

Fates of Microscopic Social Ecosystems: Keep Alive or Dead?

Haoyang Li*
Tsinghua University
lihy18@mails.tsinghua.edu.cn

Peng Cui†
Tsinghua University
cuip@tsinghua.edu.cn

Chengxi Zang†
Tsinghua University
zangcx13@mails.tsinghua.edu.cn

Tianyang Zhang
Tsinghua University
zhangty09@foxmail.com

Wenwu Zhu†
Tsinghua University
wwzhu@tsinghua.edu.cn

Yishi Lin
Tencent
yishilin14@gmail.com

ABSTRACT

A social network is an ecosystem, and one of its ultimate goals is to maintain itself sustainable, namely keeping users generating information and being informed. However, the reasons why some social ecosystems can keep self-sustaining and others end up with non-active or dead states are largely unknown.

In this paper, rather than studying social ecosystems at the population level, we analyze the fates of different microscopic social ecosystems, namely the final states of their collective activity dynamics in a real-world online social media with detailed individual level records for the first time. We find huge complexities in microscopic social ecosystems, including complex species types, complex individual interaction networks, and complex dynamics and final states. In order to capture the observed complexities in the real-world data, we propose a microscopic ecological model, which is able to capture the complex fates of heterogeneous microscopic social ecosystems accurately in both synthetic and empirical datasets. Furthermore, we analyze the driven factors of the fates of microscopic social ecosystems, including interaction networks of individuals and dynamical interaction mechanisms of species, leading to the control of microscopic social ecosystems, that is the ability to influence the temporal behaviours and their final states towards active or dead fates.

KEYWORDS

Social Ecosystem, Social Activity, Dynamic Model

ACM Reference format:

Haoyang Li, Peng Cui, Chengxi Zang, Tianyang Zhang, Wenwu Zhu, and Yishi Lin. 2019. Fates of Microscopic Social Ecosystems: Keep Alive or Dead?. In *Proceedings of The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Anchorage, AK, USA, August 4–8, 2019 (KDD '19)*, 9 pages. <https://doi.org/10.1145/3292500.3330827>

*Beijing National Research Center for Information Science and Technology (BNRist)

†Corresponding authors

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '19, August 4–8, 2019, Anchorage, AK, USA

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6201-6/19/08...\$15.00

<https://doi.org/10.1145/3292500.3330827>

1 INTRODUCTION

An ecosystem in nature is a community of plants and animals which live interdependently, ranging from the earth, a tropical rain forest, or a fox-rabbit ecosystem. What is a healthy ecosystem and how to keep itself sustainable are major issues in ecology studies [8, 9, 16, 20, 29].

Similarly, social media is a new kind of ecosystem with complex social features. However, many social ecosystems are struggling with keeping users active, such as the death of Myspace in the U.S. and Renren in China. Other social media like Facebook, Twitter, and StackOverflow are somehow losing users' activities [30]. In order to prevent social media from losing users' activities or eventually die-out, it is important to understand the mechanisms of how social ecosystems can keep active at the microscopic level, that is information producers are keeping generating user-generated content (UGC) while information consumers are kept informed and can give feedback or rewards to the producers.

Previous social ecosystems studies mainly focus on the activity of social ecosystems at the population level, including competition dynamics of two viruses/ideas [5] or membership-based websites [28]. Epidemic models based on SI or SIR model [1, 22] are proposed to capture the dynamics of different populations. These works assume that individuals in a social ecosystem are homogeneously mixing, which largely ignore the interaction networks of individuals and the heterogeneity of individuals' influence [12, 30]. In order to solve aforementioned problems, network-based ecological models are investigated to capture the activity of collaboration networks [30], or the resilience dynamics of ecosystems [12], etc. However, due to the lack of detailed datasets, the dynamics of individual behaviour and the fates of microscopic social ecosystems, namely the final states (e.g. alive, active or dead) of their collective activity dynamics are unexplored.

In this paper, we analyze a real-world dataset consisting of 5,500 microscopic social ecosystems with 675,027 users from WeChat, i.e., one of the world's largest social networks. Indeed, we find rich unexplored complexities in real-world microscopic social ecosystems, including: (1) Uncertain species¹ types: an individual in a social ecosystem may belong to multiple species types simultaneously and change its species type over time; (2) Complex interaction networks: linked individuals with different species types form different microscopic social ecosystems, and interaction relationships between

¹Species is the basic concept for classification and taxonomic rank in biology. For analogy and convenience, we borrow it into this paper to represent the "virtual species" [27] (hereinafter referred to as "species") living on the social ecosystems for classifying users into different groups/memberships/categories and quantifying them.

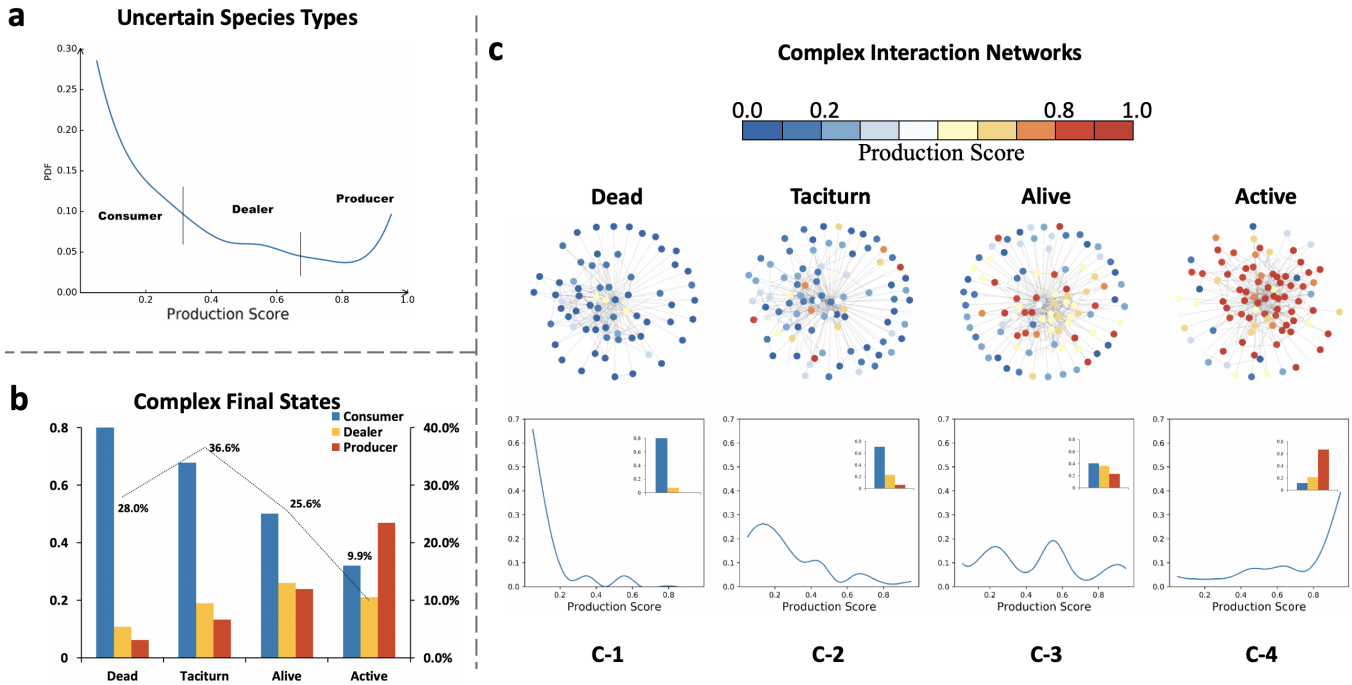


Figure 1: Complexities in Microscopic Social Ecosystems. a. Uncertain species types captured by the distribution of production score of WeChat Users. We find a bimodal distribution for the score, indicating that most users are either information producers or information consumers, and others are information dealers. b. Complex final states of microscopic social ecosystems, including dead ones (28.0%), taciturn ones (36.6%), alive ones (25.6%) and active ones (9.9%), where the blue, yellow and red bars represent the proportion of consumers, dealers and producers respectively. c. Complex interaction networks of microscopic social ecosystems. c-1 to c-4 show four typical microscopic social ecosystems in b, featuring different interaction networks and different production score distributions.

species are also heterogeneous; (3) Complex final states: final states of different microscopic social ecosystems exhibit quite different patterns, ranging from dead ones to active ones and beyond.

In order to capture the complexities in real-world microscopic social ecosystems, we propose the **SocialFate** model, an ecological model at the microscopic level. Specifically, we first define the species types of individuals in microscopic social ecosystems who are playing uncertain roles. Next, we propose our model at the microscopic individual level by incorporating uncertain species types, network structures, and species interaction mechanisms, which can accurately capture the observed complex final states of different microscopic social ecosystems. Our **SocialFate** model encompasses several traditional ecological models as special cases. Furthermore, we illustrate how aforementioned driven factors captured by our **SocialFate** model influence the activity dynamics of different microscopic social ecosystems, implying the applications of controlling the microscopic social ecosystems towards active or healthy states instead of dead ones.

In short, we summarize our contributions as follows:

- **Novel Problems and Findings:** We study the fates of social media in ecological perspectives at the microscopic level. By data-driven approaches, we uncover new and complex patterns of microscopic social ecosystems in the real world.

- **Novel Model:** We model the fates of microscopic social ecosystems with complex dynamics. Our model encompasses several population ecological models as special cases.
- **Control Applications:** We reveal the determining factors of the fates of microscopic social ecosystems, leading to the control of the fates towards active or dead.

Table 1: The Analogy between Rainforest Ecosystems and Social Ecosystems

Rainforest Ecosystems	Social Ecosystems
Species (e.g., trees/insects/birds, grains/rats/raccoons)	Species (e.g., bloggers/forwarders/followers)
Energy flow	Information flow
Energy production (e.g., trees photosynthesis)	Information production (e.g., upload videos)
Energy transmission (e.g., insects feed on leaves)	Information transmission (e.g., watch videos and forward)
Energy consumption (e.g., birds feed on insects)	Information consumption (e.g., watch videos)

2 RELATED WORK

We review related works in ecology and social ecosystem studies:

Ecological Models. Ecology studies focus on species populations and how these populations interact with each other and the environment. Ecologists first studied the growth law of single species. Malthus proposed the exponential law [26] of population growth. McKendrick extended the exponential growth to the logistic growth equation [18]. Then, ecologists began to study the population ecology of different species and their interactions. One of the most famous models is the Lotka-Volterra (LV) model [13], which describes the population dynamics of prey and predator. Interactions between more species [15, 19, 24, 25] are further modeled based on the Lotka-Volterra model. However, these models are so-called homogeneous models at the population level. Each individual has same probability in contacting others. These works ignore the interaction networks of individuals and the heterogeneity of individuals' influence.

Social Ecosystems. Social media [2, 17, 21, 31–34] is another kind of ecosystem with complex social features. We compare social ecosystems and rainforest ecosystems in Table 1. For example, species in social ecosystems can be defined functionally or behaviorally (e.g., bloggers/forwarders/followers), while species in rainforest ecosystems can be defined taxonomically or phylogenetically (e.g., trees/insects/birds). The same thing is that they both try to divide individuals into species according to their roles. In social ecosystem studies, [5] studied competing dynamics of two species and found factors impacting final states. [28] studied competing dynamics of membership-based websites to predict their growth and death trends. [27] considered online activities behaving like species in an ecological system and studied large-scale co-evolving online activities. However, all these works modeled dynamics of species at the population level. The concepts of "species" and "information flow" in social ecosystems have been proposed but how to quantify species of any user and figure out the interaction are not explored. Due to the lack of detailed datasets, the dynamics and fates (alive, active or dead) of microscopic social ecosystems are largely unknown. Driven by a real-world dataset, we illustrate the complexities discovered in microscopic social ecosystems in next section, which none of existing methods focus on specifically.

3 COMPLEXITIES IN MICROSCOPIC SOCIAL ECOSYSTEMS

We find rich complexities in microscopic social ecosystems, including: (1) Uncertain species types (Figure 1-a); (2) Complex interaction networks (Figure 1-c); (3) Complex final states (Figure 1-b).

Uncertain Species Types: An individual may belong to multiple species types simultaneously or change its species type over time in a social ecosystem. In a rainforest ecosystem, the species types are deterministic and stable. However, in a social ecosystem, an information producer can be a information consumer simultaneously and vice versa. Besides, a pure information producer may become a pure information consumer later.

We tag the species types of users in a social ecosystem according to their information niche similar with the energy niche in a rainforest ecosystem. For example, users in WeChat can produce

information by publishing a photo post and their friends can consume it through like or comment function. We classify the species of users in a social ecosystem quantitatively according to what extent a user n is a pure producer by the production score which is defined as:

$$production\ score(n) = \frac{\#production(n)}{\#production(n) + \#consumption(n)}, \quad (1)$$

where the number of production $\#production(n)$ is the number of photos the user n posted in a predefined time period and the number of consumption $\#consumption(n)$ is the number of comments or likes he/she made in the same time period. In our experiments, we set the time period as 1 month. We plot the distribution of $production\ score(n)$ of all the users as shown in Figure 1-a. We find 3 different clusters, i.e., different species types:

$$species\ types := \begin{cases} producer & , production\ score \in [2/3, 1] \\ dealer & , production\ score \in [1/3, 2/3). \\ consumer & , production\ score \in [0, 1/3) \end{cases} \quad (2)$$

- **Producers** are information sources in social ecosystems who tend to generate information frequently. Typical producers include media accounts, celebrities, or salesmen who post lots of advertisements.
- **Consumers** are users who seldom post or produce information but prefer to like or comment others' posts.
- **Dealer** are users who strike a balance between producing and consuming. Such users may convert to producers or consumers in the future.

Then we calculate the species type of users by their production scores with Equation 2 for all the users in our dataset. As shown in Figure 1-a, we plot the distribution of the continuous random variable $production\ score$ of each individual being an information producer in social ecosystems, indicating uncertain species types in the real world scenarios and difficulty of modeling uncertain species types in social ecosystems.

Complex Interaction Networks: Linked individuals with different species types form largely different microscopic social ecosystems, and interaction relationships between species are also heterogeneous. Lacking of interaction datasets of individuals, previous ecological studies can only model population dynamics [3, 11, 14, 23], which assume individuals are homogeneously mixing, namely random graphs or complete graphs. In reality, individuals are linked with their different neighbors rather than all the other individuals in a network. As shown in Figure 1-c, we illustrate four types of different microscopic social ecosystems with different network structures, indicating interaction relationships between species are also heterogeneous. More specifically, how information producers influence information consumers is different in one microscopic social ecosystem from others.

Complex Final States: Different microscopic social ecosystems exhibit quite different patterns of final states of their collective activity dynamics, namely fates. By discretizing the production score of each individual being an information producer into three types, namely producer, dealer, and consumer in each microscopic social ecosystem, we cluster the microscopic social ecosystems according to their abundance distributions of these three types. The fate of a microscopic social ecosystem is the collective final states

of each individuals in this system. By a data-driven approach, We find four typical microscopic social ecosystem patterns as shown in Figure 1-b:

- **Dead** microscopic social ecosystems consist of many users who seldom produce posts and we name these ecosystems "dead", accounting for 28.0% of all the microscopic social ecosystems.
- **Taciturn** microscopic social ecosystems have several users producing information and most of users prefer listening, serving as the the largest cluster with a 36.6% proportion of all the microscopic social ecosystems.
- **Alive** microscopic social ecosystems have relatively close number of producers and consumers, which account for 25.6% of all the microscopic social ecosystems.
- **Active** microscopic social ecosystems have a large number of users who are information producers, accounting for 9.9% of all the microscopic social ecosystems.

4 PROPOSED METHOD

In this section, based on the aforementioned complexities in the real world datasets, we propose our **SocialFate** model to capture the fates of microscopic social ecosystems.

4.1 Background - Ecological Models

Lotka-Volterra model: One of the most famous models in population ecology is the **Lotka-Volterra model** (LV) [13]. It is used to describe the dynamics of two species interacting in an ecosystem, one as predator and the other as prey. The equations are:

$$\begin{cases} \frac{d\lambda_1}{dt} = \alpha\lambda_1 - \beta\lambda_1\lambda_2 \\ \frac{d\lambda_2}{dt} = \delta\lambda_1\lambda_2 - \gamma\lambda_2 \end{cases}, \quad (3)$$

where λ_1 is the number of prey, λ_2 is the number of predator. In the equations, α represents the own growth rate of prey and γ represents the dead rate of predator. β and δ represent the interaction strength between prey and predator.

Competitive Lotka-Volterra model: Competition is another common interaction mechanism between species. The population dynamics of species competing for some common resources can be described by the **competitive Lotka-Volterra model** [4]. When there are K species competing against each other for some resources, the population of species i follows the equation:

$$\frac{d\lambda_i}{dt} = r_i\lambda_i(1 - \sum_{j=1}^K \alpha_{i,j}\lambda_j), \quad (4)$$

where λ_i is the population size of species i at a given time, r_i is inherent per-capita growth rate of species i , and $\alpha_{i,j} > 0$ represents the intensity of competition between species i and j .

Generalized Lotka-Volterra model: Besides predator-prey and competition, there are some other relationships [6, 7] between different species, such as mutualism, commensalism, amensalism, etc. So ecologists use interaction matrix A to capture all these relationships between species, in which $A_{i,j}$ describes the rate of species j impacting on species i . So the **Generalized Lotka-Volterra model**

(GLV) [18] can be written as:

$$\frac{d\lambda_i}{dt} = \lambda_i(r_i + \sum_{j=1}^K A_{i,j}\lambda_j). \quad (5)$$

Different from the competitive Lotka-Volterra model, the entry values of interaction matrix A can be either positive or negative, which can represent more kinds of interaction relationships. The solutions of the equations are λ_i^* , determined by the initial population $\lambda_i(0)$, parameters such as growth rate r_i , and interaction matrix A , solved by equations $\frac{d\lambda_i}{dt} = 0$. The solutions represent the final state of all populations and the fate of this ecosystem.

4.2 Our SocialFate Model

Table 2: Symbols and Definitions

Symbols	Definitions
N	Total number of users in a social ecosystem
K	Total number of species in a social ecosystem
X	User-Species matrix ($N \times K$)
$X_{n,k}$	Probability that user n belongs to the species k
$X_{n,:}$	Species probability distribution of the user n
M	Network structure ($N \times N$)
Φ	Local species abundance matrix
$\Phi_{n,k}$	Abundance of species k near user n
A	Species interaction matrix ($K \times K$)

The existing ecological models cannot capture the complexities observed in the real-world microscopic social ecosystems and thus fail in capturing the dynamics and the fates of them. We follow the ecological framework by modeling two parts, namely the intrinsic growth term $f(X(t); \mathbf{r})$ and the network effect term $g(X(t); M, A)$, as follows:

$$\frac{dX(t)}{dt} = f(X(t); \mathbf{r}) + g(X(t); M, A). \quad (6)$$

Specifically, we describe the uncertain species types by a User-Species matrix $X(t) \in \mathbf{R}^{N \times K}$ at time t , and each entry $X_{n,k}(t)$ in $X(t)$ represents the probability that user n belongs to the species k . We use K to represent the number of species and N for the number of users in a microscopic social ecosystem. We use $\mathbf{r} \in \mathbf{R}^{K \times 1}$ to capture the heterogeneous intrinsic growth rates of species where r_k is the growth rate of species k . In order to capture the complex individual interaction networks, we use adjacency matrix $M \in \mathbf{R}^{N \times N}$ to represent the network structure of a microscopic social ecosystem, where $M_{n,m} = 1$ indicates a bidirectional relationship between user n and user m , and $M_{n,m} = 0$ indicates that they are not linked. Furthermore, we use $A \in \mathbf{R}^{K \times K}$ to capture complex species interaction network where $A_{i,j}$ is the influence rate of species j on species i .

Thus, we propose our **SocialFate** model:

$$\frac{dX(t)}{dt} = X(t)\tilde{R} + X(t) * (\tilde{M}X(t)A^T). \quad (7)$$

\tilde{M} represents "average effect" of neighbors:

$$\tilde{M}_{n,m} = \frac{M_{n,m}}{\sum_{p=1}^N M_{n,p}}, \quad (8)$$

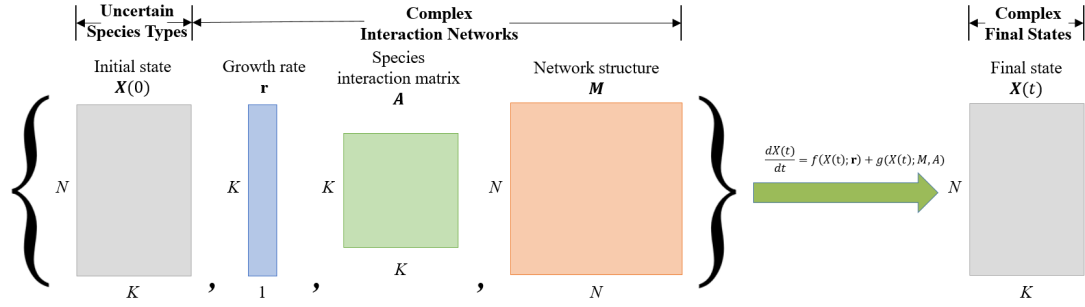


Figure 2: The framework of our SocialFate model. Our model captures the observed complexities in microscopic social ecosystems, including uncertain species types, complex interaction networks, and complex final states explicitly.

and \tilde{R} is :

$$\tilde{R} = \text{diag}(r_1, r_2, \dots, r_K). \quad (9)$$

In order to further clarify the model, we show how to update the species states of a particular user n as follows:

$$\frac{dX_{n,i}(t)}{dt} = X_{n,i}(t)(r_i + \sum_{j=1}^K A_{i,j}\Phi_{n,j}), \quad (10)$$

$$\Phi = \tilde{M}X, \quad (11)$$

$$\begin{aligned} \Phi_{n,:} &= \frac{\sum_{m=1}^N M_{n,m}X_{m,:}}{\sum_{p=1}^N M_{n,p}} \\ &= \sum_{m=1}^N \frac{M_{n,m}}{\sum_{p=1}^N M_{n,p}} X_{m,:}. \end{aligned} \quad (12)$$

$\Phi_{n,k}$ is local abundance of species k near user n , for the assumption that users are influenced by their neighbors rather than all users in the same social ecosystem. Because **SocialFate** is a probability model, we normalize each user's species probability distribution $X_{n,:}$ to satisfy $\sum_{k=1}^K X_{n,k} = 1$.

Figure 2 illustrates the framework of our **SocialFate** model. In this model, the solution of the Equation 10, derived from $dX/dt = 0$, represents the outcome that our expected stable state X^* , which is decided by not only the initial value $X(0)$ and species interaction matrix A but also the structure M of the microscopic social ecosystems. The Generalized Lotka-Volterra (GLV) model can be regarded as a degenerated case of our model when the network structure is fully connected. In this situation, the local species abundance equals to the overall abundance for each individual, and we can use the GLV model instead. Our **SocialFate** model is a general one which encompasses all the ecological models in Section 4.1 as special cases.

LEMMA 4.1. *Our **SocialFate** model degenerates to the Generalized Lotka-Volterra model in Equation 5 when individuals are fully connected, i.e., M is a matrix of ones.*

PROOF. We denote operator \mathbf{L} as:

$$\mathbf{L}(\mathbf{X}) = \frac{\sum_{n=1}^N X_{n,:}}{N}. \quad (13)$$

If individuals are fully connected in a microscopic social ecosystem, individual interaction network M is a matrix of ones, and Equation

12 leads to:

$$\Phi_{n,:} = \frac{\sum_{m=1}^N M_{n,m}X_{m,:}}{\sum_{p=1}^N M_{n,p}} = \frac{\sum_{m=1}^N X_{m,:}}{N} = \mathbf{L}(\mathbf{X}). \quad (14)$$

By replacing $\Phi_{n,j}$ in Equation 10 with $\mathbf{L}(\mathbf{X})_j$, we get:

$$\frac{dX_{n,i}(t)}{dt} = X_{n,i}(t)(r_i + \sum_{j=1}^K A_{i,j}\mathbf{L}(\mathbf{X})_j). \quad (15)$$

Because users in fully connected network are homogeneous, we can get population dynamics by summing the states of each user:

$$\frac{1}{N} \sum_{n=1}^N \frac{dX_{n,i}(t)}{dt} = \frac{1}{N} \sum_{n=1}^N X_{n,i}(t)(r_i + \sum_{j=1}^K A_{i,j}\mathbf{L}(\mathbf{X})_j), \quad (16)$$

$$\frac{d \frac{1}{N} \sum_{n=1}^N X_{n,i}(t)}{dt} = \frac{1}{N} \sum_{n=1}^N X_{n,i}(t)r_i + \sum_{j=1}^K A_{i,j} \frac{1}{N} \sum_{n=1}^N X_{n,i}(t)\mathbf{L}(\mathbf{X})_j, \quad (17)$$

$$\frac{d\mathbf{L}(\mathbf{X})_i}{dt} = \mathbf{L}(\mathbf{X})_i r_i + \sum_{j=1}^K A_{i,j}\mathbf{L}(\mathbf{X})_i \mathbf{L}(\mathbf{X})_j. \quad (18)$$

Define $\lambda_i = \mathbf{L}(\mathbf{X})_i$, and we get Generalized Lotka-Volterra model:

$$\frac{d\lambda_i}{dt} = \lambda_i r_i + \sum_{j=1}^K A_{i,j}\lambda_i \lambda_j \quad (19)$$

□

4.3 Parameter Learning

Because we model the species distribution in **SocialFate** in a probabilistic way, we can use the maximum likelihood estimation (MLE) method to learn the modeling parameters. Let $\hat{X}_{n,k}$ represents the estimated possibility that user n belongs to the species k calculated by our model, the likelihood function of a social ecosystem with N users of K species is:

$$\log \text{likelihood} = \sum_{n=1}^N \log \left[\sum_{k=1}^K \hat{X}_{n,k} \cdot \mathbf{1}_n(k) \right], \quad (20)$$

where $\mathbf{1}_n(k)$ is an indicator function, by definition,

$$\mathbf{1}_n(k) := \begin{cases} 1 & \text{,user } n \in \text{species } k \\ 0 & \text{,user } n \notin \text{species } k \end{cases}$$

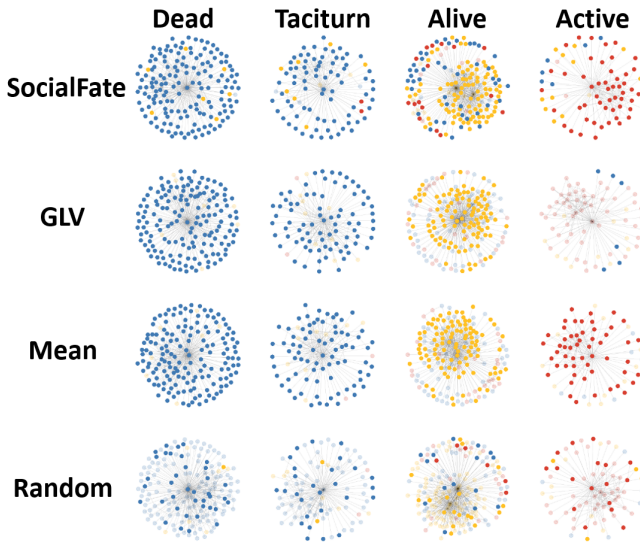


Figure 3: Illustrations of fitting results for the four typical microscopic social ecosystems by our SocialFate model and baselines. The colored nodes represent the users which are accurately captured while the shaded nodes are those with large errors. The red color, blue color and yellow color denote producers, consumers and dealers respectively.

Therefore, we can estimate the parameters by maximize the log-likelihood function with commonly-used optimization methods such as BFGS algorithm [10].

5 EXPERIMENTS

Here we introduce the dataset and the experimental settings to validate our proposed model.

5.1 Experiment Setup

Dataset. We validate our model by a real-world social network dataset - **WeChat**, which includes users' records at the microscopic level. WeChat² is the largest online social network in China with more than 600 million monthly active users. We obtain the dataset directly from WeChat via an agreement. We randomly sample 5,500 WeChat users and their ego-networks to represent different microscopic social ecosystems and their behaviour logs, including the post records, comment records and like records (a function that users can give some posts the thumbs up)³. We summarize the dataset in Table 3.

The states of microscopic social ecosystems are relatively steady in a range of time just like natural ecosystems do not change obviously without special factors. Figure 4 shows us most of the microscopic social ecosystems are steady, namely their states do not change in a month. So we can use this state as its final state.

Baselines. Given a microscopic social ecosystem consisting of user's activity like information production and consumption, we

²<https://weixin.qq.com/>

³For privacy issues, the dataset is fully anonymized and all data are collected according to the terms and conditions of WeChat.

Table 3: Dataset Description

Name	Value
Ecosystem Number	5,500
User Number	675,027
Average ecosystem population	122.7
Time Duration	30 days
Total Post Records Number	70,508,791
Total Comment Records Number	51,098,216
Total Like Records Number	119,229,171

can get network structure M and species of each user. We use \hat{X} to represent final states derived from **SocialFate** and different baselines. We compare our **SocialFate** model with 3 baselines:

- **Random model:** we randomly assign $\hat{X}_{n,k}$ following uniform distribution and then normalize it to satisfy $\sum_{k=1}^K \hat{X}_{n,k} = 1$ for each user n .
- **Mean model:** We use overall species abundance distribution of the social ecosystem as species abundance distribution of each user n . So the output of this method is $\hat{X}_{n,k} = \frac{1}{N} \sum_{n=1}^N X_{n,k}$.
- **Generalized Lotka-Volterra model (GLV):** The state-of-the-art ecological model. We can use GLV model to optimize A to get final states of species distribution for each user n .

Evaluation Metrics. For each user n , we can derive his/her ground truth final state of species distribution $X_{n,\cdot}$ from real data. $X_{n,\cdot}$ is a one-hot vector representing the species that user n belongs to. For example, $X_{n,\cdot} = (1, 0, 0)$ represents a pure consumer, while a user with $X_{n,\cdot} = (0, 0, 1)$ is a pure producer. In reality, $X_{n,\cdot}$ is a multinomial distribution, implying uncertain species types. Therefore, we can validate the goodness of fitting by mean absolute error (MAE) between user species distribution from the ground-truth $X_{n,k}$, which presents the probability that user n belongs to species k and the output results \hat{X} from different models. The MAE is:

$$MAE = \frac{1}{N} \frac{1}{K} \sum_{n=1}^N \sum_{k=1}^K \left| \hat{X}_{n,k} - X_{n,k} \right|. \quad (21)$$

5.2 Fitting the Fates of Micro-Social Ecosystems

We first validate our **SocialFate** model by answering if **SocialFate** and baselines can fit the fate of each empirical microscopic social system accurately. Given the empirical fates $X_{n,k}$, i.e., the probability that user n belongs to the species k of each microscopic social ecosystem, we fit X by our **SocialFate** and baselines by MLE framework in Equation 20 and then compare the accuracy of fitting results for all 5,500 ecosystems.

Results. Our model gets very low errors and outperforms all the baselines. Table 4 shows the fitting results of our **SocialFate** model and three baselines. With respect to MAE , our model gets smallest error value 0.092. We further investigate different sub-types, i.e., dead, taciturn, alive, active, of ecosystems. Our model hits the lowest error for all the scenarios. We illustrate the results of our model and baselines in Figure 3. For each microscopic social ecosystem, if the model fits the species distributions of user nodes accurately,

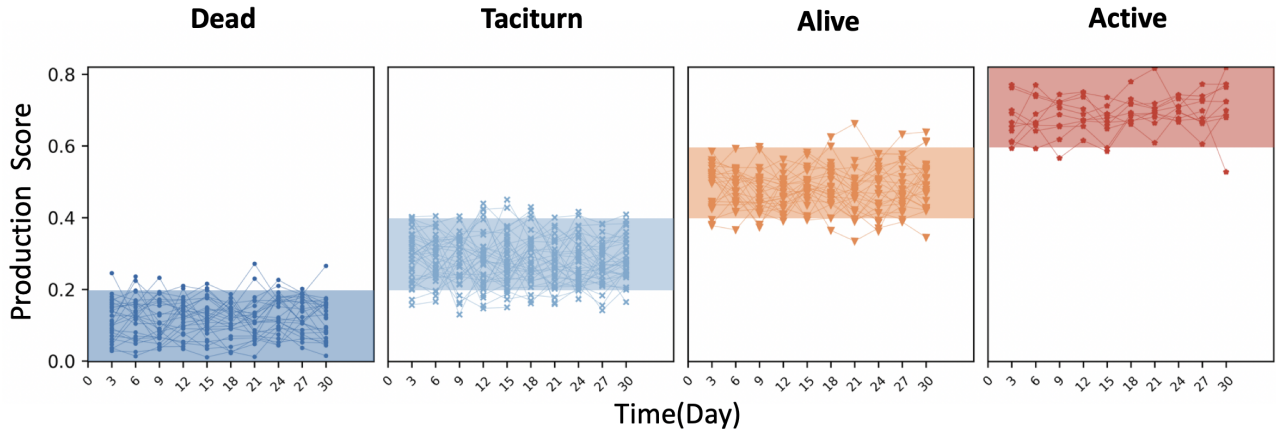


Figure 4: Illustrations of the states of four types microscopic social ecosystems in a month. We randomly selected some microscopic social ecosystems and calculated their production scores every day. Each line represents the production score of a microscopic ecosystem in 30 days. The line that is limited in one shaded area means its state not changing. The final states or fates of most microscopic social ecosystems are steady.

Table 4: The mean value of MAE of different models for fitting the final states of all microscopic social ecosystems. Our SocialFate outperforms all the baselines consistently in different types of microscopic social ecosystems.

Cluster	MAE			
	Random	Mean	GLV	SocialFate
Dead	0.444	0.185	0.295	0.046
Taciturn	0.444	0.320	0.367	0.089
Alive	0.444	0.409	0.418	0.124
Active	0.444	0.411	0.416	0.157
All	0.444	0.314	0.365	0.092

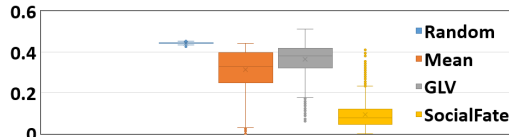


Figure 5: The distributions of MAE of different models for fitting the final states of all microscopic social ecosystems. Our SocialFate outperforms all the baselines, including the state-of-art GLV models.

we color these nodes with their corresponding colors, and shade the nodes for which the model fails. We find our **SocialFate** model recovering almost all the colors, namely species types, in the empirical data, and other models have a lot of shaded nodes, indicating worse fitting results. We also plot the distribution of MAE for all microscopic social ecosystems in Figure 5. The outliers together with the large variance of MAE indicate the heterogeneity of microscopic social ecosystems. However, our **SocialFate** model gets the best results with respect to the lowest mean error and smaller variance than baselines.

Table 5: The mean value of MAE of different models for predicting the final states of all microscopic social ecosystems. Our SocialFate outperforms all the baselines consistently in different types of microscopic social ecosystems.

Cluster	MAE			
	Random	Mean	GLV	SocialFate
Dead	0.173	0.069	0.129	0.016
Taciturn	0.173	0.138	0.155	0.036
Alive	0.174	0.215	0.196	0.086
Active	0.174	0.226	0.203	0.145
All	0.173	0.146	0.163	0.054

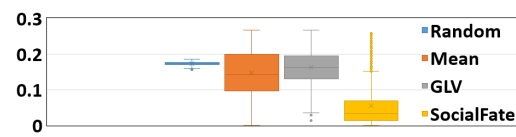


Figure 6: The distributions of MAE of different models for predicting the final states of all microscopic social ecosystems. Our SocialFate outperforms all the baselines, including the state-of-art GLV models.

5.3 Predicting the Fates of Micro-Social Ecosystems

Then, we validate our **SocialFate** model by answering if **SocialFate** and baselines can predict the missing species types distributions X of individuals in microscopic social ecosystems. Given the empirical fates $X_{n,k}$ of 60% users, i.e., we only know 60% probability that user n belongs to the species k of each microscopic social ecosystem, we predict X by our **SocialFate** and baselines by MLE framework in Equation 20 and compare the accuracy of predicting results $\hat{X}_{n,k}$ for all 5, 500 ecosystems.

Results. Our model gets very low errors and outperforms all the baselines. Table 5 shows the predicting result of our **SocialFate** model and three baselines. With respect to *MAE*, our model gets the smallest error value 0.054, while the results of GLV, Mean model, Random model are 0.163, 0.146, 0.173 respectively. For different subtypes, i.e., dead, taciturn, alive, active, of ecosystems, our model hits the lowest error for all the scenarios. The results demonstrate the predicting power of our model, implying that our model reveals the real mechanism of activity in social networks, which is that species interaction matrix A and network structure M are key factors to control the fate of microscopic social ecosystems. Figure 6 shows **SocialFate** get best performance to predict fates of microscopic social ecosystems with lower variance.

6 CONTROLLING THE FATES OF MICROSCOPIC SOCIAL ECOSYSTEMS

By capturing the complexities in microscopic social ecosystems, our **SocialFate** model can further shed light on how to control the fates of these ecosystems. We first show how to make microscopic social ecosystems extremely active or dead in synthetic data, and then extend our insights to more complex empirical datasets.

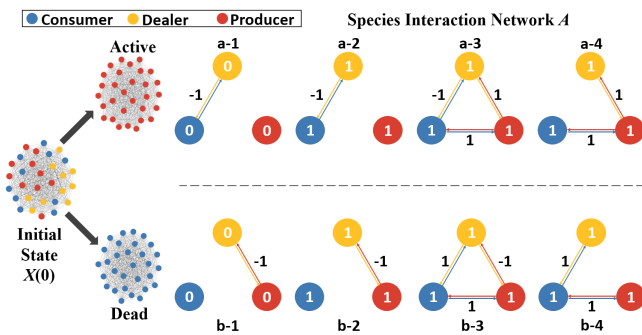


Figure 7: Typical species interaction mechanisms/networks A which control complete graph with random species towards extremely active or dead final states. The red, blue and yellow nodes represent producers, consumers and dealers respectively. The weights of edges: 1 for incentive effect, -1 for suppression effect and 0 for no correlations between nodes.

6.1 Controlling the Fates in Synthetic Data

In most cases, the underlying networks and the initial states are fixed. We try to change the species interaction networks A to control the fates of microscopic social ecosystems. For example, we can incite the information producers by changing social media policies. Here we explore what kinds of A lead to a very active or a dead microscopic social ecosystem, where A represents the relationships between information producers (red nodes), consumers (blue nodes) and dealers (yellow nodes) as shown in Figure 7. We enumerate all the $A \in \{-1, 0, 1\}^{3 \times 3}$ satisfying the condition that $A_{1,1} = A_{2,2} = A_{3,3}$ and $A_{i,j} = A_{j,i}$ ($i, j = 1, 2, 3$) and totally get $3^4 = 81$ types of species interaction networks A . We apply the **SocialFate** dynamics with each specific A on a full-connected network consisting of three species with same amounts and random

initial values. We summarize all the A which lead to very active final states or dead states in Figure 7. We find policies which incite producers (red nodes which have income links with weight 1) can lead to very active fates in complete graphs. Indeed, a microscopic social ecosystem encouraging users to generate contents can flourish, which is consistent with our intuitions. Besides, if consumers (blue nodes) and dealers (yellow nodes) suppress each other, the final states are also very active. In contrast, a microscopic social ecosystem prohibits users from generating contents (red nodes which have income links with weight -1), or only encourages consumers (blue nodes which have income links with weight 1) may lead to dead final states.

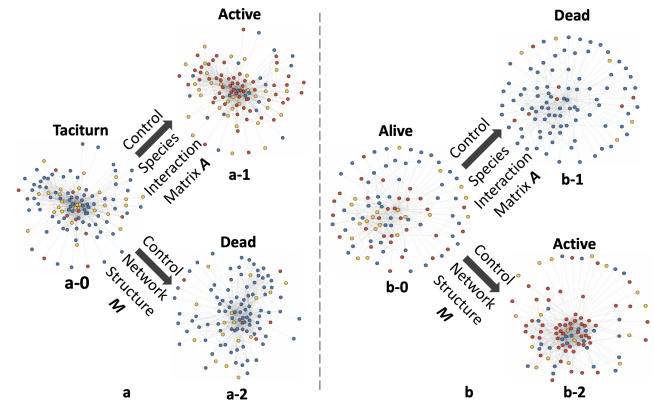


Figure 8: Controlling real-world microscopic social ecosystems towards active or dead states by tuning species interaction mechanism A or individual interaction network M .

6.2 Controlling the Fates in Empirical Data

We further explore how to control the final fates of microscopic social ecosystems in much more complex empirical scenarios. We randomly select typical taciturn or alive microscopic social ecosystems, and control their fates towards active or dead by tuning following two factors:

Species Interaction Network A : We apply the discovered typical species interaction networks A in previous Section 6.1 to the real-world microscopic social ecosystems as shown in Figure 8. Specifically, we change the empirical species interaction network in a taciturn microscopic social ecosystem (Figure 8 a-0) to a producer-incentive one (Figure 7 a-4), and change the empirical species interaction network in an alive microscopic social ecosystem (Figure 8 b-0) to a consumer-incentive one (Figure 7 b-4). Similar to our "toy" discovery in synthetic data, indeed, we change the taciturn microscopic social ecosystem (Figure 8 a-0) to an active one (Figure 8 a-1), and change the alive microscopic social ecosystem (Figure 8 b-0) to an dead one (Figure 8 b-1). Thus, by choosing specific species interaction network A , we can control the microscopic social ecosystems towards active or dead ones.

Individual Network Structure M : We change the connectivity to the hub users (the users with a relatively large number of neighbors) who are information producers. Specifically, by decreasing the connectivity to the hub information producers in Figure 8 a-0,

namely cutting the edges, we change the taciturn microscopic social ecosystem (Figure 8 a-0) to a dead one (Figure 8 a-2). In contrast, we increase the connectivity to the hub information producers in Figure 8 b-0, namely adding the edges, we change the alive microscopic social ecosystem (Figure 8 b-0) to an active one (Figure 8 b-2). Thus, by changing the individual interaction structure M , we can also control the microscopic social ecosystems towards active or dead ones.

7 CONCLUSIONS

In this paper, we figure out the reasons why some social ecosystems can keep self-sustaining and others end up with non-active or dead states. For the first time, driven by a real-world online social media with detailed individual behaviour records, we analyze the final states or fates of microscopic social ecosystems rather than at the population level. We find huge complexities in microscopic social ecosystems and then quantify them, including complex species types, complex individual interaction networks, and complex dynamics and final states. In order to capture the observed complexities in the real-world data, we propose a microscopic ecological model, which captures the complex fates of heterogeneous microscopic social ecosystems accurately in both synthetic and empirical datasets. Furthermore, we analyze the driven factors of the fates of microscopic social ecosystems, including the individual interaction networks and the species interaction mechanisms. These can be applied to the control of microscopic social ecosystems, which is the ability to influence the temporal behaviours and their final states towards active or dead fates.

ACKNOWLEDGEMENT

This work was supported in part by National Program on Key Basic Research Project (No. 2015CB352300), National Natural Science Foundation of China (No. 61772304, No.61521002, No. 61531006), National Natural Science Foundation of China Major Project (No. U1611461), and Beijing Academy of Artificial Intelligence (BAAI). Thanks for the support of Tsinghua-Tencent Joint Lab and the Young Elite Scientist Sponsorship Program by CAST. All opinions, findings and conclusions in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.

REFERENCES

- [1] Roy M Anderson and Robert M May. 1992. *Infectious diseases of humans: dynamics and control*. Oxford university press.
- [2] Sitaram Asur and Bernardo A Huberman. 2010. Predicting the future with social media. In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01*. IEEE Computer Society, 492–499.
- [3] Andrew Bakun. 1996. *Patterns in the ocean: ocean processes and marine population dynamics*. California Sea Grant, in cooperation with Centro de Investigaciones Biológicas del Noroeste, La Paz, Mexico.
- [4] Jianhai Bao, Xuerong Mao, Geroge Yin, and Chenggui Yuan. 2011. Competitive Lotka–Volterra population dynamics with jumps. *Nonlinear Analysis: Theory, Methods & Applications* 74, 17 (2011), 6601–6616.
- [5] Alex Beutel, B Aditya Prakash, Roni Rosenfeld, and Christos Faloutsos. 2012. Interacting viruses in networks: can both survive?. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 426–434.
- [6] Douglas H Boucher, Sam James, and Kathleen H Keeler. 1982. The ecology of mutualism. *Annual Review of Ecology and Systematics* 13, 1 (1982), 315–347.
- [7] Judith L Bronstein. 1994. Our current understanding of mutualism. *The Quarterly Review of Biology* 69, 1 (1994), 31–51.
- [8] Robert Costanza and Michael Mageau. 1999. What is a healthy ecosystem? *Aquatic ecology* 33, 1 (1999), 105–115.
- [9] Anthony Downs. 1972. Up and down with ecology: The issue-attention cycle. *The public* (1972).
- [10] Roger Fletcher. 2013. *Practical methods of optimization*. John Wiley & Sons.
- [11] Jean-Michel Gaillard, Marco Festa-Bianchet, and Nigel Gilles Yoccoz. 1998. Population dynamics of large herbivores: variable recruitment with constant adult survival. *Trends in Ecology & Evolution* 13, 2 (1998), 58–63.
- [12] Jianxi Gao, Baruch Barzel, and Albert-László Barabási. 2016. Universal resilience patterns in complex networks. *Nature* 530, 7590 (2016), 307.
- [13] Narendra S Goel, Samaresh C Maitra, and Elliott W Montroll. 1971. On the Volterra and other nonlinear models of interacting populations. *Reviews of modern physics* 43, 2 (1971), 231.
- [14] Kondalsamy Gopalsamy. 2013. *Stability and oscillations in delay differential equations of population dynamics*. Vol. 74. Springer Science & Business Media.
- [15] Jacopo Grilli, György Barabás, Matthew J Michalska-Smith, and Stefano Allesina. 2017. Higher-order interactions stabilize dynamics in competitive network models. *Nature* 548, 7666 (2017), 210.
- [16] J Philip Grime. 2006. *Plant strategies, vegetation processes, and ecosystem properties*. John Wiley & Sons.
- [17] Barira Hanif and Farhan Javed. 2012. Using social media data analysis for decision support. (2012).
- [18] Josef Hofbauer and Karl Sigmund. 1998. *Evolutionary games and population dynamics*. Cambridge university press.
- [19] J Nathaniel Holland, Donald L DeAngelis, and Judith L Bronstein. 2002. Population dynamics and mutualism: functional responses of benefits and costs. *The American Naturalist* 159, 3 (2002), 231–244.
- [20] Joseph Huber. 2000. Towards industrial ecology: sustainable development as a concept of ecological modernization. *Journal of environmental policy and planning* 2, 4 (2000), 269–285.
- [21] Andreas M Kaplan and Michael Haenlein. 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons* 53, 1 (2010), 59–68.
- [22] William Ogilvy Kermack and Anderson G McKendrick. 1927. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* 115, 772 (1927), 700–721.
- [23] Yang Kuang. 1993. *Delay differential equations: with applications in population dynamics*. Vol. 191. Academic Press.
- [24] Richard Law and Jerry C Blackford. 1992. Self-assembling food webs: a global viewpoint of coexistence of species in Lotka–Volterra communities. *Ecology* 73, 2 (1992), 567–578.
- [25] Jonathan M Levine, Jordi Bascompte, Peter B Adler, and Stefano Allesina. 2017. Beyond pairwise mechanisms of species coexistence in complex communities. *Nature* 546, 7656 (2017), 56.
- [26] Thomas Robert Malthus. 1878. *An essay on the principle of population: Or, a view of its past and present effects on human happiness, with an inquiry into our prospects respecting the future removal or mitigation of the evils which it occasions*. London, Reeves and Turner.
- [27] Yasuko Matsubara, Yasushi Sakurai, and Christos Faloutsos. 2015. The web as a jungle: Non-linear dynamical systems for co-evolving online activities. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 721–731.
- [28] Bruno Ribeiro. 2014. Modeling and predicting the growth and death of membership-based websites. In *Proceedings of the 23rd international conference on World Wide Web*. ACM, 653–664.
- [29] David J Schaeffer, Edwin E Herricks, and Harold W Kerster. 1988. Ecosystem health: I. Measuring ecosystem health. *Environmental Management* 12, 4 (1988), 445–455.
- [30] Simon Walk, Denis Helic, Florian Geigl, and Markus Strohmaier. 2016. Activity Dynamics in Collaboration Networks. *ACM Trans. Web* 10, 2, Article 11 (May 2016), 32 pages. <https://doi.org/10.1145/2873060>
- [31] Xin Wang, Wenwu Zhu, Chun Chen, and Martin Ester. 2018. Joint User-and Event-Driven Stable Social Event Organization. In *WWW*. 1513–1522.
- [32] Chengxi Zang, Peng Cui, and Christos Faloutsos. 2016. Beyond sigmoids: The nettide model for social network growth, and its applications. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2015–2024.
- [33] Chengxi Zang, Peng Cui, Christos Faloutsos, and Wenwu Zhu. 2018. On Power Law Growth of Social Networks. *IEEE Transactions on Knowledge and Data Engineering* 30, 9 (2018), 1727–1740.
- [34] Tianyang Zhang, Peng Cui, Christos Faloutsos, Yunfei Lu, Hao Ye, Wenwu Zhu, and Shiqiang Yang. 2016. Come-and-go patterns of group evolution: A dynamic model. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1355–1364.