Social Synchrony on Complex Networks

Qi Xuan, Zhi-Yuan Zhang, Chenbo Fu, Hong-Xiang Hu, Vladimir Filkov

***Abstract*—Social synchrony is an emergent phenomenon in human society. People often mimic others which, over time, can result in large groups behaving similarly. Drawing from prior empirical studies of social synchrony in online communities, here we propose a discrete network model of social synchrony based on four attributes: *depth* of action and *breadth* of impact, i.e., a large number of actions are performed with a large group of people involved; *heterogeneity* of role, i.e., people of higher degree play more important roles; and lastly, *emergence* of phenomenon, i.e., it is far from random. We analyze our model both analytically and with simulations, and find good agreement between the two. We find this model can well explain the four characters of social synchrony, and thus hope it can help researchers better understand human collective behavior.**

***Index Terms*—Synchronization, scale-free network, collective behavior, social network, significant emergence.**

I. INTRODUCTION

SYNCHRONIZATION is an emergent phenomenon present in many natural and artificial systems. It is mainly understood as an adjustment of the individual rhythms of actors in the systems due to their coordination with others [1], e.g., the flashing of fireflies to aid sexual selection [2] and the synchronous firing of neurons during cognitive processing [3]. In the past decade, stimulated by the discoveries of small-world [4] and scale-free [5] features in many natural and artificial networked systems, a surge of research has been focusing on the study of synchronization on networks [6], [7], [8]. The particular focus of much of that work has been on theoretical approaches to reveal and address how synchronization is coupled to, and perhaps arises from, network structure. For instance, it has been shown that small-world or scale-free networks are more easily synchronized than ER networks [9], [10], [11]. Inversely, Fu *et al.* [12] found theoretically that system synchronization can also influence the network structure, although only in an ideal scenario, with very strict conditions, making it hard to reproduce practically. E.g., theoretically, for a network of coupled oscillators to achieve global synchronization, the oscillators must be identical, but in realistic systems, this condition is difficult to achieve. A related concept to synchronization is *consensus* [13], [14], [15], [16]. First-order consensus of multi-agent systems [17], [18], [19] can be thought of as a special case of the synchronization of coupled complex dynamical systems, where the intrinsic dynamics of all agents are zero [20].

Here we are interested in a form of synchronization that occurs in groups of organisms that can communicate and mimic behavior; we call it *social synchrony*. Specifically, by *social*, we mean involving groups of humans or other animals. And by *synchrony* we mean actions performed close in time, toward achieving a common goal. A hallmark of social synchrony is non-trivial, collective dynamic behavior, characterized by self-organizing activities. Such phenomenon may have its psychological effects and further lead to the better accomplishment of certain tasks of significant complexity [21]. We note that the terms *synchrony* and *synchronization* are relatively close, but here we make a distinction: *synchrony* represents the state of two or more events occurring at the same time, while *synchronization* is more about the coordinated dynamics of many units to the same timing. Thus, synchrony is a more specialized form of synchronization, pertaining to temporal coordination. Accordingly, we will use the *synchrony*, rather than *synchronization*, in the main part of paper from Sec.II, to emphasize that our interests are in temporal social coordination.

In terms of prior work, group coordination in animals and humans has been the focus of many studies. For instance, animals moving in groups exhibit various synchronous patterns, and phenomena like birds flocking [22] and fish swimming [23] have been studied in detail. To explain movement synchronization in multiplayer scenarios among humans, Alderisio *et al.* [24] analyzed a network of non-identical Rayleigh / van der Pol (RvdP) oscillators interconnected through either diffusive or nonlinear coupling functions. There, the boundedness of the error was proven when the oscillators were coupled diffusively, under the assumption that the network is connected, simple, and undirected. Codrons *et al.* [25] found that spontaneous movement within a group was capable of creating interpersonal synchronization of motor dynamics, while the mere fact of being in a group prompted individuals to breathe in a synchronized fashion, even in the absence of shared movement. Moreover, Alderisio *et al.* [26] investigated group synchronization in a human ensemble where participants were asked to generate and coordinate an oscillatory hand motion. It was found that the coordination level of the ensemble depends on the specific way each individual moves when isolated from the others, and on the pattern of the visual coupling among group members. They further proposed a data-driven mathematical model to explain these experimental observations. Besides, Frank and Richardson [27] introduced a quantitative approach to detect phase synchronization in noisy experimental multivariate data and derived a test statistic based on the Kuramoto order parameter. This provides an objective way to measure the magnitude of group synchrony and examine whether and how the magnitude of such synchrony influences the "social dynamics" of group interaction. More examples can be found in a recent review on social synchrony and its benefits [21].

Q. Xuan and C. Fu are with the Department of Automation, Zhejiang University of Technology, Hangzhou, 310023 China. C. Fu is the corresponding author with the e-mail: cbfu@zjut.edu.cn

Z.-Y. Zhang and V. Filkov are with the Department of Computer Science, University of California, Davis, CA, 95616 USA.

H.-X. Hu is with the Department of Mathematics, Hangzhou Dianzi University, Hangzhou, 310018 China.

Although social synchrony is prevalent, large scale data collection and storage technologies have only recently caught up sufficiently to allow precise modeling. Thanks to advances in visual recording and tagging technology, GPS, mobile phone, and the Internet, we now have unprecedented opportunities to record and consequently study the dynamics of animal and human behavior. Recently, Lukeman *et al.* [28] used digital cameras to collect individual position, velocity, and trajectory of flocks of hundreds of surf scoters, based on which they revealed a distinct, concentric structure in positioning, a preference for neighbors directly in front, and strong alignment with neighbors on each side. Nagy *et al.* [29], on the other hand, used GPS devices to get the flight trajectories of pigeons and found a well-defined hierarchical structure among flock members. Song *et al.* [30] utilized mobile phones to collect human location in real time, and found that human mobility patterns are highly predictable; while Xuan *et al.* [31] collected developer commit behaviors in Apache software foundation and found recurring work patterns among software developers. They found that developers tend to shift focus along with software dependency links described by the call graphs and this tendency appears stronger with more productive developers. Choudhury *et al.* [32] defined social synchrony as the tendency of a large number of people to perform similar actions in unison, in response to a contextual trigger. They studied how humans mimic others, on the popular social media site *Digg*, and developed an evolution framework to predict social synchrony. More recently, Xuan and Filkov [33] defined synchronous software development for developers who are spatially separated. They found that developers may exhibit higher programming efficiency when they synchronize their code changes with others, which can results in project size increase despite the decreased coding effort.

Thanks to the Internet, and specifically to the social web, it has become a straightforward task to gather and mine comprehensive, big-data, data sets, about online social synchrony of people. Not surprisingly, this has resulted in an explosion of studies of social web phenomena [34], [35], [36], some on social synchrony [32], [33], [37]. However, most of the latter are empirical rather than theoretical, i.e., they reveal that real social synchrony phenomena exist, as observed in the collected data, but largely do not provide models to explain them. Choudhury *et al.* [32] developed a Dynamic Bayesian Network (DBN) model that includes an understanding of user context to predict the probability of user actions over a set of time slices into the future, and further to predict, rather than explain, the social synchrony, in a particular case. Zhao *et al.* [37] proposed a discrete model to investigate the relationship between the tie strength and information propagation in online social networks, and focused on evaluating different information propagation strategies. More recently, Alderisio *et al.* [38] designed a model-based method to study coordination in human ensembles via a computer-based set-up that enables individuals to coordinate each other's motion from a distance, which provides a good platform to investigate particular online social synchrony under a controllable environment.

In this paper, we aim to study online social synchrony in a more systematic way. Motivated by current empirical studies, we attempt to summarize typical properties exhibited by an online social synchrony phenomenon; then operationalize these properties by defining measures for them; and finally propose a mathematical model for social synchrony based on those measures. The rest of the paper is organized as follows. In Sec. II, we summarize the four typical properties of observed social synchrony phenomena, based on recent empirical studies. In Sec. III, we propose a theoretical model to explain the observed properties of social synchrony, and do a series of theoretical analysis. In Sec. IV, we give a corresponding numerical model, and validate the theoretical results by simulations. We conclude the paper in Sec. V.

II. Attributes of Social Synchrony

Various characteristics of social synchrony are noted as typical, or important, across a number of empirical studies [21], [32], [33]. Next, we identify those characteristics, and accept them as the attributes, or dimensions along which we will study social synchrony.

1) **Depth of action** *The number of actions taken in a period of time*. Choudhury *et al.* [32] found that users on the social blogging site Digg may continually "dig" news stories on a topic, and Xuan *et al.* [31] found that a developer in Apache software foundation may continually commit to certain files of a software, over a period of time.
2) **Breadth of impact** *The number of active people involved in a period of time*. Choudhury *et al.* [32] found that the continued participation of old users can impact a large number of new users in the network to participate as well.
3) **Heterogeneity of role** *People of higher degree in a social network triggering wider social synchrony, in terms of larger number of actions, and they themselves also taking more actions, in a period of time.* Choudhury *et al.* [32] found that users of higher degree as seeds are more likely to trigger social synchrony. We studied the relationship between the number of commit actions and the degree k for a developer in the social coding network of GitHub, and found that it can be well fitted by a linear function, [1] as shown in Fig. 1. The slope of the fitting line is about 5.5, indicating that on average, a developer will take 5.5 more commits when he/she gets 1 more social link in GitHub. There are two exceptions with rather huge residuals, i.e., when k is equal to 80 and 85. We checked the developers with these two particular degrees and found two outstanding contributors: Dave Abrahams, with degree 80, delineated a theory of exceptions, sit on the C++ Standards Committee, is a founding member of Boost. He made 30131 commits in four

[1] GitHub is a well-known OSS community. The data set was collected on October 12, 2013, and the social network was constructed by the following links between the developers. We get the mean value of the number of commits for the developers of the same degree in the social network, and then regress it against the degree. The *Pearson* correlation coefficient [39] between the two, defined as their covariance divided by the product of their standard deviations, equals to 0.91, indicating a statistically significant linear correlation between them. Note that here we only consider the developers with degree smaller or equal to 100, which take up about 99% of all developers in GitHub.

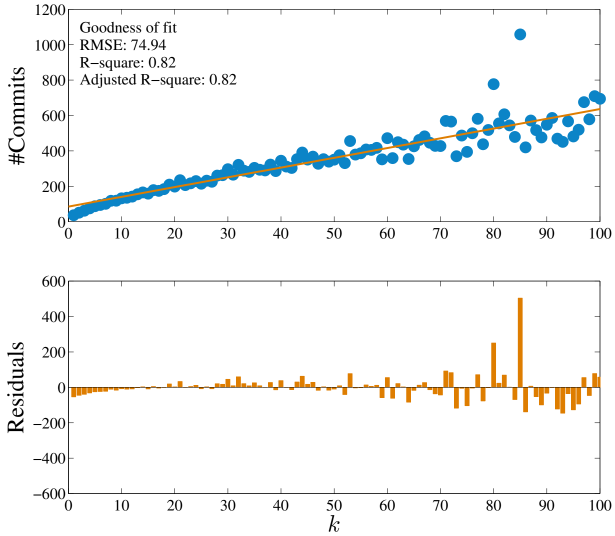

Fig. 1. (Top) The relationship between the number of commit actions and the degree k for a developer in the social network of GitHub, which is well fitted by a linear function. (Bottom) The residuals of the fit.

years; Brain Chan, with degree 85, is the founder and chief developer of Liferay. He made 49929 commits in three years. When they are removed, the average number of commits made by the developers with degree 80 drops from 777.0 to 540.3; and that made by the developers with degree 85 drops from 1058.0 to 548.9. Meanwhile, the linear fit gets much better in this case, i.e., for the goodness of fit, Root Mean Square Error (RMSE) decreases from 74.94 to 47.43, while R-square (the square of *Pearson* correlation coefficient) and adjusted R-square both increase from 0.82 to 0.91.

4) **Emergence of phenomenon** *Significantly more actions occurring in a period of time than given by chance*. When we refer to social synchrony, we always consider it as a significant emergence, which is different from a random phenomenon. Xuan and Filkov [33] identified significantly more bursts of synchronous development in reality, by comparing with a null model, where the commit activities for each developer are randomly redistributed.

When posited along those four dimensions, social synchrony, while sufficiently different, is somewhat related to a number of different concepts, *e.g.*, network epidemiology [40], [41], [42], information flow [43], [44], social cascades [45], [46], social correlation [47], and social recommendation [48], [49], when we interpret *action* as being infected, transferring information, or buying product etc. It is also tangentially related to opinion formation on social networks, which has been extensively studied recently. *E.g.*, Huang *et al.* [50] introduced a simple model to study opinion formation on networks with community structure, and found that a community may persist and never be assimilated when its cohesion reaches a certain level. They further studied a minority's opinion evolution in this setting [51], where majority rule is applied to govern the evolution, and found that a larger group size would bring more advantage to the minority. Moreover, Qian *et al.* [52] proposed an adaptive bridge control strategy, calling for controlling a special kind of nodes named bridge without any global or local knowledge, and found that the efficiency of this strategy is closely related with the clustering coefficient.

In the following, we will propose a mathematical model for social synchrony around the above four attributes, based on which we will provide rigorous measures, and then analyze the model both theoretically and numerically.

III. Theoretical Analysis

A complex network can be represented by a graph $G = (V, E)$ with nodes $V = \{v_1, v_2, \ldots, v_N\}$ and links $E \subset V \times V$. In the network, a node can be either active or inactive. Our discrete social synchrony (SS) model can then be described as follows.

Discrete SS Model: At time step $t+1$, a node in the network G is either spontaneously active with probability β, or it is activated (or stimulated) with probability α by one of its neighbors that were active at time step t.

Suppose that node v_i is active at time step t with probability $p_i(t)$, then the probability that it is active at time step $t+1$ is given by:

$$p_i(t+1) = 1 - (1-\beta) \prod_{v_j \in \pi_i} [1 - \alpha p_j(t)], \tag{1}$$

where π_i is the set of neighbors of node v_i. Different from the information propagation model in [37], here a node can be either active or inactive at each time step, which is fully determined by the active probability presented in Eq (1). In practice, we find that α is quite small, which can help to simplify this model, as we describe next.

To motivate the model simplification, we take Open Source Software (OSS) development as an example. Typically, there are a number of developers in a project, and each developer may commit to different files at different times. Two developers can be considered to program synchronously if they commit to the same files close in time [33]. Suppose a developer committed to a set of files, $\mathcal{F}$, at time t, and there are h other developers, each of whom committed to at least one file in $\mathcal{F}$ in the short time interval $[t, t+\Delta]$. Here, Δ can be considered as the time interval between two successive time steps above. Then, denoting by H the number of developers in the project before $t+\Delta$, the stimulated probability at time t can be approximated as

$$\alpha = \frac{h}{H-1}, \tag{2}$$

since all developers in an OSS project can be considered fully connected to each other, i.e., every commit is visible to all developers. Then, we can get the average stimulated probability α for the whole project by considering all commit actions. E.g., for the 31 OSS projects in our data set [33], [53] (available via Figshare at: https://dx.doi.org/10.6084/m9.figshare.

3181555), on average, we get $\alpha = 0.0021, 0.0081, 0.0117$ for time intervals $\Delta = 1, 5, 10$ (days), respectively. For $\Delta = 1$ (day), we mean a developer is stimulated by another if they committed to at least one same file within one day. This result suggests that first, in reality, the stimulated probability α is indeed very small; and second we can get a smaller α by considering shorter time intervals, i.e., the chance to observe the actions of different individuals in a short period of time is rather small. Thus, by choosing a sufficiently small time interval between two successive time steps and setting it as the unit time, Eq. (1) can be simplified to

$$\begin{aligned} p_i(t+1) &= 1-(1-\beta)\left[1-\alpha\sum_{v_j\in\pi_i}p_j(t)\right] \\ &= \eta\sum_{v_j\in\pi_i}p_j(t)+\beta, \end{aligned} \tag{3}$$

with $\eta = \alpha(1-\beta)$. Note that, based on Eq. (3), p_i is smaller than 1 only when α is small enough and β is smaller than 1. Assuming no degree correlation between linked nodes, we can use mean-field theory to transform the node presentation into a degree presentation [54], [55], [56]. Then, Eq. (3) becomes

$$\rho_k(t+1) = \eta k\rho(t)+\beta. \tag{4}$$

where ρ_k is the average active probability of nodes with degree k and ρ is the average active probability of all the nodes in the network. Denoting by $P(k)$ the degree distribution of the network, multiplying Eq. (4) by $P(k)$ and then summing it over degree k, we obtain

$$\rho(t+1) = \eta\langle k\rangle\rho(t)+\beta. \tag{5}$$

Subtracting $\rho(t)$ from both sides yields

$$\rho(t+1)-\rho(t) = (\eta\langle k\rangle-1)\rho(t)+\beta. \tag{6}$$

From this, we can derive the dynamics of the active probability in the whole network:

$$\dot{\rho}(t) = (\eta\langle k\rangle-1)\rho(t)+\beta, \tag{7}$$

under the assumption that the time interval between the two successive time steps is small enough and is set as the unit time [57]. By solving this linear differential equation, given the initial state $\rho(0)$, we get

$$\rho(t) = \left[\rho(0)-\frac{\beta}{1-\eta\langle k\rangle}\right]e^{-(1-\eta\langle k\rangle)t}+\frac{\beta}{1-\eta\langle k\rangle}. \tag{8}$$

Remark 1: Our model can be generalized by considering different spontaneous and stimulated probabilities for different nodes and links, respectively. Denoting by β_i the spontaneous probability of node v_i, and by α_{ji} its stimulated probability by neighbor v_j. Without loss of generality, we introduce weights on the links and set $\alpha_{ij} = \alpha w_{ij}$. Then, Eq (1) changes to

$$p_i(t+1) = 1-(1-\beta_i)\prod_{v_j\in\vec{\pi}_i}\left[1-\alpha w_{ji}p_j(t)\right], \tag{9}$$

where $\vec{\pi}_i$ is the set of *incoming* neighbors of node v_i. Under the same assumption that the stimulated probabilities are small enough, Eq (9) can be simplified to

$$p_i(t+1) = \eta_i\sum_{v_j\in\vec{\pi}_i}w_{ji}p_j(t)+\beta_i, \tag{10}$$

with $\eta_i = \alpha(1-\beta_i)$. If we set

$$\beta \equiv \frac{\sum_{i=1}^{N}\beta_i}{N}, \tag{11}$$

and assume that there is no *incoming weighted* degree correlation between linked nodes [55], and also that the spontaneous probability β_i is uncorrelated with the weighted degree of node v_i, corresponding to Eq. (4), we have

$$\rho_w(t+1) = \eta w\rho(t)+\beta, \tag{12}$$

where ρ_w is the average active probability of nodes with incoming weighted degree w, and $\eta = \alpha(1-\beta)$. Here, the *incoming weighted degree* of a node is defined as the sum of the weights of its incoming links. Denoting by $P(w)$ the incoming weighted degree distribution of the network, multiplying Eq. (12) by $P(w)$ and then summing it over incoming weighted degree w, we obtain

$$\rho(t+1) = \eta\langle w\rangle\rho(t)+\beta. \tag{13}$$

This is similar to Eq. (5), except that the average degree, $\langle k\rangle$, is replaced by the average incoming weighted degree, $\langle w\rangle$, in Eq. (13). Thus, since most of the equations in the following exposition are based on Eqs. (4) and (5), to get the corresponding results for the situation where the spontaneous and stimulated probabilities vary for different nodes and links, we just need to replace k by w.

A. Depth of Action

Social synchrony is a temporal phenomenon, and intuitively its effect can be characterized by the number of actions $\varphi(T)$ in a period of time T, which we call *depth of action*. It can be estimated by

$$\begin{aligned} \varphi(T) &= \int_{t=0}^{T}N\rho(t)dt \\ &= N\left[\rho(0)-\frac{\beta}{1-\eta\langle k\rangle}\right]\frac{1-e^{-(1-\eta\langle k\rangle)T}}{1-\eta\langle k\rangle} \\ &\quad +\frac{\beta}{1-\eta\langle k\rangle}NT. \end{aligned} \tag{14}$$

Remark 2: In Eq. (14), $\rho(t)$ is the average active probability of all the nodes in the network at time t. Since there are a total of N nodes in the network, $N\rho(t)$ represents the number of actions in the whole network at time t, and the integration of which, i.e., $\varphi(T)$, thus means the total number of actions in the whole network over a period of time T. As we can see, the first term of Eq. (14) is largely determined by $\rho(0)$ and the increasing ratio decays exponentially with time when $\eta\langle k\rangle < 1$, while the second term increases with time linearly. In other words, over time, the effect of the initial active probability on *depth of action* will diminish quickly, and the number of actions on the network will be mainly determined by the spontaneous and stimulated probabilities β and α, the network size N, and the average degree $\langle k\rangle$ in the second term of Eq. (14), when the observed time T is long enough. Note that the effect of $\rho(0)$ may be quite significant if we only focus on the initial short time period.

It is interesting and helpful to consider two extreme cases, i.e., $\alpha = 0$, $\beta > 0$ and $\alpha > 0$, $\beta = 0$. When $\alpha = 0$ and $\beta > 0$, $\eta = 0$, and the active probability of a node at the next time step will not be influenced by the current active probabilities of its neighbors, which means that all nodes in the network are independent from each other, i.e., our social synchrony model degenerates into a random model. Eqs. (5) and (14) become

$$\rho(t) = \beta, \quad t \geq 1, \tag{15}$$
$$\varphi(T) = N\rho(0) + \beta N(T-1). \tag{16}$$

Eq. (16) is an intuitive result in this case: the total number of actions in the network is proportional to the spontaneous probability, the network size, and the length of observation time.

On the other hand, when $\alpha > 0$ and $\beta = 0$, $\eta = \alpha(1-\beta) = \alpha$, and the active probability of a node at the next time step is fully determined by the current active probabilities of its neighbors, *i.e.*, information diffusion dominates the dynamics and there is no spontaneous actions on the network. Eqs. (8) and (14) become

$$\rho(t) = \rho(0)e^{-(1-\alpha\langle k\rangle)t}, \tag{17}$$
$$\varphi(T) = N\rho(0)\frac{1-e^{-(1-\alpha\langle k\rangle)T}}{1-\alpha\langle k\rangle}. \tag{18}$$

Eq. (17) indicates that, in this case, there is a critical point, α_c, for the stimulated probability, $\alpha_c = 1/\langle k\rangle$, below which the activity of the network will die out with time and, based on Eq. (18), there is a limited total number of actions, as estimated by

$$\varphi(\infty) = \frac{N\rho(0)}{1-\alpha\langle k\rangle}. \tag{19}$$

Based on Eq. (5), we can find that the average active probability of the nodes in the whole network is always larger than β when $\alpha > 0$, indicating that the action of an individual is stimulated by the actions of its neighbors in this case. On the other hand, it is always smaller than β if $\alpha < 0$, [2] meaning the action of an individual is inhibited by the actions of its neighbors. Note that, in a social network, different links may play different roles in social synchrony and information diffusion. That is, the action of an individual can be either stimulated or inhibited by the actions of its neighbors, and thus, it is also very interesting to study social synchrony on a layered network with different kinds of links. In this paper, however, we only consider the case when $\alpha > 0$, which is the reason why we observe significantly more bursts of social synchrony in reality than by chance.

B. Breadth of Impact

In order to investigate the sphere of influence of a social synchrony on a network, it is also important to estimate the *breadth of impact*, i.e., the number of active nodes $\phi(T)$, in a period of time T.

Firstly, the mean active probability of a node with degree k at each time point in a period T can be calculated by

$$\begin{aligned}\vartheta_{k,T} &= \frac{1}{T}\int_0^T \rho_k(t)dt \\ &= \frac{1}{T}\left[\rho(0) + \eta k\int_0^{T-1}\rho(t)dt + \beta(T-1)\right] \\ &= \frac{1}{T}\left[\rho(0) + \beta(T-1) + \frac{\eta k\varphi(T-1)}{N}\right]\end{aligned} \tag{20}$$

Then, the breadth $\phi(T)$ during a period T, can be estimated by

$$\phi(T) \approx N\sum_k P(k) f_T(\vartheta_{k,T}), \tag{21}$$

where $f_T(x)$ is a probability function defined as

$$f_T(x) = 1 - (1-x)^T, \tag{22}$$

and denotes the probability that a node is active at least once in the period T, assuming its probability of being active at any time equals to x.

Remark 3: As defined in Eq. (20), $\vartheta_{k,T}$ is the mean active probability of a node with degree k at each time in a period T, thus $1-\vartheta_{k,T}$ represents the mean probability of the node being inactive at each time in this period. Then, $(1-\vartheta_{k,T})^T$ means the probability of the node being inactive in this whole period. So, $f_T(\vartheta_{k,T})$ in Eq. (21) denotes the probability that the node with degree k is active at least once in the period T. Since there are total $NP(k)$ nodes having degree k in the network, the number of active nodes of degree k can be estimated by $NP(k)f_T(\vartheta_{k,T})$. Therefore, $\phi(T)$ is fully defined by Eq. (21), and given a network, can be used to estimate the total number of active nodes in it over a time period T.

C. Heterogeneity of Role

Different nodes and links may play quite different roles in network dynamics [58], [59], [60], [61], [62], especially in heterogeneous and modular networks. Therefore, it is also very interesting to theoretically study if nodes of different degrees may play different roles in social synchrony, i.e., study *heterogeneity of role*.

To motivate, we have found, *e.g.*, that software developers of higher degree in the social coding network of GitHub also contribute more code changes (commits) there, as evidence by the linear relationship in Fig. 1. This is consistent with our current model, in particular the node activity versus degree relationship in Eq. (4), where it is theoretically shown that the active probability of a node is in proportion to its degree k in the network. This is the starting point of our connection between social synchrony and centrality.

Choudhury *et al.* [32] found that users of higher degree as seeds are more likely to trigger social synchrony, which can be shown to be consistent with our model, as follows. We consider the special case when $\alpha > 0$ and $\beta = 0$. Suppose the event is initialized by a node with degree k and there is no degree correlation between linked nodes, then each neighbor

[2] As a probability, of course α cannot be smaller than zero, but we can still make an analogy, i.e., negative α makes the neighboring nodes inhibit, rather than stimulate, each other.

of the node is active with probability α, and, on average, there will be αk active nodes, at the next time step, i.e., we have $\rho(1) = \alpha k/N$. From this time step on, Eqs (17) and (18) become

$$\rho(t) = \frac{\alpha k}{N} e^{-(1-\alpha\langle k\rangle)(t-1)}, \tag{23}$$

$$\varphi(T) = \alpha k \frac{1 - e^{-(1-\alpha\langle k\rangle)(T-1)}}{1-\alpha\langle k\rangle} + 1, \tag{24}$$

respectively.

Remark 4: In Eqs (23) and (24), the stimulated probability α, network size N, and the average degree $\langle k\rangle$ are all constants. Thus, as we can see, both the average active probability of the network and the total number of actions in the period of time T are in proportion to the degree k of the node that initializes the event. In other words, it can be theoretically shown that, based on our model, hub nodes of higher degree play more important roles in the process and can enable larger-scale social synchrony.

D. Emergence of Phenomenon

Given an empirical data set, to test whether an observed pattern is a significant emergence, rather than a random phenomenon, we need to create a baseline, or a null model, of the pattern occurring purely by chance [33], [63]. From those null models, a corresponding simulated data set can be generated and the prevalence of the pattern in question can be then contrasted between the empirical and the simulated data, using an appropriate significance statistic, e.g., their relative difference. We and others have previously observed that social synchrony is a significant emergence [21], [32], [33]. Here we seek to derive the *emergence of phenomenon* for a social synchrony theoretically, from our statistical model, and focus on the dimension of social synchrony depth.

Assuming we have observed a system for a long time, i.e., $T \to \infty$, then, according to Eq. (14), the total number of actions in this period is close to

$$\varphi(T) = N\left[\rho(0) - \frac{\beta}{1-\eta\langle k\rangle}\right]\frac{1}{1-\eta\langle k\rangle} + \frac{\beta}{1-\eta\langle k\rangle}NT. \tag{25}$$

Note that $\varphi(T)$ is a linear function of T, i.e., $\varphi(T) \to \infty$ as $T \to \infty$. Hence, we don't let $T \to \infty$ in reality, and just assign it with a large enough number. In a null model, these actions will be uniformly distributed in the time period T. Thus, the number of actions in a short time period $\theta(\theta \ll T)$ in the null model can be estimated by

$$\begin{aligned}\varphi^o(\theta) &= \frac{\theta}{T}\varphi(T) \\ &= \frac{N\theta}{T}\left[\rho(0) - \frac{\beta}{1-\eta\langle k\rangle}\right]\frac{1}{1-\eta\langle k\rangle} \\ &\quad + \frac{\beta}{1-\eta\langle k\rangle}N\theta.\end{aligned} \tag{26}$$

When T is large enough, the first term of Eq. (26) can be ignored, and thus it can be further simplified to

$$\varphi^o(\theta) = \frac{\beta}{1-\eta\langle k\rangle}N\theta. \tag{27}$$

On the other hand, in our model, according to Eq. (14), we can get the number of actions in this short period of time by replacing T with θ.

To measure the distance of a prediction of our model to that of the null model, given same parameters, we choose the relative difference between them as our index of significant emergence:

$$\chi(\theta) = \frac{\varphi(\theta) - \varphi^o(\theta)}{\varphi^o(\theta)}. \tag{28}$$

In theory, based on Eqs. (14), (27), and (28), we have

$$\chi(\theta) = \left[\frac{\rho(0)}{\beta} - \frac{1}{(1-\eta\langle k\rangle)}\right]\frac{1-e^{-(1-\eta\langle k\rangle)\theta}}{\theta}. \tag{29}$$

Remark 5: Generally, the index defined by Eq. (29) is a decreasing function of the spontaneous probability β and the length of the time period θ, which is reasonable if we consider that increasing each of these two parameters may result in more random spontaneous actions being included in this time period, and thus may decrease the significant emergence index, which is mainly determined by the ratio of neighborhood diffusive actions over the random spontaneous ones. When $\eta\langle k\rangle$ is much smaller than one and the spontaneous probability β is extremely small, changing α will not influence the significant emergence index much. The more interesting case is when $\eta \to 1/\langle k\rangle$. In this case, we get

$$\frac{1-e^{-(1-\eta\langle k\rangle)\theta}}{\theta} \sim \frac{1-[1-(1-\eta\langle k\rangle)\theta]}{\theta} \sim 1-\eta\langle k\rangle \tag{30}$$

Substituting Eq. (30) into Eq. (29), we have

$$\chi(\theta) \sim \frac{(1-\eta\langle k\rangle)\rho(0)}{\beta} - 1, \tag{31}$$

which indicates that social synchrony depth becomes less of a significant emergence as η further increases. This is because, in this case, the actions are very dense on the time axis, i.e., there are actions at almost each time step, and thus their random redistribution is no different than the distribution of the empirical observations.

IV. Discrete model and simulation results

In this part, we conduct four experiments to study the degree to which our analytic approximations derived above agree with simulations of our discrete *SS* model in Sect. III.

We use a generative network model proposed by Catanzaro *et al.* [64] to generate uncorrelated random scale-free networks, with degree distribution satisfying $P(k) \sim k^{-\gamma}$. This network model has three parameters, the network size N, minimum degree $k_{\min}$, and exponent γ.

We carry out the discrete *SS* model simulations using a Monte Carlo method, as follows:

1) *Initialization.* A network G is given. $N\rho(0)$ of nodes in the network are active at time $t = 0$.
2) *Neighboring effect.* At every time t, each active node at time $t-1$ has an equal effect on its neighbors, i.e., they become active at time t with probability α. As a result, the more active neighbors a node has at time $t-1$, the higher its probability of becoming active.

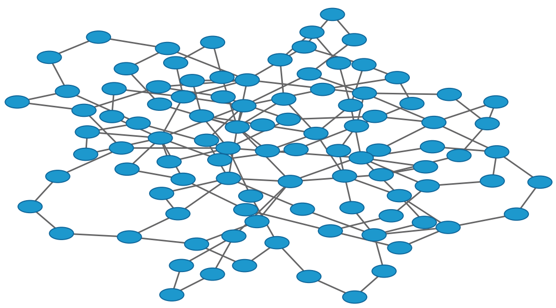

Fig. 2. The scale-free network generated by Catanzaro *et al.*'s model with parameters $N = 100$, $k_{\min} = 2$, and $\gamma = 2$.

3) *Spontaneous activity*. After the above step, at each time t, each inactive node will become active spontaneously with probability β.

Note that the probability that a node is active at time t can be calculated by Eq. (1), which must be smaller than 1 (since both α and β are smaller than 1).

Experiment 1: In this experiment, we focus on *depth of action*. First, we generate a network, as described above, with parameters $N = 100$, $k_{\min} = 2$, and $\gamma = 2$, shown in Fig. 2. Then, we get the analytic and simulated values for $\varphi(T)$ with $\rho(0) = 0.1$ (10 nodes are randomly chosen as active nodes initially) and $T = 100$, for α varying from 0 to 0.25 and β from 0 to 0.1. For the analytic values we use Eq. (14). For the simulated values we implement the above discrete model on the same network for 50 rounds, and then use the mean value of $\varphi(T)$. Note that in the discrete model, at each time step, each node is either active or inactive, and we count the number of active nodes. Then, we aggregate the numbers over $T = 100$ time steps and take those as the simulated value of the depth $\varphi(T)$ in a single round. Recall that almost all our analytic results are based on the assumption that the stimulated probability α is very small. Therefore, we choose α to range from 0 to 0.25, since 0.25 is large enough to show any difference between the analytic and simulated results. On the other hand, we do not have such restriction on β, i.e., the values for β are chosen arbitrarily, and we find that the analytic and simulated results still match well even for $\beta \geq 0.9$ in this experiment. For the second and fourth experiments, as we will see, the analytic and simulated values do not change much as β further increases from 0.1, while for the third experiment, we just set $\beta = 0$ to precisely observe the social synchrony stimulated by a single node.

Fig. 3 (Top) shows the analytic and simulated values of $\varphi(T)$ as functions of the stimulated and spontaneous probabilities α and β. To compare the two, let $Y_A(\alpha, \beta)$ and $Y_S(\alpha, \beta)$ denote the analytic and simulated values for α and β, respectively. The relative difference between $Y_A(\alpha, \beta)$ and $Y_S(\alpha, \beta)$ is given by

$$\epsilon(\alpha, \beta) = \left| \frac{Y_A(\alpha, \beta) - Y_S(\alpha, \beta)}{Y_A(\alpha, \beta)} \right| \times 100\%, \quad (32)$$

and is shown in Fig. 3 (Bottom) as a color map. We also calculate an average relative difference, over all $N_{\alpha,\beta}$ cases

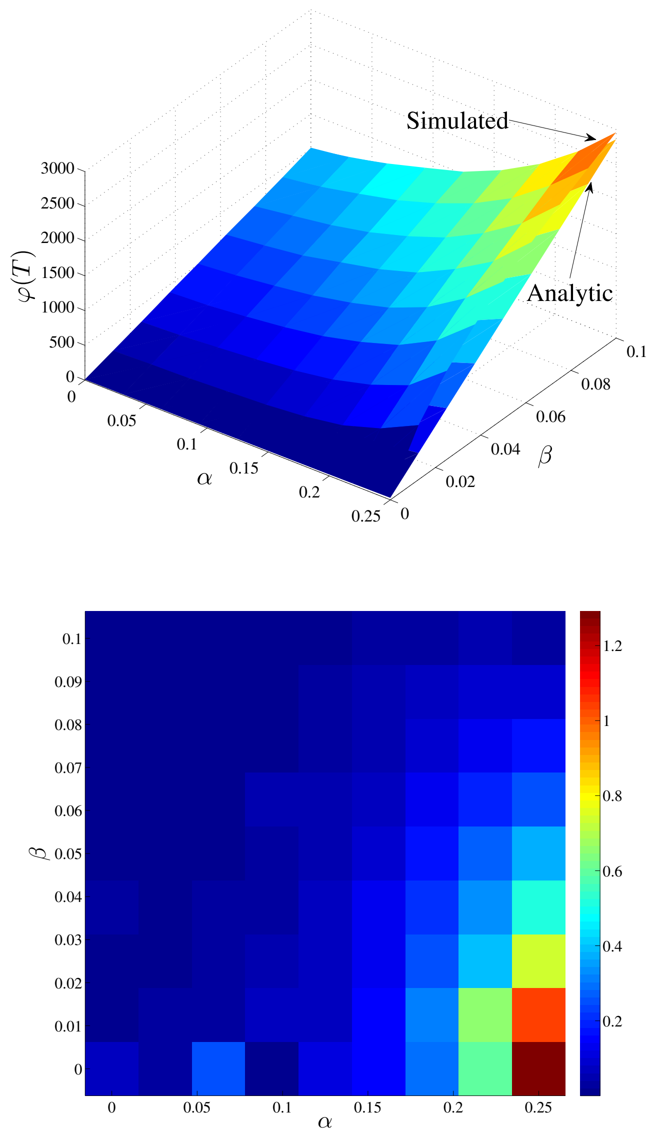

Fig. 3. (Top) The analytic and simulated values of the depth $\varphi(T)$ for a social synchrony as functions of stimulated and spontaneous probabilities α and β. (Bottom) The color map for the relative difference ϵ between analytic and simulated values of depth as a function of α and β.

of different (α, β), as

$$\epsilon = \frac{\sum_{\alpha,\beta} \epsilon(\alpha, \beta)}{N_{\alpha,\beta}}. \quad (33)$$

In this experiment, we have $N_{\alpha,\beta} = 81$ cases, and the average relative difference ϵ equals 13.6%. In Fig. 3 (Bottom), we can see that ϵ is an increasing function of α, while it is a decreasing function of β. This is because, as the spontaneous probability β increases, both the analytic and simulated numbers of actions increase linearly, while the gap between them stays relatively stable, as shown in Fig. 3 (Top). When we consider the 63 cases for $\alpha < 0.2$, we have a smaller $\epsilon = 6.1\%$, which further decreases to $\epsilon = 2.5\%$ when considering the 36 cases for $\alpha < 0.1$. These results indicate that the analytic and simulated results match well in most

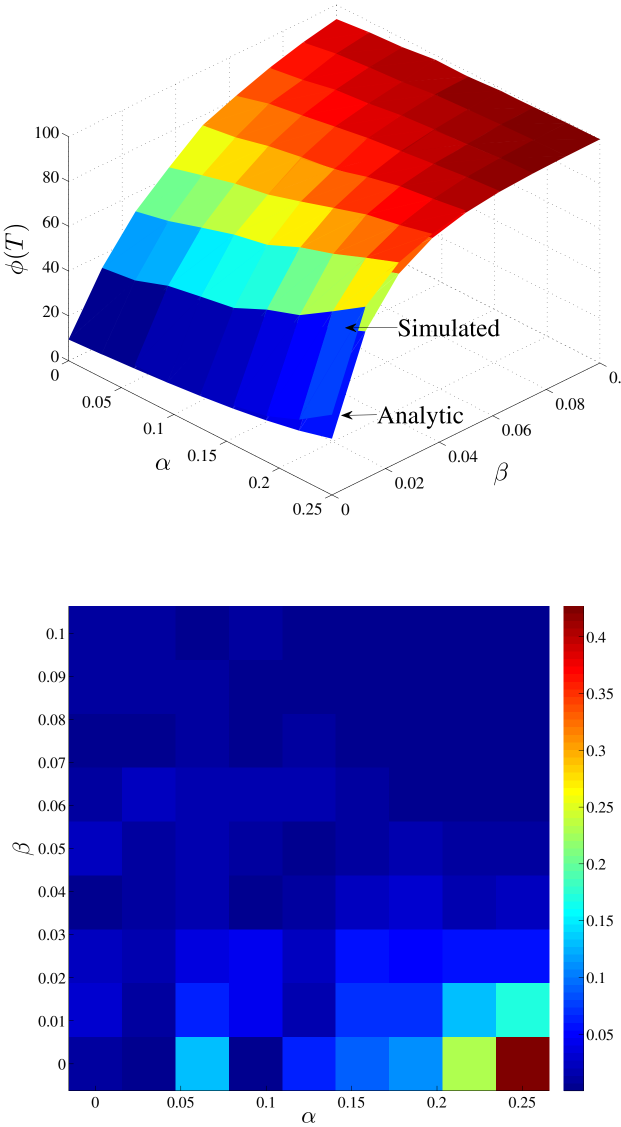

Fig. 4. (Top) The analytic and simulated values of the breadth $\phi(T)$ for a social synchrony as functions of stimulated and spontaneous probabilities α and β. (Bottom) The color map for the relative difference ϵ between analytic and simulated values of breadth as a function of α and β.

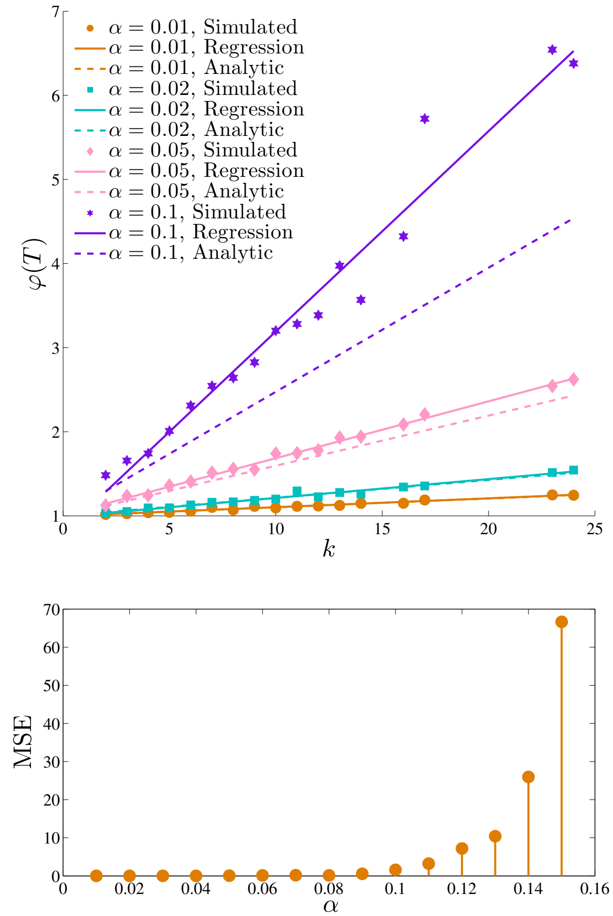

Fig. 5. (Top) The relationship between the depth $\varphi(T)$ for a social synchrony stimulated by a single node and the degree k of that node. We provide the analytic and simulated values, and also the regression line for the simulated values. (Bottom) The Mean Squared Error (MSE) between the analytic and simulated values for the cases of different stimulated probability α.

cases, especially for small α. The difference between them is considerable only when α is relatively large, which violates the assumption used to get the simplification Eq. (3) from Eq. (1).

Experiment 2: In this experiment, we focus on *breadth of impact*. We use the same network as in the first experiment, and then get the analytic and simulated values of $\phi(T)$ with $\rho(0) = 0.1$ and $T = 25$ for α varying from 0 to 0.25 and β from 0 to 0.1. Here we choose a smaller T to make sure that the number of active nodes in this period of time is smaller than the total number of nodes in the network. For the analytic values we use Eq. (21); for the simulated values, similarly to above, we implement the discrete model for 50 rounds and then record the mean value of $\phi(T)$. Note that in the discrete model, a node may be active several times in the period $T = 25$. In such a case, we only count it once. That is, we count the number of nodes that are active at least once in $T = 25$ time steps as the simulated value of the breadth $\phi(T)$ in a single round. The results are shown in Fig. 4. This time, the analytic and simulated values of $\phi(T)$ are very close to each other in almost all cases considered, as shown in Fig. 4 (Top), and the average difference between analytic and simulated values is quite small, i.e., $\epsilon = 3.1\%$, strongly validating the accuracy of our theoretical results on the breadth of impact. In Fig. 4 (Bottom), we can see that ϵ is also a decreasing function of β in this case. This is because, as β increases, both the analytic and simulated numbers for the active nodes become saturated, i.e., their values are close to the network size and, thus, the difference between the models goes to zero.

Experiment 3: In this experiment we study more precisely the relationship between *depth of action* for a social synchrony stimulated by a single node and the degree of that node. We generate a network with widely varied node degrees. To this end, we set the network parameters as $N = 1000$, $k_{\min} = 2$,

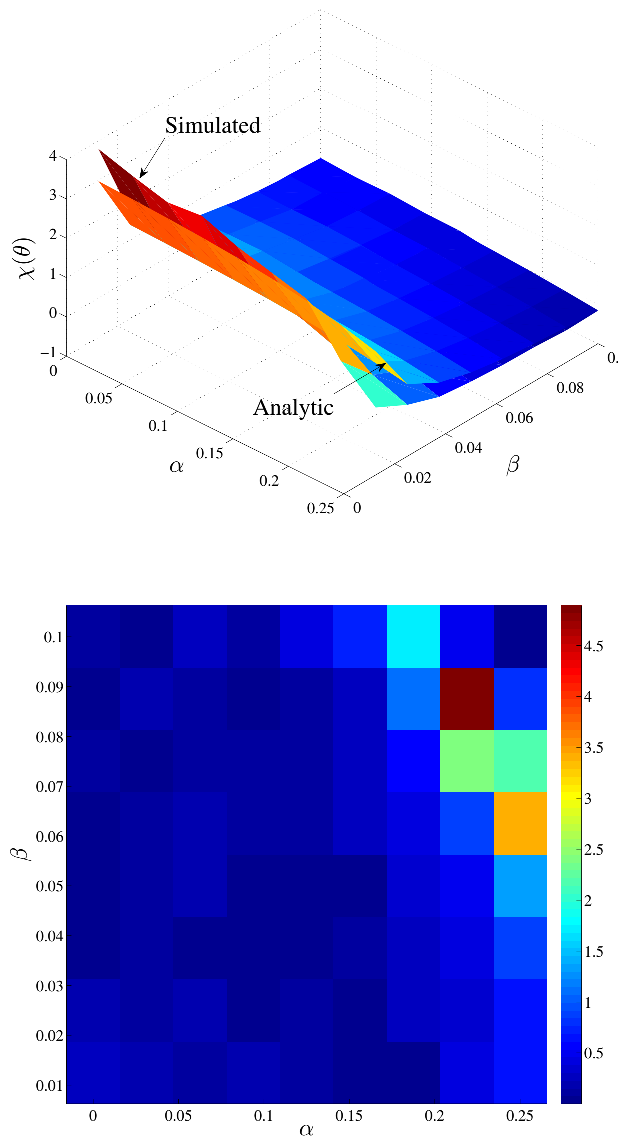

Fig. 6. (Top) The analytic and simulated values of the significant index $\chi(\theta)$ for a social synchrony as functions of stimulated and spontaneous probabilities α and β. (Bottom) The color map for the relative difference ϵ between analytic and simulated values of significant index as a function of α and β.

and $\gamma = 2$. For each degree k in the network, we randomly choose a node of that degree as the only initial active node. For the social synchrony model, we vary the stimulated probability α and set $\beta = 0$ and $T = 25$. For the analytic values, we use Eq. (24). For the simulated values, we use the mean value of $\varphi(T)$ by implementing the discrete model on the same network for 50 rounds. The relationships between $\varphi(T)$ and the degree of the initial active node for $\alpha = 0.01, 0.02, 0.05, 0.1$ are shown in Fig. 5 (Top). Moreover, we also calculate the Mean Square Error (MSE) between the analytic and simulated values for each case, and the relationship between MSE and the stimulated probability α is shown in Fig. 5 (Bottom). We find that 1) the simulated values of different α can be well fitted by lines with different positive slopes, i.e., the social synchrony stimulated by the node of higher degree tends to include larger number of actions, validating that the nodes of larger degree play more important role in driving social synchrony. 2) the MSE is relatively low for small values of the stimulated probability α, i.e., $\alpha < 0.1$. This suggests that the analytic and simulated values in this experiment match well when α is small enough; otherwise, the gap between the two increases fast as α increases, i.e., the average relative difference is $\epsilon = 3.5\%$ when $\alpha = 0.01$, and increases to 28.6% when $\alpha = 0.1$. 3) For large α, the regression lines always have larger slopes than the theoretical lines, indicating that, in this case, the hub nodes of high degree in a network play even more important roles than predicted theoretically.

Experiment 4: In this experiment we focus on identifying surprising burst of actions in a relatively short period of time. The network parameters are set the same as in the first two experiments. Then, for α varying from 0 to 0.25 and β from 0.0125 to 0.1, we get the analytic and simulated values of $\chi(\theta)$, with $\theta = 5$. Note that, based on Eq. (29), when $\beta = 0$, the analytic significance $\chi(\theta)$ goes to infinity, so it is not considered here. For the analytic values, we use Eq. (29). For the simulated values, we set $\rho(0) = 0.2$, $T = 500$, generate the actions based on the discrete model in this relatively long period of time T, and observe the number of actions $\varphi(\theta)$ in the first short period of time, θ. Then, for the null model, we uniformly redistributed these actions randomly, across the period of time T, and calculate the corresponding $\varphi^o(\theta)$. Finally, we use Eq. (28) to get $\chi(\theta)$. We run the discrete model for 50 rounds and then record the mean value. Here, we use a larger $\rho(0)$ in order to generate a larger number of actions in the beginning than in the rest of the time, so that we can observe more significant burst of actions in the short initial period.

We find that, in most cases, both the analytic and simulated values of the significant emergence index $\chi(\theta)$ are larger than 0, the values decrease as the spontaneous probability β increases, and $\chi(\theta)$ becomes slightly negative for relatively large α and β, as shown in Fig. 6 (Top). These findings suggest that most of the time we can identify surprising bursts of actions in a social synchrony, but it gets more difficult to do so when α and β are relatively large, i.e., when there are lots of actions in any given period of time. Since most of the analytic and simulated values are quite small in this case, as expected, the average relative difference ϵ is relatively large, as shown in Fig. 6 (Bottom). It is equal to 43.9% for all the $N_{\alpha,\beta} = 72$ cases, while this value decreases to 11.4% for the 32 cases when $\alpha < 0.1$. However, we can still find the same trends of analytic and simulated values as α and β increase, indicating a good match between them.

Remark 6: Here, we study social synchrony on uncorrelated random scale-free networks, with degree distribution satisfying $P(k) \sim k^{-\gamma}$, proposed by Catanzaro *et al.* [64]. We chose this network model due to the following two reasons: 1) Many real social networks are scale-free [5], [62], [65] and contain a few hub nodes of quite large degree; 2) The networks generated by this model have an ideal topological property, i.e., there is no degree correlation between linked nodes, which is the main assumption to use mean-field theory. As a result, this kind of networks have been extensively adopted to numerically study network dynamics such as reaction-diffusion process [54], [55], and thus are also adopted here since mean-field theory is the key to get Eq. (4) from Eq. (3). Even so, our theoretical and numerical results can be naturally generalized to other kinds of networks, and the only requirement for these networks is that there is no degree correlation between linked nodes. We, thus, extend the first, second, and fourth experiments onto random [66] and small-world [4] networks, and obtain very similar results. For the third experiment, we find that it is necessary to study it on networks of widely varied degrees, in order to derive clearer relationship between the depth for a social synchrony stimulated by a single node and the degree k of that node. For this purpose, neither random nor small-world networks are good candidates.

Remark 7: We chose the four attributes of social synchrony because they are present in many empirical findings. But, they are not necessarily independent from each other. Eqs (20)-(22) indicate that the *breadth of impact* $\phi(T)$ is positively correlated with the *depth of action* $\varphi(T)$, which is validated by the numerical results as well. As we can see in Figs. 3 and 4, both $\varphi(T)$ and the $\phi(T)$ increase as the stimulated probability α and the spontaneous probability β increase. And even the trends of the relative differences between analytic and simulated values for these two characteristics are quit similar, i.e., they are distinct only when α is large while β is small. The only difference between the two is that the *depth of action* can keep increasing as α, β, and T increase, while there is a limitation for the *breadth of impact*, as it cannot be larger than the network size N. The *heterogeneity of role* is also associated with the *depth of action* $\varphi(T)$, as indicated by Eq. (24) and Fig. 5. The difference is that, for the *heterogeneity of role*, we mainly focus on the social synchrony stimulated by a single node (β is thus set to zero) and study the relationship between the depth $\varphi(T)$ and the degree k of that node, rather than the overall *depth of action* as a function of α and β. The *emergence of phenomenon* is defined based on the *depth of action*, as presented in Eq. (28), and is used to measure the significance of the latter in observation versus chance. However, in contrast to the latter, it is a decreasing function of β, as shown in Fig. 6, which is reasonable considering that social synchrony tends to become more of a random phenomenon as the spontaneous probability β increases. The relative difference between the analytic and simulated results for any experiment is notable when α is large, although in practice, when α is likely very small, this should not be the case.

V. Conclusion

In this paper we sought to study social synchrony, as characterized by four typical properties: depth, breadth, heterogeneity, and emergence. We proposed a discrete social synchrony model on complex networks to explain these properties theoretically, assuming that social links play an important role in social synchrony. We performed theoretical and numerical analysis of social synchrony on complex networks, and found that the analytic and simulated results match well under the assumption that the stimulated probability is small enough. These findings validate the effectiveness of our model and may provide useful insights for better understanding of online social synchrony.

For simplicity, in this work we set spontaneous and stimulated probabilities as constants in our social synchrony model, which might not be true in reality. In fact, people may have much more passion to present and spread some interesting issues online when they are relatively new, but this passion abates with time. This suggests that it would be more appropriate to set larger spontaneous and stimulated probabilities in the early stage of the evolution of the model. Our results in the third experiment indicate that, in this case, hub nodes may play even more prominent roles than expected. Future work should consider methods that can incorporate changing trends in α and β over time, in real social systems, and allow for these two parameters to change with time in a coupled fashion. Meanwhile, more experimental studies will also be necessary to understand, and further improve, the models, which should be possible given the large amounts of data coming from large studies of social online media.

Acknowledgment

The authors acknowledge support from the National Natural Science Foundation of China (61572439, 61273212, 11505153, 61503103), and Zhejiang Natural Science Foundation (LQ15A050002).

References

[1] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: a universal concept in nonlinear sciences*. Cambridge University Press, 2003, vol. 12.

[2] R. E. Mirollo and S. H. Strogatz, "Synchronization of pulse-coupled biological oscillators," *SIAM Journal on Applied Mathematics*, vol. 50, no. 6, pp. 1645–1662, 1990.

[3] P. Fries, "A mechanism for cognitive dynamics: neuronal communication through neuronal coherence," *Trends in Cognitive Sciences*, vol. 9, no. 10, pp. 474–480, 2005.

[4] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[5] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[6] Y. Yang and J. Cao, "Exponential synchronization of the complex dynamical networks with a coupling delay and impulsive effects," *Nonlinear Analysis: Real World Applications*, vol. 11, no. 3, pp. 1650–1659, 2010.

[7] X. Yang and J. Lu, "Finite-time synchronization of coupled networks with markovian topology and impulsive effects," *IEEE Transactions on Automatic Control*, vol. 61, no. 8, pp. 2256–2261, 2016.

[8] S. Ahmadizadeh, D. Nešić, D. R. Freestone, and D. B. Grayden, "On synchronization of networks of wilson–cowan oscillators with diffusive coupling," *Automatica*, vol. 71, pp. 169–178, 2016.

[9] H. Hong, M.-Y. Choi, and B. J. Kim, "Synchronization on small-world networks," *Physical Review E*, vol. 65, no. 2, p. 026139, 2002.

[10] X. F. Wang and G. Chen, "Synchronization in small-world dynamical networks," *International Journal of Bifurcation and Chaos*, vol. 12, no. 01, pp. 187–192, 2002.

[11] J. Lu, X. Yu, G. Chen, and D. Cheng, "Characterizing the synchronizability of small-world dynamical networks," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 51, no. 4, pp. 787–796, 2004.

[12] C. Fu and X. Wang, "Network growth under the constraint of synchronization stability," *Physical Review E*, vol. 83, no. 6, p. 066101, 2011.

[13] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.

[14] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520–1533, 2004.

[15] W. Yu, G. Chen, M. Cao, and W. Ren, "Delay-induced consensus and quasi-consensus in multi-agent dynamical systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 10, pp. 2679–2687, 2013.

[16] Y. Chen, J. Lu, X. Yu, and D. J. Hill, "Multi-agent systems with dynamical topologies: Consensus and applications," *IEEE Circuits and Systems magazine*, vol. 13, no. 3, pp. 21–34, 2013.

[17] Z. Lin, L. Wang, Z. Han, and M. Fu, "Distributed formation control of multi-agent systems using complex laplacian," *IEEE Transactions on Automatic Control*, vol. 59, no. 7, pp. 1765–1777, 2014.

[18] X. Dong, J. Xi, G. Lu, and Y. Zhong, "Formation control for high-order linear time-invariant multiagent systems with time delays," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 3, pp. 232–240, 2014.

[19] G. Wen, Z. Duan, G. Chen, and W. Yu, "Consensus tracking of multi-agent systems with lipschitz-type node dynamics and switching topologies," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 2, pp. 499–511, 2014.

[20] H. Su, Z. Rong, M. Z. Chen, X. Wang, G. Chen, and H. Wang, "Decentralized adaptive pinning control for cluster synchronization of complex dynamical networks," *IEEE Transactions on Cybernetics*, vol. 43, no. 1, pp. 394–399, 2013.

[21] Q. Xuan and V. Filkov, "Synchrony in social groups and its benefits," in *Handbook of Human Computation*. Springer, 2013, pp. 791–802.

[22] J. T. Emlen, "Flocking behavior in birds," *The Auk*, vol. 69, no. 2, pp. 160–170, 1952.

[23] E. Shaw, "Schooling fishes: the school, a truly egalitarian form of organization in which all members of the group are alike in influence, offers substantial benefits to its participants," *American Scientist*, vol. 66, no. 2, pp. 166–175, 1978.

[24] F. Alderisio, B. G. Bardy, and M. Bernardo, "Entrainment and synchronization in networks of rayleigh–van der pol oscillators with diffusive and haken–kelso–bunz couplings," *Biological Cybernetics*, vol. 110, no. 2-3, pp. 151–169, 2016.

[25] E. Codrons, N. F. Bernardi, M. Vandoni, and L. Bernardi, "Spontaneous group synchronization of movements and respiratory rhythms," *PLoS One*, vol. 9, no. 9, p. e107538, 2014.

[26] F. Alderisio, G. Fiore, R. N. Salesse, B. G. Bardy, and M. di Bernardo, "Interaction patterns and individual dynamics shape the way we move in synchrony," *arXiv preprint arXiv:1607.02175*, 2016.

[27] T. Frank and M. Richardson, "On a test statistic for the kuramoto order parameter of synchronization: An illustration for group synchronization during rocking chairs," *Physica D: Nonlinear Phenomena*, vol. 239, no. 23, pp. 2084–2092, 2010.

[28] R. Lukeman, Y.-X. Li, and L. Edelstein-Keshet, "Inferring individual rules from collective behavior," *Proceedings of the National Academy of Sciences*, vol. 107, no. 28, pp. 12 576–12 580, 2010.

[29] M. Nagy, Z. Ákos, D. Biro, and T. Vicsek, "Hierarchical group dynamics in pigeon flocks," *Nature*, vol. 464, no. 7290, pp. 890–893, 2010.

[30] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[31] Q. Xuan, A. Okano, P. Devanbu, and V. Filkov, "Focus-shifting patterns of oss developers and their congruence with call graphs," in *Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering*. ACM, 2014, pp. 401–412.

[32] M. De Choudhury, H. Sundaram, A. John, and D. D. Seligmann, "Social synchrony: Predicting mimicry of user actions in online social media," in *Proceedings of the 2009 International Conference on Computational Science and Engineering*. IEEE, 2009, pp. 151–158.

[33] Q. Xuan and V. Filkov, "Building it together: Synchronous development in oss," in *Proceedings of the 36th International Conference on Software Engineering*. ACM, 2014, pp. 222–233.

[34] L. Jin, Y. Chen, T. Wang, P. Hui, and A. V. Vasilakos, "Understanding user behavior in online social networks: A survey," *IEEE Communications Magazine*, vol. 51, no. 9, pp. 144–150, 2013.

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

[36] Q. Xuan, H. Fang, C. Fu, and V. Filkov, "Temporal motifs reveal collaboration patterns in online task-oriented networks," *Physical Review E*, vol. 91, no. 5, p. 052813, 2015.

[37] J. Zhao, J. Wu, X. Feng, H. Xiong, and K. Xu, "Information propagation in online social networks: a tie-strength perspective," *Knowledge and Information Systems*, vol. 32, no. 3, pp. 589–608, 2012.

[38] F. Alderisio, M. Lombardi, G. Fiore, and M. di Bernardo, "Study of movement coordination in human ensembles via a novel computer-based set-up," *arXiv preprint arXiv:1608.04652*, 2016.

[39] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*. Springer, 2009, pp. 1–4.

[40] H.-F. Zhang, J.-R. Xie, M. Tang, and Y.-C. Lai, "Suppression of epidemic spreading in complex networks by local information based behavioral responses," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 24, no. 4, p. 043106, 2014.

[41] W. Wang, M. Tang, H.-F. Zhang, H. Gao, Y. Do, and Z.-H. Liu, "Epidemic spreading on complex networks with general degree and weight distributions," *Physical Review E*, vol. 90, no. 4, p. 042803, 2014.

[42] Q.-H. Liu, W. Wang, M. Tang, and H.-F. Zhang, "Impacts of complex behavioral responses on asymmetric interacting spreading dynamics in multiplex networks," *Scientific Reports*, vol. 6, no. 06, p. 25617, 2016.

[43] P.-Y. Chen, S.-M. Cheng, and K.-C. Chen, "Optimal control of epidemic information dissemination over networks," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2316–2328, 2014.

[44] M. Cha, F. Benevenuto, H. Haddadi, and K. Gummadi, "The world of connections and information flow in twitter," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 42, no. 4, pp. 991–998, 2012.

[45] L.-J. Kao and Y.-P. Huang, "Mining influential users in social network," in *Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2015, pp. 1209–1214.

[46] A. V. Sathanur, M. Sui, V. Jandhyala, M. D. Tyka, and N. A. Deflaux, "Simulating context-driven activity cascades in online social networks on the google exacycle platform," in *Proceedings of the 2014 IEEE International Conference on Systems, Man and Cybernetics*. IEEE, 2014, pp. 1388–1391.

[47] A. Anagnostopoulos, R. Kumar, and M. Mahdian, "Influence and correlation in social networks," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2008, pp. 7–15.

[48] Y. Fernandess and D. Malkhi, "On spreading recommendations via social gossip," in *Proceedings of the 20th Annual Symposium on Parallelism in Algorithms and Architectures*. ACM, 2008, pp. 91–97.

[49] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou, "Recommender systems," *Physics Reports*, vol. 519, no. 1, pp. 1–49, 2012.

[50] G. Huang, J. Cao, G. Wang, and Y. Qu, "The strength of the minority," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 18, pp. 4665–4672, 2008.

[51] G. Huang, J. Cao, and Y. Qu, "The minority's success under majority rule," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 18, pp. 3911–3916, 2009.

[52] C. Qian, J. Cao, J. Lu, and J. Kurths, "Adaptive bridge control strategy for opinion evolution on social networks," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 21, no. 2, p. 025116, 2011.

[53] Q. Xuan, P. Devanbu, and V. Filkov, "Converging work-talk patterns in online task-oriented communities," *PLoS One*, vol. 11, no. 5, p. e0154324, 2016.

[54] Q. Xuan, F. Du, T.-J. Wu, and G. Chen, "Emergence of heterogeneous structures in chemical reaction-diffusion networks," *Physical Review E*, vol. 82, no. 4, p. 046116, 2010.

[55] Q. Xuan, F. Du, H. Dong, L. Yu, and G. Chen, "Structural control of reaction-diffusion networks," *Physical Review E*, vol. 84, no. 3, p. 036101, 2011.

[56] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Reviews of Modern Physics*, vol. 87, no. 3, p. 925, 2015.

[57] V. Colizza, R. Pastor-Satorras, and A. Vespignani, "Reaction–diffusion processes and metapopulation models in heterogeneous networks," *Nature Physics*, vol. 3, no. 4, pp. 276–282, 2007.

[58] Q. Xuan and T.-J. Wu, "Node matching between complex networks," *Physical Review E*, vol. 80, no. 2, p. 026103, 2009.

[59] Q. Xuan, C. Fu, and L. Yu, "Ranking developer candidates by social links," *Advances in Complex Systems*, vol. 17, no. 07n08, p. 1550005, 2014.

[60] L. Lü and T. Zhou, "Link prediction in weighted networks: The role of weak ties," *Europhysics Letters*, vol. 89, no. 1, p. 18001, 2010.

[61] L. Lü, T. Zhou, Q.-M. Zhang, and H. E. Stanley, "The h-index of a network node and its relation to degree and coreness," *Nature Communications*, vol. 7, p. 10168, 2016.

[62] Q. Xuan, F. Du, and T.-J. Wu, “Empirical analysis of internet telephone network: From user id to phone,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 19, no. 2, p. 023101, 2009.

[63] Q. Xuan, M. Gharehyazie, P. T. Devanbu, and V. Filkov, “Measuring the effect of social communications on individual working rhythms: A case study of open source software,” in *Proceedings of the 2012 International Conference on Social Informatics*. IEEE, 2012, pp. 78–85.

[64] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, “Generation of uncorrelated random scale-free networks,” *Physical Review E*, vol. 71, no. 2, p. 027103, 2005.

[65] A.-L. Barabási, “Scale-free networks: a decade and beyond,” *Science*, vol. 325, no. 5939, pp. 412–413, 2009.

[66] P. Erdös and A. Rényi, “On random graphs i,” *Publ. Math. Debrecen*, vol. 6, pp. 290–297, 1959.