Inferring Online Relationships from User Characteristics

Suwimon VONGSINGTHONG[1,*], Sirapat BOONKRONG[2] and Herwig UNGER[3]

[1]Information Technology and Management Department, Faculty of Business Administration, Krirk University, Bangkok 10220, Thailand
[2]School of Information Technology, Institute of Social Technology, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand
[3]Communication Network Department, Faculty for Mathematics and Computer Science, Fern Universität in Hagen, Hagen, Germany

(*Corresponding author; e-mail: suwimonv@yahoo.com)

Abstract

A new approach to understanding human online behavior in regard to psychological functioning is proposed through the developed user's activity model incorporating the influences of social behavior, network, and content. Microscopic levels of user characteristics induced by personality traits were interpolated as interaction rules, whilst an unsupervised clustering algorithm was applied to penetrate the individual complexity. Temporal behavior of disparate users was mimicked, and streaming network data was generated and computationally analyzed. A comprehensive understanding of how individuality, friendship, and varying temperaments dramatically reshaped the networks was gained from insight synthesis of network properties characterized by small-world, scale-free, and centrality measures. Evidence illustrates that users with high extraversion possess high numbers of friends and spread massive information, while high conscientious and high intellect users are seriously discreet in accepting friends and often produce influential content. These results not only provide a wealth of challenges for product recommendation, network structure optimization, and design, but also are useful for the prediction of future network structural evolution.

Keywords: Diffusion, stochastic model, structural evolution, temporal behavior, user characteristics

Introduction

Our approach to exploring insights into the processes of Social Network Systems (SNSs) was presented by a stochastic and dynamic modeling framework integrating the 3 crucial influences: user behavior in response to content, enticement of content, and underlying network structure [1]. A user was treated as a token with attached diverse temperaments. Each had its own specific control behavior, inherited by personality and interstate transition probabilities. These actions were sequenced together as a network of complex human society. Through simulations, the interplay between information diffusion and network evolution was revealed where short-term and long-term system behavior can be hierarchically quantified. Outstanding challenges were revealed in system design and optimization, viral marketing, and network topology prediction.

Attention has been focused on explaining the effects of human behavior on SNSs through the analysis of diffusion patterns. Works on diffusion mechanisms [2-4] showed that information spreading did not depend on the size of network but on networks of friends. A user's role [5] and personality attributes [6,7] were indicated as prime influences attracting social relationships. Such connectivity could be smoothly noticed in the form of in-degree [8]. However, the user's role, social relation, flow of information, and network structure changes were activated by the user's temporal dynamics [4,9,10].

Thus, a complex phenomenon of user relationship evolution could be more accreted if behavioral interactions and the user's similarities [11,12] and individual characteristics [13] were taken into consideration. The crucial factors introducing reactions were the outcomes of messages produced by a few sparse influential users [14].

A user phenomenon in SNSs was explained by divergent parameters, such as social topology, user behavior, inherent content characteristics, and social connections through multiple techniques: data mining, network analysis, and modeling. Nevertheless, none were able to explicate the social power of the individual in reshaping the network structure. Therefore, this study takes a primer step onwards. A formal random activity model of Colored Petri-nets (CPNs) was developed. User behavior, Big Five Personality factors (BFP): a psychological functioning differentiating a user's behavior to another [13], content, and network were converged as rules to impel a future user's actions. The disorder nature of human dynamics over time was mimicked by a functional form of network model. The complexity of human dynamics, internal states of decision-making, and user connectivity were then traced and analyzed.

User processes were divided into 3 main functional components. The *lifetime* handled user life cycle; the *global clock* synchronized user activities within his lifetime; and the *online activity* manipulated specified user interconnections. Each actor, therefore, played a role within its group where this role was expressed in the form of the action it offered. Via the developed models, user interactions in the network were mimicked. Social network analysis was applied to synthesize large-scale streaming data. Subsequently, differential user characteristics, individual social power, and structural and dynamical patterns emerging from human interactions were characterized and expressed for interpretation and generalization. Unanswered questions on how information propagates through an existing network were sought out; (i) what are the effects of divergent user characteristics on the flow of information?, (ii) is social network topology (path length, clustering coefficient, etc.) unpredictably changed by patterns of relationships or does user personality regulate the network evolution? and (iii) to what extent does the spreading process depend on the social power and behavioral profiles of the individual?

The remainder of this paper is organized such that the work on propagation influences and the novel related theories is outlined as *Theoretical Foundation*s; how these specific goals were accomplished is explained in *Experimental Design & Hypotheses*; simulation algorithms and execution procedures are highlighted in *Simulation of Online Participation*; by investigating social structure, not only variant friendship patterns are disclosed, the dynamics user behavior regulated by network structure and probability distributions are evinced and disclosed in *Results and Discussion*; the limitations of the extant valuation study and possible directions for moving the literature forward appear in *Limitations and future directions*, and, lastly, this study is closed with *Conclusions*.

Theoretical foundations

Without effective and efficient contextual information in managing propagation, the effects of divergent user characteristics to the complex network topology become tiring to unveil. Hereby, related literature on users' networking behavior and its effects and theoretical concepts was reviewed. The problems related to the existing architectures were explained, following by a proposed solution.

Social network structure

Prior research had demonstrated that large real-world networks, like social networks, had been studied extensively from a structural point of view. Morales *et al.* [14] performed a quantitative analysis on the dynamic patterns of human activity, and generalized that social activities were fed by a small group of exceptionally active users, but that the majority hardly participated. Seufert *et al.* [15] simulated user communication processes, and exhibited their effects in the form of network topology evolution. Sun [16] expressed that a user's navigability introduced small-world network structural changes. A few studies [17,18] suggested that small-world degrees were interdependent to the dynamic effects of a user's spatial, temporal, and social factors.

These studies, in general, reviewed social relationships in terms of networks consisting of nodes and ties. The series of measurement deployed were size, node degree, community structure, clustering

coefficient, and path length. Though characterizing network anatomy can guide the prediction of future network expansions and give clues to human behavior diversification, its maturity can only be affirmed by the synthesis of human nature and individual social power.

User properties

Several theoretical and empirical studies on human dynamics have proven that user behaviors were dominated by inherited characteristics of personality [7], content [19], and temperament [20]. Liu *et al.* [21] exhibited that the prime factors driving users to share information were individual characteristics and social capital factors. Yuea *et al.* [19] exposed that the closeness of user connections not only depended on the social behavioral interactions, but also on the interestingness of content. Wakefield and Wakefield [20] presented that an individual's passion was the key motivator of social media use, and the role of excitement was a moderator for doing a specific activity. Wiesneth [9] experienced a user's attachment behavior through a message exchanged, and discovered that it not merely originated a preferential attachment in a dynamic network structure, but it positively correlated with user centrality.

Though user behaviors are influenced by diverse determinants, such as biological processes, needs, and temperaments, their differences can be described by BFP factors. The 5 basic dimensions; Extraversion (E), Agreeableness (A), Conscientiousness (C), Neuroticism (N), and Intellect (I) [6,13], can rigorously describe the inclination of user activities. Wang *et al.* [6] observed user communication patterns and expressed that: Extravert was tied to communicative functions; Neurotic was attached to status updating; Agreeable was allied to commenting; and Intellect preferred online games. Lönnqvist *et al.* [7] examined the association of the Five-Factor Model personality dimensions and node degree transitivity, and proffered that the transitivity demonstrated by Extravert was high, whereas that of Intellect was low. Marshall *et al.* [13] explored the traits and motives influencing user activities and presented that different personalities conveyed different types of content creation, such that Extraverts frequently updated personal data, and Intellects often developed intellectual content; however, Neuroticisms brought none.

Given the ideas that the population-based heterogeneity in personal attributes coexists with complex connectivity between individuals, questions are raised about to what degrees the content penetrates user behavior in information authoring and spreading.

Dynamics of content

As interesting and informative content was the key to increasing numbers of readers [12] and initiating further interactions, an explanation of why some content was spread rapidly and wildly has been explored. Szabo and Huberman [22] predicted the popularity of online content using numbers of votes and views, and indicated that content exposure was constrained to a small number of users and its popularity possessed a linear correlation between the logarithmically transformed popularity of content at early and later times. Lerman *et al.* [23] pointed out that a popular story was abruptly visible to many people at a fast rate; however, as the story aged, the accumulation of views slowed down and saturated. Zhou *et al.* [8] traced how a popular topic was diffused and demonstrated that the network was gradually evolved into a highly centralized network with a diffusion center defined as a super group. Kim *et al.* [24] observed the trend of topic sharing, and discovered that each content exhibited different diffusion patterns, while topic-related diffusion patterns reflected different heterogeneity between categories. Wang *et al.* [25] investigated how fast, and to whom, people forwarded information, and proposed that the efficiency of the spreading process was highly affected by departmental structure. Wang [26] suggested that the influence of any post was jointly determined by the features of the post, such as content, created time, and the author's characteristics.

The availability of massive information propagated in SNSs allows the possibility of greater perception of the relation between social behavior and network structure, including how information spreads in networks, how fast and how far it emanates, how the network structure affects information spreading, etc. Without contextual knowledge of the spreading pattern, these questions can be cumbersome.

Diffusion in SNSs

Diffusion in SNSs takes place when a user reads and copies the post of another user, and those posts are read and copied by other users again and again. These complex propagation processes have been widely observed through modeling and simulation. Yuea *et al.* [19] studied user social behavioral interactions, and discovered that the similarities between users could be obtained from interrelated social interactions, and that users with similarities created similar effects on the topological structure. Rúbio *et al.* [11] classified users with different propagation behaviors into groups, and detected that the largest group were non-participative. Yi *et al.* [3] presented that the community structure was an important indicator for information diffusion. Lee *et al.* [12] pointed out that a user's roles and characteristics could be inferred by user-generated content. Wang *et al.* [27] suggested that the differences in human ranks and roles had disparate effects on the propagation mechanism.

It is obvious that diffusion is definitely connected to human social relationships and activities. Thus, a sufficient background in network metrics and relevant statistical analysis is essential to answer the research questions.

Centrality measures

Centrality is a metric used to rank nodes. In some order, it can be used to specify which is the most connected or important node to the network as a whole. In this study, the metrics applied were "*Betweenness*" and "*Closeness*" [28]. One of the novel metrics discarded was "*Degree*", as it is a measure of immediate influence, as opposed to global effect. A brief description is provided as follows:

1. *Betweenness Centralities (B_C):* Paths are crucial in understanding how far apart 2 people (nodes) are within a network. The Betweenness centrality of an edge can be viewed as a metric to quantify the importance of an edge in the network. An edge with higher Betweenness centrality falls on many shortest paths, and is, therefore, more important to the network structure. This indicates that there is a very critical individual in the transmission, and its removal can disrupt the flow of information.

Mathematically, B_C of an edge e within a network is defined as the ratio of the total number of shortest paths between all pairs of vertices in the graph that cross e to the total number of possible shortest paths between the 2 nodes that include e;

$$B_C = \sum_{u \in V, v \in V} \frac{\phi_e(u, v)}{\phi(u, v)}$$

(1)

2. *Closeness Centrality (C_C):* This centrality captures the average distance between a given node and every other node in the network. For example, if nodes are assumed to be channels through which information are relayed, then the Closeness centrality of a node shows how fast it will take information to flow directly from that node to all other connected nodes. High Closeness centrality scores are associated with nodes at the outer perimeter of the network. The higher the scores, the longer the time it takes information to flow to the core of the network.

Mathematically, C_C can then be seen as the mean shortest path between a given node v and all other nodes that are within its reach.

$$C_C = \sum_{t \in V \backslash v} \frac{d_G(v, t)}{n - 1}$$

(2)

Statistically significant differences between groups

The conclusion whether or not user behavior variables were significantly different between groups were inferred by 2 inferential statistics; (a) ANOVA- if the shape of distributions is normally distributed, or (b) Kruskal-Wallis H Test, which is the nonparametric one- if the shape unfits to normality [29]. Although (b) is not an earnest analysis of variance, it can be described as an "*analysis of variance by ranks*" resembling (a). Additionally, a few assumptions must be met prior to a solid result being

confirmed: (1) the scale on which the dependent variable is measured has the properties of an equal interval scale; (2) the independent variable consists of more than 2 categorical of independent groups; (3) there is no relationship between the observations in each group or between the groups, and (4) the distributions of the independent variables in each group govern the same shape. Lastly, the determination whether any of the differences between the means are statistically significant shall be expressed as the *p-value* with a significance level of *0.05 ($p < 0.05$).*

Contributions

The models of realistic systems, like SNSs, were often perceived as being extremely large and complex. The previous works on human behavior often simulated a user's participation with a specific set of events. These stipulations hindered the solution-finding process and returned illusive perceptions of model functionality. Therefore, a structural approach, with sound and explicit conceptualizations, were defined, such that actors, users, and content were different kinds of autonomous elements. Each had clear rules for transitions. Transitions are links brewed from a microscopic level of social relationship. Each link was induced by diverse temperaments ruled by personality and emotion energized by content. The intricacy was disposed by placing users with similar behaviors into groups using unsupervised clustering. User activities were parameterized using cluster-based reduced order modeling, where real-world data was transformed into control variables and restated as mathematical equations. The equations were tied together into stochastic models integrating user behavior, network structure, and content as primary influences [1]. Both static and dynamics structures of actors were captured thoroughly by model analytical technique. The occurrence of user lifetime activities was simulated. The flow of information in the network was demonstrated as graphical illustrations and assured by statistical analysis. This concept has the foremost robustness of diffusion simulation, which can extend finer granularity to the existing approaches.

Experimental design & hypotheses

Investigation of the effects of user's characteristic on dynamic socializing patterns is framed and demonstrated as an experimental design framework in **Figure 1**. Expectations were defined as hypotheses and proven via the examination of network structural properties.

1. Conceptual design: For better understanding of diffusion behaviors, user social activities were broadened into events. A user's next event was reflected by differential reaction modes of neighbors, temperamental behavior dominated by personality, and subjective orientation of attitudes.
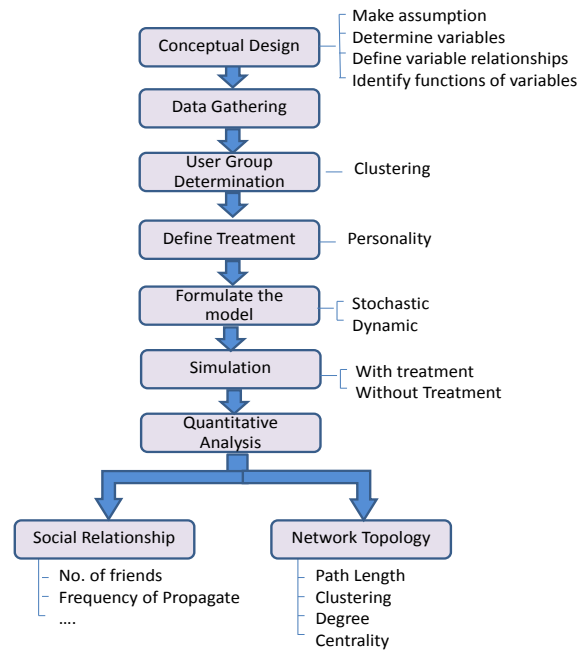
Figure 1 A proposed framework of user social relationship in SNSs.

2. Data gathering: A model which could closely represent the real-world system was focused on. Most of the user salient features were incorporated through the following steps:

- Questionnaire launched: The randomness characteristics of model behavior and user interactions in Facebook were collected by a questionnaire. Attentive information comprised the user's demographic attributes, BFP scores, reciprocating content-driven activities, and interchangeable indicators.

- Data collection and verification: For ease of access to adequate samples with inordinate cases, university students in Bangkok and metropolitan areas were recruited as input elements. Data was then verified using statistical analysis.

3. User Group Determination: The complexity, nonlinearity, and disparity of determinants derived from user characteristics were diminished by partitioning users with respect to their similarities. Correlations between social activities and personality were discovered. Such values guaranteed the empowerment of personality to a particular diffusing action, where the declaration of correlations between user groups and personality attributes is displayed in **Table 1**.

Table 1 Group characteristics.

Group	Definition	%	E	A	C	N	I
1	*Chummy Users*	6.7	Medium	High	Medium	Medium	High
2	*Inert Users*	59.0	Medium	Medium	Medium	Medium	Medium
3	*Socialize Users*	2.8	High	Medium	Medium	High	Medium
4	*Introvert Users*	14.4	Low	Medium	Medium	Medium	Low
5	*Cool Users*	9.0	Medium	Medium	High	Low	High
6	*Brainy Users*	7.8	Low	Medium	High	Low	High

4. Treatment declaration: A clear and objective comparison was achieved by setting up the experimental cases to include 6 personality treatment groups and a control group. User contexts, such as characteristics, number of friends, and behavior were assumed to be equivalent for each group.

5. Model formulation: The models were constructed according to SNS features and historical user behaviors. A user's decision to act or react depended on temperaments derived from content. A single content could be interpreted differently, depending on the differential characteristics of users. For example, a person with high Neuroticism might think the content was torturous, in contrast to one with high Intellect.

These nondeterministic and dynamic variables were characterized by a family of probability distributions ($\mathbb{P}$) [1], such that there existed some (open) subset of a finite dimensional Euclidean space, say Θ such that $\mathbb{P}_\theta = (\theta \in \Theta \subset \mathbb{R}^d)$ in other words, each parameter in Θ was associated to $\theta \in \Theta$ where θ was derived using the method of maximum likelihood estimation. Therefore, the identifiable parameters were defined according to their group activity distributions, such that $\Theta = \{create/distribute, share, response, add friend, remove friend, browse\}$.

The steps towards the formulation were classified as;

• Identify functions: the complexity of the social network functions was lowered by segregating those functions into sub-models. An unknown parameter was estimated via a model-fitting method, and the transition between states indicating user interaction was facilitated by a stochastic equation.

• Determine variables: A user's operations associated with diffusion and friendship relations were defined as dependent variables, whereas time was treated as an independent variable.

• Establish relationship between variables: Two types of relationship were determined; i) the relationship between user and content; such as browse, create/distribute, share, and respond, and ii) the relationship between users, such as add and remove friends. Somehow, both actions were the aftermath of content reciprocation.

• Modeling: The dynamic user friendship relation, an underlying formation of information diffusion, was developed using the computational, graphical, and stochastic capabilities of CPNs.

6. Simulation: Groups of users with similar contexts were assigned into the model simulator one at a time. For system behavior comparable, user transitions in the form of time varying interactions between actors and content were recorded into a log file.

7. Quantitative analysis: The differences in outcomes between groups were analyzed and observed to identify dynamic patterns of node behavior and to predict future structural changes.

By analyzing a user's behavioral interactions, how user friendship influences information flow and its impact on network topology is attained through the process of hypothesis testing.

Goals & hypotheses

The uniqueness of individual effects on network structure was prescribed by 3 crucial goals.

1. *Users behave randomly with respect to personal characteristics.*

The first aimed to provide guidance to the prediction of content flow and the trend of content distribution. Since a user's decision to act in any circumstances was not predefined by any specific rules, their socializing patterns were investigated by aggregating an individual's activities in generating social links between user and content, such as browse, share, and response.

Imagine a user reads content. He may communicate a few ideas, leading to interactions; for instance, giving a "*Like*", or responding, forwarding, or sharing that content. These interactions leave explicit traces which can be measured in terms of intensity of activities and number of friends. Such complexity, in general, governs the scale-free property [30], where predefined expectations are:

H^{11}: *High frequency of propagation positively correlates with Socialize user.*
H^{12}: *High frequency of social relation positively correlates with Socialize user.*
H^{13}: *High number of friends positively correlates with Socialize user.*
H^{14}: *Much cogitation before adding friends positively correlates with Cool user.*
H^{15}: *High number of views positively correlates with Socialize user.*

2. *Different user characteristics influence network topology differently.*

The second proposition aimed to disclose the future network topology. This could be verbally referred to from its node's dynamics, where nodes become densely-connected to their neighborhood according to the time and relationships they contributed. When more links were created, the shortest paths appeared, and node interactions formed into triangles with degree distribution stepped up in the power-law form [16,17], so called the small-world property. Thus:

H^{21}: *High declination of average shortest path length positively correlates with Socialize user.*
H^{22}: *High clustering coefficient positively correlates with Socialize user.*
H^{23}: *High degree distribution positively correlates with Cool user.*

3. *Different user behavior introduces different social power in the network.*

The last aimed to rank the node's position in proportion to the others and to the whole network using the metric of structural centrality. By quantifying the importance of nodes, their connections with others could be predicted. This metric could further explain content diffusion patterns and the specific characteristics of users in a group [31]. By understanding the social involvement of users, advertisers can suggest a vigorous plan for effective product recommendation. Hence:

H^{31}: *High Betweenness centrality positively correlates with Chummy user.*
H^{32}: *High Closeness centrality positively correlates with Brainy user.*

The processes towards the proof-of-concepts are elucidated in the following section.

Simulation of online user participation

Understanding of how a user traversed the network was accomplished by simulating the user's sessions in an SNS. A complete profile of user interactions was traced. Types, frequency, and sequences of user activities were analyzed to examine if the hypotheses defined were precise.

Stochastic replication algorithms

A framework of a Discrete-time Event-driven Simulation model (DES) is presented as a replication algorithm in **Figure 2**. Users with similar characteristics functioning in content reciprocating (Θ) were assigned into the network, one group at a time. A user might traverse from one state to another according to the probabilities derived from real-world samples. This probability-driven state gave rise to the nondeterministic and asynchronous activation of their neighborhoods. In one step a single event could be progressed with respect to interaction rules. With a focus on the temporal structure instructed by the occurrence of user interconnectedness, these processes were automatically repeated with the configurable parameters, as demonstrated in **Table 2**.
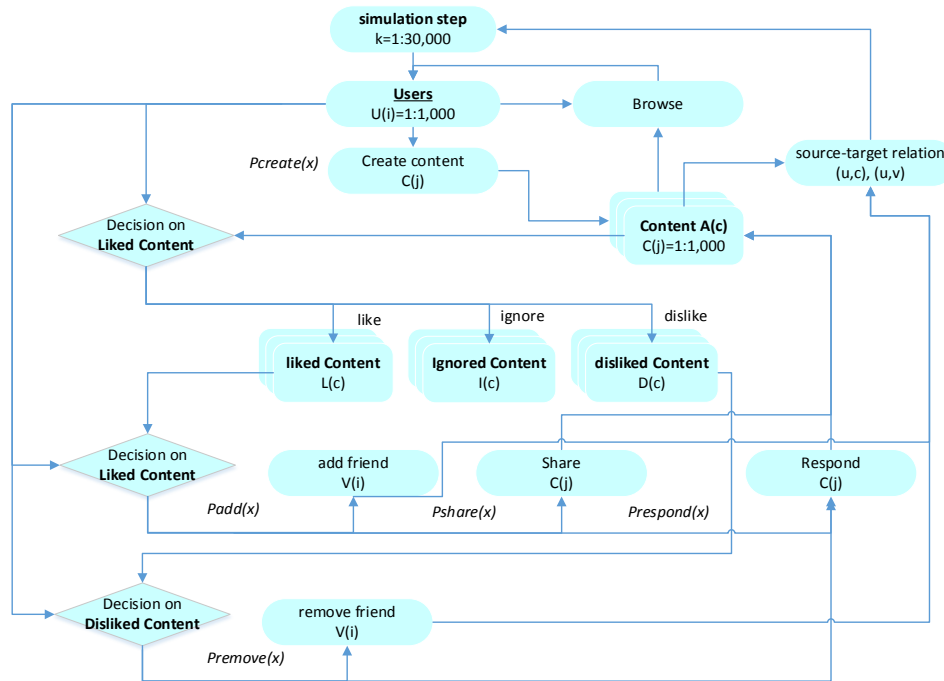
Figure 2 Reproduction of user activities using model parameters derived from real-world data.

Table 2 Configurable parameters.

Parameter	Type	Definition	Description
U	User	$u_i \in U\ (i=1..1000)$	1000 user nodes
C	Content	$c_i \in C\ (i=1..1000)$	1000 content nodes
K	Step	$k \in K\ (k=1..30000)$	Simulation step
T	Time Unit	$t_i = (t_{i-1}+\Delta t) \in T\ (\Delta t = 15\ min)$	Time to switch action
$A(c)$	Content	$c_i \in A(c)$	All Content
$L(c)$	Content	$c_i \in L(c) \subset A(c)$	Liked Content
$D(c)$	Content	$c_i \in D(c) \subset A(c)$	Disliked Content
$I(c)$	Content	$c_i \in I(c) \subset A(c)$	Ignored Content

At $t = 0$, a thousand users (u_i) in one of the predefined groups were assigned to an online in idle state. No individual user knew another upfront. When the simulator started, a user was randomly selected to be active. An event a user conducted was declared as a stochastic formulation of activity within the boundary of the social network constraint. The sequence of events was sorted by the simulated time. Its occurrence could only begin when content existed in the content stack $(0 < c_i \leq 1000)$. When the content was recognized, a user temperament was triggered by the rhythm of the content. This relationship was annotated as (u_i, c_i).

As an example, a user might give a '*Like*' to the content and share it with friends, dislike it and respond negatively, or disregard it. Each decision forced the content to be virtually transferred to a predefined storage, such that $I(c)$-ignored content (no transmission ability), $L(c)$-liked content (prepared to be forwarded, responded, or shared) and $D(c)$-disliked content (repulsed), as shown in **Figure 3**.
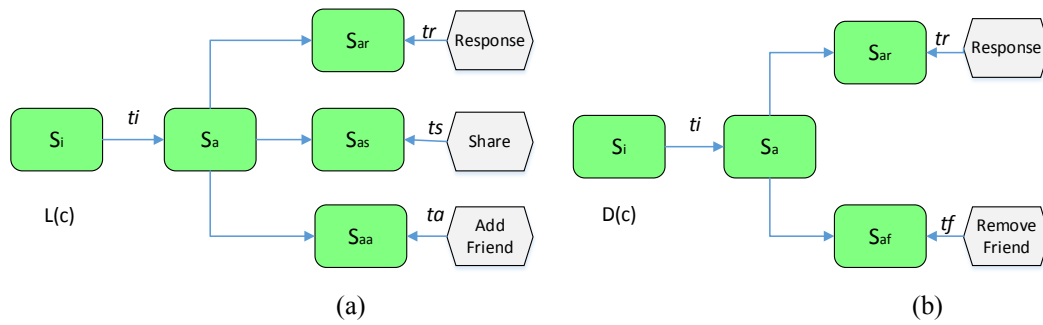
Figure 3 Transition probability trees for user node state changes from S_i (idle) to S_a (active): (a) is the state driven by liked content, and (b) is the state driven by disliked content.

$$S_i \xrightarrow{\alpha} S_a \qquad S_a \xrightarrow{1-\alpha} S_i \tag{3}$$

In each simulation step, an inactive node could interact with their neighborhood nodes if a content node existed in *L(c)* or *D(c)*. However, if the content node of its neighbor was removed, such node no longer had ability to make other nodes active. By this assumption, content in *L(c)* and *D(c)* governed high potency to drive a user to change a state from inactive (S_i) to active (S_a), with probability (α) as illustrated in Eq. (3). The probability (α) originated from a user's decision (node) on content, such as beneficial, annoyed, or discouraged. On the other hand, a node could be in an active state with α, and moved to an inactive state with *1-α* depending on rules and contextual information.

Table 3 Transition states of information propagation.

Transition probability	Description
t_i	a node will change state from inactive to active
t_r	a node will respond to content
t_s	a node will share content
t_a	a node will add a content owner as a friend
t_f	a node will remove a content owner from friends

User transitions (**Table 3**) are illustrated, such that: if the content status is *liked*, a user node can be positioned in one of 3 different states: *response (S_{ar}), share (S_{as}),* and *add friend (S_{aa})*. If its status is *disliked*, a user node can be positioned in one of 2 states: *response (S_{ar}),* and *remove friend (S_{af})*.

Execution and preliminary observation

The simulator was successively executed in fast mode until *k* reached its maximum. The simulation time was conserved by discarding real-world user lifetime synchronization and declaring *Δt* as *0*. At each sampling level, the simulation was run over 20 rounds. In each round, a synthetic user-content relationship file was generated and examined for unexceptional cases. This data was then loaded into the database for ease of data manipulation and complex enquiry. Through database analytics, high dimensional data was transfigured into a source-target mapping and an adjacency matrix.

It must be noted that, at lower sampling proportions (10 - 30 % of data), a few measures were not eligible, as the data was too sparse to be gauged. Nevertheless, it was always the case that some of the 20 samples provided enough ties for the measures. Eventually, via the analysis of networking data consisting of sequences of user, interaction, reciprocating content, its neighbor, and emotional state, 3 essential user properties could be recognized; the correlation between user characteristics and behavior, the influence of user characteristics on the network, and the social power of distinguishable user characteristics.

Results and discussion

The network of user and content nodes at low simulation steps *(k = 500)* is demonstrated in **Figure 4(a)**. The links between nodes were exiguous. As time increased, several nodes were more active in introducing content. Through the processes of friendship relation, a denser network with a large number of nodes and links appeared, in **Figure 4(b)**. Different colors and node sizes indicated distinctive node degrees. The larger a node size, the higher the connections the node dominated. Taking a closer look into the graph structural features, this network demonstrated the small-world property. Its path length and diameter started with ∞ and decreased to 3.23 and 6.00, respectively. Its graph density and clustering coefficient started with 0 and accreted to 0.01 and 0.11, respectively. Such preliminary diffusion results indicated that this network captured many properties of real-world complex networks.

The following experiments were set up to verify if information flowing in SNSs was randomly formless or predominated by certain personality traits. Insights into how traits harness SNS participation not only guide the prediction of future network structures, but also energize buzz ideas to launch a product.
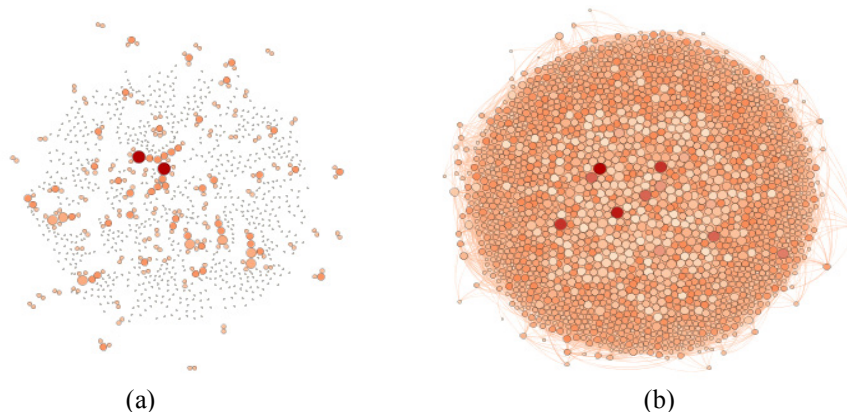

(a)   (b)

Figure 4 Network graph of the simulated network with autonomous simulation (a) at *k = 500* and (b) *k = 30000*.

Influences of user characteristics to social interactions
Social interactions of focus were the number of posts, shares, likes, comments, number of friends, and friend acquisition process. Graphical representation was adopted, and the scale-invariant property of the generic distributions was analyzed and affirmed by statistical hypothesis testing.

Propagation behavior
Users' actions, such as create/distribute, share, and respond were extracted as propagation behavior. **Figure 5(a)** presents the propagation phenomenon of imbalance. Very large number of users were found to propagate just once during their lifetime, and a few distributed over 20 times. This behavior gave rise to a power-law property of ($1.796 \leq \gamma \leq 2.311$) similar to [11]. Gaining insight into individuality, the groups with high Extraversion or high Agreeableness often ranked high in spreading, as shown in **Figure 5(b)**. This attribute is a positive sign for high energy in social relationships, similar to [6].
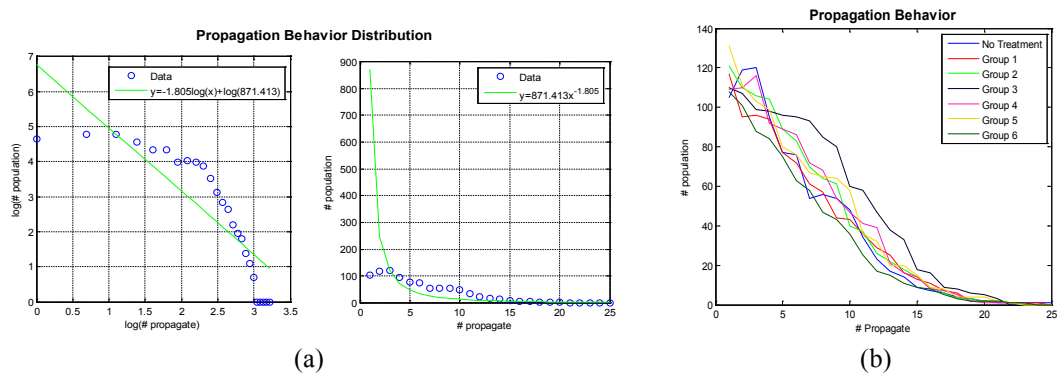
Figure 5 (a) Propagation behavior distribution and (b) propagation frequency per group.

When it came to the question on how long it took any single content to spread to everyone in a group, individual characteristics were found to be directly associated. For example, if an action was assumed to be switched every 15 min, the single content could be spread to everyone in 270 days for no treatment group, 198 days for Group 1, 244 days for Group 2, 125 days for Group 3, 239 days for Group 4, 203 days for Group 5, and 250 days for Group 6. Intuitive investigation indicates that this nature relates to one's characteristics. Thus, if an advertisement is expected to be fed in at a fast rate, companies should focus on customers with high Extraversion and high Agreeableness.

Socializing pattern

Referring to the concept of traffic, the active nodes are nodes who act as social hubs, while the inactive nodes are nodes without neighbors. Hereby, the indicator for socialization is declared as the disappearance of isolated nodes as time evolves. **Figure 6(a)** depicts that a similar power-law phenomenon with scale-invariant of $2.16 \leq \gamma \leq 2.36$, which complies well with [32]. This behavior suggests that such variations are controlled by different interactions among groups, which is not normal distribution [33]. It implies that the higher the number of users' interrelations, the lower the number of isolated nodes.
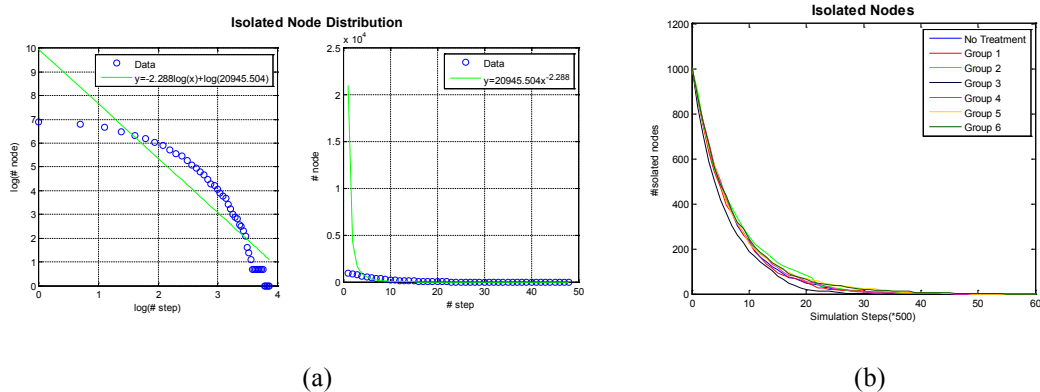

Figure 6 (a) Isolated node distribution and (b) isolated node disappearance per group.

The number of node connections grows fast during $k = 0 - 20,000$ and moves towards a steady state afterwards, as illustrated in **Figure 6(b)**. One interesting finding was that, when new content appeared, neighbor nodes were curious, and rapidly interacted with it. As the content aged, the more unattractive the

content was. The number of interactions faded out. It can be claimed that the motivation for a node to create a link not only depends on the interestingness of content, but also on its freshness.

It is also of merit for a company to know how long it will take a user to interact with at least one content during his lifetime. Thus, the disappearance of isolated nodes is shown in **Figure 6(b)**. The time taken for everyone to read the content was 250 days for no treatment group, 281 days for Group 1, 244 days for Group 2, 177 days for Group 3, 239 days for Group 4, 198 days for Group 5, and 265 days for Group 6. As different personality traits induce different level of socialization, viral marketing should focus on people with high Extraversion, high Intellect, and high Conscientiousness, in order for the campaign to reach everyone within the community rapidly.

Number of friends
In order to answer a quantitative question on how big these networks can grow, it is perspicuous to explain it in the context of the number of friends. Users with high numbers of friends tend to spend more time, and have more activities, in the network; thus, they govern high ability to make significant changes to the network. The number of friend distribution possessed a power-law shape which is agreeable to [34], with scale-invariant of $(1.81 \leq \gamma \leq 2.51)$, as shown in **Figure 7(a)**. This relationship is the signature of a complex social relationship that leads to a state of self-organized criticality.
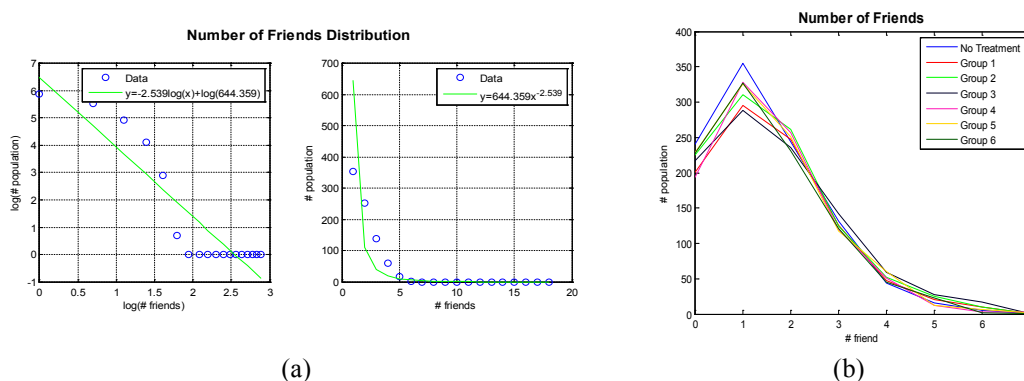

Figure 7 (a) Number of friend distribution and (b) number of friends per group.

When considering how many friends are typical for SNS users, **Figure 7(b)** demonstrates that many people have only a single friend, while a few have as many as 7. Though these values were not able to represent the actual number of friends in Facebook (with a 5000-friend limitation), it could explain the real nature of these samples. Many users have only a few friends in their communication network, which is lower than average [34]. Taking a closer look, the Socialize users are found to carry the highest numbers of friends, which is a result of the higher the Extraversion they govern.

Decision to be friends
Though the 6 degrees of separation concept states that people in this world can be connected within 6-7 degrees of each other [35], it is informative to know how long it will take 2 people to become friends. Hence, the amount of information exchanged between a pair of people before a strong relationship could be initiated was investigated. **Figure 8(a)** demonstrates that its pattern possesses a power-law property similar to [22], with scale-invariant of $2.37 \leq \gamma \leq 2.79$. This form indicates that many users will add a friend upon reading a single complacent content, whereas a few proceed with more interactions. This phenomenon may slightly be different from real-world behavior. They usually exchange a certain number of messages before finalizing anyone as a friend. However, their neighbors in this study are current classmates, and a fast decision is conceivable. **Figure 8(b)** presents that an add-friend cycle varies significantly with regards to individuality.
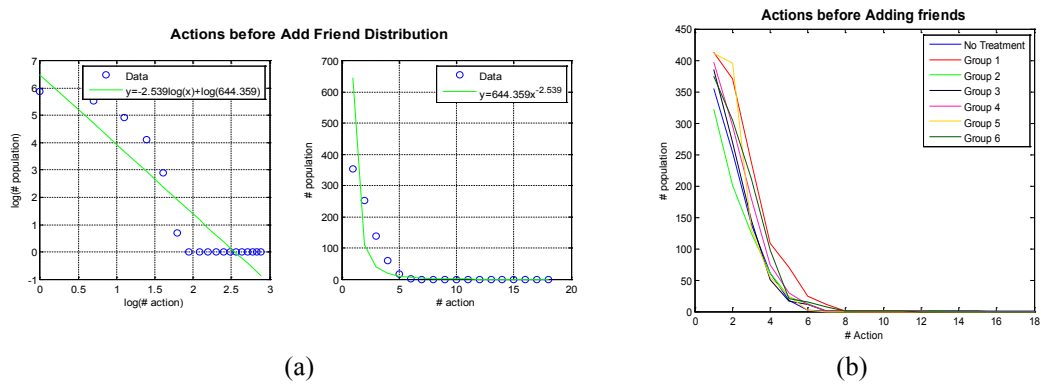
Figure 8 (a) Number of actions before adding friend distribution and (b) number of actions per group.

Content popularity

People often categorize content based on past popularity and user appeal [22], which is informative to content providers and advertisers. This study experienced the popularity of content through the frequency of browsing, giving 'Likes', commenting, and redistributing. **Figure 9(a)** reflects the distribution of the average normalized popularity for all messages over time. Its popularity possesses a power-law with scale-invariant of $2.16 \leq \gamma \leq 2.87$ similar to [25,36].
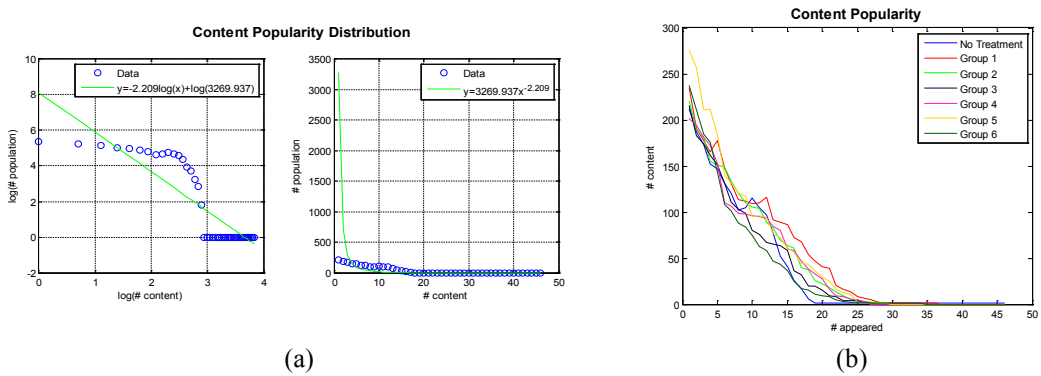

Figure 9 (a) Content popularity distribution and (b) number of content viewed/shared per group.

It can be noticed that most content had less than 10 views/shares, whereas very few had over 200, as depicted in **Figure 9(b)**. During $0 \leq k \leq 20000$, high numbers of views were detected, and those views moved quiescently afterwards. Thus, a great deal of content incessantly disappeared, and only a few appealed to users. Therefore, publishers should be aware that it is not easy to make attractive content. This phenomenon not only indicates that only a small fraction of content achieves high popularity and, thus, has potential to spread widely through SNSs [37], but the wisdom of users also provide collaborative filtering to select flavored submissions. Moreover, high numbers of views are associated with people with high Extraversion. Implying that, though SNSs are good media for commercials, they are one of the most essential hindrances, leads to questions on how to design content that can draw many peoples' attention.

Though group behaviors in the form of graphs, illustrated in **Figures 4(b) - 9(b)**, provided meaningful information, they were not able to reveal differences between groups explicitly. Statistically significant differences in those non-Gaussian (power-law) distributed variables were, thus, accessed by One-way analysis of variance, a Kruskal-Wallis H Test.

Table 4 Kruskal-Wallis H test of statistically significant differences in user behavior.

Group	Propagation		Isolated Node		Number of Friends		Actions before Add		Content Pop.	
	N	Mean Rank	N	Mean Rank	N	Mean Rank	N	Mean Rank	N	Mean Rank
0	27	75.31	61	208.43	796	2592.22	8	33.28	46	172.33
1	24	99.19	61	224.89	746	2744.71	6	45.55	46	201.84
2	22	93.57	61	213.02	783	2740.83	15	36.27	46	195.85
3	21	107.67	61	172.62	769	2867.91	18	33.28	46	265.05
4	27	77.50	61	210.85	774	2635.97	13	41.58	46	205.61
5	22	94.88	61	228.29	775	2667.71	13	62.50	46	219.05
6	29	67.67	61	236.80	750	2633.00	11	58.44	46	213.77
	Chi-Square =12.883		*Chi-Square =14.483*		*Chi-Square =18.997*		*Chi-Square =11.146*		*Chi-Square =20.314*	
	df =6		*df =6*		*df =6*		*df =6*		*df =6*	
	p-value =.045		*p-value =.048*		*p-value =.004*		*p-value =.041*		*p-value =.002*	

0 = no treatment group, 1 - 6 = groups with personality treatment, p-value < 0.05

Through the analysis of mean ranks and p-values shown in **Table 4**, it can be assured that the predefined essential behavior showed statistically significant differences between groups. The number of sample sizes (*N*) in a particular behavior was different according to the pair of nodes originating such activity. For all observed cases, the asymmetric significant differences (*p-value < 0.05*) can clearly indicate the stochastic dominance between the samples.

- Propagation Behavior: $\chi^2(6) = 12.883$, $p = 0.045$

Group 3 governs the highest mean rank. This phenomenon indicates that the highest propagation rate is initiated by the *Socialize users* while the *no-treatment* group propagate the least. Thus, H_{11}(High frequency of propagation positively correlates with Socialize user.) is supported and complies with [26].

- Isolated Node: $\chi^2(6) = 14.384$, $p = 0.048$

Group 3 shows the lowest mean rank in contrast to Group 6. The lower the value, the higher the number of isolated node disappearances, implying that the *Socialize users* are the most active and the *high Intellect* are the least active. Hereby H_{12} *(High frequency of social relation positively correlates with Socialize user)* is accepted and is agreeable with [6,7].

- Number of Friends: $\chi^2(6) = 18.977$, $p = 0.004$

Group 3 shows the highest number of friends. This attribute shows successive events of propagation. As this group occupies the highest propagation rate, they tend to create more friends along with such a process. Hereby, H_{13} *(High number of friends positively correlates with Socialize user)* is accepted and is well matched with [3,4].

- Actions before Adding Friends: $\chi^2(6) = 13.146$, $p = 0.041$

The highest mean rank belongs to Group 5, indicating that its members are more discreet when they reach the state of making the decision to add friends. Hence, neighbors of this group need to gain more

trust before they can be friends. Therefore, H_{14} *(Much cogitation before adding friends positively correlates with Cool user)* is confirmed.

- Content Popularity: $\chi^2(6) = 13.146$, $p = 0.002$

Group 3 occupies the highest mean rank of in-degree, indicating that people with high Extraversion govern high curiosity. They are eager to know what is new in their timeline. Hence, H_{15}: *(High number of views positively correlates with Socialize user)* is proven.

The presented friendship paradigm is an interesting sociological phenomenon which can clearly explain user's activities in SNSs. The results comply well with empirical results where many human activities govern the power-law distribution [30], in which the tendency of activities is high as their scale-invariants increase. The distinguishable capability of this model is that its process can start from scratch. Beginning with a disconnected network, the edges can still be rewired through the attractiveness of content.

Focusing onto one's personality trait, the higher the Extraversion a user governs, the higher the number of friends, the higher the chance of information spreading, and the higher the social relations. One of the most amazing findings is that a user generally adds friends when he is satisfied with just one congenial thought or idea. However, vast amounts of content spread in this network is meaningless; in consequence, the chance of a user adding a new friend is low. In addition, as the number of friends has often translated to social status, the demonstrated friendship paradox can provide interesting applications for immunization and influence maximization as well.

Inferring network topology evolution from user characteristics

A strong characterization of the relationship between a user and neighbors can be quantified as a network topology [33]. Generic properties of the network structures within spreading processes were observed; the clustering coefficient, the degree distribution, and the path length. These values are the basic time-dependent network characteristics which can quantify the real-world network.

Path length

Path length is generally used as an indicator of node connectedness. When nodes are tightly connected with the interactions they have with neighbors, the link formation that bridges the disparate nodes of the network can naturally be explained by low average path length. **Figure 10(a)** shows that the path length distribution is well fitted with the Pareto distribution. This phenomenon was slightly different from the previous study which modeled with Weibull [38]. However, both distributions governed some similarities, a long righted-tail that indicated a non-Markovian process. This process is not only a stochastic process with memoryless property but a sign of a certain complexity to the process of estimating a "typical value".
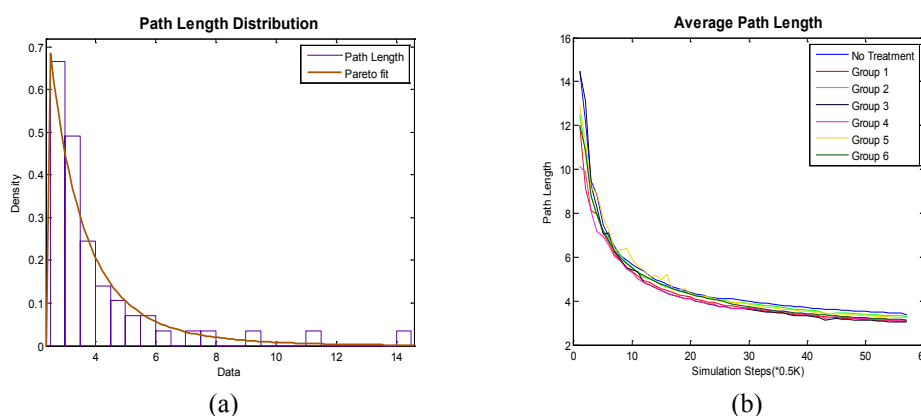

(a)   (b)

Figure 10 (a) Path length distribution and (b) average path length per group.

Figure 10(b) illustrates that the path lengths of all groups abruptly decreased from infinity to approximately 3, which is relatively low. As *k* increased from *0* to *5,000*, over 80 % of users in the network started exchanging content with neighbors. Afterwards, the network quickly gathered edges, and its average path length drastically decreased and moved towards a steady state when *3< path length<4*. This property may again have resulted from a high degree of reciprocity within the network.

Clustering coefficient

The clustering coefficient of a graph is the average clustering coefficient of all its nodes, which is the measure of the extent to which one's friends are also friends of each other. If *G* represents friendships among people, the clustering coefficient $C(i)$ measures the ratio of the number of friendships between any 2 of *i*'s friends, relative to a situation where all these friends induce a complete subgraph of *G* [33]. **Figure 11(a)** shows that its distribution does not fit with any parametric distribution. The whole pattern consists of 3 different scaling regions at $C(i) = 0.002, 0.04$, and *0.11*.

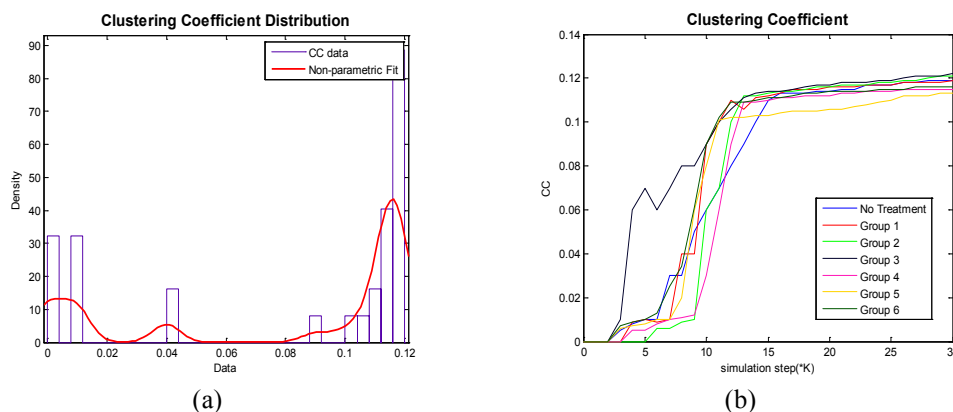(a)   (b)

Figure 11 (a) Clustering coefficient distribution and (b) clustering coefficient per group.

The network average clustering coefficient was *0.11*, smaller than those of other real-world networks, implying that the friend relationships in this network are more loosely connected. This aftermath is a result of a classmate crisis where people are more connected in real life. Their favorite contacts are more likely to be face-to-face or conducted via WhatsApp or mobile phone; thus, SNSs are not their first choice.

Figure 11(b) shows that the clustering coefficient immensely increases during *k = 5,000 - 10,000* for most groups. In particular, when *k* reached *8,000*, the number of user connections in the network exceeded 80 %. The group with the highest jump of triangular formations (friends of friends connections) during *k = 2,000 - 3,000* was Group 3, as it occupied the highest number of propagandists, implying that some content may be able to draw a neighbor's attentions, leading to a high jump in the node linking process. However, after *k = 10,000*, the clustering coefficient moved towards stabilization, because edges were already added for all nodes in a neighborhood and there were no more connections available.

Degree distribution

The degree distribution of the friendship network indicates that the number of link points in or out of a node. Hereby, a node degree was defined as the number of content viewed or created. Their degrees possess a power-law property similar to [35], as depicted in **Figure 12(a)**. Their scale-invariants were $1.62 \leq \gamma \leq 2.22$, meaning that only a few nodes had high interactions. **Figure 12(b)** provides more detail of node activity per group. There are many nodes with degrees less than 10, and there are a few nodes with degrees more than 50.
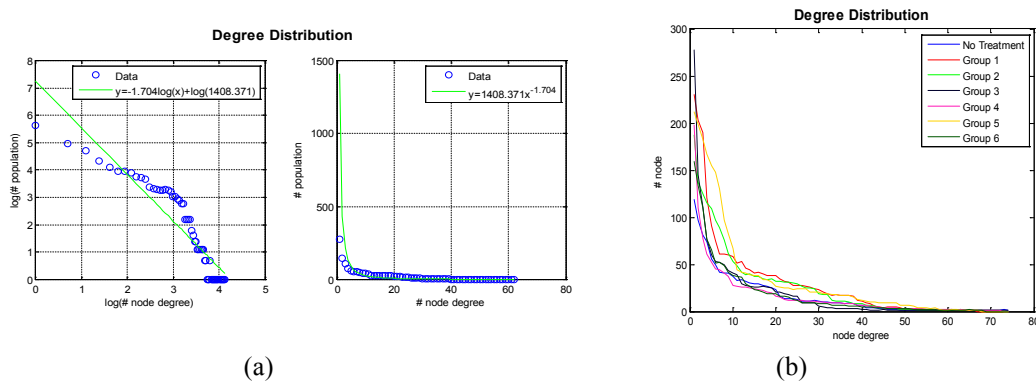
(a)   (b)

Figure 12 (a) Node degree distribution and (b) Node degree per group.

Though some of the networks created by this model had relatively small power-law behaviors in comparison to those of most networks in nature and human societies ($2 \leq \gamma \leq 3$), it resembles that of email messages [35] and the Cyworld network [39]. An explanation of the lower scale-invariants is that a user's motivation for content spreading is rather low and occurs infrequently because they prefer to connect by other routes.

Figures 10 - 12 may be able to express the evolution of network topology graphically; however, statistically significant differences between groups can provide more legible guidelines and, therefore, **Table 5** is presented.

Table 5 Kruskal-Wallis H test of statistically significant differences in network properties.

Group	Path Length		Clustering Coefficient		Degree Distribution	
	N	Mean Rank	N	Mean Rank	N	Mean Rank
0	57	238.26	31	108.74	74	221.91
1	57	180.04	31	115.15	66	286.98
2	57	218.23	31	114.76	72	253.78
3	57	171.28	31	129.63	74	198.98
4	57	181.46	31	92.60	72	238.35
5	57	214.10	31	78.84	67	297.60
6	57	196.32	31	113.36	68	239.88
	Chi-Square =15.361		*Chi-Square =13.596*		*Chi-Square =25.279*	
	df =6		*df =6*		*df =6*	
	p-value =.018		*p-value =.034*		*p-value =.001*	

0 = no treatment group, 1 - 6 = groups with personality treatment, p-value < 0.05

Table 5 demonstrates statistically significant differences in network properties of interest, including path length, clustering coefficient, and degree distribution. *N* of each group in the specific networks was different, originating from the frequency of content exchange.

- Path length: $\chi^2(6) = 15.361$, $p = 0.018$

Group 3 demonstrates the lowest mean rank, indicating that the longer time a user spends in the network, the more conversation he makes. As a result, the path length is depreciated. Hence, H_{21}: *(High declination of average shortest path length positively correlates with Socialize user)* is granted.

- Clustering Coefficient: $\chi^2(6) = 13.596$, $p = 0.034$

Group 3 occupies the highest clustering coefficient, indicating that the more responsive the people are, the higher chance for triangular relationships (friends of friends). Thereby, H_{22} *(High of clustering coefficient positively correlates with Socialize user)* is accepted.

- Degree Distribution: $\chi^2(6) = 25.279$, $p = 0.001$

The group with the highest degree is Group 5; though one may think that a user with few friends will be less active than a user with more friends, the result can be observed in the case where activities are defined as both sending and receiving. There are nodes with large degree yet where very few messages are exchanged, and there are nodes with very small degree yet where a lot of messages are posted. This result agrees with [2,3], where users with moderate numbers of friends and moderate Extraversion occupy the highest degree distribution. Then, H_{23}*(High degree distribution positively correlates with Cool user)* can be referred to.

The aforementioned network properties have no potency to be saturated, but they tend to move slower as the number of connections reach the full capacity of the network. In addition, the efficiency of the model in generating a complex network with real-world properties is affirmed via these signatures.

Inferring social power of user characteristics

Centrality metrics are important for understanding the trends in relationships and information spreading. In network analysis, each centrality can tell different stories of a group's role. By investigating the characteristics of centrality, the pattern of information flow introduced by users with differential attributes can be ascertained.

Betweenness centrality

In this study, a group's behavior is of more interest, as demonstrated by the Betweenness distribution of each user group in **Figure 13(a)**. This fitted well with the lognormal distribution, which was different from the previous power-law [31]. However, both demonstrated a long-tail property, meaning that there was an inverse relation between the frequency and the number of events. **Figure 13(b)** demonstrates the social power of all nodes in descending order as log-log plots. Their curves slowly declined as their Betweenness decreased from 10^4 to 10^3 and abruptly declined after the centrality were lower than 10^3. An explanation for this is that most nodes exchange or view at least one content during a lifetime, and only very few nodes remain disconnected. The higher the Betweenness, the more well-connected the nodes are. The group with higher Betweenness have considerable influence within the network by means of controlling information passing to others. Therefore, advertisers should focus on this standard social science measure to discover how important one's individuality to the information flow within SNSs is.
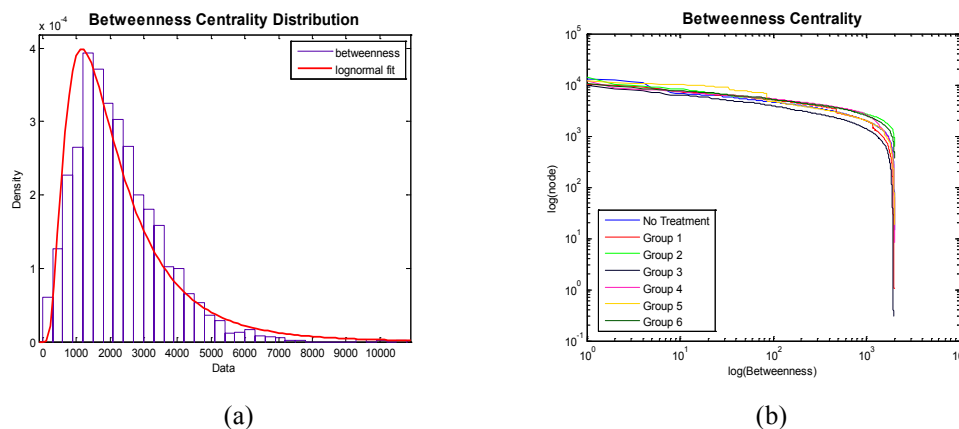

(a)   (b)

Figure 13 (a) Betweenness centrality Distribution and (b) Betweenness centrality per group.

Closeness centrality

Closeness centrality describes the extent of influence a node has on the network. Though the direct metric focuses on the geodesic distance from each node to all others, it as well can be regarded as a measure of how long it will take information to spread from a given node to others. **Figure 14(a)** illustrates that its distribution occupies the Poisson distribution fit resembling [31]. The population mean of all groups seems to show, however, that the highest belongs to Group 6, as shown in **Figure 14(b)**. This scenario implies that people in the Brainy group are potential information spreaders in contrast to their propagation behavior. Such value also signifies that neighbors are more favorable to follow high intellectual content and more comfortable to deal with people with stable emotions.
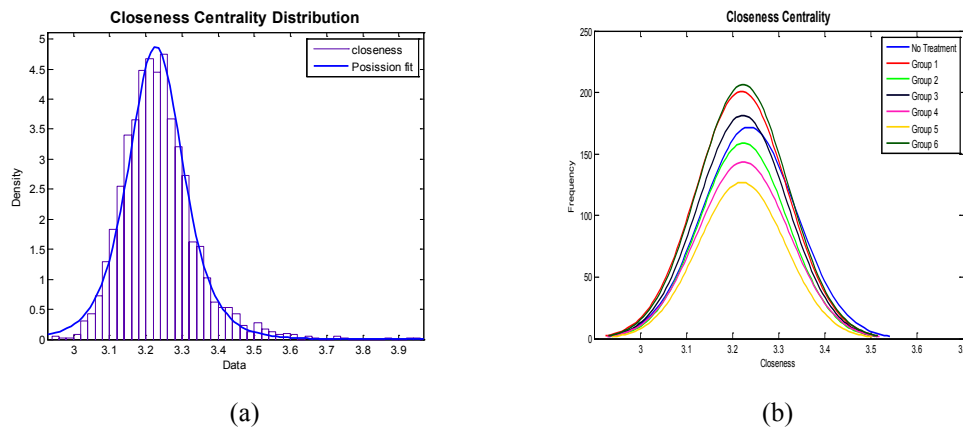

(a)   (b)

Figure 14 (a) Closeness centrality distribution and (b) Closeness centrality per group.

Although the differences between groups are graphically illustrated in **Figures 13** and **14**, quantitative analysis is essential to clearly indicate the difference between groups. As the Kruskal-Wallis H test fits well with nonparametric distribution, it was applied for Betweenness, and as ANOVA fits with Poisson distribution, it was deployed for Closeness. Their statistically significant differences between groups are evinced in **Table 6**.

Table 6 Kruskal-Wallis H test of statistically significant differences in Betweenness centrality.

Kruskal-Wallis H test (Betweenness)			Anova Test (Closeness)					
Group	**N**	**Mean Rank**	**Group**	**Mean**	**S.D.**			
0	1877	6287.76	0	171.25	9.83			
1	2000	8828.54	1	172.34	9.51			
2	2000	4720.19	2	192.71	9.64			
3	1912	7758.98	3	158.22	9.83			
4	2000	6474.98	4	141.19	9.39			
5	2000	8156.35	5	132.11	9.74			
6	2000	6596.12	6	171.25	9.83			
Chi-Square = 1390.81			**Group**	**SS.**	**Df**	**Mean Square**	**F**	**Sig.**
df=6			Between Groups	.361	6	.060	6.343	.013
p-value = .001			Within Groups	132.659	13993	.009		
			Total	133.020	13999			

0 = no treatment group, 1 - 6 = groups with personality treatment, p-value < 0.05

The mean rank and mean values of both centralities demonstrate a strong support for statistically significant differences of social power in the observed groups, such that:

- Betweenness Centrality: $\chi^2(6) = 1390.81$, $p = 0.001$

Group 1 occupies the highest mean rank. It indicates that people with high Intellect but medium Extraversion have larger numbers of shortest paths passing through a node than others. In fact, the high Extraverts (Group 3) were expected to govern highest the Betweenness with respect to their high number of friends and interactions. Through scrupulous investigation, they were found to possess high Neurotic characteristics, one of the hindrances for friendship environments. Thus, H_{31} *(High Betweenness centrality positively correlates with Chummy user)* is accepted.

- Closeness Centrality: $F(6,13993) = 6.343$, $p = 0.013$

The users with high influence in the network are Group 5 users, as they possess the lowest mean value. Thus, H_{32} *(High Closeness centrality positively correlates with Brainy user)* is rejected. When analyzing in detail the related personality attributes, people in Group 5 and 6 (*Brainy users*) are found to possess similar characteristics, such that they both have high intellect and high conscientiousness, which are essential for being empowered. However, people in Group 5 surpass those of Group 6, because they display higher Extraversion, a strong sign for being outgoing and empowering in information spreading.

Focusing on the personality attributes of users with high centrality, the high Extraverts are not considered as users with high influence, but the high Intellects are. Another determinant that may impact the efficiency of information spreading is related to the individual characteristics of those participating in the spreading process.

Limitations and future directions

This study examined the variances of SNS activities between users in association with personality factors, networks, and content. Three outcomes were focused on; "*individual behavior*", "*network topology*", and "*social power*". For ease of inference, hypotheses on how individual traits predicted specific forms of SNS participation were elucidated. The findings clarified the motivational underpinnings of the links between emergence diffusion behavior and network topology, which were consistent with previous research.

Though this study provides significant evidence for relationships between personality traits and propagation mechanism, there exist a few limitations that should be addressed for future studies. First, control variables, such as the topic of content and types of content (stock, political, disaster, celebrity) were neglected. These variables could increase understanding as to why a user passed along the content to their neighbors. Second, a few important results were drawn from user online activity; however, an individual's offline social environment could provide a strong influence on decisions on whether a user would actively participate in SNSs. Last, demographic information was discarded; including it into the model could provide more informative data on friend relations.

Conclusions

A new approach to understand how user characteristics affected the flow of information, how the network topology evolved from user interactions, and how people in these networks were connected was proposed. A framework for modeling user interactions was presented by integrating 3 essential influences: the underlying network topology which determined the basic and important forms of social interactions among users, the user behavior which motivated the system of interactions, and the attractiveness of content which impelled user's decisions. By characterizing the simulated user social interaction data, the statistical and distributional properties of most of the important variables of SNS sessions can effectively be inferred: (i) patterns of content flow and trends of content based on the dynamic and temporal behavior of users,;(ii) future network structural evolution, and (iii) the importance of nodes and behavior transitions of user associated with characteristics. Evidence indicates that user activities display the models of power-law distributions. Insights indicate that users with high Extravert tend to be the major players in propagation-related activities. Users with high Intellect are more likely to be the central propellants. Finally, universal inferences across several SNS interactions and other groups

of users can be drawn from the cross-personality comparisons and can be assured by small-world, scale-free properties and centrality measures.

References

[1] S Vongsingthong, S Boonkrong and H Unger. Modeling network evolution by Colored Petri Nets. *Walailak J. Sci. & Tech.* 2018; **1**, 41-61.

[2] J Yang and J Leskovec. Modeling information diffusion in implicit networks. *In*: Proceedings of the 10th IEEE International Conference on Data Mining, Sydney, Australia, 2010, p. 599-608.

[3] C Yi, Y Bao, S Sun and Y Xue. A novel method for information propagation model perceiving. *In*: Proceedings of the International Conference on Computing, Networking and Communications, Communications and Information Security Symposium, Honolulu, Hawaii. 2014, p. 6-10.

[4] S Garg and S Kumar. Modeling and analyzing information diffusion behaviour of social networks. *In*: Proceedings of the International Conference on Issues and Challenges in Intelligent Computing Techniques. Ghaziabad, India, 2014, p. 566-72.

[5] Z Wang. Social media distribution: A data-driven approach. *In:* Proceedings of the 7th International Conference on Internet Multimedia Computing and Service, Hunan, China, 2015.

[6] JL Wang, LA Jackson, DJ Zhang and ZQ Su. The relationships among the Big Five Personality factors, self-esteem, narcissism,vand sensation-seeking to Chinese University students' uses of socialvnetworking sites (SNSs). *Comput. Human Behav.* 2012; **28**, 2313-9.

[7] J-E Lönnqvist, JVA Itkonen, M Verkasalo and P Poutvaara. The Five-Factor Model of personality and Degree and Transitivity of Facebook social networks. *J. Res. Pers.* 2014; **50**, 98-101.

[8] Y Zhoua, B Zhang, X Suna, Q Zhenga and T Liua. Analyzing and modeling dynamics of information diffusion in microblogging social network. *J. Network Comput. Appl.* 2017; **86**, 92-102.

[9] K Wiesneth. Evolution, structure and users' attachment behavior in enterprise social networks. *In*: Proceedings of the 49th Hawaii International Conference on System Sciences, Hawaii, USA, 2016, p. 2038-47.

[10] RA Rossi, J Neville, B Gallagher and K Henderson. Modeling dynamic behavior in large evolving graphs. *In*: Proceedings of the 6th ACM International Conference on Web Search and Data Mining. Rome, Italy, 2013, p. 667-76.

[11] T Rúbio, R Gomes and A Pereira. Behavioral analysis in social networks: An approach based on intelligent system. *In*: Proceedings of the 18th Brazilian Symposium on Multimedia and the Web. São Paulo, Brazil, 2012, p. 193-6.

[12] AJT Lee, FC Yang, HC Tsai and YY Lai. Discovering content-based behavioral roles in social networks. *Decis. Support. Syst.* 2014; **59**, 250-61.

[13] TC Marshall, K Lefringhausen and N Ferenczi. The Big Five, self-esteem, and narcissism as predictors of the topics people write about in Facebook status updates. *Pers. Individ. Dif.* 2015; **85**, 35-40.

[14] TJ Morales, J Borondo, JC Losada and RM Benito. Efficiency of human activity on information spreading on Twitter. *Soc. Networks* 2014; **39**, 1-11.

[15] M Seufert, S Lange and T Hoßfeld. More than topology: Joint topology and attribute sampling and generation of social network graphs. *Comput. Comm.* 2016; **73**, 176-87.

[16] X Sun. Evaluating structure of complex networks by navigation entropy. *In*: Proceedings of the 8th International Conference on Semantics, Knowledge and Grids, Beijing, China, 2012, p. 229-32.

[17] E Ozkan-Canbolat and A Beraha. A configurational approach to network topology design for product innovation. *J. Bus. Res.* 2016; **69**, 5216-21.

[18] CT Li and SD Lin. Social flocks: A crowd simulation framework for social network generation, community detection, and collective behavior modeling. *In*: Proceedings of the 17th International Conference on Knowledge Discovery and Data Mining, California, USA, 2011, p. 765-8.

[19] K Yuea, H Wua, X Fub, J Xua, Z Yina and W Liua. A data-intensive approach for discovering user similarities in social behavioral interactions based on the bayesian network. *Neurocomputing* 2016; **219**, 364-75.

[20] R Wakefield and K Wakefield. Social media network behavior: A study of user passion and affect. *J. Strat. Inf. Syst.* 2016; **25**, 140-56.

[21] L Liua, CMK Cheungb and MKO Lee. An empirical investigation of information sharing behavior on social commerce sites. *Int. J. Inform. Manag.* 2015; **36**, 686-99.

[22] G Szabo and BA Huberman. Predicting the popularity of online content. *Comm. ACM* 2010; **53**, 80-8.

[23] K Lerman, R Ghosh and T Surachawala. Social contagion: An empirical study of information spread on Digg and Twitter follower graphs. *ACM Trans. Embed. Comput. Syst.* 2018 (*in press*).

[24] M Kim, D Newth and P Christen. Macro-level information transfer in social media: Reflections of crowd phenomena. *Neurocomputing* 2016; **172**, 84-99.

[25] D Wang, Z Wen, H Tong, CY Lin, C Song and AL Barabási. Information spreading in context. *In*: Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work. Hyderabad, India, 2011.

[26] D Wang. A study of the relationship between narcissism, extraversion, drive for entertainment, and narcissistic behavior on social networking sites. *Comput. Hum. Behav.* 2017; **66**, 138-48.

[27] Y Wang, M Iliofotou, M Faloutsos and B Wuc. Analyzing Communication Interaction Networks (CINs) in enterprises and inferring hierarchies. *Comput. Netw.* 2013; **57**, 2147-58.

[28] L Muchnik, S Pei, LC Parra, SDS Reis, J Jose, S Andrade, S Havlin and HA Makse. Origins of power-law degree distribution in the heterogeneity of human activity in social networks. *Nature* 2013; **3**, 1783.

[29] AL Barabasi, H Jeonga, Z Nedaa, E Ravasza, A Schubertd and T Vicsekb. Evolution of the social network of scientific collaborations. *Physica* 2002; **311**, 590-614.

[30] E Yan and Y Ding. Applying centrality measures to impact analysis: A coauthorship network analysis. *J. Assoc. Inf. Sci.* 2009; **60**, 2107-18.

[31] K Tanaka, M Takahashi and K Tsuda. Comparison of centrality indexes in network Japanese text analysis. *Int. J. e-Educ. e-Bus. e-Manag. e-Learn.* 2013; **3** 37-42.

[32] DC Howell. *Fundamental Statistics for the Behavioral Sciences*. 7th Eds. Linda Schreiber, USA, 2011.

[33] H Liu, A Nazir, J Joung and CN Chuah. Modeling/Predicting the evolution trend of OSN-based applications. *In*: Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 2013, p. 771-80.

[34] S Lattanzi and Y Singer. The power of random neighbors in social networks. *In*: Proceedings of the 8th ACM International Conference on Web Search and Data Mining, Shanghai, China, 2015, p. 77-86.

[35] MEJ Newman. The structure and function of complex networks. *SIAM J. Appl. Math.* 2003; **45**, 167-256.

[36] C Zhou, Q Zhao and W Lu. Impact of repeated exposures on information spreading in social networks. *Plos One* 2015; **10**, e0140556.

[37] H Li, X Cheng and J Liu. Understanding video sharing propagation in social networks: Measurement and analysis. *ACM Trans. Multimedia Comput. Comm. Appl.* 2014; **10**, 33.

[38] C Bauckhage, K Kersting and B Rastegarpanah. The Weibull as a model of shortest path distributions in random networks. *In*: Proceedings of the 8th Workshop on Mining and Learning with Graphs. Illinois, USA, 2013.

[39] Y Ahn, S Han, H Kwak, S Moon and H Jeong. Analysis of topological characteristics of Huge online social networking services. *In*: Proceedings of the International World Wide Web Conference Committee, Canada, 2017, p. 835-44.