this study, we limit the scope of the target topics to the main topic of Apple related products. The counted number of blog post titles is then normalized by the total number of posts for each respective blogger to give the *PL-score*. A high percentage of target topic or related feature postings would indicate that the target blog site is of *Participating* type, and conversely, a low percentage shows a *Listening* type. The *Broad-Focused (BF)* type represents the scope of the topics discussed within the blog site. A wide range of topics and their related features discussed would indicate a *Broad* type, while a limited number of topics and their related features discussed refer to a *Focused* type. The *BF-score* is measured by the number of unique target topics and their related features found in the blog post titles and normalized by the total number of topics in the network.

TABLE 1. INFUSE model – Influence style analysis.

Influence Style	Type	Description	Corresponding model
Engagement	Participating-Listening	Describes the blogs' or bloggers' participation level in posting specific target topics.	*Authority* (Cialdini, 2001)
	○ Participating	Refers to blogs or bloggers that actively post articles on target topics and share information readily.	E.g., *Feeder* from Klout influence matrix
	○ Listening	Refers to blogs or bloggers that do not share much but could follow the blog posts with a low level of posting activities on target topics.	E.g., *Observer* from Klout influence matrix
	Broad-Focused	Describes the scope of the topics discussed in the blog posts.	*Authority* and *Scarcity* (Cialdini, 2001)
	○ Broad	Refers to bloggers that post a wide range of topics.	E.g., *Broadcaster* from Klout influence matrix
	○ Focused	Refers to bloggers that concentrate on specific topics which they are normally good at.	E.g., *Specialist* from Klout influence matrix
	Creating-Sharing	Describes the level of blog posts' content originality.	*Informational influence* (Cialdini & Goldstein, 2004), *Scarcity* (Cialdini, 2001), *Objective consensus* approach (Mackie, 1987)
	○ Creating	Refers to blogs or bloggers that are original in their blog post content.	E.g., *Thought Leader* from Klout influence matrix
	○ Sharing	Refers to blogs or bloggers that tend to share by restating the content of the linked blog posts.	E.g., *Curator* from Klout influence matrix
	Consistent-Casual	Describes the consistency and duration of blogs' or bloggers' ability to garner in-links.	*Consistency* and C*ommitment* (Cialdini & Trost, 1998), *Authority* and *Reciprocity* (Cialdini, 2001)
	○ Consistent	Refers to blogs or bloggers that exhibits continual ability to garner in-links over a duration of time.	E.g., *Pundit* from Klout influence matrix
	○ Casual	Refers to blogs or bloggers that is irregular and intermittent in their ability to garner in-links.	E.g., *Explorer* from Klout influence matrix
Persuasion	Subjective-Objective	Describes the types of Persuasion style	*Emotional influence* (Brehm, 1966); *Affect and Arousal* (Cialdini & Goldstein, 2004)
	○ Subjective	Refers to blogs or bloggers who are opinionated and expressive in stating their viewpoints	
	○ Objective	Refers to rationale blogs or bloggers who express their opinions through appeals to reasons.	
Persona	Positive-Negative-Controversial	Describe the types of Persona	
	○ Positive	Refers to blogs or bloggers with a strong degree of compliance from their followers	*Compliance and Conformity* (Kelman, 1958)
	○ Negative	Refers to blogs or bloggers with strong degree of disagreement and dissent from their followers.	*Dissent* (Milgram, 1963)
	○ Controversial	Refers to blogs or bloggers that creates controversies from their postings and are both likely to be challenged or supported by many.	*Controversy* (Cai et al., 2011)

The *Creating-Sharing (CS)* type measures the originality of the content posted by the blog site, and is evaluated through analyzing the similarity between the target blog posts content and its out-link blog posts content based on the Jaccard coefficient. The Jaccard coefficient is defined as the size of the intersection divided by the size of the union of the two linked posts content (A and B) as shown in equation (1):

$$J(A\_(linking\ blog\ post), B\_(linked\ blog\ post)) = (|A \cap B|)/(|A \cup B|) \qquad -(1)$$

A high similarity value would mean that the target blog post shares most of the out-link blog post content. On the other hand, the target blog post contains more original content if the similarity value is low. The *CS-score* is given as (1 minus Jaccard coefficient) to reflect the propensity to influence for bloggers who are creative and original in their content. The *Consistent-Casual* type indicates the consistency and duration of the target blog post's ability to garner in-links, and is determined by the number of target blog site postings in an arbitrarily chosen near-term (within 1 month), mid-term (between 1 to 3 months), and long-term (beyond 3 months) duration linked by the in-link posts within the analyzed time period. The *CC-score* is computed by assigning a nil score for near-term posts, a score of 0.5 for mid-term posts, and a score of 1.0 for long-term posts. This gives consideration for the consistency of the blogger to harness links for its mid and long term postings. The assumption is that bloggers with links to their long-term posts tend to be more consistent in their influence. The assigned scores are summed and normalized over the number of posts for each blogger.

Persuasion (PES) Style Analysis. We detect the persuasion style of bloggers by analyzing the subjectivity expressed by the bloggers in their blog posts through counting the number of subjective terms in the target blog posts. This is done by matching the subjectivity terms from Wilson, Wiebe, and Hoffmann (2005) with the target blog posts terms. The matched number of subjectivity terms is then normalized with the length of the blog post. The *PES-score* is derived from taking the average score of the blogger's total postings.

Persona (PER) Style Analysis. The blogger's persona is identified by detecting the sentiments expressed between the in-link and target blog posts. It is possible to analyze comments and feedback as an observation of readers' response to the blog posts. However, we limit the scope of our study to detecting only the influence styles between linked blog posts. In our model, the sentiment analysis rules were adapted from Tan, Na, Theng, and Chang (2012) with the *PER-score* of each blogger given as the ratio of number of similar sentiments posts over the total number of posts for each blogger. Sentiment analysis is performed on clauses at the aspect level of the target topics using the typed dependency polarity pattern rules. This is because aspect level analysis can provide in-depth sentiment analysis results (Thet, Na, & Khoo, 2010). For example, in the sentence "*I absolutely love the new retina display of the iPhone, but not the Wi-Fi connectivity*", an aspect level analysis could determine the difference in opinion for the "*display*" and "*connectivity*" aspects (properties) of the iPhone for the respective clauses "*I absolutely love the new retina display of the iPhone*" and "*but not the Wi-Fi connectivity*". The typed dependency polarity pattern rule is denoted as "*typed-dependency(governor-term:[polarity], dependent-term:[polarity])→[polarity]*". For example, in the adjectival modifier pattern rule "AMOD(phone:[**neutral**], great:[+])→[+]" yields a positive sentiment polarity for the phrase "great phone". The other typed dependency polarity pattern rules used include the adverbial modifier (ADVMOD), direct object modifier (DOBJ), and nominal subject (NSUBJ) rules. The subjectivity lexicon from Wilson, Wiebe, and Hoffmann (2005) is used to tag the prior polarity of the governor and dependent terms. The typed dependency polarity pattern rules are evaluated using a bottom up approach based on the phrase structure tree as shown in Figure 8 for the clause "*I absolutely love the new retina display of the iPhone.*", with the generated typed dependencies shown in Table 2. The typed dependencies are mapped to the respective phrases through matching the governor term of the typed dependencies to the key (head) term in the phrases. Starting from the bottom phrase in the phrase structure tree, the initial noun phrase (NP) (indicated by P1) would give a neutral output as there are no typed dependency polarity pattern rules evaluated. For the subsequent noun phrase (P2), the adjectival modifier pattern rule "AMOD(display:[**neutral**], new: [**neutral**])→[**neutral**]" would also give a neutral output. In the recursive evaluation, the preceding lower level sentiment polarity output is used as the governor term polarity of the subsequent typed dependency polarity pattern rule. Hence, the governor term "display" in the subsequent direct object pattern rule would inherit the neutral polarity of the preceding evaluation output. With this, the direct object pattern rule "DOBJ(love:[+], display:[**neutral**])→[+]" gives a positive polarity sentiment output within the verb phrase indicated by P3. The adverb modifier pattern rule "ADVMOD(love:[+], absolutely:[**intensify**])→[-]" in the adverbial phrase (P4) will similarly yield a positive polarity. In addition, the intensifier term "absolutely" further increases the positive value of the polarity, which is recursively input into the governor term in the nominal subject pattern rule "NSUBJ(love:[+], I:[**neutral**])→[+]" to give the eventual positive sentiment polarity output for the clause in the final recursive step.

We manually created a lexicon for the list of specific target topics with the related aspects and features for the identification of target topics with examples shown in Table 3. We identify the target topic by matching the target terms and feature terms in the lexicon list with terms found in the noun phrases of the sentence. For example, the identified "*retina*" feature for the clause "*I absolutely love the new retina display of the iPhone.*" is mapped to the display aspect of the "*iPhone*" target topic, where the predicted triplet output (*target topic, aspect, sentiment polarity*) for the clause is given as ("*iPhone*", "*Display*", "*Positive*"). The overall target topic sentiments found in the blog post is derived by aggregating the individual clause sentiments for the target topic's aspects, where the eventual result may show the blog post is positive on the "*iPhone*" target topic.

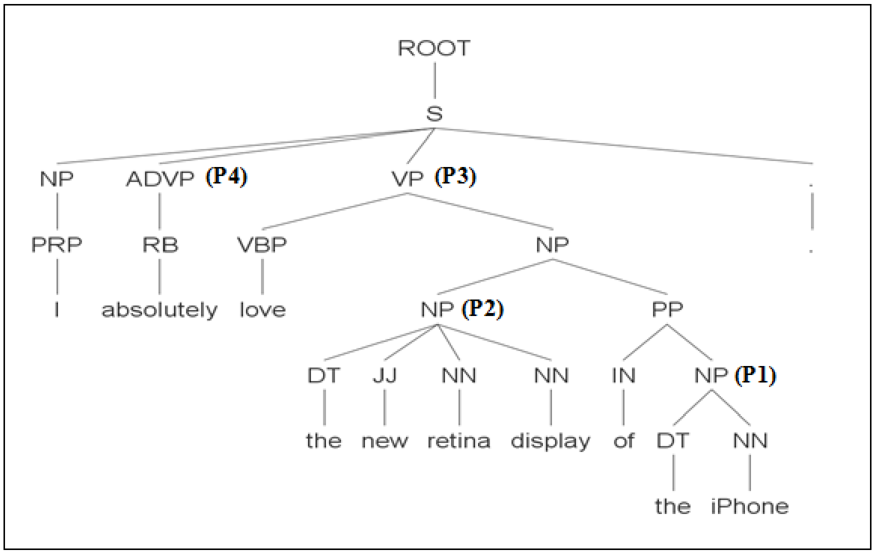

FIG. 8. Phrase structure tree for the clause "*I absolutely love the new retina display of the iPhone*."

TABLE 2. Generated typed dependencies.

nsubj(love-3, I-1)
advmod(love-3, absolutely-2)
root(ROOT-0, love-3)
det(display-7, the-4)
amod(display-7, new-5)
nn(display-7, retina-6)
dobj(love-3, display-7)
det(iPhone-10, the-9)
prep_of(display-7, iPhone-10)

TABLE 3. Examples of target topics, aspects and features.

Target Topics	Overall Aspects	Features
iPad, iPhone, Mac	Connectivity	LTE, network, Wi-Fi, etc
	Display	Screen, pixel, retina, etc
	Application	Safari, iTunes, Siri, iCloud, Facebook, etc
	Operating system	iOS

Influence Detection - Independent Cascade approach

We apply the *Independent Cascade* (IC) approach (Goldenberg, Libai, & Muller, 2001) in the INFUSE-IC model to introduce the concept of influence styles for detecting bloggers' influence diffusion in a network. In the IC approach, the aim is to study the use of influence styles in detecting influence diffusion that is initiated from a single linked blogger versus the *Linear Threshold* (LT) approach (Granovetter, 1978) that considers influence from all active neighbors. The bloggers' network is defined as a directed network (or equivalent graph) $G = (V, E)$, where V is the set of nodes (referred as bloggers) and E a set of links, where each link is denoted by $e = (v, w) \in E$ and $v \neq w$, meaning there exists a directed link from a node v to a node w. For each directed link, a real value $\kappa_{v,w}$ with $0 < \kappa_{v,w} < 1$ is specified as the diffusion probability through link (v, w). In the IC model, diffusion process proceeds from a given initial active set $D(t=0)$ in the following way. When a node $v(\in D(t))$ first becomes active at time-step t, it is given a single chance to activate each currently inactive child node w, and the attempt succeeds with probability $\kappa_{v,w}$. If v succeeds, then w becomes active at time-step $t+1$, i.e., $w(\in D(t+1))$. If multiple parent nodes of w first become active at time-step t, then their activation attempts are sequenced in an arbitrary order, but all the attempts are performed at time-step t. Whether or not v succeeds, v will not make any further attempts to activate w in subsequent rounds. However, our study focus only on the influence diffusion between immediate linked bloggers and influence propagation throughout the network was not considered. Nevertheless, influence propagation across a blogosphere network could still be inferred through analyzing the influence diffusion through subsequent linked bloggers. We derive the influence diffusion probability of each link through introducing the *Influence Style Function (F_{IS})* of a blogger, *i*, given in equation (2):

$$F_{IS}(i) = \left\{\left[W_{PL} \times \frac{PL_{i-score} - PL_{min\,score}}{PL_{max\,score} - PL_{min\,score}}\right] + \left[W_{BF} \times \frac{BF_{i-score} - BF_{min\,score}}{BF_{max\,score} - BF_{min\,score}}\right] + \left[W_{SC} \times \frac{SC_{i-score} - SC_{min\,score}}{SC_{max\,score} - SC_{min\,score}}\right] + \left[W_{CC} \times \frac{CC_{i-score} - CC_{min\,score}}{CC_{max\,score} - CC_{min\,score}}\right] + \left[W_{PES} \times \frac{PES_{i-score} - PES_{min\,score}}{PES_{max\,score} - PES_{min\,score}}\right] + \left[W_{PER} \times \frac{PER_{i-score} - PER_{min\,score}}{PER_{max\,score} - PER_{min\,score}}\right]\right\} / 6 \qquad \text{- (2)}$$

The Influence Style Function (F_{IS}) is computed by taking the normalized average of the six influence scores. In this study, we assumed a simple approach by taking all weight (*W*) values to be 1 in the model evaluation. Future studies would attempt to identify optimal weight values to improve performance. We compute the influence diffusion probability from blogger A to the respective linked blogger B as: $\kappa_{A,B}$=*Prob(Blogger A to influence) x Prob(Blogger B is influenced).* The probability of Blogger A to influence is inferred through the *(F_{IS})* score of the influencing Blogger (A), while the probability of Blogger B is influenced is derived from the similar sentiment ratio expressed when Blogger B is linked to Blogger A as shown in equation (3):

$$\kappa_{A,B} = F_{IS}(A)\ x\ \{number\ of\ similar\ sentiment\ posts_{A,B} / number\ of\ linked\ posts_{A,B}\} \qquad \text{- (3)}$$

, where $F_{IS}(A)$ is the normalized influence style function score of Blogger A, while *{number of similar sentiment posts*$_{A,B}$*/ number of linked posts*$_{A,B}$*}* is the similar sentiment ratio between Blogger A and Blogger B. To illustrate how influence style is used to detect influence in the IC approach, the influence style scores for Blogger A (F_{IS}(A)=0.73) and the similar sentiment value between Blogger A and Blogger B ($\Omega_{A,B}$=0.75), with the computed influence diffusion probability of Blogger A towards Blogger B given as ($\kappa_{A,B}$= F_{IS}(A)(=0.73) * $\Omega_{A,B}$ (=0.75) = 0.55) are shown in Figure 9. Blogger B will be activated if ($\kappa_{A,B}$=0.55) is greater than an influence threshold value referred as the Independent Cascade (IC) weight value, which we determined for optimal performance of the INFUSE-IC model to be equal 0.025 through extensive experiments. We will discuss how we get the IC weight value in the Experiments and Evaluation Results section. As seen in equation (3), the INFUSE-IC model attempts to relate influence diffusion probability to influence diffusion by comparing influence diffusion probability against the IC weight as an influence threshold value.

Influence Detection – Linear Threshold approach

In addition, we apply the *Linear Threshold* (LT) approach in the INFUSE-LT model to study the use of influence styles in detecting influence diffusion that is initiated from all the active neighbors of a blogger versus the Independent Cascade approach which only considers influence diffusion from a single neighbor at each step of the evaluation process. The Linear Threshold model conceptualizes a more realistic situation in a blogosphere network as bloggers are usually linked to multiple bloggers, hence they could be influenced by their neighboring bloggers instead of only one blogger. In the Linear Threshold model approach described by (Kempe, Kleinberg, & Tardos, 2003), a node *v* is influenced by each neighbor *w* according to a weight $\kappa_{w,v}$ such that the total weight of its neighbors is less than 1. Each node *v* chooses a threshold θ_v uniformly at random from the interval [0, 1], which represents the weighted fraction of *v*'s neighbors that must become active in order for *v* to become active. Given a threshold, and an initial set of active nodes A_0 (with all other nodes inactive), the diffusion process flows in discrete steps where in step t, all nodes that were active in step t-1 remain active and a node v is activated when the total weight of its active neighbors is at least θ_v :

$$\sum_{w\ active\ neighbor\ of\ v} \kappa_{w,v} \geq \theta_v \qquad \text{- (4)}$$

Thus, the Linear Threshold model intuitively represents the different latent tendencies of nodes to be influenced when their neighbors do. In our study, the weight $\kappa_{A,B}$ of a node B taken as Blogger B that is linked to Blogger A is computed as the influence diffusion probability from Blogger A to the linked Blogger B given in equation (3). Concurrently, Blogger B is linked to Blogger C, which has a weight similarly computed as the influence diffusion probability from Blogger C to Blogger B given as $\kappa_{C\,B}$.

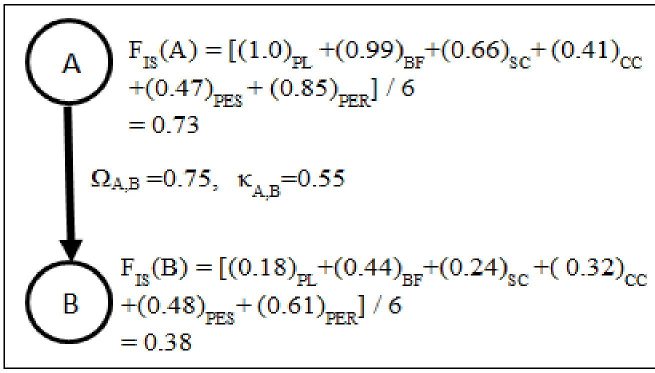

FIG. 9. Independent Cascade Scoring for linked Blogger A and Blogger B.

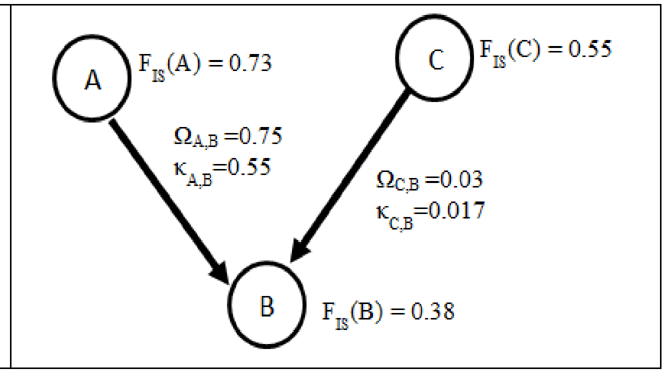

FIG. 10. Linear Threshold Scoring for Blogger A and Blogger C linked to Blogger B.

Blogger B is activated if the combined normalized weights ($\kappa_{A,B}$ + $\kappa_{C\ B}$)/2 of the linked Bloggers A and C to Blogger B is greater than an arbitrary threshold value, assuming equal weightage of both linked bloggers in influencing the linking blogger. Similarly, the threshold value was determined to be 0.015 through extensive experiments. From Figure 10, where both Blogger A and Blogger C have influence links to Blogger B, the influence style score for Blogger C (F_{IS}(C) = 0.55) and the similar sentiment value between Blogger C and Blogger B ($\Omega_{C,\ B}$ = 0.03) are computed to give the influence diffusion probability of Blogger C towards Blogger B ($\kappa_{C,B}$= F_{IS}(C)(=0.55) * $\Omega_{C,B}$(=0.03) = 0.017). Similarly, the influence diffusion probability of Blogger A towards B is given as $\kappa_{A,B}$= F_{IS}(A)(=0.73) * $\Omega_{A,B}$ (=0.75) = 0.55. The Linear Threshold approach further considers the overall influence of a blogger's neighbors in computing the influence diffusion probability, hence the normalized influence diffusion probability towards Blogger B is {$\kappa_{A,B}$(=0.55) + $\kappa_{C,B}$(=0.017)}/2 = 0.28 based on an assumption that the blogger's neighbors have equal weightage in the influence diffusion probability. Thus, Blogger B is activated as the total influence diffusion probability is greater than the threshold value of 0.015. In our study, we consider all neighboring blogger link pairs (Blogger A-to-Blogger B and Blogger C-to-Blogger B) to be active when the evaluated blogger (Blogger B) is activated as the neighboring bloggers have contributed to influencing the blogger. It can be seen that Blogger C could activate Blogger B when the Linear Threshold model is used as compared to the Independent Cascade model where Blogger C's influence diffusion probability alone would be lower than the required threshold value. This illustrates the ability of the Linear Threshold model to consider the overall influence effect of the linked bloggers in a networked blogosphere.

Experiments and Evaluation Results

We identified seven blog sites (www.9to5mac.com, www.macrumors.com, www.tuaw.com, www.engadget.com, www.techcrunch.com, www.slashgear.com, and www.theverge.com) from Technorati.com for the evaluation. The "search for more reactions" feature found in Technorati.com was used to extract the in-link posts to the target blog site. A total of 107 bloggers with 9107 in-link posts, 8879 target posts, and 19317 out-link posts were extracted from the blog sites from 11 Feb 2013 to 5 April 2013 for our data set. From the dataset, we select and describe a chain of five arbitrarily chosen linked bloggers to illustrate and analyze the influence style of individual bloggers using the INFUSE model.

Influence Style Evaluation

Table 4 shows the influence style scores of the five linked bloggers with their blog sites given in brackets respectively. The Participating-Listening type results show Blogger1 (PL-score=0.99) was actively posting on the target topics with minimal discussions on other topics, while Blogger4 was less participating with a PL-score of 0.27 with fewer postings on the target topics. Blogger2, Blogger3, and Blogger5 were the least participating on the target topics with PL-scores less than 0.2. In evaluating the scope of topics discussed, Blogger1 and Blogger4 had a broad scope (BF-score=0.99) amongst the five bloggers by covering most of the target topics. On the other hand, Blogger2 (BF-score=0.20) was more focused on specific target topics in relation to the other bloggers. Both Blogger1 and Blogger2 had higher CS-scores (>0.60), which shows them to be original in their blog post content. Blogger3 and Blogger5 were less original in their content with CS-scores of (>0.50) each, while Blogger4 was seen to share the most linked post content with a CS-score of 0.32. Both Blogger1 and Blogger5 were considered to be consistent with CC-scores of 0.40 and 0.47 respectively.

TABLE 4. Bloggers' influence style scores.

Influence style	Blogger1 (Macrumors.com)	Blogger2 (Engadget.com)	Blogger3 (Macrumors.com)	Blogger4 (Tuaw.com)	Blogger5 (Techcrunch.com)
Engagement style					
○ Participating-Listening (PL)	0.99	0.10	0.14	0.27	0.17
○ Broad-Focused (BF)	0.99	0.20	0.64	0.99	0.41
○ Creating-Sharing (CS)	0.66	0.62	0.51	0.32	0.52
○ Consistent-Casual (CC)	0.40	0.10	0.33	0.18	0.47
Persuasion style (PES)	0.47	0.53	0.55	0.54	0.52
Persona style (PER)	0.75	0.75	0.62	0.79	0.63

While Blogger3 was mildly consistent, both Blogger 2 and Blogger4 were casual in engagement style with scores of 0.10 and 0.18, meaning few bloggers were linking to them throughout the analyzed time period. The bloggers except for Blogger1 had similar Persuasion Style scores within the range of (0.52 to 0.55) showing a similar persuasion style. On the other hand, Blogger1 was more objective in style as shown in his lower PES-score (0.47) indicating fewer subjective terms were used in the blog post content. From the Persona score results, it could be seen that Blogger1, Blogger2, and Blogger4 had a high level of agreement from their linking bloggers as shown in the high PER-score range (0.75 to 0.79). On the other hand, Blogger3 and Blogger5 were deemed to be controversial with their mid-ranged PER-score range (0.62 to 0.63). In the persona analysis, the similar sentiments and dissimilar sentiments ratios between the in-link blog posts and the target blog posts towards the respective target topics for the bloggers were evaluated. As observed in Tan, Na, Theng, & Chang (2012), neutrality expresses agreement between the linked blog posts. We considered neutral polarity as similar to positive or negative polarity where opinionated and neutral pairs would give similar sentiment outputs. Compliance could be inferred from the similar sentiments expressed on common target topics and related features between the linked blog posts, even if the sentiments are negative. This can be seen from the linked blog post that exhibited negative sentiments on the common feature term (iOS) as shown in Figure 11. On the other hand, high dissimilarities in sentiments show disagreement towards the target blog posts, which in turn indicate the linked bloggers' negative persona. This is seen in Figure 12, where the disagreement is shown in the positive sentiments and negative sentiments expressed by the in-link blog post and the target blog post respectively on the feature term "LTE".

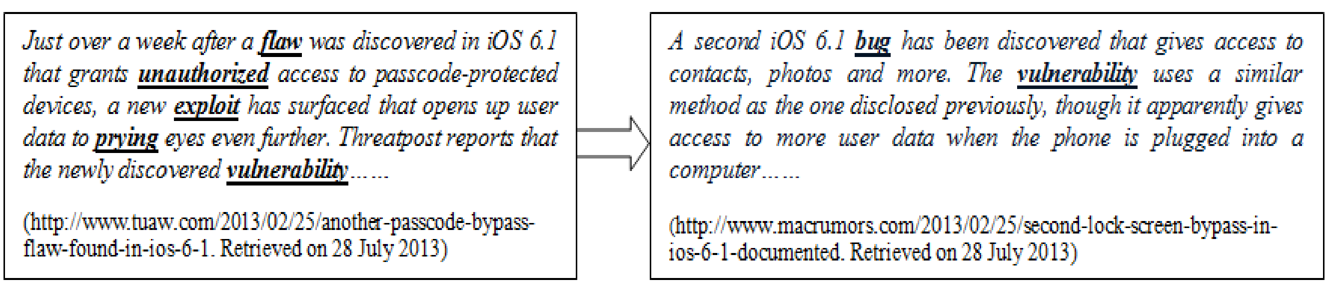

FIG. 11. Example of compliance between linked blog posts with negative sentiments on the feature "iOS".

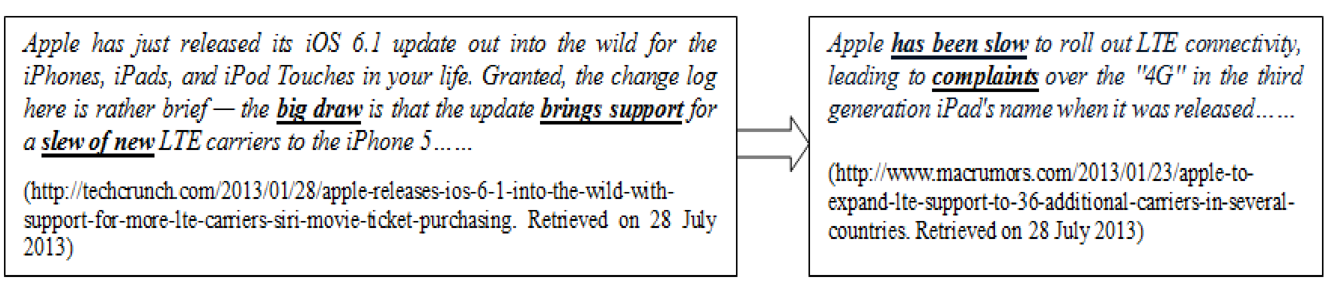

FIG. 12. Example of non-compliance with positive sentiments on the feature "LTE" in the linking blog post linked to target blog post with negative sentiments.

Influence Style Profiles. Figure 13 shows the graph of the aggregated engagement style scores, as well as the persuasion and persona scores of the five linked bloggers which indicates the relative degree of each influence style amongst the bloggers. The influence style profiles of the bloggers are given in Table 5. For example, it can be seen that Blogger1 is highly engaging, less subjective, and positive in persona. Each blogger had an influence style which differs between bloggers from different blog sites as well as bloggers from within the same blog site. This is seen in Blogger1 and Blogger3 who were from macrumors.com, but Blogger3 displayed a less engaging, more subjective, and lower positive persona as compared to Blogger1. By further analyzing the influence styles at blogger level, the proposed INFUSE model could provide in-depth details of the influence exerted by the individual bloggers further differentiating the bloggers in terms of influence style.

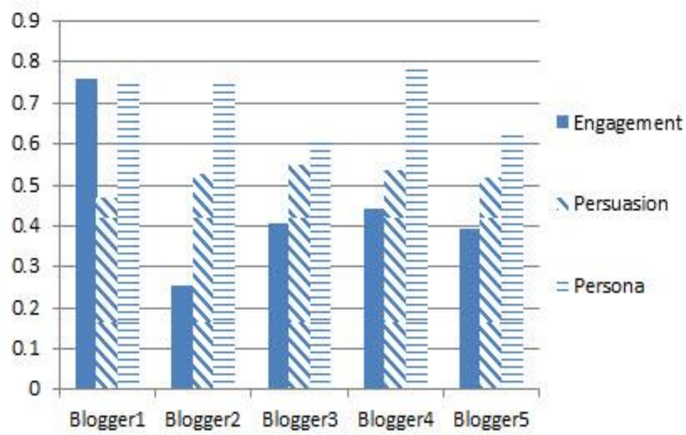

FIG. 9. Independent Cascade Scoring for linked Blogger A and Blogger B.

TABLE 5. Bloggers' influence Style Profiles.

	Engagement Style	Persuasion Style	Persona Style
Blogger1	Highly Engaging (0.76)	Less Subjective (0.47)	Positive (0.75)
Blogger2	Less Engaging (0.25)	Subjective (0.53)	Positive (0.75)
Blogger3	Engaging (0.41)	Subjective (0.55)	Controversial (0.62)
Blogger4	Engaging (0.44)	Subjective (0.54)	Positive (0.79)
Blogger5	Less Engaging (0.39)	Subjective (0.52)	Controversial (0.63)

Influence Diffusion Detection Evaluation

Influence diffusion detection was evaluated between each blogger link pair to assess the performance of the extended INFUSE models based on the Independent-Cascade (INFUSE-IC) and Linear Threshold (INFUSE-LT) approaches. Each influencing blogger node is activated (taken as the initial node) at least once in the iterative process steps to evaluate the probability of influence occurring between the linked bloggers. The in-degree weight, defined as the number of in-links over the total number of links for a blogger, as well as the similar sentiment value between two linked bloggers were used in the baseline comparison to the extended INFUSE models. For the In-degree baseline approach, a node is activated when the in-degree weight of the influencing blogger is higher than that of the influenced blogger. In the second baseline approach, only similar sentiment values (SENT-only) between the linked bloggers were used as indication of influence diffusion. In addition, a third baseline comparison using the similar sentiment values between linked bloggers as the diffusion probability values in a Sentiments-Independent Cascade (SENT-IC) approach was used in the evaluation. Extensive testing showed optimal performance is achieved when the IC weight value for the SENT-IC model equals 0.05. In the evaluation testing, 514 linked blog posts of 362 unique blogger pairs were manually classified into 265 influencing and 97 non-influencing blogger pairs. The two coders' Cohen's kappa index (Cohen, 1960) was computed to be 0.68, with the conflicting tags reviewed and then manually re-classified and used as the answer keys. We measured the performance of the models in detecting influence diffusion using the F1-score measure given as F1 = (2•Precision•Recall)/(Precision+Recall). In the INFUSE-IC and INFUSE-LT models, a blogger-pair consisting of a neighboring blogger and the evaluated blogger is deemed to be influencing when the evaluated blogger is activated. The results of the evaluation testing are shown in Table 6. It can be seen that the INFUSE-IC and INFUSE-LT models, both with F1-score=0.76, performed better than the In-Degree (F1-score=0.45), SENT-only (F1-score=0.63), and SENT-IC (F1-score=0.68) baseline approaches. The poor performance in the In-Degree approach could be attributed to that links alone may not indicate influence as the linking blogger's post could also express disagreement towards the linked blogger's posts. Both the sentiment value approaches (SENT-Only and SENT-IC) performed better than the In-Degree approach as similar sentiment expressions between linked bloggers provide clearer indication of influence through the common stance on the topics discussed. The higher performance of the SENT-IC approach versus the SENT-Only approach shows the capability of the Independent Cascade method in modelling influence diffusion as compared to using only the sentiment values between linked bloggers. Both INFUSE-IC and INFUSE-LT models performed well against the baseline methods in detecting influence diffusion within a blogger network. This is because the extended INFUSE models relates closely to influence diffusion by further considering the bloggers' influence styles, where each of the influence style characteristic could have provided clearer indication of the blogger's propensity to influence. The similar F1-scores between the INFUSE-IC and INFUSE-LT indicate on par performance in detecting influence diffusion flow for both approaches. However, recall for INFUSE-LT is marginally higher and this could be because the INFUSE-LT model considers the influence probability from all active neighbours of a blogger resulting in higher probability of activating the neighbour-blogger pair compared to the INFUSE-IC model where only a single neighbour is evaluated at each step. The scores for the INFUSE-IC and INFUSE-LT models at respective threshold values are shown in Figure 14 where it can be seen that INFUSE-LT has an overall higher recall value. The corresponding threshold values for the highest F1-scores of the INFUSE-IC model and INFUSE-LT model are 0.025 and 0.015 respectively. These numbers were subsequently used as the threshold values in our evaluation of the models based on the best performance given for the models. The threshold values range of {0.01, 0.1} was chosen because there was no significant improvement in the F1-scores of the models beyond this range after extensive experiments.

We further analyze the relation between influence styles scores and bloggers' influence to determine the significant influence styles that could detect influence diffusion. The independent sample t-test results between influence styles and influence diffusion indicator given in Table 7 shows that the influence styles, except for Consistent-Casual (CC) style are significantly different ($p<=0.05$) in their *μ(yes)* mean score and *μ(no)* mean score, indicating their ability to detect influence. The difference in mean scores for CC-style is insignificant ($p>0.05$), which reveals both influential and non-influential bloggers to have similar consistency in their blog posts' links. This also indicates CC-style is not able to differentiate a blogger's ability to influence and could be omitted from the Influence Style Function computation. Further to that, the *μ(yes)* mean score and *μ(no)* mean score for Subjectivity style show bloggers with lower PES-scores to diffuse more influence. Additional experiments conducted on the INFUSE-IC model, excluding CC-style and reversing the polarity of Subjectivity style (1 minus PES-score) in the Influence Style Function, yielded results of Precision=0.75, Recall=0.71, F1-score=0.73, which is marginally higher in precision but lower in recall performance versus the initial INFUSE-IC results given in Table 6. Further experiment on the effect of topic on the use of influence style to detect influence diffusion shows Persona Style to be topic dependent. When the PER-scores were computed based on both Apple and Android related topic posts, performance of the INFUSE-IC model was lower at (Precision=0.72, Recall=0.64, F1-score=0.68) compared to the initial INFUSE-IC model, which is based only on Apple related topics. In addition, t-test results between influence styles and influence diffusion indicator shows Persona to be insignificant ($p>0.05$) when both Apple and Android topics were considered in the PER-scores. This could be because a blogger who is positively recognized in the Apple topic domain is usually not equally agreed upon for Android related postings, thus rendering Persona style not suitable for general topic influence diffusion detection.

TABLE. 6 Influence diffusion detection evaluation test results.

	In-Degree	SENT-Only	SENT-IC	INFUSE-IC	INFUSE-LT
Precision	0.63	0.72	0.73	0.73	0.72
Recall	0.35	0.56	0.64	0.79	0.81
F1-score	0.45	0.63	0.68	0.76	0.76

TABLE. 7 Independent sample t-test results.

Influence Styles	p	Influence diffusion	
		μ(yes)	μ(no)
Participating-Listening	0.001	0.31	0.17
Broad-Focused	0.013	0.59	0.49
Creating-Sharing	0.000	0.49	0.42
Consistent-Casual	**0.182**	0.41	0.44
Subjectivity	0.013	0.49	0.51
Persona	0.002	0.79	0.76

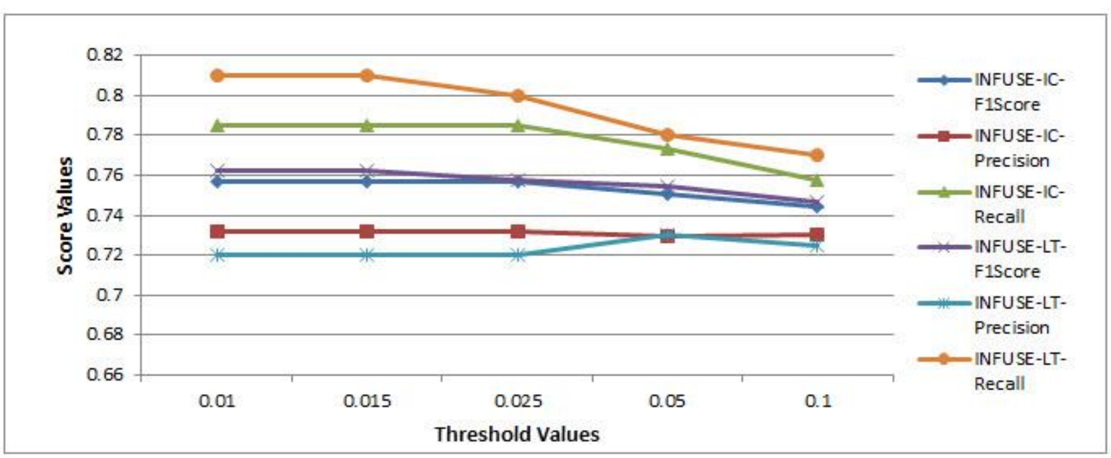

FIG. 14. Scores for the INFUSE-IC and INFUSE-LT model at respective threshold values.

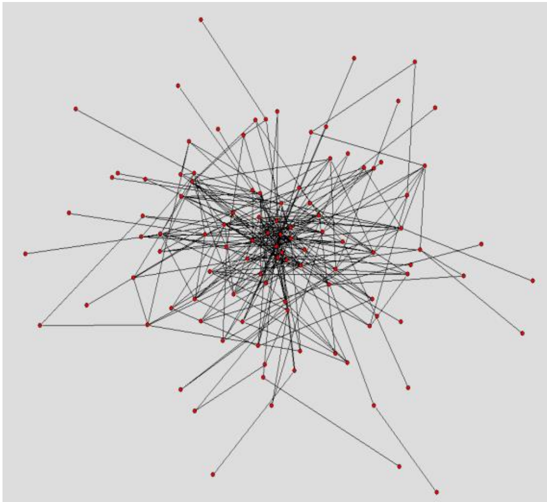

FIG. 15. Bloggers Influence Network.

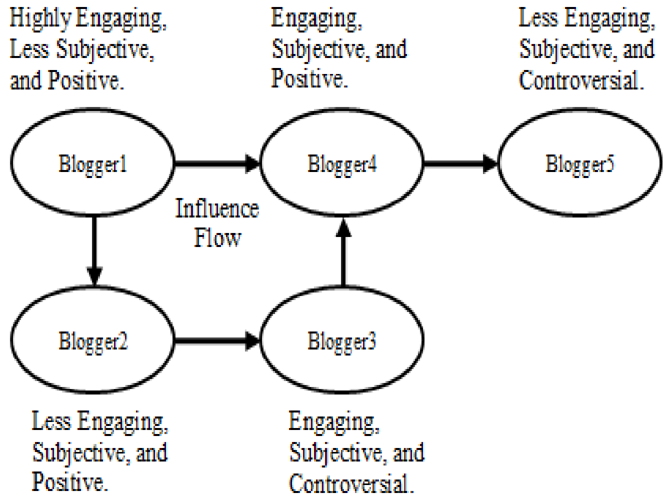

FIG. 16. Influence diffusion path.

Discussion

Influence Style Analysis

While previous studies (Adar & Adamic, 2005; Agarwal & Liu, 2008) that used a limited influence scope based on link frequency could only detect the presence of influence for actively linked bloggers, our proposed INFUSE model further describes the influence exerted by the bloggers by providing a clear profile of the bloggers' influence styles as shown in Table 5. In addition to indicating influence presence, a blogger could be described through his engagement style, persuasion style, as well as persona style in the manner in which influence is exerted using the INFUSE model. The model analyzed blog posts at the blogger-level to give an in-depth evaluation of the individual blogger influence style. While bloggers could exhibit similar influence style traits, it is observed that individual bloggers could differ in varying degree for each influence style type. This is true even if the bloggers are from the same blog site as in the case for Blogger1 and Blogger3 as seen in Tables 4 and 5 where Blogger1 was highly participative in posting targeted topics as compared to Blogger3. On the other hand, Blogger1 and Blogger3 were broad in their target topic scope in comparison with the other bloggers indicating wider coverage on the target topics in their blog postings. Being able to differentiate the participation level and topic focus of individual bloggers allows blog readers to determine the bloggers' depth and scope of knowledge in respective topics. Though bloggers could differ in influence style, there would be certain rules, such as the non-infringement of copyrights commonly observed by bloggers, which restricts them in sharing other blog post content resulting in a similar Creating-Sharing influence style for the bloggers. Influence consistency is also observed to differ between bloggers, where Blogger5 was observed to be the most consistent blogger with constant in-links to blog postings over a timeline. However, our results showed consistency of links to blog postings is not significant in detecting influence diffusion meaning bloggers who had only recent in-links could be influential as well. The five bloggers were evaluated to be subjective in their persuasion style based on their subjectivity score. The homogeneity in persuasion style is indicative of a product review blog site, which is generally subjective in nature as compared to news reporting blog sites which tend to be more objective. We would attempt to improve subjectivity detection by taking into account the grammatical relations between words in our future work in consideration of the linguistic constraints in matching subjectivity terms for semantic orientation. For example, the subjectivity value of the phrase "*very mad*" would increase by considering the adverb modifier relationship "ADVMOD(mad, very)", instead of just matching "*mad*" as a subjective term. The INFUSE model further identified Blogger1, Blogger2 and Blogger4 to be positive in persona through the linguistic sentiment analysis, which is in-line with the observation that in-link blog posts tend to be in agreement with popular bloggers. On the other hand, Blogger3 and Blogger5 were evaluated to be controversial, indicating both agreement and disagreement were expressed by the in-link blog posts towards them. By analyzing the influence style of individual bloggers, we are able to provide more fine-grained descriptions of the bloggers' influence. Previous studies which assumed a monolithic definition of influence would not be able to differentiate the varying influence styles of the individual bloggers. The difference in influence styles shows that bloggers exert influence in various ways, and this could determine the different level of influence exerted by the bloggers.

Influence Diffusion Detection

The knowledge on individual blogger's influence style was used in the INFUSE model to detect influence diffusion in the blogger network. The extended INFUSE models (INFUSE-IC and INFUSE-LT) performed better than the baseline methods as the model further maps the influence characteristics of the bloggers to the links between bloggers. This provides a closer match to the influence patterns in the blogger network which gives better performance for detecting influence diffusion as compared to the use of the presence of links and their sentiment polarity. In comparison, the INFUSE-IC had higher precision, while INFUSE-LT had higher recall. The higher number of influencing blogger-pairs in the dataset could have resulted in a higher recall value for the INFUSE-LT model due to the Linear Threshold approach's propensity to classify bloggers as influencing through combining the influence probabilities of neighboring bloggers. On the other hand, precision performance of the INFUSE-LT model was affected because there was a possibility that non-influencing pairs were unwantedly activated as a result of the influence probability scores of active neighbors. For example in Figure 10, where there could be no influence diffusion from Blogger C to Blogger B due to the low similar sentiments value between them, but nonetheless the Blogger C to Blogger B pair was activated due to the overall influence probability scores being higher than the threshold value. The use of the Consistent-Casual influence style type was found to be not significant in detecting influence diffusion, which means that the number of in-links to past blog postings has no effect on influence diffusion detection. This shows that the timeline of blog postings is not a clear indicator of influence diffusion in a blogger network. From the lower mean PES-scores indicating influence, the results also revealed that objective bloggers exert more influence. Our study also revealed influence diffusion detection to be topic specific and performance would suffer if the Persona influence style is computed based on generic topics.

In addition to detecting influence diffusion, the extended INFUSE models further describe the manner and style in which influence is diffused through the network path by providing the bloggers' influence styles. From the results, we could see that influential bloggers within the blogger network tend to be participating, broad, creative, objective, and have positive persona. We could plot the influence styles of each blogger along the influence path to visualize the pattern of influence styles in the influence flow within the network. Figure 15 gives the bloggers' influence network with Figure 16 showing the influence path of the five evaluated linked bloggers. It could be seen there is influence flow from Blogger1 who is highly participating, less subjective, and positive in persona towards bloggers who have influence style that is listening, more subjective and less positive in persona.

Additional experiment results show performance for the INFUSE-IC model, excluding Consistent-Casual influence style type and reversing the polarity of Subjectivity style, was marginally higher in precision but lower in recall performance than that of the initial INFUSE-IC model. This aspect will be further investigated in future work with a larger dataset. Our study assumed equal weightage of the neighboring bloggers in influencing a blogger in the INFUSE-LT model as seen in the combined normalized weights $(\kappa_{A,B} + \kappa_{C\,B})/2$ of the linked Bloggers A and C to Blogger B. In this equation, the effect of a large weight value could have been reduced by a small weight value. We would further determine the optimal weight function in our future work.

It is noted that influence style could be specific to blog domains as seen in a reputed political blogger that posts infrequently, but holds much weight compared to an influential product review blogger who is actively engaging the readers. Nevertheless, the INFUSE model could be generalized across blog domains to detect influence diffusion through identifying the influence styles of individual bloggers of each domain. Though, our study has shown the INFUSE model to detect and describe influence diffusion effectively, it is noted that bloggers' influence styles are not static and would change over time. Hence, the effectiveness of the INFUSE model is specific to the period of analysis. A different period of analysis could have identified the influence style differently since the measurement values would have changed. The INFUSE model is also dependent on the performance of the sentiment analysis approach. Any performance changes to the sentiment analysis would inadvertently affect the results of the INFUSE model.

Our study propose a novel approach of analyzing bloggers' influence styles in the INFUSE model, and using the influence styles as features in the INFUSE-IC and INFUSE-LT models to improve the performance of influence diffusion detection between linked bloggers. Identification of bloggers' influence style would give a better description and understanding of the influence exerted by the bloggers through their engagement style, persuasion style, and persona. Possible applications of the INFUSE model include detecting bloggers' influence styles to determine the type of influence flow in the blogosphere, for example companies may want to market their products through bloggers who are original, objective, and received positively by the readers versus bloggers who are subjective and controversial.

Conclusion

In contrast to previous studies which focused on detecting the presence of influence between bloggers, our proposed INFUSE model further describes the influence of bloggers in terms of engagement, persuasion, and persona style. The detailed analysis of influence could differentiate the influence style of bloggers, and provide a more comprehensive description of the influence exerted. A more comprehensive and accurate portrayal of the blogger's influence through the evaluated influence profiles would in turn improve influence detection within the blogosphere. Previous studies that used graph-based blog features to detect influence in the blogosphere had assumed that influence exists between linked blogs, and had not considered the manner and style in which influence was diffused in the network. Our study applies the Independent Cascade and Linear Threshold approaches to the INFUSE model using influence style as features to detect influence diffusion in a blogger network. The extended INFUSE models relate closely to the influence characteristics of the bloggers and performed well compared to the in-degree and sentiment values baseline approaches. In addition to detecting influence diffusion in a blogger network, the INFUSE model describes the influence style of each blogger in the influence path, providing a fine-grained description of the manner in which influence is propagated.

References

Adar, E., & Adamic, L.A. (2005). Tracking information epidemics in blogspace. In A. Skowron, R. Agrawal, M. Luck, T. Yamaguchi, Morizet-Mahoudeaux, J. Liu, & N. Zhong (Eds.): Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (pp. 207-214). Washington, DC: IEEE Computer Society.

Agarwal, N., & Liu, H. (2008). Blogosphere: Research issues, tools, and applications, SIGKDD Explorations Newsletter, 10(1), 18-31.

Agarwal, N., Liu, H., Tang, L., & Yu, P.S. (2008). Identifying the influential bloggers in a community. In M. Najork, A. Broder, & S. Chakrabarti (Eds.): Proceedings of the International Conference on Web Search and Web Data Mining (pp. 207-218). NY: ACM.

Agarwal, N., & Liu, H. (2009). Modeling and Data Mining in Blogosphere. San Rafael, CA: Morgan & Claypool.

Barbieri, N., Bonchi, F., & Manco, G. (2012). Topic-Aware Social Influence Propagation Models. In B. Goethals, & G. Webb. (Eds.): Proceedings of the IEEE 12th International Conference on Data Mining (pp. 81-90). Washington, DC: IEEE Computer Society.

Brehm, J. W. (1966). A theory of psychological reactance. Academic Press.

Cai, K.K., Bao, S.H., Yang, Z., Tang, J., Ma, R., Zhang, L., & Su, Z. (2011). OOLAM: An Opinion Oriented Link Analysis Model for influence persona discovery. In I. King, W. Nejdl, & H. Li (Eds.): Proceedings of the 4th International Conference on Web Search and Data Mining, (pp. 645-654). NY: ACM.

Chen, W., Wang, C., & Wang, Y. (2010). Scalable influence maximization for prevalent viral marketing in large-scale social networks. In B. Rao, B. Krishnapuram, A. Tomkins, Q. Yang (Eds.): Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, (pp. 1029-1038). NY: ACM.

Cialdini, R.B. (2001). Influence: Science and practice (4th ed.). Boston: Allyn & Bacon.

Cialdini, R.B., & Goldstein, N. J. (2004). Social Influence: Compliance and conformity. Annual Review of Psychology, 55, 591-621.

Cialdini, R.B., & Trost, M.R. (1998). Social influence: Social norms, conformity and compliance. In D.T. Gilbert, & S.T. Fiskel (Eds.): The handbook of social psychology (4th ed.), 21, 151-192. New York: McGraw-Hill.

Cohen, J.A. (1960). Coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37-46.

Costa, P. T. Jr., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. Psychological Assessment, 4(1), 5-13.

Ghosh, R., & Lerman, K. (2008). Community detection using a measure of global influence. In L. Giles, J. Yen, H. Foley, M. Smith, & H. Zhang (Eds.): Proceedings of the 2nd SNA-KDD Workshop on Social Network Mining and Analysis (pp. 20-35). Berlin, Heidelberg: Springer-Verlag.

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.

Goyal, A., Bonchi, F. & Lakshmanan, L. V. (2010). Learning influence probabilities in social networks. In B. D. Davison, T. Suel, N. Craswell, B. Liu (Eds.): Proceedings of the third ACM international conference on Web search and data mining (pp. 241-250). NY: ACM.

Granovetter, M.S. (1973). The strength of weak ties. The American Journal of Sociology, 78(6), 1360–1380.

Granovetter, M. (1978). Threshold models of collective behavior. The American Journal of Sociology, 83(6), 1420-1443.

Guadagno, R.E., Okdie, B.M., & Eno, C.A. (2008). Who blogs? Personality predictors of blogging. Computers in Human Behavior, 24, 5.

Hassan, A., Qazvinian, V., & Radev, D.R. (2010). What's with the attitude?: identifying sentences with attitude in online discussions. In H. Li. & L. Màrquez (Eds.): Proceedings of the Conference on Empirical Methods in Natural Language Processing (pp. 1245-1255). Stroudsburg, PA: ACL.

Hogg, M.A., & Vaughan, G.M. (2005). Social psychology. Harlow: Pearson/Prentice Hall.

Kelman, H.C. (1958). Compliance, identification, and internalization: Three Processes of attitude change. Journal of Conflict Resolution, 2(1), 51-60.

Kempe, D., Kleinberg, J., & Tardos, E. (2003). Maximizing the spread of influence through a social network. In L. Getoor, T. Senator, P. Domingos, C. Faloutsos (Eds.): Proceedings of the 9th ACM International Conference on Knowledge Discovery and Data Mining (pp. 137-146). New York, NY: ACM.

Kimura, M., & Saito, K. (2006). Tractable models for information diffusion in social networks. In J. Fürnkranz, T. Scheffer, M. Spiliopoulou (Eds.): Proceedings of the 10th European conference on Principle and Practice of Knowledge Discovery in Databases (pp. 259-271). Berlin: Springer-Verlag.

Krackhardt, D. (1992). The strength of strong ties: The importance of philos in organizations, 216–239. Harvard Business School Press, Boston, MA.

Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2010). Predicting positive and negative links in online social networks. In M. Rappa, P. Jones, J. Freire, & S. Chakrabarti (Eds.): Proceedings of the 19th International Conference on World Wide Web (pp. 641-650). New York, NY: ACM.

Li, H., Bhowmick, S.S., & Sun, A. (2011). CASINO: Towards conformity-aware social influence analysis in online social networks. In B. Berendt, A. Vires, W. Fan, C. Macdonald, I. Ounis, I. Ruthven (Eds.): Proceedings of the 20th ACM International Conference on Information and Knowledge Management (pp. 1007-1012). New York, NY: ACM.

Lim, S-H., Kim, S-W., Kim, S., & Park, S. (2011). In W. Chu, & W.E. Wong (Eds.): Proceedings of the 2011 ACM Symposium on Applied Computing (pp. 937-941). New York, NY: ACM.

Mackie, D.M. (1987). Systematic and nonsystematic processing of majority and minority persuasive communications. Journal of Personality and Social Psychology, 53, 41-52.

Matsumura, N., Yamamoto, H. & Tomozawa, D. (2008). Finding influencers and consumer insights in the blogosphere. In E. Adar, M. Hurst, T. Finin, N. Glance, N. Nicolov, & B. Tseng (Eds.): Proceedings of International Conference on Weblogs and Social Media (pp. 76-83). California CA: AAAI Press.

Milgram, S. (1963). Behavioral study of obedience. Journal of Abnormal and Social Psychology, 67, 371-378.

Moilanen, K., & Pulman, S. (2007). Sentiment Composition. In R. Mitkov, & G. Angelova (Eds.): Proceedings of the Recent Advances in Natural Language Processing (pp. 378-382). Borovets, Bulgaria.

Saito, K., Nakano, R., & Kimura, M. (2008). Prediction of Information Diffusion Probabilities for Independent Cascade Model. Knowledge-Based Intelligent Information and Engineering Systems Lecture Notes in Computer Science, 5179, 67-75.

Shaikh, M.A.M., Prendinger, H., & Ishizuka, M. (2008). Sentiment Assessment of Text By Analyzing Linguistic Features And Contextual Valence Assignment. Applied Artificial Intelligence, 22(6), 558-601.

Song, X., Chi, Y., Hino, K., & Tseng, B. (2007), Identifying opinion leaders in the blogosphere. In A.H.F. Laender, A. O. Falcão, Ø.H.Olsen, M. J. Silva, R. Baeza-Yates, D.L. McGuiness, & B. Olstad (Eds.): Proceedings of the 16th ACM Conference on Information and Knowledge Management (pp. 971-974). New York, NY: ACM.

Tan, L.K.W., Na, J.-C., Theng, Y.-L., & Chang, K.Y. (2012). Phrase-level sentiment polarity classification using rule-based typed dependencies and additional complex phrases consideration. Journal of Computer Sciences and Technology. 27(3), 650-666.

Tan, L.K.W., Na, J.-C., & Theng, Y.-L. (2013). Influence Detection Using Blogger's Influence Style. In S.R. Urs, J.-C. Na, & G. Buchanan (Eds.): Proceedings of the International Conference on Asian Digital Libraries. (pp. 132-142). NY: Springer.

Thet, T.T., Na, J.-C., & Khoo, C.S.G. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. Journal of Information Science, 36(6), 823-848.

Wilson, T., Wiebe J., & Hoffmann P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In R.J. Mooney, C. Brew, L-F. Chien, K. Kirchhoff (Eds.): Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (pp. 347-354). Stroudsburg, PA: ACL.

Wilson, T., Wiebe J., & Hoffmann P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Computational Linguistic. 35(3), 399-433.

Xiang, R., Neville, J., & Rogati, M. (2010). Modeling relationship strength in online social networks. In M. Rappa, P. Jones (Eds.*): Proceedings of the 19th international conference on World Wide Web* (pp. 981-990). NY: ACM.

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. Journal of Research in Personality. 44(3), 363-373.