Received August 25, 2017, accepted December 19, 2017, date of publication January 3, 2018,
date of current version February 28, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2789334

Modeling Content and Membership Growth Dynamics of User-Generated Content Sharing Networks With Two Case Studies

RONG-HUEI CHEN, (Student Member, IEEE), AND SHI-CHUNG CHANG, (Member, IEEE)
Department of Electrical Engineering, National Taiwan University, Taipei 10617, Taiwan

Corresponding author: Shi-Chung Chang (e-mail: scchangee@ntu.edu.tw)

This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant NSC-98-2221-E002-138-MY3, Grant NSC-99-2219 -E-002-004, Grant NSC-100-2218-E-002-027-MY3, Grant MOST-102-2221-E-002-206, and Grant MOST-103-2221-E-002-220-MY2, in part by the Ministry of Education under Grant 102R3401-2, and in part by the U.S. Department of Commerce under Grant BS123456.

ABSTRACT User-generated content sharing networks (UGCSNets), in which members are content contributors as well as users, have had a significant impact on the sharing economy and on society via the sharing and reuse of contents. In a UGCSNet, managing for growth requires a quantitative grasp of how individual members' participation and sharing affect and are affected by the membership and content volume; these interactions form a dynamic loop. In this paper, a quantitative modeling approach for the loop dynamics of UGCSNet growth is developed by exploiting limited empirical data. A teaching material sharing network serves as a baseline case study, and Wikipedia serves as a validation case for the modeling approach design. The novel modeling approach consists of 1) set of generalized bass diffusion model-embedded stochastic difference equations (GBDSDEs) of the loop dynamics and 2) a quasi-bootstrap-based nonlinear least square method to extract from the limited empirical data and periodically update the model parameters as the UGCSNet evolves. In GBDSDEs, two difference equations describe the number of members and content volume evolution. The stochastic drives consist of measures of individual participation and content uploading. The drive models are an innovative generalization of the bass diffusion model as probabilistic models of known qualitative descriptions regarding how the individual willingness to participate and share is affected by the total membership and content volume. Analyses of the coefficients of determination show good fits between model predictions and actual outcomes for both Smart Creative Teachers Net and Wikipedia growths. Applications of the modeling approach to what-if analyses demonstrate its value to predict and assess the effects of specific managerial strategies—such as the initial content volume and the number of founding altruistic members—on the growth of a UGCSNet.

INDEX TERMS User-generated content, sharing network, growth dynamics, Bass diffusion model, network state-dependent generalization, positive feedback to individual, quantitative model.

I. INTRODUCTION OF GROWTH DYNAMICS MODELING

User-generated content sharing networks (UGCSNets) over the internet have become common worldwide [1], [6]–[9] and have had significant effects on the sharing economy and on society via the sharing and reuse of content. User-generated content is ''any form of content that was created by users of an online system or service'' [1], and a UGCSNet is an online platform for people to both contribute and exploit user-generated content. In particular, UGCSNets are creating new content sharing patterns, empowering users to be more creative, and leading to the development of new business opportunities [12]. Representative UGCSNets include Flickr [6], YouTube [7], Wikipedia [8], and Facebook [9]. Flickr provides a platform on which to share photographs. YouTube is one of the most popular UGCSNets, with over 1 billion daily visits, over 100 million videos watched every day and over 100 hours of video uploaded to YouTube every minute [7]. Wikipedia is a free encyclopedia written collaboratively by the people who use it; this process allows users to contribute and edit content pages in a web environment [8]. Facebook provides social networking services in the form of online meeting places in which it's over

2 billion monthly active users can share photos and videos, send messages, play games, and more [9]. Although these and other UGCSNets have achieved success, many more UGCSNets have failed to grow.

There are many forms of user-generated content: content sharing sites, such as YouTube, allow users to post their media; blog services allow users to post about many topics; wikis, such as Wikipedia, allow users to edit the content; and education knowledge sharing sites allow users to share their own knowledge and material. Other forms of user-generated content include internet forums where people talk and discuss about different topics and social networking sites, such as Facebook, Twitter, and Instagram, where users interact with other people by chatting, writing messages, or posting images or links. In this paper, we address a specific aspect of content sharing in UGCSNets. Our focus is on loop dynamics, including the evolution of content volume and membership. Social aspects and characteristics such as creating a profile page and publishing messages and connecting to other users and creating social relationships are not addressed in this paper.

The life cycle of a UGCSNet spans the following four stages [11]: introduction, growth, saturation and decline. In the introduction stage, network founders establish an infrastructure, recruit new members, install the initial content and develop operations and management rules. Network growth refers to a sustaining phase in which the growth rate of membership or content volume steadily exceeds a certain level, such as 5%. Network saturation is a phase in which the growth rate steadily remains below a certain level, such as less than 1%, whereas network decline refers to a sustained decrease in membership or new content sharing. A UGCSNet may skip any of these four stages in its evolution.

In the content sharing aspect of UGCSNets, individual members share texts, images, photos and content to other members and are providers of new content. Through content sharing processes, shared content becomes useful to individuals [12]. Users of a large UGCSNet have a high probability of finding what they are looking for [13]. The usefulness of the content in turn attracts and retains members. At the core of the process, the feedback loop links members, content sharing, useful content and the ability to attract and retain members [12]. However, the size of the network can negatively impact the process. Larger networks are more likely to be subject to free-riding [14]. Individuals may not contribute their own content because they expect that others will provide it. If the content contributions become inadequate, then the UGCSNet will not be able to continue attracting and retaining members [12]. The salient feature of UGCSNets is a loop through which individual participation and sharing and UGCSNet membership and content volume affect one another [12]–[18], as shown in Fig. 1. *Membership increases when the number of joining non-members is higher than the number of leaving members, and the cumulative content volume increases with new uploads. The willingness of a non-member to participate and the willingness of a member to stay and upload new content may increase, (positive feedback) or decrease (negative feedback) with an increase in the cumulative membership and content.* The overall impact on the dynamics of UGCSNet is a combination of positive and negative effects.

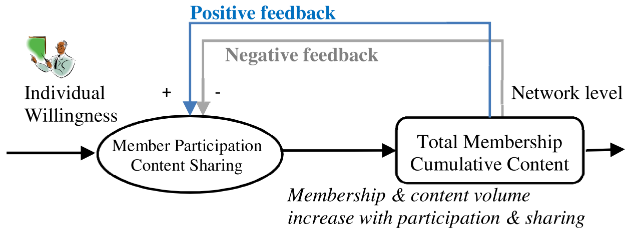


FIGURE 1. **Positive & negative feedback effects in a UGCSNet.**

The network lifecycle of a UGCSNet is affected by these positive and negative feedback effects. The positive feedback plays a major role in the dynamics from the introduction to saturation stages [12], [13]. The lifecycle from the introduction to saturation stages of a UGCSNet is expected to go through very rapid [11]. In the introduction stage, both the content volume and membership are low. In the growth stage, both the membership and content volume growth will speed up, and the positive feedback will increase the willingness of members to join and upload new content [13]. Positive feedback in the saturation stage reduces the willingness of members to leave; this effect leads to a higher saturation value of the membership and content volume. Conversely, negative feedback has a potentially significant impact on network decline [12] once more users become free riders. This study focuses on the positive feedback loop, leaving the negative feedback loop for future research.

As discussed above, many new UGCSNets did not grow and instead failed quickly. Only a few have grown to become a major UGCSNet with many members. Managing a UGCSNet to sustain healthy growth involves many strategic decisions/questions, such as predicting i1) the growth dynamics, including the point at which growth begins and the growth rate, and ii) the saturation time and saturation level (network size) in the saturation stage. The problem studied herein is to quantitatively model the loop dynamics between an individual and a network and to determine their effect on the growth of a UGCSNet based on qualitative descriptions from the literature and the empirical data. A major challenge is to quantitatively model the evolutions in the membership and content development as well as the interactions between them based on collective empirical data.

To develop a quantitative modeling approach that exploits empirical data from network evolution, this paper will utilize a teaching material sharing network (TMSN) as a baseline case study and Wikipedia as a validation case for modeling the approach design. Four research subproblems in the broader problem of developing a quantitative modeling approach will be addressed:

P1) how to model the effects of individual participation and sharing on membership and content volume;

P2) how to model the positive feedback effects of membership and content volume on the willingness to participate and upload new content;

P3) how to extract and update model parameters from limited empirical data; and

P4) how to validate the approach and apply the approach for predicting managerial strategies for management.

To address the above subproblems of P1 to P3 and to achieve our goals, our novel approach consists of two parts:

Part I: Generalized Bass diffusion model-embedded stochastic difference equations (GBDSDEs) of the loop dynamics to solve P1 and P2. It includes two submodels.

M1) A set of individual participation and sharing model-embedded stochastic difference equations describes the number of memberships and content volume evolution with individual participation and sharing as stochastic drives, respectively.

M2) The drives consist of measures of individual participation and content uploading. The drive models are an innovative generalization of the Bass diffusion model (BDM) [29] as probabilistic models of known qualitative descriptions of how an individual's willingness to participate and share is affected by the total membership and content volume.

Part II: A Quasi-bootstrap-based nonlinear least square (QBNLS) method [42] that estimates and periodically updates the GBDM parameters using limited empirical data as the UGCSNet evolves to address P3.

By analyzing the coefficients of determination between the predicted outcomes and actual ones in two real case scenarios (namely, Smart Creative Teachers Net (SCTNet) [10] and Wikipedia [8] services), a proof of the goodness of fit of the model is provided. A comparative simulation demonstrates the effectiveness of GBDSDEs in capturing the positive feedback effects on the network growth dynamics. Applications of this approach to what-if analyses demonstrate the ability to predict and assess the effects of specific managerial strategies, such as the initial content volume and number of founding altruistic members (members who do not expect to receive payment for publishing content), on the growth of a UGCSNet.

The remainder of this paper is organized as follows. Section II first summarizes the theoretical background of the existing quantitative and quantitative models. Section III then identifies the challenges in designing quantitative models. Section IV presents models of the evolution of the individual members and network states based on the empirical data. Section V then presents models for the participation and sharing willingness of a member with the feedback from network states. Section VI illustrates the extraction and updating processes of the model parameters and model validation based on the limited data. Applications of the proposed approach are shown in Section VII. The conclusions are presented in Section VIII.

II. THEORETICAL BACKGROUND

The previous literature has explored various aspects of UGCSNet dynamics. Butler [12] presented a qualitative resource-based model that links members, benefit provision, and member attraction and retention. Cha *et al.* [26] investigated the largest UCG video system, YouTube, and studied the popularity life cycle of videos, the intrinsic statistical properties of requests and their relationship with the age of the video to understand the efficiencies and inefficiencies of UGCSNets. Guille *et al.* [35] investigated the information diffusion phenomenon of online social networks including popular topic detection, information diffusion modeling and influential spreaders identification. Kumar *et al.* [27] considered user interactions in social networks and modeled the evolution of these networks. Numerous studies have stressed that network externalities appear to be important in the development of UGCSNets [13]–[19]. Other studies have adopted the BDM to capture and analyze the dynamics of the user acceptance of online UGCSNets [13], [34]. In addition, several studies have adopted epidemiological models to mathematically capture and explain the member adoption and abandonment of online social networks [30]–[33]. These studies have provided valuable insights into UGCSNet dynamics modeling and are an important foundation for further membership and content sharing analyses.

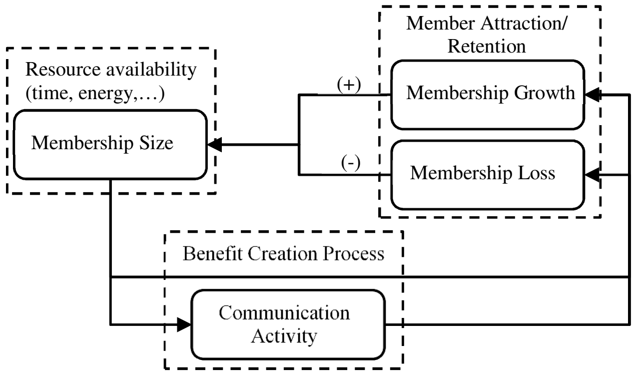


FIGURE 2. Resource-based model of sustainable social structures [12].

A. RESOURCE-BASED MODEL OF SUSTAINABLE SOCIAL STRUCTURES

Butler presented a resource-based model for telecommunication networks that focuses on a feedback process at the core of the internal dynamics of online social structure sustainability, as shown in Fig. 2 [12]. Members contribute time, energy, and other resources to enable a social structure that provides benefits for individuals. These benefits, which include information and influence, are the basis of the ability of a social structure to attract and retain members.

There are several blocks in the model:

1) *Membership Size and Resource Availability:* Current members are key providers of resources, and the membership size provides a measure of the resource availability. Larger networks tend to have access to more resources than smaller networks and are expected to more sufficiently provide valuable benefits to members

because resource availability is an aspect of benefit provision.

2) *Communication Activity and Benefit Provision*: Communication activity is a key factor of the dynamics model. Through communication processes, the shared resources will become useful for individuals. Regardless of the nature of the available resources within a network, without communication activity, those resources will remain dormant, and no benefits will be provided to individuals.
3) *Member Attraction and Retention*: The benefits in turn attract and retain members, thereby developing and sustaining the resource base of the network.

The resource-based model provides valuable insights into the loop dynamics of communication networks. However, the model focuses less on web-based UGCSNets and the relationship between user-generated content and users. In addition, the model does not focus on quantitative relations among each component.

B. NETWORK EXTERNALITIES

Katz and Shapiro first highlighted that network externalities refer to "the value or effect that users obtain from a product or service will bring about more values to consumers with the increase of users, complementary products, or services [20]". The classic example is a telephone network, in which the utility of the network to the individual increases with the number of users the individual can talk with. Network externalities have been studied in many areas, such as business firms [21]–[24], consumer electronics equipment [25] and UGCSNets [13]–[14], [17]–[19].

Positive network externalities also appear to be important to developing UGCSNets, in which the user-generated content increases with the number of members in the network [13]–[14], [17]–[18]. Researchers believe that an individual intends to join and use a social network when the number of members of the network reaches a critical mass [19], and they become more willing to use and upload content when more peers join [15]–[17]. This work uses the number of members of a UGCSNet to represent peer network externalities. In addition, other researchers have highlighted how the degree to which users perceive complementary items or services, such as user-generated content, influences a user's intent to join and use the network [13]; examples of such user-generated content include photo sharing, knowledge sharing, and video sharing. These services help increase the availability of complementary products perceived by users and further enhance a user's continued intention to use the network [16]–[17]. This study uses the amount of content to represent the indirect network externalities.

However, instances may also occur in which a network with a size reduces the value of the network for users, resulting in negative network externalities [14]. For example, when more users login to a web network (e.g., an internet access service), the network can become slow for all users. Asvanund *et al.* [14] focused on how positive and negative network externalities can influence the group size. Although their research describes the importance of network externalities on the growth and decline of UGCSNets, the concepts that they presented to describe the values and features were mainly qualitative, with little emphasis on quantitative methods for network lifecycle management.

C. BDM

To capture the dynamics of the user acceptance of online UGCSNets, many studies have adopted the BDM [13], [34]. This model, developed by Frank Bass [29], describes the process of how new products become adopted as a function of the level of product innovation and imitation between adopters and potential adopters using difference equation (1a), as shown below, where $P(t)$ is the number of adopters at time t and m is the total number of potential adopters in the market. The basic notion underlying the BDM is that adopters can be classified as innovators or as imitators, and the speed of adoption depends on the level of innovation and imitation among the adopters.

$$\begin{aligned} P(t+1) &= P(t) + \alpha \times (m - P(t)) + \beta \times \frac{P(t)}{m} \times (m - P(t)) \\ &= P(t) + [\alpha + \beta \cdot \frac{P(t)}{m}] \times (m - P(t)). \qquad \text{(1a)} \end{aligned}$$

The BDM demonstrates that product life cycles follow an S-curve pattern. An S-curve pattern implies that new product sales are initially slow, and the sales grow at a rapid rate before the rate of growth tapers off. According to the classical BDM (1a), the likelihood of adoption of a new product at time t, given that it has not yet been adopted, depends linearly on two forces.

1) In ***innovation term*** modeling, individuals decide to adopt an innovation independently of the decision of other individuals [29]. In the BDM, the constant coefficient α is an innovation coefficient that describes the constant attractiveness of a product/innovation to a user. In this term, the rate of diffusion is proportional to the number of people who have not yet tried the product (m-$P(t)$).
2) In ***imitation term*** modeling, adopters are influenced to adopt based on the adoption of a previous adopter [29]. This term from the equation shows that the rate of diffusion is proportional to the number of people who have purchased the product multiplied by those who have yet to purchase the product. The constant coefficient β is an imitation coefficient that describes the effects of imitation. The behavioral rationale for the model is that diffusion occurs among people who have not purchased the product imitating the people who have.

In particular, for $\beta > \alpha$, the user adoption will increase to a maximum before decreasing to zero [34], which becomes explicit by writing (1a) as

$$\begin{aligned}p(t) &= P(t+1) - P(t)\\ &= [\alpha + \beta \cdot \frac{P(t)}{m}] \times (m - P(t))\\ &= \alpha \cdot m + (\beta - \alpha)P(t) - \frac{\beta}{m}P(t)^2\\ &= a_1 + a_2P(t) - a_3P(t)^2, \end{aligned} \tag{1b}$$

which shows that the adoption rate ($p(t)$) is the result of two antagonistic processes: a propensity to grow ($a_1 + a_2P(t)$) countered by a propensity to decline ($a_3P^2(t)$). The BDM can often be adopted for dynamic analyses because it can model product sales or market acceptance and, thus, may also be able to explain the effect of collective attention dynamics on adoption motivations. Nonetheless, the BDM does not address the evolution of user-generated content or the interaction between content and user adoption.

D. EPIDEMIOLOGICAL MODELS

To mathematically characterize and explain user adoption and abandonment of online social networks, many studies have adopted epidemiological models [30]–[33]. Online social networks, which are one class of UGCSNets, can be depicted as a society infected by an epidemic disease. The process of adopting and abandoning a network resembles the process of infection and recovery. The typical epidemiological model is an SIR model that considers a fixed population and three compartments: susceptible people, $S(t)$; infected people, $I(t)$; and recovered people, $R(t)$. In the SIR model, the infectious disease spreads by contact with infected people. The infected people recover by acquiring immunity naturally. Similarly, the main motivation of a person joining a network might be influenced by the adoption of a previous adopter; this process is similar to the imitation effect in the BDM. Membership growth in a network is similar to that of an infectious disease because it is driven by contact with existing members. Similarly, becoming a network user corresponds to becoming infected, and quitting corresponds to recovering from infection. Therefore, the SIR model can represent the growth and decline of a social networking service. Becoming a social networking service user corresponds to infection, and quitting the social networking service corresponds to recovering from infection.

Compared to the BDM, the epidemiological model may model not only the adoption but also the abandonment of UGCSNets. In terms of adoption, the epidemiological model addresses the relationship between new members and existing members (the imitation effect in the BDM) but not the relationship between new members and the resources of the network (the innovation effect in the BDM).

III. THE CHALLENGES OF QUANTITATIVELY MODELING UGCSNET GROWTH DYNAMICS

The objective of this section is to identify the challenges of quantitatively modeling the loop dynamics of the growth of UGCSNets based on data collected from a baseline study case, SCTNet, and a validation case, Wikipedia. Then, we propose a framework for a modeling approach to overcome these challenges.

A. THE CHALLENGES OF QUANTITATIVELY MODELING GROWTH DYNAMICS

SCTNet [10] is a TMSN that is popular among elementary school teachers in southern Taiwan; the network was introduced in January 2000 and is operated by the Education Bureau of Kaohsiung city and National Sun Yat-Sen University. SCTNet provides members with basic sharing functions with regard to teaching materials (TMs), such as searching, uploading, and downloading TMs. The network also provides advanced functions, such as forums, workshops, and e-learning programs, which enable digital content sharing and participation in discussions of interest to instructors. Member teachers contribute TMs to SCTNet on a voluntary basis. In the content sharing aspect of the TMSN, individual member activities include *participation activities*, such as joining and leaving, and *sharing activities*, such as downloading and uploading TMs. The total number of members and the total TM volume are two *states of network* development that are the cumulative effects of individual participation and uploading.

Wikipedia [8] is a free-content internet encyclopedia that is supported and hosted by the non-profit Wikimedia Foundation, and it was launched in 2001. The usability and functionality of Wikipedia makes it easy for users to create new content pages and to edit and share knowledge with page viewers across the world [2], [3]. In addition, Wikipedia generates large data sets regarding its evolution and dynamics. Historically, Wikipedia grew from hundreds of active editors in 2001 to thousands in 2004 and peaked in 2007 [8]. The contributions of editors have propelled Wikipedia to a high level of quality and completeness [4]. Suh et al. noted that the tremendous growth of Wikipedia attracted more contributors to increase its value, and in turn, Wikipedia became valuable [5]. In regard to sharing aspects in growth, Wikipedia has similar characteristics to a TMSN.

To address the problem of quantitatively modeling growth dynamics, the collective statistics of a study case, SCTNet, were collected monthly from January 2001 to December 2007, covering the periods of introduction, growth, and saturation. The collective statistics of the validation case, Wikipedia, were collected monthly from 2001 to 2007. Such collective statistical data are available for analysis by researchers. Based on the collective statistics, there are four research challenges to overcome when quantitatively modeling UGCSNet growth dynamics, one for each subproblem mentioned in Section I.

C1) Modeling individual participation and sharing and their effects on membership and total content volume using only collective statistics

Differences exist among the individual users of a network, but only collective and network state statistics are available

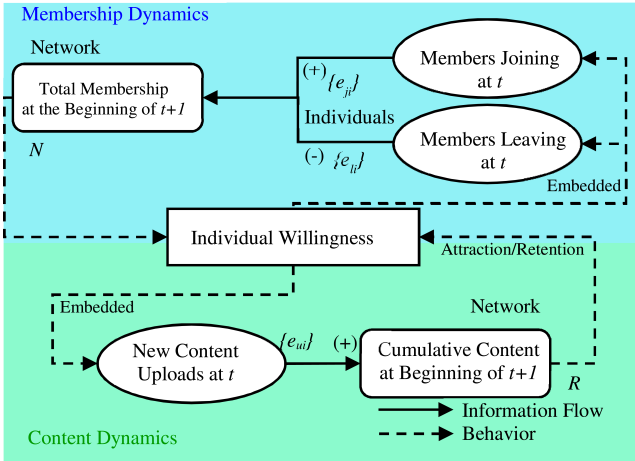


FIGURE 3. **Proposed modeling framework of loop dynamics.**

from a UGCSNet, such as a TMSN. Modeling individual differences with collective empirical data is challenging.

C2) Modeling the positive feedback effect from membership and total content volume to individual sharing and participation

Developing this quantitative model requires the innovative integration of qualitative descriptions from the literature with a positive feedback effect extracted from empirical data.

C3) Extracting and updating model parameters with the evolution of a UGCSNet *given limited data availability*

Only limited data and statistics are available for modeling, particularly during the early development stage of a UGCSNet. Extracting and updating the dynamic model parameters that fully exploit the available data as a UGCSNet evolves are essential and challenging tasks.

C4) Validating the resultant approach for modeling the loop dynamics of a TMSN and Wikipedia

The membership and total content volume are the superposition of an individuals' willingness to participate and share, and the willingness of each participation to share is affected by the membership and total content volume. It is quite challenging to determine how to evaluate network dynamics with sufficient fidelity and to validate the resulting approach using the available empirical data from the introduction through the saturation stages. In addition, demonstrating the potential of applications that can manage the growth of a new UGCSNet is also essential but challenging.

B. MODELING FRAMEWORK OF QUANTITATIVE GROWTH DYNAMICS

To address C1 and C2, we develop a set of GBDSDEs to model the network growth dynamics with the aim of capturing the feedback loops of the UGCSNet. Fig. 3 illustrates the proposed modeling framework. The upper side depicts the membership dynamics, and the lower side describes the content dynamics. Members contribute content to the UGCSNet. The members' willingness to share content is affected by the existing members and current content. The cumulative content is the aggregation of the individual contributions. Then, the shared content and existing members constitute the basis from which to attract and retain members. In GBDSDEs, two difference equations describe the membership and content volume evolutions separately using individual participation and content uploading as stochastic drives. The drive models are a generalization of the BDM and innovatively model how the activities of individuals are affected by the feedback of the network states of the existing members and cumulative content.

To address C3, a QBNLS method is developed to extract and update the model parameters from limited empirical data as the UGCSNet evolves. To address C4, we develop what-if scenarios to analyze how the number of altruistic members among the founding members and the number of initial content entries may influence UGCSNet growth dynamics such that the models may help a manager find cost-effective strategies for network growth.

IV. DIFFERENCE EQUATION MODELING OF THE LOOP DYNAMICS WITH INDIVIDUAL ACTIVITIES AS STOCHASTIC DRIVES

Now we consider modeling the individual activity events and their effects on the network states based on the collected statistics of $n(t)$, $N(t)$, $r(t)$, $D(t)$ and $L(t)$. A set of individual participation and sharing model-embedded stochastic difference equations (SDEs) [36], [37] are developed to describe the network state evolutions of the membership and content volume that are driven by the probabilistic occurrences of individual participation and sharing events. Limited by having access to collective data exclusively, the differences among the individuals are captured by modeling the occurrence of each activity event as 0-1 indicator random variables, which are independent and identically distributed among the members. We then define how the event occurrence probabilities are estimated from available empirical data.

A. DESIGN OF THE PARTICIPATION AND SHARING MODELS

After developing the network growth dynamic models, a few symbols are defined in Table 1. The assumptions of the model are as follows.

AS1) This paper adopts the content volume to represent content richness in the following discussions.

AS2) Event occurrences during $(t, t+1]$ are probabilistic and depend on the network states of the membership $N(t)$ and content volume $R(t)$ and member i's participation state $x_{pi}(t)$. Because only collective data are available, the occurrence probability of each type of event is assumed to be independent and identically distributed among the individuals. During $(t, t+1]$, a member has at most one occurrence of each event type.

AS3) A member will access the TMSN/Wikipedia at least once during each time slot $(t, t+1]$.

TABLE 1. Symbols for the participation and sharing models.

t	time slot index, $t=0, 1,2,3$ ...
$n(t)$	number of new members during (t-1, t]
$r(t)$	number of new content entries uploaded during (t-1, t]
$D(t)$	total number of downloads during (t-1, t]
$L(t)$	total number of logins by members during (t-1, t]
TN	total number of people.
S_a	a set of joining, leaving, uploading and downloading activity events, and S_a={j, l, u, d}.
M	s set of members.
NM	a set of non-members.
Individual participation and sharing states	
$x_{pi}(t)$	participation state variable of user i at time t, where $x_p(t)$=1 if user i is a member at time t and $x_p(t)$=0 otherwise.
$x_{si}(t)$	cumulative number of uploading/downloading events by member i at time t, $x_s(t) \in \{0\} \cup N$ and $s \in \{u, d\}$.
Network states	
$N(t)$	total membership at time t with $N(0)$ given.
$R(t)$	number of contents at time t.
Probabilistic event occurrence	
$e_{ai}(t+1, x_{pi}(t), N(t), R(t))$	random variable of the occurrence of activity event a given member i's participation state, $x_{pi}(t)$, and network states, $N(t)$ and $R(t)$, during (t, $t+1$], where $e_{ai}(t+1, x_{pi}(t), N(t),$ and $R(t))$=1 if an activity event a occurs and $e_{ai}(t+1, x_{pi}(t), N(t)$, and $R(t))$=0 otherwise; the admissible $a \in \{l, u, d\}$ if $x_{pi}(t)$ = 1 and $R(t)$ > 0, and $a \in \{l, u\}$ if $x_{pi}(t)$=1 and $R(t)$=0 and $a \in \{j\}$ if $x_{pi}(t)$=0.
$p_{ai}(x_{pi}(t), N(t), R(t))$	occurrence probability of activity a given user i's participation state, $x_{pi}(t)$, and network states, $N(t)$ and $R(t)$, during (t,$t+1$], where $a \in S_a$ and is admissible under the given $x_{pi}(t)$, $N(t)$ and $R(t)$.

1) PARTICIPATION MODELS

Fig. 4(a) depicts the state transition of the membership state of an individual i through the occurrence of member participation (join and leave) events. The state space $S_p = \{0, 1\}$ describes the membership states of a non-member and member. A participation transition occurs if event joining/leaving occurs. A non-member that joins the network becomes a member. In contrast, once a member leaves, the member then becomes a non-member. Equations (2a)–(2c) describe the transition between the member and non-member states using a stochastic difference equation, in which (2a) and (2b) define random variables of *joining* and *leaving* events, and (2c) describes the participation state equation of user i.

$$e_{ji}(t+1, x_{pi}(t), N(t), R(t)) \equiv \begin{cases} 1, \text{ with } p_{ji}(x_{pi}(t), N(t), R(t)), & \text{if } x_{pi}(t) = 0; \\ 0, \text{ with } 1 - p_{ji}(x_{pi}(t), N(t), R(t)), & \text{if } x_{pi}(t) = 0; \\ 0, & \text{otherwise.} \end{cases} \tag{2a}$$

$$e_{li}(t+1, x_{pi}(t), N(t), R(t)) \equiv \begin{cases} 0, \text{ with } 1 - p_{li}(x_{pi}(t), N(t), R(t)), & \text{if } x_{pi}(t) = 1; \\ 1, \text{ with } p_{li}(x_{pi}(t), N(t), R(t)), & \text{if } x_{pi}(t) = 1; \\ 0, & \text{otherwise.} \end{cases} \tag{2b}$$

$$x_{pi}(t+1) = x_{pi}(t) + e_{ji}(t+1, x_{pi}(t), N(t), R(t)) - e_{li}(t+1, x_{pi}(t), N(t), R(t)), \quad x_{pi}(0) = 0. \tag{3}$$

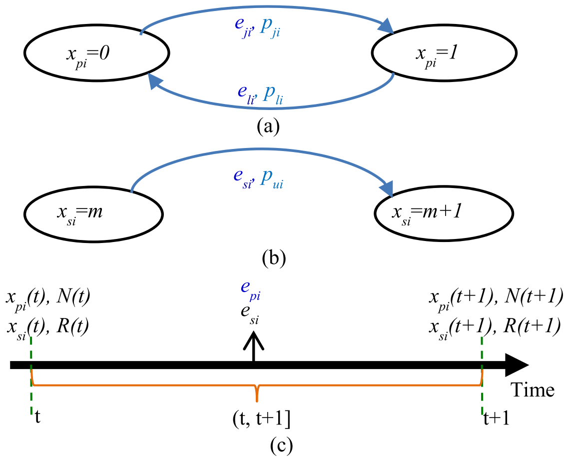


FIGURE 4. (a) Participation model and (b) sharing model of a member i and (c) individual and network state transitions.

The space of the total membership state, $N(t)$, is $S_N = \{0, 1, \ldots, TN\}$. As shown in Fig. 4(c), membership change is a combination of the joining and leaving events of individuals. The maximum number of members leaving during $(t, t+1]$ is $N(t)$, and the maximum number of non-members joining during $(t, t+1]$ is $TN - N(t)$. Thus, we express the total membership evolution using an individual participation model-embedded stochastic difference equation as follows:

$$N(t) \equiv \sum_{i=1}^{TN} x_{pi}(t). \tag{4a}$$

By combining (2), (2c) and (4a), we obtain

$$N(t+1) = N(t) + \sum_{k=1}^{TN-N(t)} e_{jk}(t+1, x_{pk}(t), N(t), R(t)) - \sum_{i=1}^{N(t)} e_{li}(t+1, x_{pi}(t), N(t), R(t)), \tag{4b}$$

where k represents the non-members ($k \in NM$) and i represents the members ($i \in M$).

2) SHARING MODELS

Fig. 4(b) depicts the state transition of the sharing (uploading and downloading) states of a member i through the occurrence of member sharing (uploading and downloading) events. The state space $S_s = \{0, 1, 2, \ldots\}$ describes the amount of content that is uploaded/downloaded, where $s \in \{u, d\}$. Equations (5a)–(5c) describe the sharing state transitions as the SDEs below, where (5a) and (5b) define the random variables of the *uploading* and *downloading* event occurrences and (5c) and (5c) describe the sharing state equations of member i.

$$e_{ui}(t+1, x_{pi}(t), N(t), R(t)) \equiv \begin{cases} 1, \text{ with } p_{ui}(x_{pi}(t), N(t), R(t)), & \text{if } x_{pi}(t) = 1; \\ 0, \text{ with } 1 - p_{ui}(x_{pi}(t), N(t), R(t)), & \text{if } x_{pi}(t) = 1; \\ 0, & \text{otherwise.} \end{cases} \tag{5a}$$

$$e_{di}(t+1, x_{pi}(t), N(t), R(t)) \equiv \begin{cases} 1, \text{ with } p_{di}(x_{pi}(t), N(t), R(t)), & \text{if } x_{pi}(t)=1, R(t)>0; \\ 0, \text{ with } 1-p_{di}(x_{pi}(t), N(t), R(t)), & \text{if } x_{pi}(t)=1, R(t)>0; \\ 0, & \text{otherwise.} \end{cases} \tag{5b}$$

$$x_{ui}(t+1) = x_{ui}(t) + e_{ui}(t, x_{pi}(t), N(t), R(t)), \quad x_{ui}(0) = 0; \tag{6a}$$

$$x_{di}(t+1) = x_{di}(t) + e_{di}(t, x_{pi}(t), N(t), R(t)), \quad x_{di}(0) = 0. \tag{6b}$$

Similarly, the state space of the content volume, $R(t)$, is $S_R = \{0, 1, 2, \ldots\}$. The maximum number of members uploading during $(t, t+1]$ is $N(t)$. Then, we express the total content volume evolution using an individual sharing model-embedded corresponding stochastic difference equation as follows:

$$R(t) \equiv \sum_{i=1}^{TN} x_{ui}(t), \tag{7a}$$

which can be combined with (5a) as follows:

$$R(t+1) = R(t) + \sum_{i=1}^{N(t)} e_{ui}(t+1, x_{pi}(t), N(t), R(t)). \tag{7b}$$

B. ESTIMATING EVENT OCCURRENCE PROBABILITIES BASED ON THE AVAILABLE EMPIRICAL DATA

The key to the models of the individual activity events and state transitions involves three event occurrence probabilities $p_{ai}(x_{pi}(t), N(t), R(t)), a \in \{j, u, d\}$. By exploiting the monthly available statistics of $n(t)$, $N(t)$, $r(t)$, $D(t)$ and $L(t)$ and the notation of the relative frequency, the empirical data-based estimates of the event occurrence probabilities are defined as follows.

1) OCCURRENCE PROBABILITY OF JOINING EVENT

$$\hat{p}_{ji}(t+1, x_{pi}(t), N(t), R(t)) \equiv \frac{n(t+1)}{TN - N(t)}, \tag{8a}$$

which corresponds to the ratio of the number of new members to the number of non-members at time t.

In practice, for the SCTNet and Wikipedia, there is no periodic renewal requirement to maintain membership. Thus, UGCSNet management cannot be used to determine whether a member has left or simply has become inactive. Hence, the probability of leaving is not estimated.

2) OCCURRENCE PROBABILITIES OF THE UPLOADING AND DOWNLOADING EVENTS

$$\hat{p}_{ui}(t+1, x_{pi}(t), N(t), R(t)) \equiv \frac{r(t+1)}{L(t)} \tag{8b}$$

$$\hat{p}_{di}(t+1, x_{pi}(t), N(t), R(t)) \equiv \frac{D(t+1)}{L(t)} \tag{8c}$$

which correspond to the ratios of the number of uploaded/downloaded content entries to the total number of logins, respectively. The occurrence probability of uploading/ downloading describes the individual uploading/ downloading willingness of a member per login [36] and assumes AS3. In current practice, a member may download more than one unit of content per login. However, individual data are not available for estimating amount of downloaded content per individual login.

V. GENERALIZED BDMS FOR EMBEDDING POSITIVE FEEDBACK INTO STOCHASTIC DRIVES

The qualitative research results presented in [12]–[18] suggest that the event occurrence probabilities of individual participation and content uploading in (2) and (5) are functions of the total membership and content volume. This section will first summarize the qualitative models of the UGCSNet loop and then develop quantitative models to determine how one's individual willingness to participate and share is affected by the membership and content volume in the growth stage. An analysis of the empirical data from the SCTNet reveals that the event occurrence probabilities increase according to an S-shaped curve with respect to the values of the membership and content volume from the introduction through the saturation stages. Taken together, these results motivate extending the use of the BDM to describe technology acceptance within a population [29] over time to a baseline model of how the probability of each event occurrence of a member evolves with network states instead of time. The results from references [13] and [39] further indicate that the attractiveness of a UGCSNet is positively correlated with the content that is shared among the network members. Thus, instead of adopting the constant innovation coefficient, as in BDM, we construct probability models for innovative event occurrence that integrate the aforementioned qualitative descriptions into a content volume-dependent innovation coefficient to capture the positive feedback effect of the content volume on the two drives expressed in (2) and (5). The event occurrence probability models are then embedded into the SDEs to capture the loop dynamics between the network states and individual events.

A. EXTENSION OF THE BDM TO MODEL AN INDIVIDUAL'S WILLINGNESS TO PARTICIPATE AND SHARE

References [12] and [13] show that the amount of content and the number of community members are two main motivations for individual participation and sharing. A high existing membership and rich content tend to attract new members and keep members from leaving [12]. Powell and Tapscott [15], [16] also noted that an individual becomes more willing to share and join when more peers share content and join; this conclusion implies that an individual's willingness changes with respect to the value of the two network states. The effects of membership and content development on an individual's willingness are considered as the imitation and innovation effects, respectively, and these effects inspire us to extend the BDM [29] as a baseline model of how each event occurrence probability of a teacher evolves along with a state variable instead of time.

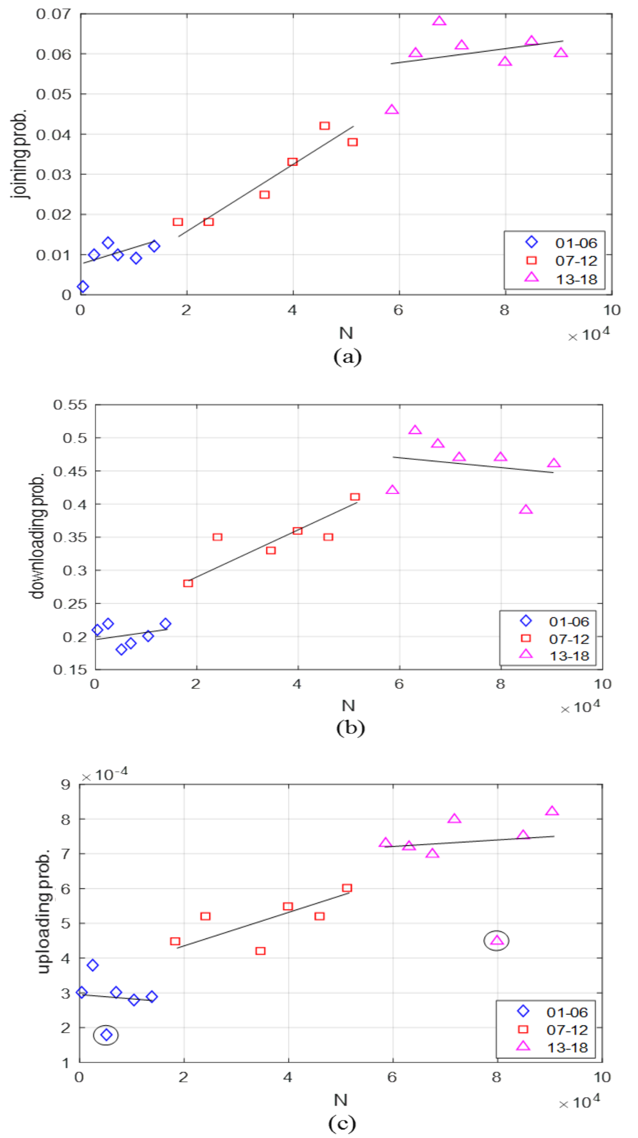


FIGURE 5. Scatter plots: (a) joining, (b) downloading, and (c) uploading probability trends vs. total membership.

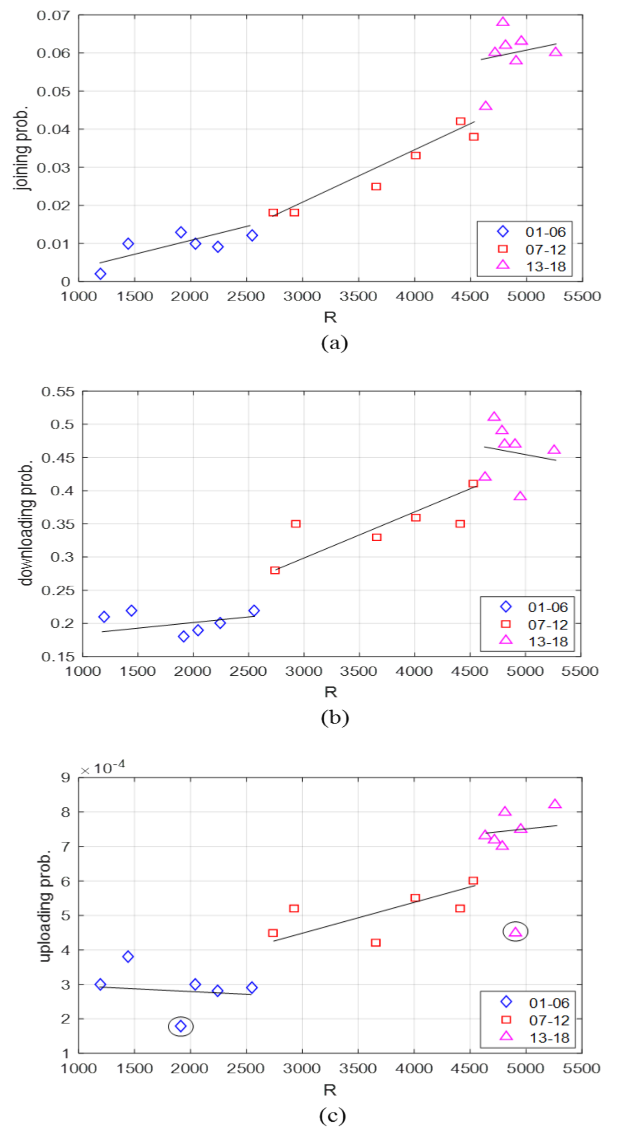


FIGURE 6. Scatter plot: (a) joining, (b) downloading, and (c) uploading probability trends vs. cumulative content volume.

In addition, observations from Chen and Chang [37] revealed that the development trajectories of the joining event occurrence probability in (2) are S-shaped curves as a function of the membership (N). A further analysis of the empirical data from the SCTNet reveals that the event occurrence probabilities of joining, uploading and downloading ((2) and (5)) are S-shaped curves with respect to the values of the membership (N) and content volume (R), as shown in Figs. 5 and 6. The SCTNet data were collected from the period between 2001 and 2005. Each data point in the scatter plot corresponds to the statistics of $p_{ai}(x_{pi}(t), N(t), R(t))$, $a \in \{j, u, d\}$ following the definitions of (8a)–(8c). As shown in Figs. 5(a)–5(c), we can characterize the evolutions of the probability of joining, downloading, and uploading with respect to the membership using three segments: i) the slow start from data points 01–06 (diamond shaped), ii) the fast growth from data points 07–12 (square shaped), and iii) the saturation in the data points from 13–18 (triangle shaped). Similar characterizations regarding the evolutions of the three event occurrence probabilities with respect to the content volume can be observed in Figs. 6(a)–6(c). There are two abnormally low points in Figs. 5(c) and 6(c) that are indicated by a circle at which sharp decreases in the uploading probability occurred when school semesters ended in July, 2001 and July, 2005. Except for the two abnormally low points, the trends in the event occurrence probabilities with respect to $N(t)$ and $R(t)$ basically fit S-curves. Therefore, the BDM in [29] can be naturally extended to model the S-shaped curves of the individual event occurrence probabilities over state variables N and R.

The manner in which each of the three event occurrence probabilities, $p_a, a \in S'$ and $S' \equiv \{j, u, d\}$, varies with respect to $N(t)$ under given a $R(t)$ and $x_{pi}(t)$ can be modeled in (1a) by replacing argument t with $N(t)$, P with probability p_{ai} and m with 1, as shown in (9a),

$$\begin{aligned} &p_{ai}(x_{pi}(t), N(t)+1, R(t)) \\ &\quad = p_{ai}(x_{pi}(t), N(t), R(t)) + \alpha_{ai} \cdot (1 - p_{ai}(x_{pi}(t), N(t), R(t))) \\ &\qquad + \beta_a \cdot p_{ai}(x_{pi}(t), N(t), R(t)) \cdot (1 - p_{ai}(x_{pi}(t), N(t), R(t))), \end{aligned} \tag{9a}$$

where the second term on the right-hand side of the equation is an innovation term, α_a reflects how the content volume yields a constant effect on the event occurrence probability in a UGCSNet, and the third term captures how imitation among members affects the event occurrence probability of a member.

As for modeling the leaving members, direct observations of a member leaving in the case of the SCTNet and Wikipedia were not available. However, members who leave a UGCSNet can be viewed as joining the community of non-members; this definition leads to the notion of modeling the probability of leaving as an S-curve. Therefore, the leaving occurrence probability is modeled by extending the BDM as follows:

$$\begin{aligned} &p_{li}(x_{pi}(t), N(t)+1, R(t)) \\ &\quad = p_{li}(x_{pi}(t), N(t), R(t)) - \alpha_{li} \cdot (p_{li}(x_{pi}(t), N(t), R(t))) \\ &\qquad - \beta_l \cdot p_{li}(x_{pi}(t), N(t), R(t)) \cdot (1 - p_{li}(x_{pi}(t), N(t), R(t))). \end{aligned} \tag{9b}$$

B. GENERALIZATION TO MODEL THE POSITIVE FEEDBACK OF THE CONTENT

Although (9a) and (9b) are innovative extensions of the BDM, EBDMs, for modeling how the activity occurrence probabilities evolve with the membership value from the introduction through the saturation stages, the two equations do not yet capture the content volume feedback of an individual's willingness to participate. The results in references [13] and [39] further indicated that the attractiveness of a UGCSNet is positively correlated with the content that is shared among the network members. A member's willingness to participate, use resources and upload content increases with the increasing content volume, and this result inspires us to further generalize the EBDM by replacing the constant coefficient of the innovation term with the content-dependent coefficient to capture the evolution of the event occurrence probability.

Let the level of innovation, I, be defined as

$$I \equiv f(R(t)), \tag{10}$$

where function $f(\cdot)$ describes the content-dependent attractiveness of the UGCSNet and is assumed to monotonically increase and marginally decrease with respect to the content volume. Such assumptions are based on the following research results from the literature:

i) I monotonically increasing with R: Belvaux [13] claimed that the probability of a member finding what they are looking for increases as the size of the shared content grows. In a UGCSNet, members can find and download the content they need more easily than without a UGCSNet. An increase in the content volume can enhance UGCSNet attractiveness.

ii) I marginally decreasing with R: The innovation of UGCSNets is a type of "sustaining innovation." Christensen [39] claimed that sustaining innovation has the features of f1) low-cost innovation and f2) eventual saturation, which imply a marginally decreasing innovation rate.

Based on the two assumed properties regarding $f(\cdot)$, we adopt a specific form of the log function for $f(\cdot)$, i.e.,

$$I_a(R(t)) = \alpha_a(\log R(t) + 1), \tag{11}$$

where α_a is a positive constant. We then generalize the EBDM in (9a) to capture the positive content feedback effect using a content volume-dependent innovation coefficient as follows:

$$\begin{aligned} &p_{ai}(x_{pi}(t), N(t), R(t)+1) \\ &\quad = p_{ai}(x_{pi}(t), N(t), R(t)) + I_a(R(t)) \cdot (1 - p_{ai}(x_{pi}(t), N(t), R(t))) \\ &\qquad + \beta_a \cdot p_{ai}(x_{pi}(t), N(t), R(t)) \cdot (1 - p_{ai}(x_{pi}(t), N(t), R(t))), \end{aligned} \tag{12a}$$

where $a \in S'$. Similarly, the generalization of the leaving occurrence probability model is defined as follows:

$$\begin{aligned} &p_{li}(x_{pi}(t), N(t), R(t)+1) \\ &\quad = p_{li}(x_{pi}(t), N(t), R(t)) - I_l(R(t)) \cdot (p_{li}(x_{pi}(t), N(t), R(t))) \\ &\qquad - \beta_l \cdot p_{li}(x_{pi}(t), N(t), R(t)) \cdot (1 - p_{li}(x_{pi}(t), N(t), R(t))). \end{aligned} \tag{12b}$$

Fig. 7 presents a block diagram of the GBDSDEs which model the loop between the network and the individual and provide a foundation to evaluate UGCSNet growth dynamics. GBDSDEs capture the previously modeled qualitative descriptions in the literature including the following: QD1) the positive feedback loop whereby an individual member's willingness to participate and upload new content is positively related to the membership and content volume [12]–[18] and QD2) The lifecycle from the introduction to saturation stages of a UGCSNet is expected to go through very rapid [11].

VI. QUASI-BOOTSTRAPPING MODEL PARAMETER EXTRACTION AND MODEL VALIDATION BY SIMULATION

This section addresses the problem of extracting and periodically updating the GBDM model parameters in (11) and (12a) by exploiting the limited empirical data increments as the network evolves. The two specific model parameters to extract are the innovation and imitation coefficients, α_a and β_a in (11) and (12a), respectively. Once α_a and $\beta_a$are extracted from the available data, equations (11) and (12a) provide the complete dynamics of the event occurrence probabilities and, in turn, the complete membership and content volume

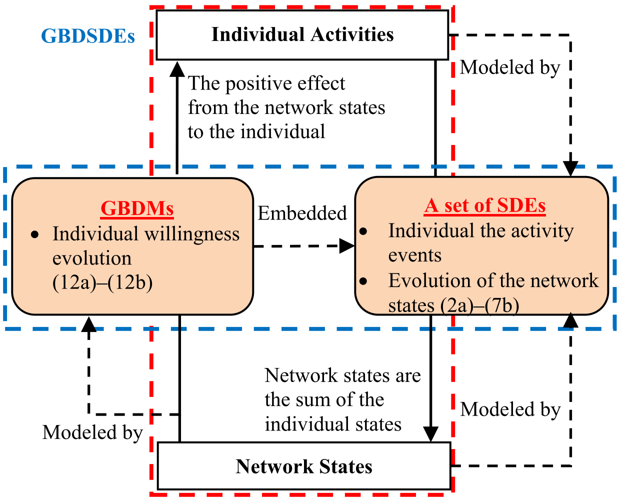


FIGURE 7. Block diagram of the GBDSDEs.

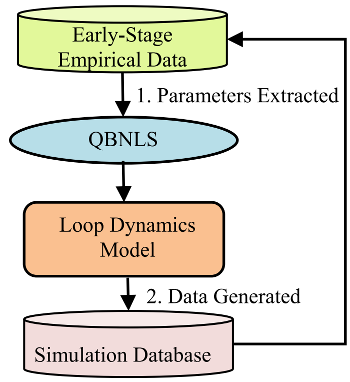


FIGURE 8. Schematic diagram of the parameter extraction and updating process.

dynamics. Such dynamics are the foundation of predicting the network state evolution. In validating our quantitative modeling approach, the comparative simulation first demonstrates the effectiveness of the GBDSDEs in capturing the positive feedback effect on the growth dynamics of the networks. Then, comparisons between the simulated evolution of the membership and content volume using the SCTNet and Wikipedia data finally validate the modeling approach.

A. GBDM PARAMETER EXTRACTION AND PERIODIC UPDATING BY USING THE QBNLS METHOD TO EXPLOIT EARLY STAGE EMPIRICAL DATA

Now we consider how to extract and update the embedded parameters of the innovation and imitation coefficients (α_a, β_a) in (11) and (12a) for each activity event $a \in \{j, u, d\}$ by exploiting the limited empirical UGCSNet data ($n(t)$, $N(t)$, $r(t)$ and $R(t)$) increments as the network evolves. The calculations of the event occurrence probability of $p_{ai}(x_{pi}(t), N(t), R(t))$ follow the definitions in (8a)–(8c). Among the available BDM parameter estimation methods in the literature [40]–[42], the QBNLS method designed by Lin and Yang [42] extends and outperforms the nonlinear least square methods presented by [40] and [41] for estimating the imitation and innovation coefficients (p, q) and market size (m).

Fig. 8 presents the schematic diagram of the GBDM parameter extraction and updating process. If we consider starting a new UGCSNet and collecting the data items of $n(t)$, $N(t)$, $r(t)$, and $R(t)$, parameter extraction and updating includes four steps:

1) The QBNLS method is applied to extract the innovation and imitation parameters (α_a, β_a) of the event occurrence probabilities of joining, uploading and downloading from the existing trajectories: $n(t)$, $N(t)$, $r(t)$, *and* $R(t)$; $t = 1, 2, 3 \ldots k1$.
2) Based on the extracted innovation and imitation parameters (α_a, β_a), we adopt a simulator of the GBDSDEs to predict the evolutions of the membership and content development from time slot *k1+1* to time slot *k2*.
3) We update the data set to enrich the database ($n(t)$, $N(t)$, $r(t)$, *and* $R(t)$, $t = 1, 2, 3 \ldots k2$.)
4) We proceed through step 1) to step 3) to periodically update the GBDM parameters and then finally complete the predictions of the membership and content volume dynamics.

B. VALUE OF CAPTURING THE POSITIVE FEEDBACK EFFECT USING GBDMS

GBDSDEs [(2a)–(7b) and (12a)] are stochastic nonlinear difference equations for which analytic solutions are difficult to obtain. As the models of individual members constitute the cornerstones of GBDSDEs and network states are simply sums of individual states, a discrete time agent-based simulation ([43]–[44]) is adopted to simulate and analyze the UGCSNet growth dynamics. An agent simulates the participation and sharing behavior of a member, and the environment of the agent simulates the evolution of the membership and content volume. In each time slot $(t, t+1]$, a member agent decides whether or not to participate and share. This probabilistic decision is based on the network states of the membership $N(t)$ and content volume $R(t)$ and member i's state $x_{pi}(t)$ using (12a). Then, the environment simulates the network state transitions of the content volume and membership using (4b) and (7b) during $(t, t+1]$.

Comparative simulations of the GBDSDEs and BDSDEs, in which the former contains a content-dependent innovation coefficient and the latter contains only a constant coefficient, are used to evaluate how the content volume affects an individual's willingness to participate and share and further the UGCSNet growth dynamics. Table 2 lists the initial simulation settings. A comparative simulation analysis shows that in the introduction stage, both models indicate low uploading and joining probabilities leading to a low content volume and low total membership. Figs. 9(a) and 9(b) show that the GBDSDEs predict faster growth for the membership and content volume during the growth stage and higher saturation values during the saturation stage than those predicted by the

TABLE 2. Simulation settings of THE captured positive feedback effect.

Parameter	Value	
Total population	200,000	
Initial membership	0	
Initial content volume	0	
Number of simulation replications	1,000	
Independent Variables of the Simulations:		
Parameter	Value	
Models	{GBDSDEs, BDSDEs}	
Coefficients of the models		
Model	Innovation	Imitation
GBDSDEs	0.03	0.38
BDSDEs	0.03	0.38

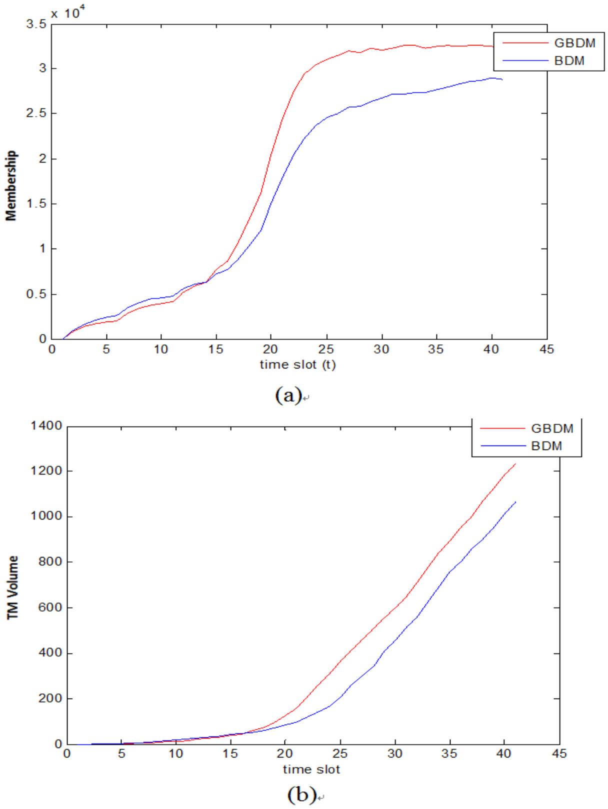


FIGURE 9. Comparison between the GBDSDEs and BDSDEs: (a) membership; (b) content volume.

BDSDEs. Such differences demonstrate that the GBDSDE has a slower start in the introduction stage, faster growth in the growth stage and higher saturation values in the saturation stage than the BDSDE model due to the positive feedback effects of the added content; these results are consistent with the qualitative descriptions of QD1 and QD2 ([11]–[18]) in Section V.

C. GBDSDE VALIDATION USING THE SCTNET AND WIKIPEDIA GROWTH DATA

To validate the proposed quantitative modeling approach using the GBDSDE and QBNLS method, the empirical data of two UGCSNets, the SCTNet, Wikipedia and an

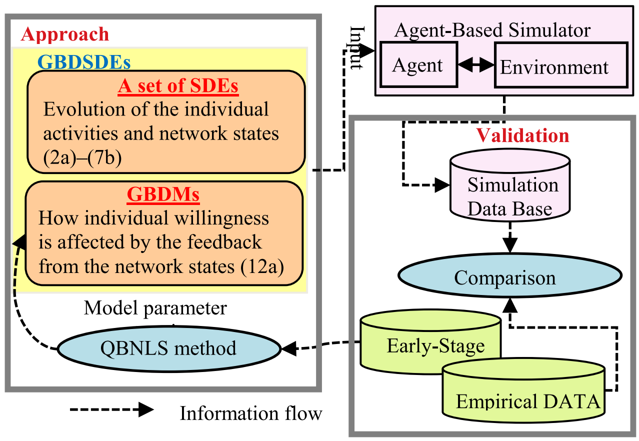


FIGURE 10. Simulation architecture for the model validation.

agent-based simulation study are utilized. Fig. 10 presents the schematic diagram of the agent-based simulation architecture, which includes three parts: the approach of the GBDSDEs and QBNLS method, an agent-based simulator, and validation by model extraction and testing. The validation procedure first applies the QBNLS method to extract parameters from the early-stage data of the SCTNet and Wikipedia.

The specific data items collected are as follows:

i) $n(t)$: number of new members (new Wikipedians in Wikipedia) during $(t-1, t]$, i.e., the t-th month;

ii) $N(t)$: total membership (Wikipedians in Wikipedia) at time t; that is, $N(t) = \sum_{k=1}^{t} n(k)$, given $N(0)=0$;

iii) $r(t)$: number of new content entries (new pages in Wikipedia) uploaded during $(t-1, t]$;

iv) $R(t)$: cumulative number of uploaded content entries at time t; that is, $R(t) = \sum_{k=1}^{t} r(k)$, given $R(0) = R_0$;

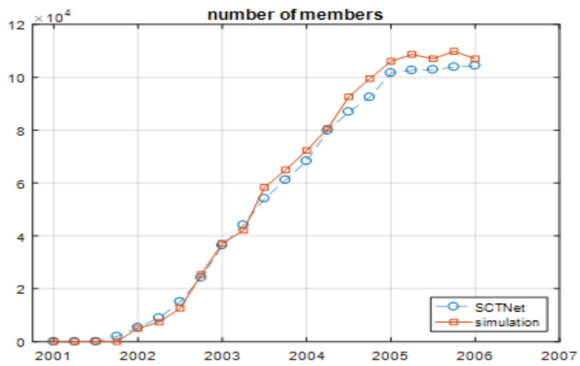
v) $D(t)$: total number of downloads (page views in Wikipedia) during $(t-1, t]$; and

vi) $L(t)$: total number of logins by members during $(t-1, t]$.

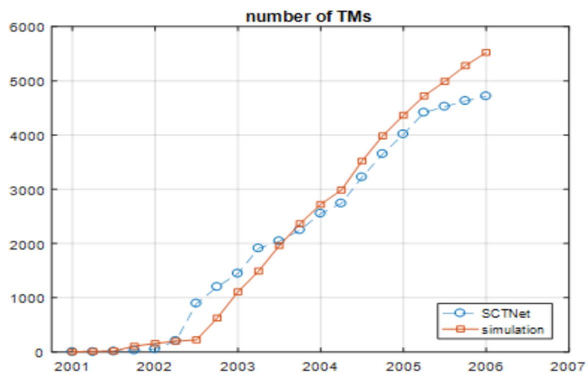
For the SCTNet and Wikipedia, there is no periodic renew requirement to maintain membership, so we assume that the leaving probability is zero. Then, the validation procedure adopts an agent-based simulator for the GBDSDEs and finally compares how well the simulation results of the GBDSDEs match with the subsequent performance of SCTNet and Wikipedia by examining the directly measured statistics.

1) GBDSDE VALIDATION WITH SCTNET: NETWORK STATES OF MEMBERSHIP AND TM VOLUME

In preparing the validation data set, in addition to the SCTNet data and statistics, the calculations of the event occurrence probabilities of $p_{ai}(x_{pi}(t), N(t), R(t))$, $a \in \{j, u, d\}$ follow (8a)–(8c), in which each data point represents three months. The GBDSDE validation separates the empirical data obtained from the SCTNet over a period of 20 data points



(a)



(b)

FIGURE 11. GBDSDE validation with SCTNet statistics: (a) membership; (b) TM volume.

(60 months) into the training and testing components. The data of the first 6 data points are for the GBDM parameter extraction, and data points 7–12 are for updating. The latter data is for comparison with the GBDSDE simulation results. In each simulation case, one hundred replications are performed for statistical analysis.

Next, we validate the membership and content volume evolution using the SCTNet data. In Figs. 11(a) and 11(b), the horizontal axis represents time, and the vertical axis represents the total membership and cumulative TM volume. The lines of the squares in Figs. 11(a) and 11(b) represent the simulation predictions.

In Fig. 11(a), the membership growth predicted by simulating the GBDSDEs is characterized by three segments with different slopes: i) a slow start with a slope of 0.62, ii) fast growth with a slope of 3.25 beginning in 2002 (19^{th} month), and iii) saturation with a slope of 0.48 and an approximate saturation value of 107,032 beginning in 2005 (52^{nd} month), where the growth time, saturation time and saturation value follow the definitions in Section I. These three slopes are calculated using a linear regression of the simulation data. Table 3 lists and compares the characteristic data of the membership growth between the simulation prediction and SCTNet data. Similar simulation results regarding the evolution of the TM volume can be observed in the SCTNet data

TABLE 3. Membership comparison between the simulation predictions and SCTNet data.

	Start	Growth		Saturation		
	Slope	Slope	Time	Slope	Time	Value
simulation	0.62	3.25	19	0.48	52	107,032
SCTNet	0.76	2.33	16	0.33	49	101,671

TABLE 4. TM volume comparison between the simulation predictions and SCTNet data.

	Start	Growth		Saturation		
	Slope	Slope	Time	Slope	Time	Value
simulation	0.22	8.63	22	x	61	5,523
SCTNet	0.27	7.33	19	0.41	52	4,415

in Fig. 11(b). The evolution of the TM volume, as predicted by the simulation using GBDSDEs, can be characterized by three segments with different slopes: i) a slow start with a slope of 0.22, ii) fast growth with a slope of 8.63 starting at month 22 (2002), and iii) saturation with an approximate saturation value of 5,523 beginning at month 61 (2006). Table 4 presents a comparison of the simulation predictions and SCTNet data. Special events, such as promotions for content contribution and membership with a prize or reward, that were organized after the training data period by the Education Bureau might have caused the occurrence of higher actual values than those simulated in the growth phase. The effects of incentives, such as special promotion events, extend beyond the scope of this paper. The coefficient of determination for the test between the simulation and empirical data [45] analyzes how well the GBDSDE models fit with the SCTNet data. The coefficient of determination, denoted R^2, provides a measure of how well the membership evolution is replicated by the GBDSDEs. An R^2 of 1 indicates that the regression line fits the data perfectly, whereas a value of 0 indicates that the line does not fit the data at all. As a whole, the GBDSDEs with parameters extracted using the QBNLS method with the SCTNet data generate membership and TM volume growth trajectory predictions with an R^2 of over 0.85 and capture the features of the slow start, fast growth, and saturation from the introduction through the saturation stages.

2) GBDSDE VALIDATION WITH WIKIPEDIA: NEW WIKIPEDIANS AND NEW CONTENT PAGE VOLUME

In preparing the validation data set, the GBDSDE validation separates the empirical data obtained from Wikipedia over a period of 28 data points (84 months) into training and testing components, in which each data point represents three months. Data for the first 6 data points (2001–2002) are for the GBDM parameter extraction process, and data for the 7–12 data (2002–2003) points are for the GBDM parameter updating process. The latter data are meant for comparison with the GBDSDE simulation results.

In Figs. 12(a)–12(b), the horizontal axis represents time (each period is 3-months), and the vertical axis represents the

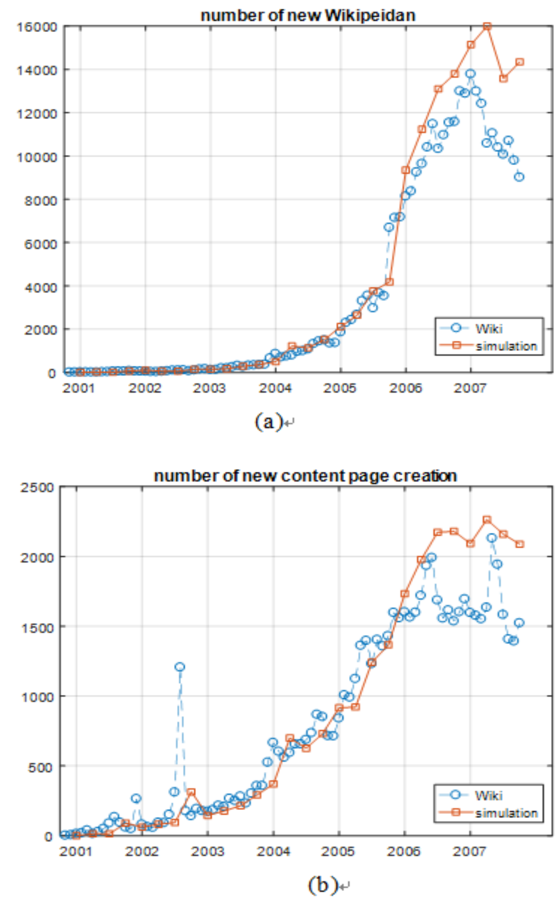


FIGURE 12. GBDSDE validation with Wikipedia statistics: (a) new Wikipedieans; (b) new content page volume.

number of new Wikipedians and the number of new content pages. The lines marked by the squares represent the simulation predictions. The line marked by the circles represents the Wikipedia statistics data. In Fig. 12(a), the evolution of new Wikipedians predicted by simulating the GBDSDEs is characterized by three segments: i) a slow start, ii) fast growth beginning in 2004 (37th month), and iii) saturation, with an approximate saturation value of 17,580, beginning in 2007 (76th month). Table 5 lists and compares the characteristic data of the new Wikipedians between the simulation prediction and Wikipedia data. Similar simulation results regarding the evolution of the new content page can be observed in Fig. 12(b). The evolution of new content pages is characterized by three segments: i) a slow start, ii) fast growth beginning in 2003 (31st month), and iii) saturation, with an approximate saturation value of 2,536, beginning in 2006 (69th month). Table 6 presents a comparison of the simulation prediction and the Wikipedia data. At the beginning of 2007, both the number of new content pages and the number of new Wikipedians began to decline [5]. In other words, the UGCSNet of Wikipedia was entering a decline phase after a very short saturation period. In fact, the GBDSDEs do not capture the decline dynamics which is a subject of future research. So the growth prediction values of 2007 shown in Figs. 12(a) and 12(b) are higher than the actual decline trajectories of Wikipedia. As a whole, the prediction results from the GBDSDEs match well with the actual Wikipedian and content page evolution trajectories, with an $R^2 > 0.8$ for predictions over the 2004–2007 period.

TABLE 5. Comparison of the simulation predictions of The Wikipedia data with regard to the new Wikipedean values.

	Start	Growth		Saturation		
	Slope	Slope	Time	Slope	Time	Value
simulation	0.57	4.35	37	-0.43	76	17,580
Wikipedia	0.41	3.03	37	-1.48	73	12,996

TABLE 6. New content page volume comparison between the simulation predictions and Wikipedia data.

	Start	Growth		Saturation		
	Slope	Slope	Time	Slope	Time	Value
simulation	0.31	6.15	31	-0.37	69	2,536
Wikipedia	0.6	5.32	31	-0.31	68	1,990

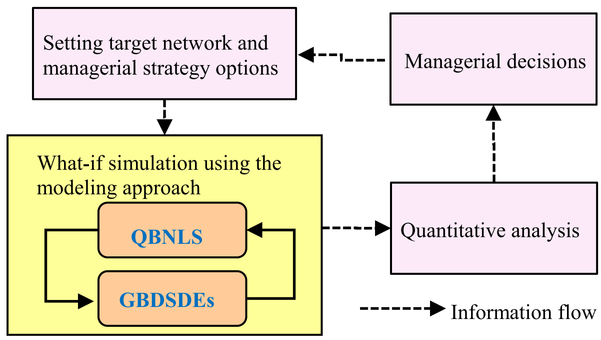


FIGURE 13. Applications of the modeling approach to the what-if analysis.

VII. POTENTIAL APPLICATIONS TO PERFORMANCE PREDICTIONS AND MANAGERIAL STRATEGY EVALUATION

Applications of the quantitative modeling approach to what-if analyses in this section are motivated by the findings of [38], [46] and [47]. Two of the many managerial strategy design issues requiring what-if analysis include: i1) the initial content volume and i2) how the altruistic founding members affect individual behaviors and network growth dynamics. Agent-based simulations of two what-if scenarios with 1,000 replications of each scenario are conducted to assess how the modeling approach can be used to facilitate management applications.

The what-if analysis using the modeling approach:

Fig. 13 presents the schematic diagram of the what-if analysis architecture, which considers starting up a new UGCSNet and planning for its growth. There are four steps:

1) The network management first plans for a target network, its target development trajectories for the membership and cumulative content volume, its target time

TABLE 7. Simulation settings for controlling the initial content volume.

Parameter	Value
Total population	200 000
Initial membership	50
Simulation replications	1 000
Time periods	36
Independent Variables of the Simulations:	
Parameter	Value
Initial content volume	{10, 21, 50, 100, 213, 457,…, 1000}

frame and managerial strategy options in the planning stage.

2) Applications of the approach to what-if analyses evaluate the effects of managerial strategies on the growth of the network. Management first applies the QBNLS method to extract the innovation and imitation parameters (α_a, β_a) of the event occurrence probabilities of joining, uploading and downloading using target trajectories. In the following study, we assume the SCTNet trajectories are target trajectories and set the assumed innovation and imitation parameters (α_a, β_a) equal to 0.01 and 0.38, respectively. Management then runs agent-based simulations of the GBDSDEs within the target time frame.
3) Network management then performs quantitative what-if analyses to predict the growth trends and evaluate how threshold values of the strategies affect growth.
4) The management finally identifies effective strategies based on the quantitative analysis.

A. ASSESSING THE EFFECTS OF INITIAL CONTENT VOLUME

Akaka *et al.* [47] noted that content resourced and sharing are salient factors required to boost network growth. Gummesson and Mele [46] noted that sharing networks commonly evolve and grow because of the added value that content sharing brings. In the GBDSDE models, the content volume provides a positive feedback effect on the event occurrence probabilities, particularly through the innovation term.

If there is a lack of initial content, the probabilities of joining and uploading will be low because of the small positive content gain, and then, the number of new members and new uploads will also remain low, and the network will not grow within the target time frame.

The hypothesis of this what-if assessment is that

H1) A UGCSNet will grow within the target time frame when the initial content volume is higher than a certain threshold.

The control variable is the amount of initial content volume. The dependent variables include the total membership and the cumulative content volume over time. One time period represents one month, and the time horizon of the simulation is 3 years (36 time periods), which is assumed as the target time frame for management. The initial settings are listed in Table 7.

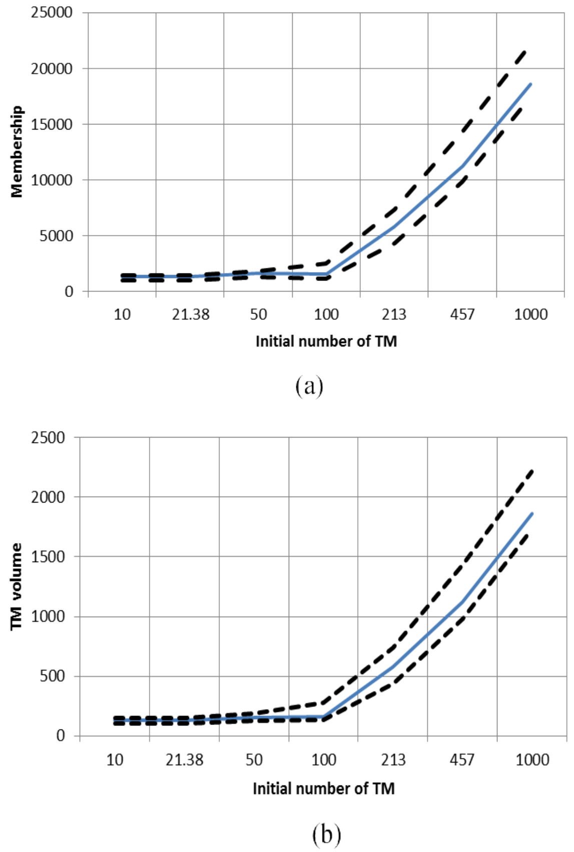


FIGURE 14. Evolution of (a) the membership and (b) content volume given various initial amounts of content.

Figs. 14(a) and 14(b) illustrate the developments of the membership and content volume after 36 time period over various values of the initial content volume. The difference between the upper and lower dotted lines indicates the 95% confidence interval. When the initial content volume is below *100*, the individual probabilities of joining and sharing stay low, and the UGCSNet does not develop in either membership or content volume within the target time frame. Conversely, when the initial content volume exceeds these *100*, both membership and content volume of the network grow higher than the initial ones after 36 time periods. This quantitative finding supports H1, and is consistent with the qualitative findings reported in the literature [47]. This what-if analysis demonstrates that the quantitative modeling approach can be used to determine the initial content volume given the target network plan.

B. ASSESSING THE EFFECTS OF INITIAL NUMBER OF ALTRUISTIC MEMBERS

Vassilakis and Vassalos [38] reported that altruistic members actively contribute to a sharing network and that sharing networks grow quickly when members are altruistic. Altruistic behavior is costly to the actor and beneficial to the recipient and the altruistic members in this simulation is

TABLE 8. Simulation settings for controlling the initial altruistic members.

Parameter	Value
Total population	50 000
Initial membership	50
Initial content volume	50
Number of simulation replications	1 000
Time periods	36
Independent Variables of the Simulations:	
Parameter	Value
Initial number of members with high altruism	{0, 10, 20, …,50}

defined as the members who actively contribute contents and promote sharing without expectation of receiving monetary returns [48]. The initial number of altruistic members is an important factor that affects the behavior of individuals and the network growth dynamics. In this assessment, there are two types of initial members: altruistic and general. In terms of GBDSDEs, altruistic members are modeled as individuals with a higher initial value of content uploading probability than general members. In the simulation, initially altruistic members are assumed to remain altruistic over the entire time horizon while late joining members are all general members.

It is intuitively clear that a higher number of altruistic founding members will lead to a higher content volume uploading and in turn a higher membership increase at the early development stage; these will then be amplified through the positive feedback effect as time evolves. The hypotheses of this what-if analysis are as follows:

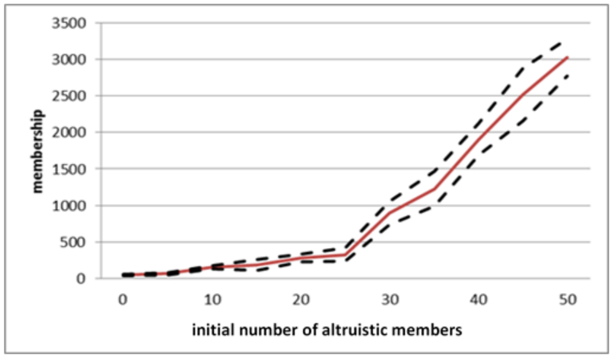
H2) A UGCSNet will grow when the initial number of altruistic members is higher than a certain threshold.

H3) An increase in the initial number of altruistic members moves the growth start time and saturation time forward and increases the saturation level.

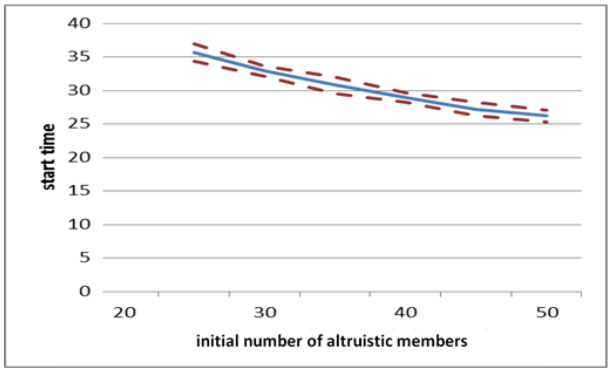
In our simulations, the number of initial altruistic members varies from 10–50 among the 50 founding members and the target time frame of management is again 36 time periods. The simulation settings are listed in Table 8. Figs. 15(a)–(c) shows what-if simulation assessment results.

In Fig. 15(a), the UGCSNet has significant membership growth after 36 months when the initial number of altruistic members is more than 25, i.e., more than 50% of the 50 founding members. Figs. 15(b) and 15(c) illustrate how the start time and saturation time monotonically decrease with the increase of the initial number of altruistic members. Again, the difference between the upper and the lower dotted lines indicates the 95% confidence interval. Our quantitative results thus support both H2 and H3 and are consistent with previous qualitative findings reported in the literature [38] that altruistic founding members has significant effects on membership development. This analysis demonstrates the application of the quantitative modeling approach in this paper to determining the initial number of altruistic members for a given target network growth plan.

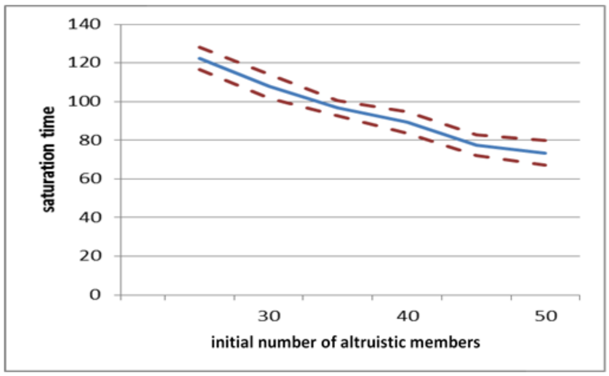
In summary, after setting a target network growth plan, the network management may apply the dynamic modeling



(a)



(b)



(c)

FIGURE 15. (a) Membership developments, (b) start time, and (c) saturation time given various initial numbers of altruistic members.

approach and perform quantitative analysis to evaluate design parameters of strategies such as initial content volume and initial altruistic members. Based on the quantitative analysis, the management may further design managerial strategies to achieve desired growth, e.g., purchasing outsourced content to increase the initial content volume or holding a competition to attract altruistic members. It can further serve as a part of cost and performance trade-off tool for strategy decision.

VIII. CONCLUSION

In this paper, we have considered the important class of content sharing and reuse networks, UGCSNets, in which members are both content contributors and users. To manage for growth of a UGCSNet, we have developed a quantitative approach to model the dynamic loop of how individual members' participation and sharing affect and are affected by the membership and content volume. Our novel modeling approach consists of i) GBDSDEs of the loop dynamics and ii) a QBNLS method to extract from limited empirical data and periodically update model parameters as the UGCSNet evolves. The skeleton of the GBDSDE is a set of individual participation and sharing model-embedded SDEs that capture the evolution of the individual member and network states with individual participation and content uploading as stochastic drives. The drive models are innovative generalization of the BDM (1969) to probabilistic models of known qualitative descriptions regarding how an individual's willingness to participate and share is affected by total membership and content volume. Our approach generalizes the constant innovation coefficient of BDM as being content-volume dependent to capture, in the growth stage, the positive feedback effect of content from a network to the individuals. The modeling approach development first used the teaching material sharing network of SCTNet as a baseline case and then validated by using the empirical data of Wikipedia, which is also a UGCSNet but has many characteristic differences with SCTNet. Analyses of the coefficients of determination have shown good fits between model predictions and actual outcomes for both SCTNet and Wikipedia growths. Applications of the modeling approach to what-if analyses of growth management strategies, for example, the initial content volume and the number of founding altruistic members, have shown the value of our modeling approach in facilitating growth management of a UGCSNet.

REFERENCES

[1] M.-F. Moens, J. Li, and T.-S. Chua, *Mining User Generated Content*. Boca Raton, FL, USA: CRC Press, 2014.

[2] S. Tirunillai and G. J. Tellis, "Does chatter really matter? Dynamics of user-generated content and stock performance," *Marketing Sci.*, vol. 31, no. 2, pp. 198–215, 2012.

[3] C.-M. Chiu, M.-H. Hsu, and E. T. G. Wang, "Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories," *Decis. Support Syst.*, vol. 42, no. 3, pp. 1872–1888, 2006.

[4] J. Giles, "Internet encyclopaedias go head to head," *Nature*, vol. 438, no. 7070, pp. 900–901, 2005.

[5] B. Suh, G. Convertino, E. H. Chi, and P. Pirolli, "The singularity is not near: Slowing growth of Wikipedia," in *Proc. 5th WikiSym. ACM*, 2009, p. 108.

[6] *Flickr Photo Sharing Home Page*. Accessed: Feb. 1, 2017. [Online]. Available: https://www.flickr.com

[7] *YouTube Home Page*. Accessed: Feb. 1, 2017. [Online]. Available: https://www.youtube.com

[8] *Wikipedia Home Page*. Accessed: Feb. 1, 2017. [Online]. Available: https://www.wikipedia.org

[9] *Facebook Home Page*. Accessed: Feb. 1, 2017. [Online]. Available: https://www.facebook.com

[10] *Smart Creative Teachers Net Home Page*. Accessed: Feb. 1, 2017. [Online]. Available: http://ds.kh.edu.tw/sites/sctnet.htm

[11] P. Adrian and K.-L. Nicole, "An experiential, social network-based approach to direct marketing," *Direct Marketing, Int. J.*, vol. 3, no. 3, pp. 162–176, 2009.

[12] B. S. Butler, "Membership size, communication activity, and sustainability: A resource-based model of online social structures," *Inf. Syst. Res.*, vol. 12, no. 4, pp. 346–362, 2001.

[13] B. Belvaux, "The development of social media: Proposal for a diffusion model incorporating network externalities in a competitive environment," *Recherche Appl. Marketing*, vol. 26, no. 3, pp. 7–22, 2011.

[14] A. Asvanund, K. Clay, R. Krishnan, and M. D. Smith, "An empirical analysis of network externalities in P2P music-sharing networks," *Inf. Syst. Res.*, vol. 15, no. 2, pp. 155–174, 2004.

[15] J. Powell, *33 Million People in the Room: How to Create, Influence, and Run a Successful Business With Social Networking*. Indianapolis, IL, USA: Que Publishing, 2009.

[16] D. Tapscott, *Grown Up Digital: How the Net Generation is Changing Your World*. New York, NY, USA: McGraw-Hill, 2008.

[17] K.-Y. Lin and H.-P. Lu, "Why people use social networking sites: An empirical study integrating network externalities and motivation theory," *Comput. Hum. Behav.*, vol. 27, no. 3, pp. 1152–1161, 2011.

[18] C.-P. Lin and A. Bhattacherjee, "Elucidating individual intention to use interactive information technologies: The role of network externalities," *Int. J. Electron. Commun.*, vol. 13, no. 1, pp. 85–108, 2008.

[19] D. Sledgianowski and S. Kulviwat, "Using social network sites: The effects of playfulness, critical mass and trust in a hedonic context," *J. Comput. Inf. Syst.*, vol. 49, no. 4, pp. 74–83, 2009.

[20] M. L. Katz and C. Shapiro, "Network externalities, competition, and compatibility," *Amer. Econ. Rev.*, vol. 75, no. 3, pp. 424–440, 1985.

[21] S. J. Liebowitz and S. E. Margolis, "Are network externalities a new source of market failure?" *Res. Law Econ.*, vol. 17, pp. 1–22, 1995.

[22] N. Economides, "Network externalities, complementarities, and invitations to enter," *Eur. J. Political Econ.*, vol. 12, no. 2, pp. 211–233, 1996.

[23] K. R. Conner, "Obtaining strategic advantage from being imitated: When can encouraging 'clones' pay?" *Manage. Sci.*, vol. 41, no. 2, pp. 209–225, 1995.

[24] R. J. Kauffman, J. McAndrews, and Y.-M. Wang, "Opening the 'black box' of network externalities in network adoption," *Inf. Syst. Res.*, vol. 11, no. 1, pp. 61–82, 2000.

[25] D. Dranove and N. Gandal, "The Dvd-vs.-Divx standard war: Empirical evidence of network effects and preannouncement effects," *J. Econ. Manage. Strategy*, vol. 12, no. 3, pp. 363–386, 2003.

[26] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "I Tube, You Tube, everybody tubes: Analyzing the world's largest user generated content video system," in *Proc. 7th ACM SIGCOMM Conf. Internet Meas.*, 2007, pp. 1–14.

[27] R. Kumar, J. Novak, and A. Tomkins, "Structure and evolution of online social networks," in *Link Mining: Models, Algorithms, and Applications*. New York, NY, USA: Springer, 2010, pp. 337–357.

[28] V. Mahajan, E. Müller, and F. M. Bass, "New product diffusion models in marketing: A review and directions for research," in *Diffusion of Technologies and Social Behavior*. Berlin, Germany: Springer, 1991, pp. 125–177.

[29] F. M. Bass, "A new product growth for model consumer durables," *Manage. Sci.*, vol. 15, no. 5, pp. 215–227, 1969.

[30] M. Tanaka, Y. Sakumoto, M. Aida, and K. Kawashima, "Study on the growth and decline of SNSS by using the infectious recovery SIR model," in *Proc. IEEE APSITT*, Aug. 2015, pp. 1–3.

[31] X. Zhu, Y. Nie, and A. Li, "Demographic prediction of online social network based on epidemic model," in *Web Technologies and Applications*. Cham, Switzerland: Springer, 2014, pp. 93–103.

[32] F. Jin, E. Dougherty, P. Saraf, Y. Cao, and N. Ramakrishnan, "Epidemiological modeling of news and rumors on twitter," in *Proc. 7th SNAKDD ACM*, 2013, p. 8.

[33] J. Cannarella and J. A. Spechler. (Jan. 2014). "Epidemiological modeling of online social network dynamics." [Online]. Available: https://arxiv.org/abs/1401.4208

[34] C. Bauckhage and K. Kersting. (Jun. 2014). "Strong regularities in growth and decline of popularity of social media services." [Online]. Available: https://arxiv.org/abs/1406.6529

[35] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, "Information diffusion in online social networks: A survey," *ACM SIGMOD Rec.*, vol. 42, no. 2, pp. 17–28, 2013.

[36] R.-H. Chen and S.-C. Chang, "Generalized bass diffusion model (GBDM) for co-creation networks—A teaching material sharing network application," in *Proc. IEEE SMC*, Oct. 2013, pp. 1531–1536.

[37] R.-H. Chen, S.-C. Chang, Y.-R. Chiou, C.-C. Lai, and L.-W. Yeh, "Empirical data-based modeling of teaching material sharing network dynamics," in *Proc. IEEE SMC*, Oct. 2011, pp. 152–158.

[38] D. K. Vassilakis and V. Vassalos, "Modelling real P2P networks: The effect of altruism," in *Proc. IEEE P2P*, Sep. 2007, pp. 19–26.

[39] C. M. Christensen, *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Boston, MA, USA: Harvard Business Review Press, 2013.

[40] V. Srinivasan and C. H. Mason, "Nonlinear least squares estimation of new product diffusion models," *Marketing Sci.*, vol. 5, no. 2, pp. 169–178, 1986.

[41] R. M. Heeler and T. P. Hustad, "Problems in predicting new product growth for consumer durable," *Manage. Sci.*, vol. 26, no. 10, pp. 1007–1020, 1980.

[42] F.-C. Lin and Y.-N. Yang, "The impact of sample size on diffusion model—The application of bootstrap method," in *Proc. WEAI 8th Biennial Pacific Rim Conf.*, 2009.

[43] H. Hattori, Y. Nakajima, and T. Ishida, "Learning from humans: Agent modeling with individual human behaviors," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 1, pp. 1–9, Jan. 2011.

[44] M. J. Lanham, G. P. Morgan, and K. M. Carley, "Social network modeling and agent-based simulation in support of crisis de-escalation," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 44, no. 1, pp. 103–110, Jan. 2014.

[45] A. Ramakrishnan, "Probability and stochastic processes," in *Prinzipien der Thermodynamik und Statistik/Principles of Thermodynamics and Statistics*. Heidelberg, Germany: Springer, 1959, pp. 524–651.

[46] E. Gummesson and C. Mele, "Marketing as value co-creation through network interaction and resource integration," *J. Bus. Market Manage.*, vol. 4, no. 4, pp. 181–198, Dec. 2010.

[47] M. A. Akaka, S. L. Vargo, and R. F. Lusch, "An exploration of networks in value cocreation: A service-ecosystems view," *Rev. Marketing Res.*, vol. 9, pp. 13–50, Jun. 2012.

[48] S. A. West, A. S. Griffin, and A. Gardner, "Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection," *J. Evol. Biol.*, vol. 20, no. 2, pp. 415–432, 2007.

RONG-HUEI CHEN (S'11) received the B.S. degree in mechanical and electro-mechanical engineering from Tamkang University, Taiwan, in 2007, and the M.S. degree in electrical engineering from National Taiwan University, Taiwan, in 2009. He is currently pursuing the Ph.D. degree with the Graduate Institute of Electrical Engineering, National Taiwan University. His research interests include data analysis, mechanism design, modeling methodology development, network dynamics modeling, and agent-based simulation.

SHI-CHUNG CHANG (S'82–M'83) received the B.S.E.E. degree from National Taiwan University, Taiwan, China, in 1979, and the M.S. and Ph.D. degrees in electrical and systems engineering from the University of Connecticut, USA, in 1983 and 1986, respectively. He is currently a Professor jointly appointed by the Electrical Engineering Department, Institute of Industrial Engineering, and the Graduate Institute of Communication Engineering, National Taiwan University, and a Research Fellow of Smart Network System Institute, Institute for Information Industry, Taiwan. He was a commissioner of the National Communications Commission, Taiwan from 2010 to 2012, and led the execution of digital terrestrial TV switchover. His research interests include optimization and control with applications to communication network, production and power systems, distributed decision making, network management and economics, and shared wireless accesses.