WALAILAK JOURNAL
http://wjst.wu.ac.th Information Technology

Modeling Network Evolution by Colored Petri Nets

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

[1]*Faculty of Information Technology, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand*
[2]*FernUniversität in Hagen, Hagen, Germany*

([*]Corresponding author's e-mail: suwimonv@yahoo.com)

Received: 5 July 2016, Revised: 4 November 2016, Accepted: 3 December 2016

Abstract

Discovering how information was distributed was essential for tracking, optimizing, and controlling networks. In this paper, a premier approach to analyze the reciprocity of user behavior, content, network structure, and interaction rules to the interplay between information diffusion and network evolution was proposed. Parameterization and insight diffusion patterns were characterized based on the community structure of the underlying network using diffusion related behavior data, collected by a developed questionnaire. The user roles in creating the flow of information were stochastically modeled and simulated by Colored Petri Nets, where the growth and evolution of the network structure was substantiated through the formation of the clustering coefficient, the average path length, and the degree distribution. This analytical model could be used for various tasks, including predicting future user activities, monitoring traffic patterns of networks, and forecasting the distribution of content.

Keywords: Information diffusion, network evolution, stochastic, user behavior, Colored Petri Nets

Introduction

With the characteristics of interactivity, simplicity, instantaneity, and accessibility, Social Network Systems (SNSs) have become a new social phenomenon that people use to broadcast information or maintain their social and business relationships with others. Though SNSs have been used as a primitive medium to spread thoughts, opinions, products, and innovations in technology in this society for decades, little was known about how user behavior, content, and networks harness information dissemination, or how the flow of information affects the network structure. The characterization of the evolving dynamic user behavior and its dependence on the underlying network structure could contribute greatly to the success of SNS providers, education institutes, and businesses.

Several attempts have been made to discover the causes and effects of information propagation in SNSs. Zhao *et al.* [1] employed a model to explore American university networks, and found that tie strengths impacted how information was propagated. Nui *et al*. [2] proposed a word of mouth propagation model to investigate network structure, and observed that user behavior influenced the scope and speed of propagation. Hu *et al*. [3] explored the emergency public health network, and found that information dissemination was immensely influenced by user interactive activities and group behaviors. Zhou *et al.* [4] presented a Twitter-like microblog model to trace repeated information propagation, and discovered that network structure had little effect on the diffusion, while personal behavior was an underlying driving factor. Sun and Yao [5] applied a model of a random-graph, scale-free, and small-world network to explain how individual personality and mentality differences impacted information adoption and diffusion, and realized that information was best disseminated through scale-free networks. Subbian *et al.* [6] introduced an information flow mining algorithm based on content propagation to locate key propagandists and patterns of information flow; it was found that the node linking process was considerably dependent on content and period of time.

Most works have put forward many models, with several parameters, to describe the behavioral mechanisms and theoretical characteristics of the diffusion process. A clear similarity among these was that all adopted user characteristics, group behavior, emotions, social relationships, network structure, or topics of content as influence parameters. However, the diffusion process in SNSs was not simply explained by any particular parameters, invariable distribution probabilities, or general mathematical models. The social network was user-centric and accessed by massive number of users with distinctive characteristics and different socialization patterns; the incorporation of only a few specific variants could make the models susceptible to unpredictable changes of individuals.

By integrating the underlying dependences of user behavior, content, network, and interaction rules into the diffusion process, the reflection of user phenomena in SNSs should be more naturalistic. Herein, the stochastic modeling and synthesizing mechanism of Colored Petri Nets was introduced to sturdily visualize and anatomize the complicated nature of human behavior. Three steps of analysis were brought up. First, user behavior data was accumulated by using a developed questionnaire, for insight investigation of the spreading patterns and the obscured community structure. Second, optimal parameters were extracted and estimated for propagation probabilities. Third, various real user propagation activities were smoothly simulated, visualized, and mathematically analyzed through the executable models. Lastly, how information diffusion influenced the evolution mechanisms was explicitly reflected through the transfiguration of the clustering coefficient, the average shortest path length, and the degree distribution. This approach is the first in-depth analysis of the interplay between information diffusion and network evolution, which can best describe divergent and pluralistic user, content, and social events.

The remainder of this paper is organized as follows. Important aspects on user roles in SNSs and general theories behind modeling are introduced in “The Modeling Framework”. The data preparation, user state definitions, and transitions are elucidated in “Parameterization Methodology”. In “Content Propagation Models with CPNs”, the model components and functionalities are explained. In “Simulations”, the simulation configuration and results are discussed, and the last section provides the “Conclusions” of this work.

The modeling framework

Developing a model to explore user behavior requires the identification of the series of events and transitions. Events refer to user interactions with content and other users within the scope of the SNS interface design. In other words, a user can appear in any state of SNS environment and traverse from one state to another via an action or transition. In this framework, the nondeterministic variation of transitions were superseded by a probabilistic interpretation derived from the reciprocity of the user’s decision, the underlying network structure, and the enticement of content.

Data description and statistical inferences

The focus group was university students. As usage amongst them was extremely high, both regular and extreme cases could be reasonably expected. Due to its richness features and large user base, the chosen study platform was Facebook. A questionnaire was designed to acquire user’s diffusion behavior and its dependent interactions. It covered 2 anticipated parts, *A*: the background and distinctive characteristics of the individual, and *B*: online activities. A pre-test was applied, and internal consistency and reliability was acceptable with a Cronbach’s alpha of 0.945. Using the corporation of university management, 1500 Thai students were recruited during October-November 2014. The final validated dataset was comprised of 1,287 respondents, with an 86 % return rate, 516 males (mean age = 22.84 years; SD = 6.46) and 771 females (mean age = 23.37 years; SD = 6.30). The majority (81.7 %) were undergraduate students, 14.5 % were high vocational certificate students, and 3.8 % were graduate students [7].

Central role of user behavior in social networks

Behavioral analysis of user interactions varies tangibly over the user population and social media structure. Lang and Wu [8] explored the user's motivation determinants (to be active and to remain online), and explained that a high degree and a high clustering coefficient indicated a long lifetime. Chen *et al*. [9] studied user lifetime and discovered that temperamental users had fewer number of friends and did not introduce much communication, contrasting with outgoing users. Benevenuto *et al.* [10] characterized user behavior in association with content, and realized that it had a power law distribution, and that essential motivations for users to post were solicitation and spontaneity. Yin *et al.* [11] observed users' behavior in tagging content, and figured out that the behavior was not stable and changed over time. Uesugi [12] revealed that personality traits influenced both usage trends and the type of SNS services used. Ljepava *et al.* [13] examined characteristics of online users and disclosed that human personality influenced decisions to create social relationships and to distribute information. Ortigosa *et al.* [14] suggested that a crucial motivation for user interactions and social relationships was that of personality traits. Koo *et al.* [15] analyzed the effects of user characteristics on the usage of SNSs, and perceived that individual personality introduced different usage patterns.

These findings were synthesized and interpolated into the model framework, as follows:

- *Lifetime*: A lifetime was measured as the duration of time a user joined and left the networks, which was a roadmap for individuals to have relationships with others. Studies on user lifetime [8,9] have highlighted that prime influences on user lifetime were personality, motivation, and friendship relations.
- *Personalities*: Personality and motivation were found to be what drive users to act. Generally, each user displayed unique interaction patterns [12], e.g., an extravert could create and diffuse information to friends regularly, while a neurotic could easily get moody with unpleasant content and end up unfriending a sender.
- *Friendship Relations*: Connection between neighbors referred to sharing, exchanging, and delivering information depending on interpersonal ties. These ties not only indicated friendship relations, but also instructed types of information exchanged and dissipated [14,15], e.g., people with business relationships might pass business content to others.
- *Information Flow*: Information could flow in any direction; however, information passing from one to another generally occurred according to their social relationships, e.g., work instruction flew from supervisors to employees.
- *Content*: Each example of content maintained unique access patterns according to its importance [10,11]. In general, content of interest would repeatedly be dissipated and, vice versa, stale content would be discarded and isolated.

During a user's lifetime, the combination of friendship relations and content characterized a particular interaction between individuals, as well as determined the flow of information from source to destination. Though the determinants that activated users to act differently to different phenomena were multitudinous, there existed a few common stimuli, such as individual characteristics, interests, emotions, and the content itself. Through combining those aspects with a modeling paradigm of Petri nets, gaining insight into how user dynamic behavior, network structure, and content influence the growth of the network became possible.

Colored Petri Nets (CPNs)

In the context of SNSs, once a user registered with the system, full participation was enabled. Users could log in from different places. They might broadcast their existing status, update, share photos, forward information, or start a conversation, without necessarily knowing each other. Anyone in the same network was eligible to generate content and spread content out to everyone. By recognition of content, unpredictable human behavior was initiated, leading to an independent unknown choice of reactions. These actions did not totally depend on a latent construct of individual characteristics, but also relied on a prompt emotion, e.g., being happy with the post, being annoyed with the response, or enjoying seeing nice videos. Inevitably, they inherited nondeterministic, concurrent, conflicting, synchronous, and asynchronous processes that introduced a complex structure and autonomous evolution to the network.

With the special characteristics of distributed environments and dynamic systems of Petri nets, events that occurred with constraints, precedence, or distinguishable frequency could be facilitated. The Petri nets' concept matches well with the intricacy of user interactions from two outstanding strengths: first, the explanatory power of a directed bipartite graph facilitates the modeling and visualizing of the complex system, and second, the flexibility of mathematical language encourages diverse quantitative analytical results in formalism representation [16]. However, modeling distinguishable users with differential characteristics introduces complications from the demand to differentiate one user from another. Each user must be declared as a token in a separate place. The modeling of user interactions could quickly become not only extremely complex but also potentially incomprehensible from a large number of places. Fortunately, CPNs fulfils the gaps on overloaded and indistinguishable state spaces by extending the classic Petri nets into a language for the modeling and validation of systems. With different types of expressions, i.e. guards, arc expressions, and expressions to define initial markings or rate functions, each token can be configured to be distinctive from others using a different token color [19]. Combining mathematical models and good graphical tools of Petri nets with a programming language of CPNs, the distinguishable users and diverse content of large and complex systems can easily be defined, as follows.

Definition 1: (*A formal definition of CPNs*) [20]. A nondeterministic, real-time and multidimensional relationship of user behavior is an 8-tuple $\prec P,T,F,\Sigma,C,g,f,m_0\succ$, where:

- P is a finite set of places; users, content, and motivations, with color set to specify the type of residing tokens. Data type in the color set can be integer, character, or multiset.
- T is a finite set of transitions (actions). Each action can be identified in Guard, a Boolean expression of the variables.
- F is a finite set of direct arcs between a user and content. Each has an expression which can define the algorithm of the variables.
- Σ is a finite, non-empty set of types, or color sets defined for places.
- $C: P \dashrightarrow \Sigma$ is a color function that is assigned to each place $p \in P$, a color set $C(p) \in \Sigma$.
- $g: T \dashrightarrow EXP$ is a guard function of Boolean type that is assigned to each transition $t \in T$.
- $f: F \dashrightarrow EXP$ is an arc function that is assigned to each arc $a \in F$, an arc expression of a multiset type $C(p)_{MS}$, where p is the place connected to the arc a.
- $m_0: P \dashrightarrow EXP$ is an initialization function that is assigned to each place $p \in P$, an initialization expression of a multiset type $C(p)_{MS}$.

The analysis of CPNs is based on multi-set handling, where the collection of all multisets (a set in which there can be several occurrences of the same element) over S is denoted by S_{MS}.

Transitions and stochastic firing processes

The relationships between users not only originate patterns of information exchange or indicate what kinds of information to be spread, between whom, and to what limits, but also engender the system to change states. The determination of next state directions can either be to 'continue' in the present state, to 'move' to another state in the cycle, or even to completely 'start' a new cycle. The choice depends on various kinds of cognitive and informational forces, such as the content, the propagandists, and knowledge about the past. Explaining these variants by fixed geometric rules or cycle occurrences will be too inflexible, since these transitions govern stochastic properties. Hence, firing transitions are denoted as:

Definition 2: (*Firing Rules)* [21]. Let N be a colored net and $m \in \mathbb{M}_N$. A transition $t \in T$ is the fireable at m with color c_t (denoted by $m\,[(t, c_t)\rangle$) if and only if: $\varphi(t)(c_t) \wedge \forall p \in P, m(p) \geq W^{-}(p, t)(c_t)$. The marking m' obtained is defined by: $\forall p \in P, m'(p) = m(p) + W(p, t)(c_t)$. In this case, $m[(t, c_t)\rangle m'$.

The set of transitions enabled at a marking m is the set $E_n(m)$. $Reach(N) \subseteq \mathbb{M}_N$, and the set of reachable markings of N is recursively defined as $\{m0\} \cup \{m \in \mathbb{M}_N \mid \exists m' \in Reach(N), t \in T, c_t \in C(t) \mid m' [(t, c_t) \rangle m'\}$.

The definition of operations on color mappings is described where the mappings are extended from C to $Bag(C')$ to mappings from $Bag(C)$ to $Bag(C')$ by the following rules:

$- f(\lambda.c) \quad = \lambda.f(c)$
$- f(c_1 + c_2) \quad = f(c_1) + f(c_2)$

Definition 3: *(Mapping)*. If f is a mapping from $Bag(C'')$ to $Bag(C')$, and g is a mapping from $Bag(C)$ to $Bag(C'')$ then $f \circ g$ is a mapping from $Bag(C)$ to $Bag(C')$ defined by $\forall c \in C, c' \in C', (f \circ g)(c)(c') = \sum c'' \in C'', f(c'')(c').g(c)(c'')$.

Definition 4: *(Mapping)*. If f is a mapping from $Bag(C)$ to $Bag(C')$, then ${}^{t}f$ is a mapping from $Bag(C')$ to $Bag(C)$ defined by $\forall c \in C, c\_ \in C', {}^{t}f(c')(c) = f(c)(c')$.

From the view of discrete event system, the aforementioned concepts have been put into an effective example of CPN formalism to explain the way in which information is propagated in the network.

Notions of user model with CPN approach

Figure 1 shows a predefined algorithm of CPN theoretical definition. Let i_j be the number of input tokens in *Place 1,* containing a set of users who were simultaneously online in an SNS. *Place 2* was the output repository. *Place 3* held input tokens m_i, emotion determinants. The transition t_j was a user interaction mastered by an inscription defined in the arc expression. Rate of transition was embedded by equation specifying a particular distribution equation derived from the probability of the occurrence. The enabling and firing of the transitions, which determined the kind of data for the successive places, was based on the evaluation of its precedence, associated arc expressions, and associated places. The most general firing rule defined for each transition t_j was a function f_j of i_j that produced an o_j-tuple of output tokens. Before the firing could be start, the conditions must be true, and the input places must contain sufficient suitable tokens to fire. If criteria were met, a token would be imparted and bounded as a multiset of a varying number of tokens to the output places identified in the output arc expressions.

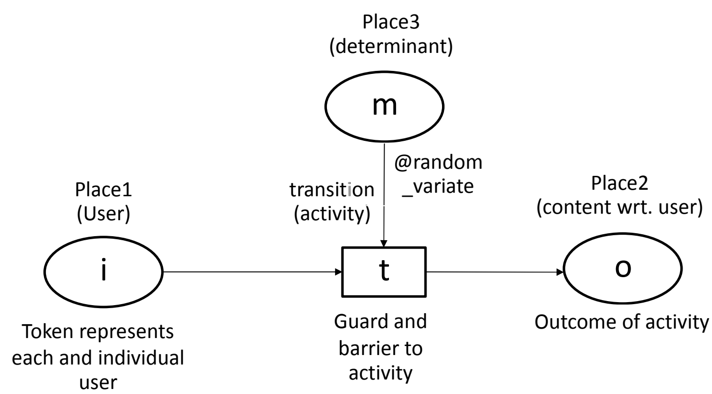

Figure 1 CPN building block.

Throughout this paper, the important components of the algorithm representing the states of the modeled system, such as users and resources, were defined as *places* with or without initial marking of tokens. All the *tokens* in each place were annotated by different sets of *coloring spaces*, defined by a color set with distinguishable data values. Transformation activities were defined as *transitions* which specify what computational step to be executed once the transition was enabled. The condition for an activity to occur was specified in the *arc inscriptions*. Important tokens for testing the preconditions were defined by *guards*. If the conditions were true, changes in user actions were converted to tokens in different places, specified by *output arc inscriptions*. The provocation of activities could be depicted as stochastic transitions between users, friends, and content, in which the concurrent disseminating of a heap of content reflected the racing of tokens flowing from one place to another. Combining the given descriptions with the concept of place types, guards, and arc inscriptions, the model components, the control algorithm, the system environment, and the flow of data objects could be hierarchically specified by CPN tools, developed by Aarhus University, Denmark [19].

Parameterization methodology

Though user processes such as authoring and sharing, and sometimes redistribution of content, regulate the flow of information at the top of the system environment, these interaction values were not applicable for used as model inputs. Therefore, an essential mechanism to reform the values into mathematic equations for process activation was parameterization.

Phases of parameterization

Several phases of refinements were employed to determine the probability annotations of stochastic patterns, as illustrated in **Figure 2**. Firstly, in order to profoundly track the pattern of information flow, and to accelerate the computational mechanism, data was collected, validated, and pre-processed by the method of clustering [3]. Propagation-related activities of user groups were extracted to determine user states and transitions. Through iterative fitting, the discreteness of data values was screened. The most probable distribution that maximizes the probability density functions was identified using the Maximum Likelihood Estimation (MLE). Probabilistic transitions between states indicated by stochastic equations were then investigated for fitness, as well as compared with the expected distributions. Finally, the simulation-based computations were executed in order to validate the evolutionary hypotheses.

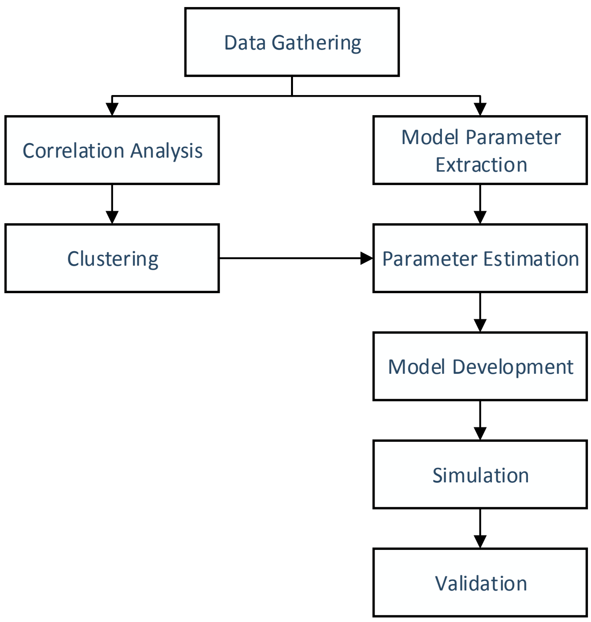

Figure 2 Parameterization process.

Clustering

The correlation measure of interrelations between all pairs of online behavior cases was explored. Personality attributes were found to implicitly correlate with propagation related behaviors. A disjoint cluster analysis was applied on the basis of the associated strength of the data point to the cluster. The users were divided into clusters; every user belonged to one and only one cluster, such that $C = C_1, ..C_6$. of U where $U \subset U_{i=1,k}$ and $C_i \cap C_j = \emptyset$ for $i \neq j$.

A 3-dimensional latent space of user clusters is shown in **Figure 3** as *1: Chummy Users, 2: Inert Users, 3: Socialize Users, 4: Introvert Users, 5: Cool Users, 6: Brainy Users* [7]. Users in each cluster were inherently assumed to possess similar behavior, and this consequently governed the same set of states and transition probabilities. By group estimation, the individual complexity and environmental variability in terms of overloaded number of parameters and models were not only systematically obliterated, but the modeling process also remained accurate and simple to investigate significant system behaviors.

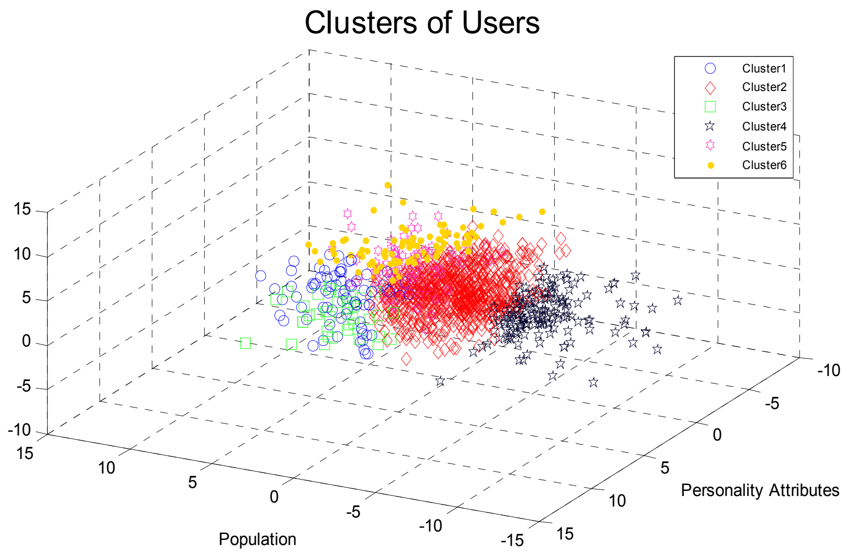

Figure 3 Preprocessed data into clusters of users.

Parameter estimation

The characteristics of model parameters were delegated by probabilistic interpretation of the selected actions, defined as; direct actions: *create and distribute*, *share*, and *response*, and side effect actions: *add friend* and *remove friend*. The likelihood that a user performed a given action was observed in terms of event distribution per cluster, using relative frequency estimation. The estimation of distribution parameters was ensured by a Chi-square goodness of fit test, and compared with the series of theoretical family of distributions. Lastly, the best fitted probability distribution of actions were summarized in **Table 1**, and the equations in **Table 2** explained each set of actions.

Table 1 Different Probability Density Functions of user activities, classified by cluster [3,9-11].

Activity	Probability density distribution					
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**
Create/Distribute	*Power Law*	*Power Law*	*Power Law*	*Power Law*	*Power Law*	*Power Law*
Share	*Power Law*	*Power Law*	*Power Law*	*Power Law*	*Power Law*	*Power Law*
Respond	*Poisson*	*Poisson*	*Poisson*	*Poisson*	*Poisson*	*Poisson*
Add friend	Long-tail distribution					
- Like	*Gamma*	*GEV**	*Weibull*	*GEV**	*GEV**	*GEV**
- Comment	*Exponential*	*GEV**	*Exponential*	*Weibull*	*GEV**	*Weibull*
- Chat	*GEV**	*Gamma*	*Beta*	*Beta*	*Exponential*	*GEV**
- Play Game	*Power Law*	*Power Law*	*Power Law*	*Power Law*	*Power Law*	*Power Law*
- Post	*Power Law*	*Gamma*	*GEV**	*Beta*	*GEV**	*Weibull*
Remove friend	*Binomial*	*Binomial*	*Binomial*	*Binomial*	*Binomial*	*Binomial*

GEV*- Generalized Extreme Value

Table 2 Transformation of distribution fits to parametric definitions.

Distribution	Parametric equation	
Beta	$P(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} I_{(0,1)}(x)$ where *B(.)* is a beta function, *a* is a first shape parameter, *b* is a second shape parameter and $I_{(0,1)}(x)$ is an indicator to ensure that x in the range of (0,1) has non-zero probability	(1)
Binomial	$P(x) = \binom{N}{x} p^x (1-p)^{(N-x)}$ $x = 0,1,2, \dots .N$ where *N* is number of trials, *x* is the number of successes in *n* trials of a Bernoulli process with probability of success *p*	(2)
Exponential	$P(x) = \frac{1}{\mu} e^{\frac{-x}{\mu}}$ where μ is mean	(3)
Gamma	$P(x) = \frac{1}{b^a \tau(a)} x^{a-1} e^{\frac{-b}{a}}$ where $\tau(.)$ is the Gamma function, *a* is a shape parameter, *b* is a scale parameter	(4)
Generalized Extreme Value	$P(x) = \frac{1}{\sigma} exp\left(-\left(1 + k\frac{(x-\mu)}{\sigma}\right)^{-\frac{1}{k}}\right)\left(1 + k\frac{(x-\mu)}{\sigma})\right)^{-1-\frac{1}{k}}$ for $1+ k\frac{(x-\mu)}{\sigma} > 0$ where *σ* is the scale parameter, *k* is a tail shape parameter and *μ* is a location parameter	(5)
Poisson	$P(x) = \frac{e^{-\mu}\mu^{-x}}{x!}$ where μ is the mean of the distribution	(6)
Power Law	$P(x) \sim x^{-\gamma}$ where *γ* is a scale invariance	(7)
Weibull	$P(x) = \frac{b}{a}\left(\frac{x}{a}\right)^{b-1} e^{\left(\frac{-x}{a}\right)^b}$ where *a* is a scale parameter, *b* is a shape parameter	(8)

**x* - variable, *e* - the base of natural logarithms

Most previous studies identified that human behavioral patterns were more probable in demonstrating heavy-tailed or long-tailed distributions [10]. Therefore, those fittings were primarily taken into consideration. However, after being deliberately verified by the Chi-Square Goodness-of-Fit test, not all actions could be mapped with these well-tried theories.

The actions of *'share'* and '*create/distribute'* possessed a power law distribution ($1.82 < \gamma_{share} < 2.21$ and $1.71 < \gamma_{create/distribute} < 2.34$), suggesting that only a few users disseminated information, while most were inert. The randomness of '*respond*' was represented by a Poisson distribution ($3.00 < \mu_{respond} <$

4.12). This pattern was different from that of typical communities, and not well-matched with the theoretical expectations. Detailed exploration was conducted and detected that the majority of users were classmates, having frequent interactions, and they, in fact, intended to respond to any messages from peers.

For '*add friend'*, various ranges of probability functions appeared. The differences spread from cluster to cluster. However, all governed a long tail property with a right skew, implying that only a few users considered adding friends from impressions gained through the development of content or actions. Lastly, '*remove friend'* occupied the Binomial distribution ($0.003 < p_{remove} < 0.140$). Though this distribution did not harmonize well with most previous studies, it agreed with the analysis of users' tweet [23] where it could be inferred that there was less chance that a user would remove friends once they decided to add them.

The next section provides information on how concepts and findings were put together and analyzed as independent defined components. The interfaced algorithms for functional separation of concerns were introduced to enable the modular instantiation of the framework. Lastly, the composition of information diffusion components was actuated via a well-defined interface and synchronized by the system clock.

Content propagation model with CPNs

Many different social relationships between users enabled by content can finally express themselves in a friendship matrix as illustrated in **Figure 4**. This process started when a user logged in and noticed the new content in timeline. As a routine, he/she ended up reading and processing. His/her next action appeared and evolved according to an immediate decision, e.g., he/she could reply to the third party content by posting, messaging or giving a "Like". How and what actions stimulated the content to flow from one to another were explained using the concept of network anatomy, as follows.

Users were thought of as nodes, and content as another type of node that links 2 or more people together. How a user responds, and what action he/she would do next, was dominated by his/her feeling raised from content interpretation. These actions introduced links. These links brought up an alteration to network structure. By incorporating the significance of the users, content, and social interactions as determinants, the friendship matrix were elucidated as:

- *Add friend*: an active link was created between 2 existing nodes. The out degree was increased by the number of links created.
- *Remove friend*: a link was deleted from 2 existing nodes. The out degree was decreased according to the link deleted.
- *Respond*: a new node was created, with a link to the existing node.
- *Share*: the new nodes were created, with links to the origin node. The out degrees were increased proportional to the number of friends.
- *Create/Distribute*: the new nodes were created, with links to the origin node. The out degrees of the node are increased proportional to the number of friends.

The presented diffusion scenario might not inherently match well with real-world phenomena, since all users participating in the network must be known to the system beforehand. However, combining the activity rules with the above assumptions, the explicit friendship should certainly be reflected by investigating its network topology.

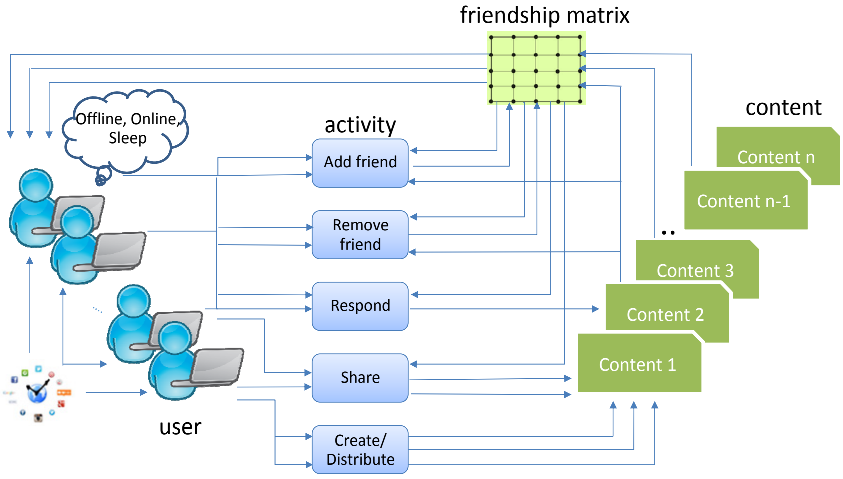

Figure 4 Pattern of user social relationships.

Model block diagram

The dynamics of the social networks of users was graphically represented as a block diagram in **Figure 5**. The system was modularly designed according to its own unique functions as 3 main components. The interface between each was signaled by vectors containing the state variables and the inputs, such as *online* and *offline*. *Global Clock* was the incipient component, which initiated consecutive processes by sending a state variable to *Lifetime*. If the rules of transformation and the compilation conditions locally interfaced between components were met, events in *Online Activities* could be activated. At every computation cycle, the next state was determined based on logical inputs and progresses, depending on a list of prioritized transition conditions. As it was nondeterministic, it was solely independent to the previous action. However, transitions between users and their social environment could possibly be traced from the data repositories of *friend and content buffers*.

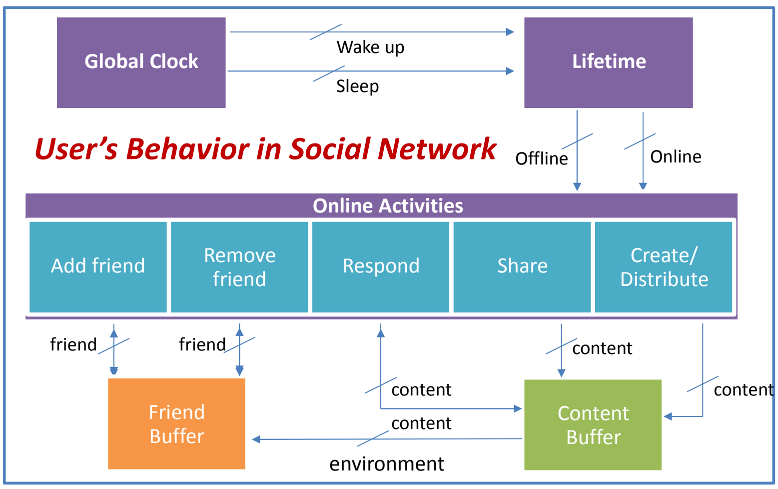

Figure 5 Simplified Petri net model block diagram.

The functions of each model component were prescribed as follows.

- *Global Clock:* a centralized authority on the passage of time. It was a repository of future user actions responsible for executing the sequence of those actions. *Global clock* would advance the simulation clock to one of the 2 primary states acquired from the survey; *Sleep* and *Wakeup*. To simplify the operation, clock counter was defined as;

$$clock\ counter \ = \frac{clock\_time * hour\_unit}{t} \tag{9}$$

where $clock\_time$ = 24 hour clock,
t = clock trigger duration,
$hour\_unit$ = minutes in an hour

- *Lifetime*: an aid to describe the availability of individual's behavior. The three naturalistic observed states were *Online, Offline*, and *Sleep*. However, the only state in which users were allowed to conduct activities was *Online*.
- *Online Activities*: the information diffusion related activities among users. Chosen activity primarily depended on content interpretation, an individual's characteristics, and existing information in data repositories.

In order to ensure the correctness of this design, the activities associated with diffusion were transformed into a bipartite graph with state-transforming functions. The net structure and the semantics, such as states, transitions, firing transitions, and rules, were then defined, and the verification of temporal properties was conducted using a model checking technique.

Global clock model

Figure 6 demonstrates the global clock component, with different activity time gaps for weekdays and weekends. Overall model synchronization was specifically managed with 2 control parts: day type and clock. The first was manipulated by a place *control* with timed type color set with 2 possible allowed values: *WD-Weekday* and *WN-Weekend*. The maximum number of tokens in *control* was limited to one, using a place *anti* with color set *'E'* along with transitions *send* and *receive*. At every unit of time, a token symbolled "t" was automatically generated, and was passed through the transition *clock*.

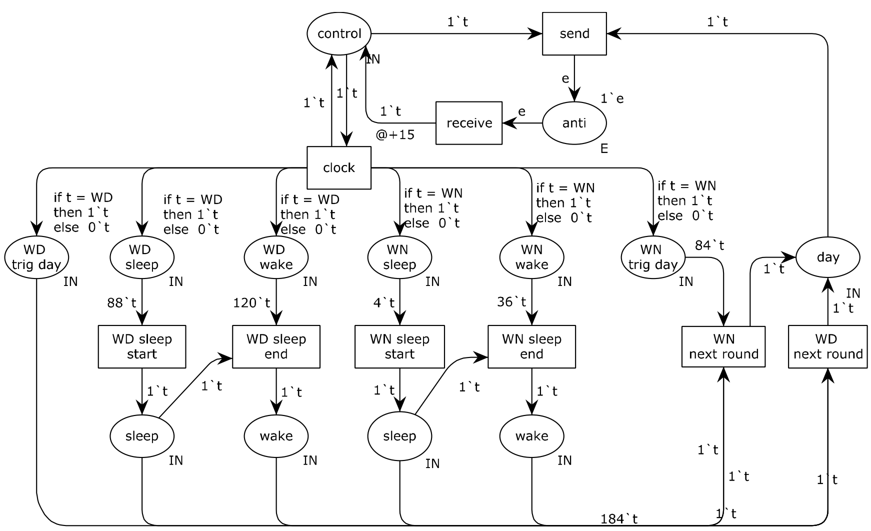

Figure 6 Global clock.

The enabling of *Weekday* starts when *t= 'WD'*. In every time unit, a token was successively generated from *clock*, and passed to all WD repositories: *WD sleep, WD wake*, and *WD trig day.* If the number of tokens in *WD sleep* was accumulated to 88 (10 p.m.), transition *WD sleep start* was activated; then, all tokens in *WD sleep* were consumed, and a new token was generated in the place *sleep.* When the number of tokens in *WD wake* was summed up to 120 (6 a.m.), and a token existed in *sleep*, *WD sleep end* was enabled, and tokens in both places would be consumed and dispatched to the place *wake*. Besides, when the number of tokens in *WD trig day* were assembled to184 (the end of wake time), *WD trig day* was activated. Then, day type was enabled to trig from *WD* to *WN*.

Similar logic was applied to weekend *(WN), t='WN'.* The differences were duration of time and data repositories, such that *WN sleep* was enabled when the number of tokens reaches 4, (1 a.m.)*, WN wake* was enabled when the number of tokens reaches 36, or 9 a.m., and *WN trig day* was enabled when the number of tokens reaches 84. These processes run continually throughout the simulations.

User lifetime model

Figure 7 illustrates the user lifetime process. The initial marking residing in the place *User Lifetime* is activated by *Global Clock*. Symbol "*n, p"* represented a state id and state description, with 3 possible values *(1:Offline, 2:Online, 3:Sleep).* A place *user state* and a place *control user state* with color set *'E'* were defined to manipulate the change of user state. The transition *change* was a binding transition to control the maximum limit of tokens residing in the place *User Lifetime*. If the token received from *Global clock* was *sleep*, the next transition determined by the transition state remained *sleep*, but if the token received was *wake*, the next possible user state could either be *Offline* or *Online*.

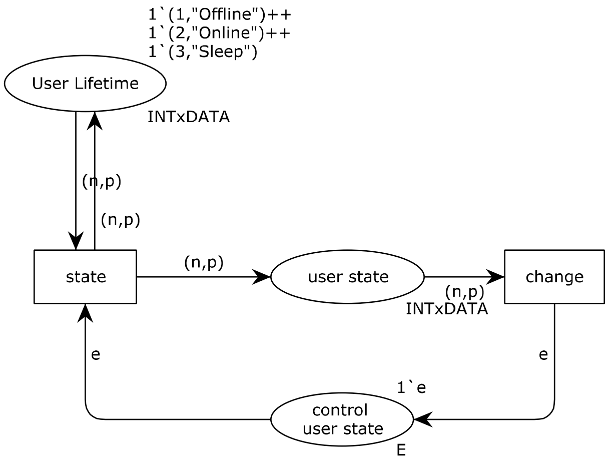

Figure 7 User lifetime.

Online activity model

The social online processes are presented in **Figure 8**. Each user in the network was represented by a distinguishable token in the place *User*. The model was enabled after *Online* received a token from *Lifetime*, indicating that users in the *User* data repository were ready to interact. Six basic Facebook functions were transformed into transitions. Five of those related to information diffusion were highlighted as a relationship matrix in **Table 3**. The other (browse) was a utility function. All *motive* (motivation) data repositories stored the probability functions to enable transition of actions.

Annotations used throughout this paper were: user as U, $u_n \in U$, content generated by users as C, $c_n \in C$, and friends as V, $v_n \in V$, and $V \subset U$. The occupation vectors of the social links indicating relationships between user, friend, content, and motivation were provided as a part of the matrix identification in the upper right corner. In the initial state, a user in the *User* repository occupied 2 possible transition options; *create content* or *browse content*. A pre-condition for the transition *browse* was the coexistence of tokens in *User* and *content buffer*. In order to *create content*, tokens were required from 2 places; *User* and *create motive*. If the conditions identified in the output arc descriptions were met, a *content* token was created and positioned in the *content buffer*. Each token in the *content buffer* contained a content id and a user id indicating the content and content owner. For example; u_1 created content c_1, the coexistence of a user, and content was declared as a multi-set of colors (u_1, c_1).

Table 3 User activity matrix identification.

Behavior	Process description	Occupation vector				Relation vector
		u	c	v	m	
Add Friend	• A user token tied with content token was added to *friend list*. • Two user tokens were bound with the existing content token and a motivation token.	✓	✓	✓	a	(u,c,v,a)
Remove Friend	• A user token in *friendlist* was fired to *nonfriend list*. • Two user tokens were bound with the existing content token and a motivation token.	✓	✓	✓	b	(u,c,v,b)
Respond	• A content token was created and moved to the content buffer. • A content token was bound with a user token and a motivation token.	✓	✓	×	r	(u,c,r)
Share	• A content token was created and fired to the content buffer. • A content token was bound with a user token and a motivation token.	✓	✓	×	s	(u,c,s)
Create/Distribute	• A content token was created and fired to the content buffer. • A content token was bound with a user token and a motivation token.	✓	✓	×	e	(u,c,e)

Matrix identification
(u,c) = user and content
(u,c,v) = user, content, and friend
(u,c,v,m) = user, content, friend, and motivation

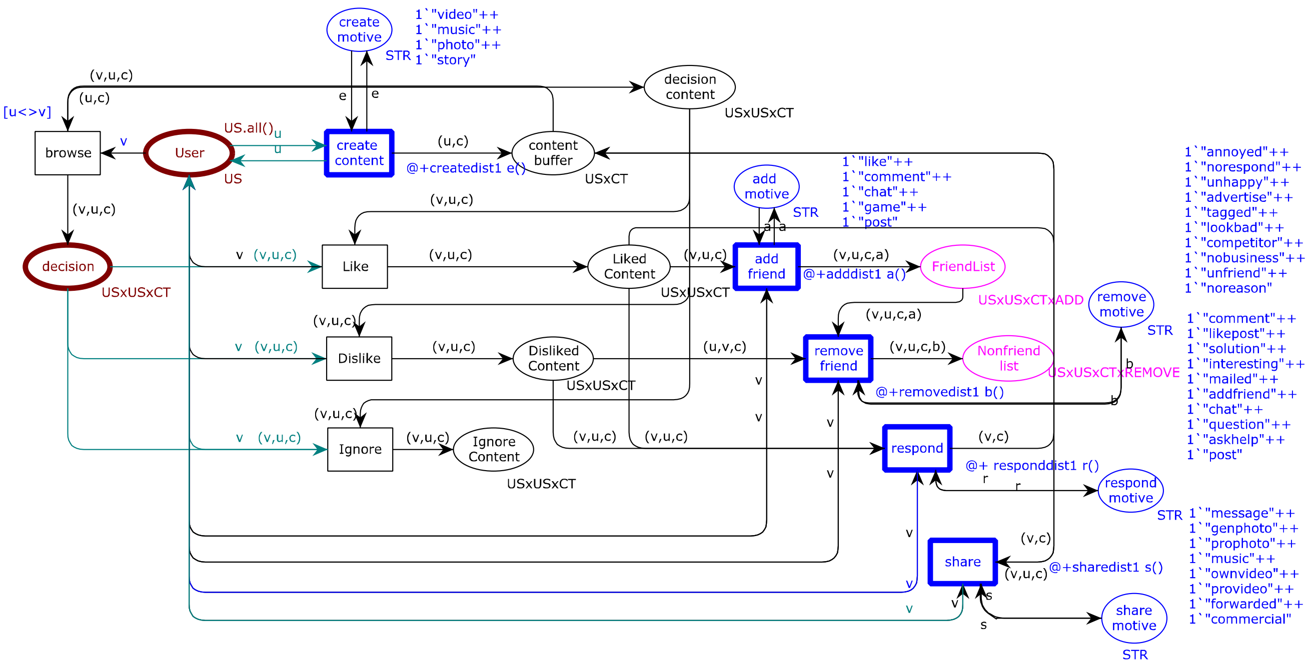

Figure 8 User friendship relations.

To enable *browse*, 2 tokens were required from the places *User* and *content buffer*. After the content was browsed, a *user* token would be positioned in the place *decision* while the content token was placed in *decision content*, meaning that a user was making a decision on the content. Determination of the next action depended on how a user interpreted the content stored in *decision content.* Sentiments could be *like, dislike*, or *ignore*. This scenario introduced a vector of (u_2, c_1, u_1), meaning that u_2 had read a content c_1 created by u_1 and made a decision on the content. The subsequent place when the transition was fired could be *liked content*, *disliked content*, or *ignored content*. If the decision was *ignore*, the content token was passed to *ignored content*, and no further transition was proceeded. If the decision was *like*, the content token was passed to *liked content*. When there was a user in the place *decision*, the next possible transition could be *add friend, respond*, or *share*. If *add friend* was selected, the content creator would be moved to the place *friend list*. For instance, u_2 read the content c_1, posted by u_1 and liked (a_1), he/she might add u_1 as a friend, u_1 was then added to *friend list*. This scenario introduced a set of entangled objects as a vector of (u_2, c_1, u_1, a_1), with symbols indicating both deterministic and stochastic properties. This vector was created in the form of links between a user and content when a user had interacted with the environment. For *respond* and *share,* if the criteria were met, a new content token would be generated and bound with the content owner in *content buffer.* Vice versa, content was existed in *disliked content*, the next possible transition could either be *remove friend* or *respond*. If an action of *remove friend* was selected, a *friend* token would be removed from *friend list* and positioned in *nonfriend list*. As simulation steps increased, the number of records were correspondingly increased to form elements of the occupation vectors.

Simulations

To capture the dissemination pattern of tokens in the model, user behavior in the real world was the imitated. The interaction rules and stochastic behavior were determined as transition firing probabilities of occurrence. The dissemination pattern of tokens was the key attribute used to monitor characteristics and behaviors of the physical propagation system.

Goals and configuration

Insight into understanding how the topological and structural properties of evolving networks grew, in particular, in the social interaction counterpart of the node and edge creation process was accomplished by the method of simulations. One hundred users belonging to the same cluster were assigned an online state simultaneously. Each was eligible to create and globally disseminate a maximum of 100 pieces of content. The simulation steps were 0 - 1000, with 100 steps incremented in each cycle. During the simulation, a random user was selected to conduct a single operation in a discrete time unit (Δt). Its transition rate was tied with the probability of an action with respect to motivation specified by cluster (**Table 2**). For each simulation step, a link between a user and content, or between users, was created. The flow of tokens, representing nodes and social links, was written into a state-transition log file. Firing transitions were then extracted and transformed into an adjacent matrix. A complete network structure evolution was analyzed from the matrix by tracing the transfiguration of its clustering coefficient, average shortest path, and degree distribution. The main structural characteristics of network evolution were explored via the given hypotheses;

H_1: The clustering coefficient of the nodes increases from triangle formation during link formation.

H_2: The average shortest-path length is small even if the networks are large.

H_3: The degree of distribution follows the power law form.

For each phenomenon, the simulations were independently executed repeatedly for 100 rounds. Results were observed and compared with the basic features of the evolving network for verification, extension, and generalization of the hypotheses.

Simulation algorithm

The objectives of this simulation were to demonstrate how the interactions of a user changed over his/her lifetime, how the flow of content was enabled by the user's behavior, and how a social link was created during the interaction. The performance of the algorithm was not of interest, because there existed a few number of uncontrolled factors, such as the system capacity, the number of transitions the token passing through, and the size of the loops.

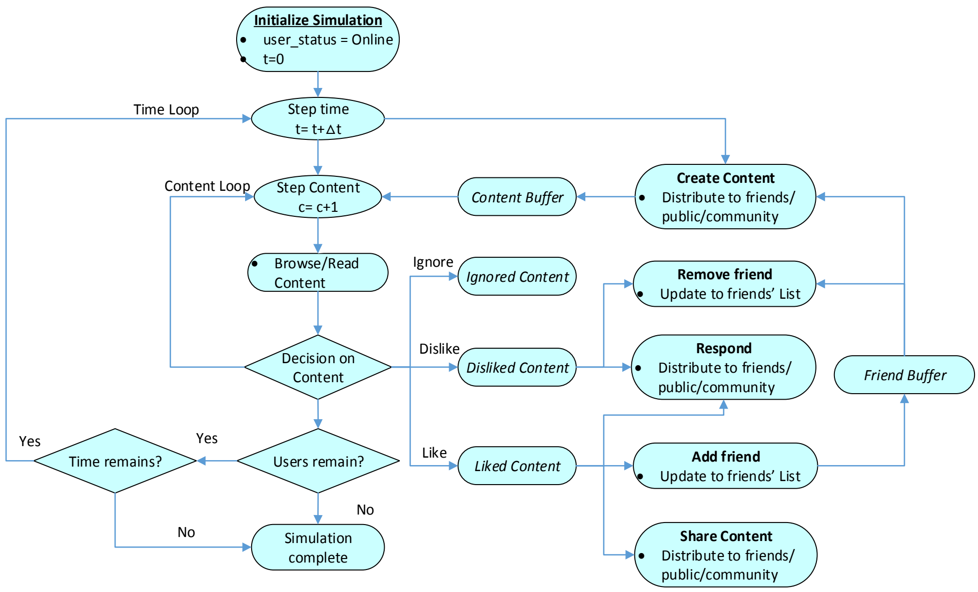

Figure 9 Simulation algorithm.

According to the simulation in **Figure 9**, initially, users were assigned an *online* state. Clock was a common control over the simulated time of each simulation task. When the simulation started, each user action was switched according to the activated transition, with regulated firing transitions identified by the stochastic property *P(x)* of motivations.

The simulation algorithm with discrete simulation steps was summarized as:

- Initialization: Users were online at time *t=0*.
- At time step *t+Δt*, a user *create content*. Content was transferred to content repository.
- Other users in the network browsed and read through the content.
- Consequence activities were provoked with regard to *decision* made on content:
  - *Ignore:* Content was unimportant, no consecutive action was performed.
  - *Like:* Content was interesting. Next possible actions could be *share content, add friend* and *respond*.
  - *Dislike:* Content was unsatisfied. Next possible actions could be *remove friend* from *friend list* or *respond*.
- Processes were continued as long as there existed content, users, and non-excessive time limit *(users* existed, *user_status* = *online* and *time* > *0*).

A meaningful relationship of co-occurrence interactions between users and content could be obtained from the simulation tracking. For scalability purpose, the number of users and content could be incremented by increasing the number of colors in a color set or the adaptation of color definitions.

Clustering coefficient evolution analysis

Since clustering coefficient was a measure of the existence of ties between individuals which enables the circulation of content, this measure was applied. In the early steps of simulation, clustering coefficient obscurely fluctuated on account of the edges of a few isolated nodes that did not form links between the nodes within their neighborhood, as shown in **Figure 10**. The probability that the 2 nearest neighbors of a vertex became the nearest neighbors of one another was in the range of 0.018 - 0.171. As time increased, clustering coefficient was gained, and could possibly be closed to one. Though its values were rather low, they remained larger than $O(n^{-1})$ (where *n* was number of vertices), and also agreed with the clustering coefficients of SNSs in the real world [24]. This phenomenon implied that nodes tended to be more tightly connected and localized cliques with their immediate neighbors through information exchanged. Therefore, H_1 was satisfied.

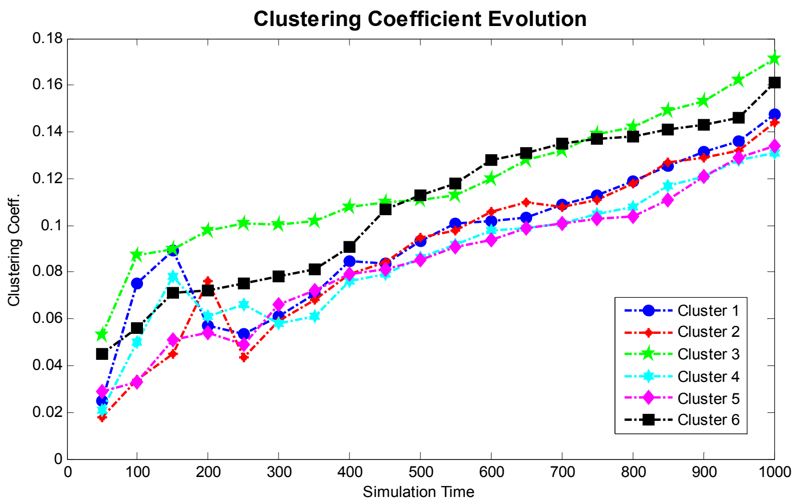

Figure 10 Clustering coefficient evolution.

Average shortest path length evolution analysis

The evolution of the interconnectedness of a network, which was an average distance between any pair of nodes, is demonstrated in **Figure 11**. In the early simulation steps, the path length was unable to be calculated because the network contained many disconnected components, making the series of data march towards infinity. As time increased, the average path length decreased from 3 to approximately 1.5, implying that a message could be transferred very fast through this network. On the other hand, if a user wanted to get in touch with others, he/she just needed to pass through 2 other persons, which agreed with the concept of 6 degrees of separation [24]. The more time users spent in the network, the links between existing nodes became denser. As time evolved, the distance between nodes was decreased. This phenomenon benefited the indirect relationships of individuals, and the result was consistent with the famous principle of [25], that the real networks exhibited a densification trend and their diameters shrunk over time. Hereby, H_2 was accepted.

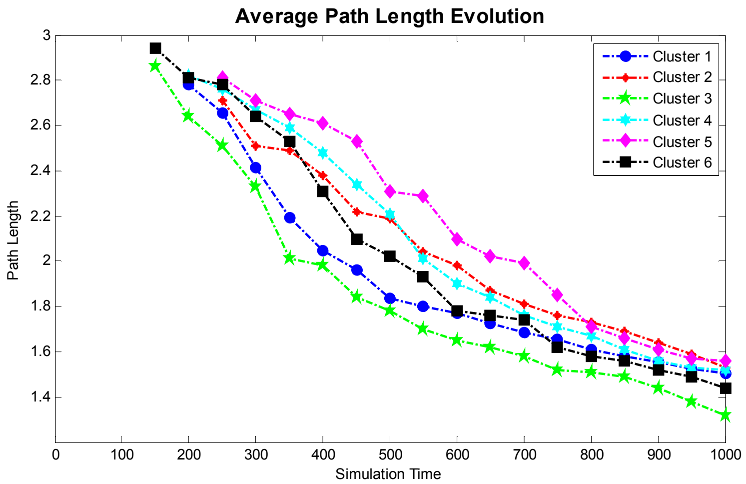

Figure 11 Average path length evolution.

Degree distribution evolution analysis

The degree distribution, gauged by fitting the frequency count of node degrees with the empirical probability distributions, is depicted in **Figure 12**. The initial and final degree of the node linking process at the 50 and 1000 simulation steps were observed. **Figure 12(a)** shows the distribution of node degrees. The possession of the power law property was indicated through a long right-tail, which was clearly far above the mean. Considering the overall distribution of clusters at each time step, **Figure 12(b)** shows the stationary values of scale variance, with an average of 2.006. During each simulation step, their values fluctuated considerably, because new links were sparsely added to disparate existing nodes. Though these values were lower than other studies, they governed the property of most real scale-free networks ($2<\gamma<3$) [24], conformed to the study of [25], and complied well with the theory of heavy tails in human behaviors [26]. It could be claimed that the number of propagandists with high interaction rate in this network were rare. Thus, H_3 was satisfied.

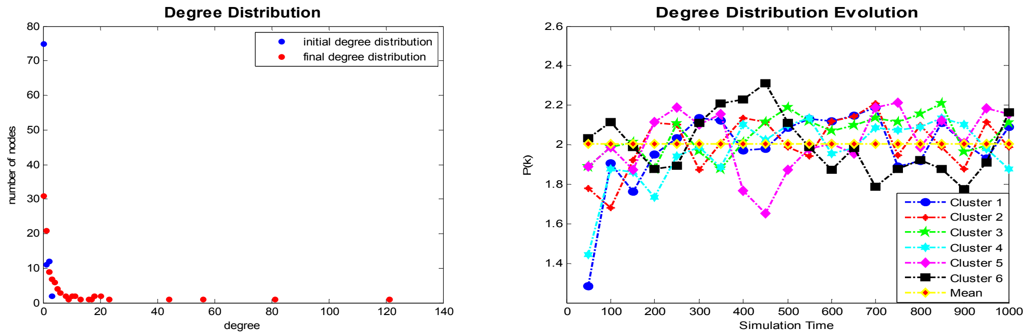

Figure 12 (a) Degree distribution, (b) Degree of Scale Invariance Evolution.

The complexity of the node linking process illustrated was imposed by the dependence between user-user, user-content, and content-reaction. The reciprocation to any content could be totally different through the user's personality, attitudes, and impressions. These events not only introduced distinguishable actions, but also diverse probabilities. As time evolved, interesting patterns of network structural characteristics were empirically appeared in the forms of: the higher the node degrees, the higher the clustering coefficient, and the lower the average path length resemblance to [8]. Based on the results, information tended to be propagated quickly in this network through high clustering coefficient and average shortest path length. The power law degree distribution also highlighted that the number of users who manipulated the flow of information was scarce. Referring to Watts and Strogatz's definition [24], the set of structural characteristics of this network indicated the small-world effect.

Discussions on network structural evolution

This work provides a big step forward in fulfilling the existing gaps in the knowledge of modeling user propagation behavior. The influences of user behavior, content, and network were integrated into the propagation processes, and their effects were demonstrated through a stochastic model of Petri nets. Three tactile contributions were the implementation of; (1) a stochastic diffusion model, which could illustrate the evolution of SNSs, (2) a dynamic parameterized model which best described each and every user in the network, and (3) the visualization and simulation of user interactions in SNSs. The minutiae of each discovery are described as follows.

First, how information diffuses can be synthesized and then systematized using a powerful approach of model decomposition and stochastic and deterministic abilities of CPNs. The major benefit is the number of users, and content can easily be escalated by adding more colors into their data repositories. Such flexibility allows the model to be useful in a wide range of applications, with potential advantages to SNS providers and businesses who have access to more precise data about user behavior. They can, therefore, make better use of the models, or adapt the models, for example, to accurately predict link creation to content and determine usage trends.

Second, how social interactions influence the network structure can be mathematically analyzed from model parameterization. Even if individual users inherently govern different levels of interaction dynamics, they can still be characterized by the dominant attributes of their personality traits [13]. The introduction of this model can release the burden on model parameter space, and increase the efficiency on future expansion of social interactions for other social networks. The likelihood of the prediction on the next activities can also be increased through the probabilistic model parameters identified by cluster. Specifically, by adding more determinants of user behavior to the future design of the systems, a higher accuracy in activity prediction is obtained.

Last, how user interactions affect the network topology is delineated by examining the inheritance of small world properties [1,22]. Compelling evidence is provided, such that; (1) the average path length agrees with the 6 degrees of separation. This indicates that, though the network composes of a very large population, each person can speedily communicate with others through a short path of intermediaries, (2) the clustering coefficient, the quantification of connection level between individuals, strongly indicates that the number of connections between users often transformed into transitivity. In other words, friends of friends can become friends without effort, and (3) the degree of distribution of node connectivity occupies the power law scaling. Such phenomenon indicates that the network is influenced by a few users who produce high rates of content distribution. Some content is also found to repeatedly appear in the network according to its popularity, which agrees with [27]. Moreover, the longer the lifetime, the connection tends to be densified, with more proliferation of clustering coefficient, more shrinkage to the average path length, and a higher value of degree scale invariant.

This analytical model and its simulations can specifically lead to the prediction of system behavior under different conditions, which is not easily studied experimentally in a live system. The growth and optimization of connections is pointed out using 2 interdisciplinary theories, human behavior and social network analysis. It is considered to be an alternative model to study the dynamic evolution process of SNSs, based on the human relation mechanism and the distribution of power law of complex networks. Intricate experimental data can also be simplified for ease, in order to advance the understanding of the systems of interest. The mechanisms of user behaviors that lead to a substantial delay and high content locality in information propagation can also be learnt from the simulation. Through a number of experiments with data describing user activity traced from Facebook, the proposed approach is proven to predict future user relationships and interactions, with a better performance both in terms of explaining observed data model fitting and of understanding the network evolution.

Conclusions

For the first time, the reciprocity of user behavior, interaction rules, content, and network structure to the interplay between information diffusion and network evolution was graphically and mathematically analyzed by a stochastic Petri net modeling method. Its merits contribute to an opportunity to explore how a community emerges from a network structure, how links are formed from user interactions, and how a network evolves from the flow of information. Three steps beyond the descriptive analysis were taken. First, a developed questionnaire was introduced to collect user behavior and characteristics. Then, a clustering method was applied to reduce the complexity of the analysis. Second, behaviors with respect to the diffusion process were extracted and composed into a model of CPNs. Last, simulations were used for efficient visualization and synthesis in order to demonstrate the connection mechanisms of user-content linkage creation with variable probabilities of motivation factors. The dynamic evolution of user interactions was explained through the known effects of the small-world principle via the empirical properties of high clustering coefficient, low average path length, and power law degree distribution. This model was proven to have better explanatory and predictive power over existing baseline models through the incorporation of individual characteristics and behaviors, network, and content into the diffusion mechanism.

Despite the meaningful contribution of this study, there are a few limitations that should be addressed, such that these findings are not able to be generalized beyond the context in which the user and structural network exist for 2 reasons. First, user behavior, user states, and transition probabilities were derived from a unique group of university students who qualified as extreme users and were technically oriented. Second, the study site was Facebook; assumptions and findings might not be applicable to other type of SNSs. For extension, acquiring behavioral data from a diverse range of population and over different SNSs can lead to a more robust inference applicable to typical real-world phenomena.

Acknowledgements

This work was partially supported by grants from the Memorial Foundation of King Rama VII and Queen Rambhaibarni and from King Mongkut's University of Technology, North Bangkok. The collaboration of Krirk University and its affiliates in providing data access was also appreciated.

References

[1] J Zhao, J Wu, X Feng, H Xiong and K Xu. Information propagation in online social networks: A tie-strength perspective. *Knowl. Inf. Syst.* 2012; **32**, 589-608.
[2] L Niu, X Han and Y Xu. On charactering of word-of-mouth propagation in heterogeneous online social networks. *J. Multimed.* 2014; **9**, 668-76.
[3] H Hu, H Mao, X Hu, F Hu, X Sun, Z Jing and Y Duan. Information dissemination of public health emergency on social networks and intelligent computation. *Comput. Intell. Neurosci.* 2015; **2015**, 181038.
[4] C Zhou, Q Zhao and W Lu. Impact of repeated exposures on information spreading in social networks. *Plos One* 2015; **10**, e0140556.
[5] Q Sun and Z Yao. Evolutionary game analysis of competitive information dissemination on social networks: An agent-based computational approach. *Math. Probl. Eng*. 2015; **2015**, 679726.
[6] K Subbian, C Aggarwal and J Srivastava. Mining influencers using information flows in social streams. *ACM Trans. Knowl. Discov. Data.* 2016; **10**, 26.
[7] S Vongsingthong, S Boonkrong and H Unger. Refining user lifetime model by big five characteristics. *In:* Proceedings of the 8th GI Conference, Autonomous Systems. Marollca, Spain, 2015, p. 342-57.
[8] J Lang and SF Wu. Social network user lifetime. *Soc. Netw. Anal. Min.* 2013, **3**, 285-97.
[9] X Chen, R Geng and S Cai. Predicting microblog users' lifetime activities: A user-based analysis. *Electron. Commer. Res. Appl*. 2015; **14**, 150-68.
[10] F Benevenuto, T Rodrigues, M Cha and V Almeida. Characterizing user navigation and interactions in online social networks. *Inf. Sci*. 2012; **195**, 1-24.
[11] H Yin, B Cui, L Chen, Z Hu and X Zhou. Dynamic user modeling in social media systems. *ACM Trans. Inf. Syst.* 2015; **33**, 1-44.
[12] S Uesugi. Effects of personality traits on usage of social networking service. *In*: Proceedings of the 2011 International Conference on Advances in Social Networks Analysis and Mining. Kaohsiung, Taiwan, 2011, p. 629-34.
[13] N Ljepava, RR Orr, S Locke and C Ross. Personality and social characteristics of Facebook non-users and frequent users. *Comput. Human. Behav.* 2013; **29**, 1602-7.
[14] A Ortigosa, JI Quiroga and RM Carro. Inferring user personality in social networks: A case study in Facebook. *In*: Proceedings of the 11th International Conference on Intelligent Systems Design and Applications. Córdoba, Spain, 2011, p. 563-8.
[15] Y Koo, S Lim, K Kim and Y Cho. Analysis of user characteristics regarding social network services in South Korea using the multivariate probit model. *Tech. Forecast Soc. Change* 2014; **88**, 232-40.
[16] PJ Haas. *Stochastic Petri Nets: Modeling, Stability, Simulation*. *In*: PW Glynn and SM Robinson (eds.). Operations Research and Financial Engineering. Springer, 2002.
[17] Z Wang, T Wen and W Wu. Modeling and simulation of rumor propagation in social networks based on Petri net theory. *In*: Proceedings of the 12th IEEE International Conference on Networking, Sensing and Control, Howard Civil Service International House. Taipei, Taiwan, 2015, p. 492-7.
[18] O Basak and YE Albayrak. Petri net based decision system modeling in real-time scheduling and control of flexible automotive manufacturing systems. *Comput. Ind. Eng*. 2015; **86**, 116-26.
[19] K Jensen. *An Introduction to the Theoretical Aspects of Coloured Petri Nets. In*: JW de Bakker, WP de Roever and G Rozenberg (eds.). A Decade of Concurrency, Lecture Notes in Computer Science. Springer-Verlag, 1994, p. 230-72.
[20] F Liu, M Heiner and M Yang. An efficient method for unfolding colored Petri nets. *In:* Proceedings of the 2012 Winter Simulation Conference. Berlin, Germany, 2012, p. 1-12.
[21] S Evangelista and JF Pradat-Peyre. *An Efficient Algorithm for the Enabling Test of Colored Petri Nets*. Aarhus University, Denmark, 2004, p. 137-56.
[22] DR Bild, Y Liu, RP Dick, ZM Mao and DS Wallach. Aggregate characterization of user behavior in Twitter and analysis of the retweet graph. *ACM Trans. Inter. Tech.* 2015; **15**, 1-24.

[23] S Boccalettia, V Latorab, Y Morenod, M Chavezf and DU Hwanga. Complex networks: Structure and dynamics. *Phys. Rep.* 2006; **424**, 175-308.
[24] J Leskovec, J Kleinberg and C Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. *In*: Proceedings of the 11th International Conference on Knowledge Discovery and Data Mining. Chicago, Illinois, USA. 2005, p. 177-87.
[25] AL Barabási. The origin of bursts and heavy tails in human dynamics. *Nature* 2005; **435**, 207-11.
[26] H Coltzau and H Unger. A model-based approach on superpopular contents in online social networks. *In:* Proceedings of the 8th GI Conference, Autonomous Systems. Marollca, Spain, 2015, 145-56.
[27] J Leskovec, L Backstrom, R Kumar and A Tomkins. Microscopic evolution of social networks. *In:* Proceedings of the 14th International Conference on Knowledge Discovery and Data Mining. Las Vegas, Nevada, USA, 2008, p. 497-506.