

Stylized facts in social networks: Community-based static modeling

Hang-Hyun Jo,^{1,2,3} Yohsuke Murase,⁴ János Török,^{5,6} János Kertész,^{6,5,3} and Kimmo Kaski³

¹*Asia Pacific Center for Theoretical Physics, Pohang 37673, Republic of Korea*

²*Department of Physics, Pohang University of Science and Technology, Pohang 37673, Republic of Korea*

³*Department of Computer Science, Aalto University School of Science, Espoo FI-00076, Finland*

⁴*RIKEN Advanced Institute for Computational Science, Kobe, Hyogo 650-0047, Japan*

⁵*Department of Theoretical Physics, Budapest University of Technology and Economics, Budapest H-1111, Hungary*

⁶*Center for Network Science, Central European University, Budapest H-1051, Hungary*

(Dated: August 1, 2018)

The past analyses of datasets of social networks have enabled us to make empirical findings of a number of aspects of human society, which are commonly featured as stylized facts of social networks, such as broad distributions of network quantities, existence of communities, assortative mixing, and intensity-topology correlations. Since the understanding of the structure of these complex social networks is far from complete, for deeper insight into human society more comprehensive datasets and modeling of the stylized facts are needed. Although the existing dynamical and static models can generate some stylized facts, here we take an alternative approach by devising a community-based static model with heterogeneous community sizes and larger communities having smaller link density and weight. With these few assumptions we are able to generate realistic social networks that show most stylized facts for a wide range of parameters, as demonstrated numerically and analytically. Since our community-based static model is simple to implement and easily scalable, it can be used as a reference system, benchmark, or testbed for further applications.

I. INTRODUCTION

Characterizing the social networks is of crucial importance to understand various collective dynamics taking place in them [1–3], as exemplified by disease spreading and diffusion of innovation and opinions. In recent years, the characterization of social networks in the unprecedented detail has become possible because of the availability of a number of large-scale digital datasets, e.g., face-to-face interactions [4–6], emails [7, 8], mobile phone communication [9, 10], online forums [11, 12], Social Networking Services (SNSs) like Facebook [13] and Twitter [14], and even massive multiplayer online games [15, 16]. However, these datasets capture only a part of the entire social network, implying that any conclusions derived from such datasets cannot be simply extrapolated to the whole society. Here the entire social network indicates a comprehensive picture of human social relationships with complex community structure due to today’s multiple communication channels, and can be called a multi-channel weighted social (MWS) network. This raises a series of questions: How can one translate conclusions from partial datasets to the MWS network? More importantly, what does the MWS network look like? The first question has been investigated in terms of sampling biases [17–20], while the second question is largely unexplored mainly due to the lack of comprehensive datasets.

Characteristics of the MWS network are expected to be partially reflected in the empirical findings from some aspects of the network. By collecting such findings from diverse sources, we find several commonly observed features or *stylized facts* of social networks [21–23]. These include broad distributions of local network quantities [1, 24], community structure [25], ho-

mophily [26, 27], and intensity-topology correlations [9], etc. More recently, geographical and/or demographic information of social networks have also been explored [28–30], which are beyond the scope of this paper. One can find the previous efforts of modeling social networks: The global picture for social networks has been described by the Granovetter’s hypothesis of “strength of weak ties” [31], indicating that communities of strongly connected nodes are weakly connected to each other. This picture has been empirically confirmed [9, 32] and subsequently produced with computational modeling by considering cyclic and focal closure mechanisms in tie formation [22, 33, 34]. However, it was recently suggested that communities could be overlapping [35, 36] in contrast to the picture of separate communities. This overlapping behavior is mostly due to the multilayer nature of social networks [37, 38], in which each layer may correspond to a certain type of human relationship or context. This means that an individual can belong to one community in one layer but simultaneously to another community in another layer. Accordingly, dynamical models for multilayer, overlapping community structure have been introduced, while reproducing other stylized facts for local network quantities [39]. There are also several other dynamical models that partially reproduce stylized facts [40–42].

As many models mentioned above are dynamic and evolutionary in nature, they tend to take considerable amount of computational time. For relatively simpler and easier implementation, we take an alternative approach of static modeling to reproduce the stylized facts in social networks. For our model, we randomly assign a number of communities to a given set of isolated nodes using a few reasonable assumptions such that the community size is heterogeneous, and larger communities are as-

signed with smaller link density and smaller link weight. As we assign communities by hand rather than grow the network by means of some link formation mechanisms, our model can be called static. We also remark that the community size distribution is an input rather than an output of our model, although it is one of stylized facts. With the above mentioned few assumptions about communities, apparently realistic social network structures are generated showing most stylized facts for a wide range of the parameter space. Furthermore, thanks to the random nature of assigning communities to nodes, we can to some extent analytically calculate various local network quantities, e.g., for the assortative mixing, local clustering coefficient, and neighborhood overlap.

This static modeling approach of ours is comparable to other static modeling studies, which can be classified, but not exclusively, into four categories: (i) Erdős-Rényi (ER) random graphs, (ii) configuration models (CMs), (iii) stochastic blockmodels (SBMs), and (iv) exponential random graph models (ERGMs). The ER random graphs [43] are the simplest kind of static models, and its variants have been studied, such as graphons [44, 45], weighted random graphs [46], or ER random graphs with community structure [47]. In the simplest form of CMs a binary network is constructed only by using the pre-determined degree sequence of nodes, without any other correlations [48]. It has been extended for containing the arbitrary distributions of subgraphs [49], to weighted networks [50, 51], or to networks with overlapping community structure [52] or with hierarchical community structure [53]. Next, the SBM was originally suggested for the community structure, characterized by a matrix consisting of the linking probabilities within communities and between communities [54–56]. As the traditional SBMs are not comparable with the empirical degree heterogeneity, the degree-corrected SBMs, by which the degree heterogeneity can be properly considered, have been studied [57, 58]. The SBMs have also been extended to incorporate the overlapping communities by considering the mixed membership [59, 60] or to the weighted networks, for which see Ref. [61] and references therein. Finally, the family of ERGMs has been extensively studied in social sciences [44, 62] as well as in terms of statistical mechanics [63]. Here an ensemble of networks with given network features is considered according to the probability in the form of Boltzmann factor. The ERGMs have been extended for weighted networks [64] or for networks with community structure [58].

In our work, we will be exploring a different static modeling approach by explicitly considering the communities with various sizes, linking probabilities, and link weights. This way we arrive at a simple and scalable static model, which may serve as a reference system, benchmark, or testbed for further applications.

Our paper is organized as follows: In Section II, we summarize the observed stylized facts for social networks from diverse sources. Then we introduce the community-based static model in Section III. In Section IV, by per-

forming large-scale numerical simulations, we find a wide range of the parameter space in which the stylized facts are reproduced. In Section V, we present the analytical results for local network quantities. Finally, we conclude our work in Section VI.

II. STYLIZED FACTS

In Table I we present a summary of the commonly observed features or *stylized facts* in many digital datasets for social networks. These include the statistics of local network quantities and results for the global structure, both of which can be either topological or intensity-related.

Let us first consider topological quantities. The degree k of a node is the number of its neighbors. Degree distributions $P(k)$ in most datasets are found to be broad and overall decreasing [1, 12, 13, 24, 65]. This implies that the most probable degree or the mode of $P(k)$, denoted by m_k , is of the order of 1, leading to the fact that m_k is much smaller than the average degree $\langle k \rangle$. This stylized fact is however clearly not consistent with our common sense that it is unlikely to find a majority of individuals with only one or few relationships in a society. This discrepancy can be explained by the channel selection mechanism [20]: Most datasets for social networks represent only one communication channel or even a part of it, while social interactions generally take place on multiple communication channels, namely, the multiplexity of social networks [23, 37, 38]. Since individuals with low degree in one channel can have high degree in another channel, by considering the multi-channel feature of social networks, $P(k)$ is expected to have a peak at the degree larger than 1, implying that m_k is of the comparable order of $\langle k \rangle$ [20, 66]. Such $P(k)$ can be called overall peaked, see Fig. 1(a) in Ref. [67] for example. This does not necessarily exclude the broadness of $P(k)$.

Homophily, the tendency that alike people are attracted to each other, is one of the governing principles of social network formation [26]. One manifestation of homophily is degree assortativity, i.e., the tendency that high degree nodes are linked together. This correlation has been quantified in terms of the assortativity coefficient ρ_{kk} , which is a Pearson correlation coefficient (PCC) between degrees of neighboring nodes [27]. Many social networks are found to be assortative with $\rho_{kk} \approx 0.09\text{--}0.4$ [68]. This is also shown by the increasing behavior of the average degree of neighbors for nodes with degree k [13, 24]. This quantity is denoted by $k_{\text{nn}}(k)$ and is defined as follows:

$$k_{\text{nn}}(k) \equiv \langle k_{i,\text{nn}} \rangle_{\{i|k_i=k\}}, \quad (1)$$

$$k_{i,\text{nn}} \equiv \frac{1}{k_i} \sum_{j \in \Lambda_i} k_j, \quad (2)$$

where Λ_i denotes the set of i 's neighbors and $k_i = |\Lambda_i|$ is the node i 's degree.

TABLE I. Stylized facts derived from various datasets with the expected behaviors for the MWS network [22, 23]. The symbol \nearrow (\searrow) implies that the overall trend of a quantity is monotonically increasing (decreasing) according to its argument. The initially increasing and then decreasing behavior is denoted by $\nearrow\searrow$. For the Granovetterian community structure, see the main text for the details.

Category	Property or measure	Stylized fact (expectation)
Topological	Degree distribution, $P(k)$	\searrow ($\nearrow\searrow$)
	Average degree of neighbors as a function of degree, $k_{\text{nn}}(k)$	\nearrow
	Local clustering coefficient as a function of degree, $c(k)$	\searrow
	Community size distribution, $P(g)$	\searrow
Intensity-related	Strength distribution, $P(s)$	\searrow ($\nearrow\searrow$)
	Weight distribution, $P(w)$	\searrow
	Strength as a function of degree, $s(k)$	\nearrow
	Neighborhood overlap as a function of weight, $o(w)$	\nearrow
	Granovetterian community structure, Δf_c	> 0

High clustering is evident in social networks, implying that your neighbor's neighbor is likely to be also your neighbor. For a node i , the local clustering coefficient c_i is defined as the number of links between i 's neighbors, denoted by e_i , divided by the possible maximal number of links between them, i.e., $\frac{k_i(k_i-1)}{2}$, as follows:

$$c_i \equiv \frac{2e_i}{k_i(k_i-1)}. \quad (3)$$

Then one can measure the average local clustering coefficient for nodes with degree k as

$$c(k) \equiv \langle c_i \rangle_{\{i|k_i=k\}}. \quad (4)$$

The quantity $c(k)$ is found to be a decreasing function of k [13, 15, 24, 69]. This finding can be explained by considering the case that the effect of making new neighbors, corresponding to $\sim k_i^2$ in Eq. (3), is typically stronger than that of finding new links between neighbors, in relation to e_i in Eq. (3). For example, if every new neighbor of a node i creates a new link to one of node i 's existing neighbors, then $e_i \sim k_i$, leading to $c(k) \sim k^{-1}$. This behavior can be measured in terms of the PCC between c_i and k_i , which is denoted hereafter as ρ_{ck} .

At the larger scale, social networks have rich community structure: Nodes in communities are densely connected, while nodes between different communities are sparsely connected [25]. It has been shown that the community size distribution $P(g)$ has a heavy tail or power-law form, e.g., as evidenced in Refs. [11, 35, 36, 70, 71]. We remark that in our model $P(g)$ will be used as an input rather than as an output of our model.

Next, we consider the intensity-related quantities, i.e., strength s and weight w . The weight of a link quantifies the interaction activity between two nodes [72], e.g., the frequency of contact in communication. The strength of a node, denoting the activity of the node, has been defined as the sum of weights of links involving the node:

$$s_i \equiv \sum_{j \in \Lambda_i} w_{ij}, \quad (5)$$

where w_{ij} is the weight of the link ij . The distributions of these quantities, $P(s)$ and $P(w)$, are also found to be broad and overall decreasing, implying that both individual and interaction activities are heterogeneous [24]. The overall decreasing behavior of $P(w)$ can be interpreted as the prevalence of weak links [73] in social networks.

In addition, the average strength of nodes as a function of the degree, $s(k)$, is found to be overall increasing [24, 72]. Here $s(k)$ is defined as

$$s(k) \equiv \langle s_i \rangle_{\{i|k_i=k\}}. \quad (6)$$

This behavior can be measured in terms of the PCC between s_i and k_i , denoted by ρ_{sk} . Combining the peaked $P(k)$ and the increasing $s(k)$, $P(s)$ is expected to be also peaked. The overall peaked $P(s)$ can be tested in terms of the mode of $P(s)$, denoted by m_s .

These intensities are correlated with topological properties, which can be called intensity-topology correlation or weight-topology correlation as used in Ref. [9]. A link-level consequence of weight-topology correlation can be measured by the average neighborhood overlap for links with weight w , denoted by $o(w)$. The neighborhood overlap of a link is the fraction of common neighbors of neighboring nodes, say i and j , among all neighbors of those nodes:

$$o(w) \equiv \langle o_{ij} \rangle_{\{ij|w_{ij}=w\}}, \quad (7)$$

$$o_{ij} \equiv \frac{e_{ij}}{k_i+k_j-2-e_{ij}}, \quad (8)$$

where e_{ij} denotes the number of common neighbors of nodes i and j , whose degrees are k_i and k_j , respectively. It has been empirically found that the stronger links tend to show larger neighborhood overlap [9, 15, 69, 74–76], implying that closer friends tend to have more common friends. We also note that $o(w)$ begins to decrease for very large w in some cases [9, 69], and that the overall decreasing $o(w)$ is found in some collaboration networks [69, 77]. The overall increasing behavior of $o(w)$ can be measured in terms of the PCC between o_{ij} and w_{ij} , and it is denoted by ρ_{ow} .

The weight-topology correlation emerges at the global scale such that communities of strongly connected nodes are weakly connected to each other [9, 32]. This is a consequence of the Granovetter’s hypothesis of “strength of weak ties” [31]. As the Granovetterian structure is maintained by weak links, the network will be disintegrated by removing weak links rather than by removing strong links. Precisely, a link percolation analysis can be applied such that links are removed one by one either from the weakest links (ascending link removal) or from the strongest links (descending link removal) to see when the network gets disintegrated. In both cases, we may observe the percolation transition at some value of the fraction of removed links f . If the percolation threshold for ascending (descending) link removal is denoted by f_c^a (f_c^d), we expect that

$$\Delta f_c \equiv f_c^d - f_c^a \quad (9)$$

is significantly larger than 0 as shown in Ref. [39]. It is because the Granovetterian structure will be disintegrated earlier for ascending link removal than for descending link removal. Note that $\Delta f_c > 0$ does not guarantee the Granovetterian structure [46]. The percolation threshold can be determined by measuring the fraction of the largest connected component (LCC) R_{LCC} and the susceptibility χ as functions of f . Here the susceptibility for the network size N is defined as $\chi = \frac{1}{N} \sum_s n_s s^2$, where n_s is the number of connected components of size s and the summation is taken over all connected components but the LCC. At the percolation threshold f_c , R_{LCC} vanishes and χ diverges in the thermodynamic limit [78].

Finally, we remark that these stylized facts have been deduced mostly from datasets of single communication channels, while here we are interested in the multi-channel weighted social (MWS) network representing a multiplicity of communication channels. Empirical findings in single channel datasets may reflect some properties of the reality, but they may also introduce biases [20]. Thus, our aim in this paper is to reproduce, in addition to the stylized facts, the expected peaked behavior of $P(k)$ and $P(s)$ for the MWS network, as listed in Table I with expected behavior in parenthesis.

III. MODEL

In order to devise a community-based static model for stylized facts in social networks, let us consider an undirected weighted network of N nodes and C communities. For each community, the community size g is drawn from a distribution $P(g)$. The minimum size of the community is set as $g_0 = 3$, i.e., a triangle. The maximum size of the community g_{\max} is set to be 10^3 that is sufficiently large but not too large for being realistic. Then g different nodes are randomly chosen to form a community, where links between them are created with linking probability $p(g)$. Here $p(g)$ is a decreasing function of g to represent the tendency of link density being sparser for

larger communities. Each of these links is assumed to have a positive weight randomly drawn from a distribution $P_g(w)$ with a characteristic weight $w(g)$. Here $w(g)$ is a decreasing function of g to represent the tendency of weaker ties in larger communities. For these reasons, we assume the following functional forms in the range of $g_0 \leq g \leq g_{\max}$:

$$P(g) \equiv A g^{-\alpha}, \quad (10)$$

$$p(g) \equiv \left(\frac{g_0}{g}\right)^\beta, \quad (11)$$

$$P_g(w) \equiv \frac{\theta[w(g)-w]}{w(g)}, \quad w(g) \equiv w_0 \left(\frac{g_0}{g}\right)^\gamma \quad (12)$$

with non-negative exponents α , β , and γ . Here $A \equiv (\sum_{g=g_0}^{g_{\max}} g^{-\alpha})^{-1}$ denotes a normalization constant, and $\theta(\cdot)$ is the Heaviside step function. We take a power-law form for $P(g)$ based on the empirical findings in Refs. [11, 35, 36, 70]. The power-law form for $p(g)$ is also inspired by the observations [11]. Finally, we take the uniform distribution for $P_g(w)$ in the range of $(0, w(g)]$ as no evidence is known for its shape. Other candidates for $P_g(w)$ will be shortly discussed in Appendix A. These settings immediately imply that the average number of neighbors due to the community of size g is $(g-1)p(g)$, i.e., of the order of $g^{1-\beta}$. Since a membership to the larger community could imply more ties created, it is assumed that $\beta \leq 1$. Further, the sum of weights due to those neighbors in the community of size g scales with $g^{1-\beta-\gamma}$. This sum decreases with g for sufficiently large γ , meaning that people may spend less time in communication with the members of the larger communities. See Fig. 1 for the schematic diagram.

Since in social networks each node may belong to multiple communities, the network has an overlapping community structure and this is assured here by construction. Moreover, such network can be interpreted in the frame of a multiplex network [37, 38]. In our model each *layer* indexed by g is defined as a set of links created in communities of the same size g [79]. Then, a pair of nodes may be connected by multiple links when they belong to multiple communities, irrespective of whether those communities are in the same layer or not. However, we reduce multiplex weighted networks generated by our model to single layer weighted networks by assigning a unique weight to each pair of nodes. Here we use the rule of weight aggregation such that a weight w_{ij} for a link ij is given by

$$w_{ij} \equiv \sum_{g \in G_{ij}} w_{ij,g}, \quad (13)$$

where G_{ij} denotes a set of gs for communities in which the link ij is created, and $w_{ij,g}$ is the weight of link ij created in the layer g . Note that alternative rules for weight aggregation can be used, e.g., by taking a maximum weight from a set of $w_{ij,g}$.

We remark that it is very unlikely for a pair of nodes to belong to multiple communities, which is indeed asymptotically the case as shown in Appendix B. Hence, links

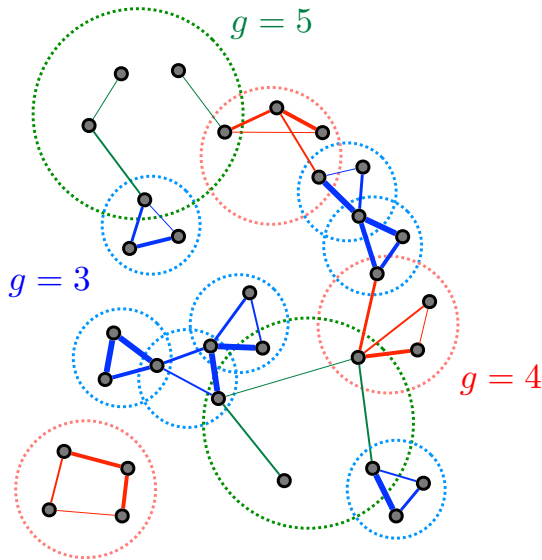


FIG. 1. (Color online) Schematic diagram of the community-based static model with overlapping communities of various sizes with $g = 3, 4,$ and 5 . The larger dotted circle indicates the larger community size, and thicker lines indicate larger link weights.

can be divided into exclusive sets according to the community size g in which the link is created, implying that $|G_{ij}| = 1$ for all links.

IV. NUMERICAL RESULTS

For numerical simulations, we generate networks of size $N = 3 \cdot 10^4$ with $g_0 = 3$ and $g_{\max} = 10^3$. We also set $w_0 = 1$ without loss of generality. We begin with N isolated nodes, to which communities are added sequentially until the average degree reaches a predetermined value, e.g., $\langle k \rangle = 100$. This in turn determines the total number of communities, C . We first study the effects of α and β on the topological properties of the generated networks to find the parameter region for reproducing the stylized facts. Then, for a fixed value of β within the region for the topological stylized facts, we study the effects of α and γ on the intensity-related properties of generated networks.

A. Topological properties

We obtain the simulation results of topological quantities, i.e., m_k , ρ_{kk} , and ρ_{ck} , for the wide range of α and β . In Fig. 2(a), m_k is found to overall increase with α and β . As $\langle k \rangle = 100$ in the simulations, we take a threshold value for m_k as 10 to distinguish the region of overall decreasing $P(k)$ for small α and β from that of overall peaked $P(k)$ for large α and β . This threshold value is

arbitrary yet reasonable for the qualitative description. In Fig. 2(b), we find that $\rho_{kk} \geq 0$ for the entire range of the parameter space, and that ρ_{kk} overall decreases with α and β . Here we take 0.05 as the threshold value to separate the assortative region for small α and β from the uncorrelated region for large α and β . Finally, we observe the non-monotonic behavior of ρ_{ck} according to α and β , as shown in Fig. 2(c). For small β , there exists an intermediate region of α with $\rho_{ck} > 0$. As β increases, this region becomes narrower and finally disappears. Summarizing the results, we find the wide region of α spanning over $(2, 4.5)$ and of β spanning over $(0.1, 0.6)$, where all topological stylized facts are reproduced, i.e., relatively large m_k and ρ_{kk} as well as negative ρ_{ck} . This region is depicted as shaded in Fig. 2(d).

In order to understand these results, we remind that the larger value of α leads to more frequent triangles as $g_0 = 3$, and that the larger value of β leads to sparser communities, especially for relatively large communities. We first consider the limiting case with large α and β , and then the other limiting case with small α and β .

In the limiting case when $\alpha \rightarrow \infty$, all communities are triangles, irrespective of the value of β . The network is then a random graph but consisting of triangles, namely a random 3-uniform hypergraph in graph theory [80]. As $P(k)$ is peaked [81], m_k is much larger than 1 for sufficiently large $\langle k \rangle$. We numerically find that $m_k \approx \langle k \rangle$ for $\alpha \geq 5$ is independent of β , as seen in Fig. 2(a). It is also expected that there is no degree-degree correlations, i.e., $\rho_{kk} \approx 0$. Furthermore, the local clustering coefficient may be negatively correlated with k , i.e., $\rho_{ck} < 0$, as explained in Section II. We find similar behaviors in the limiting case when $\beta \rightarrow \infty$, as communities of size larger than g_0 practically have no links in them, irrespective of α . These arguments are consistent with simulation results already when $\beta = 1$, as depicted in Fig. 2(a-c).

On the other hand, when $\beta = 0$, we have $p(g) = 1$ for the entire range of g , leading to the network of overlapping cliques of various sizes. If the value of α decreases from infinity, more and more large cliques are created in the network. Since the total number of links is fixed, links are then concentrated more in large communities, leading to a number of isolated nodes. The value of m_k then decreases from $\langle k \rangle$ as α decreases. Degrees of nodes in the same community tend to be similar, which along with the heterogeneous community size for finite α leads to the positive ρ_{kk} . As for the local clustering coefficient c , it must be 1 for all connected nodes unless communities are overlapping. Since communities overlap in general, the value of ρ_{ck} is expected to be negative as nodes belonging to multiple communities tend to have higher k but smaller c . However, in the intermediate range of α the positive ρ_{ck} is found in Fig. 2(c). It is because nodes belonging to even larger communities tend to have higher k and to find more links between their neighbors, hence the larger value of c . If α becomes too small, most nodes have c of the order of 1, while some nodes with high k have smaller values of c due to the overlapping commu-

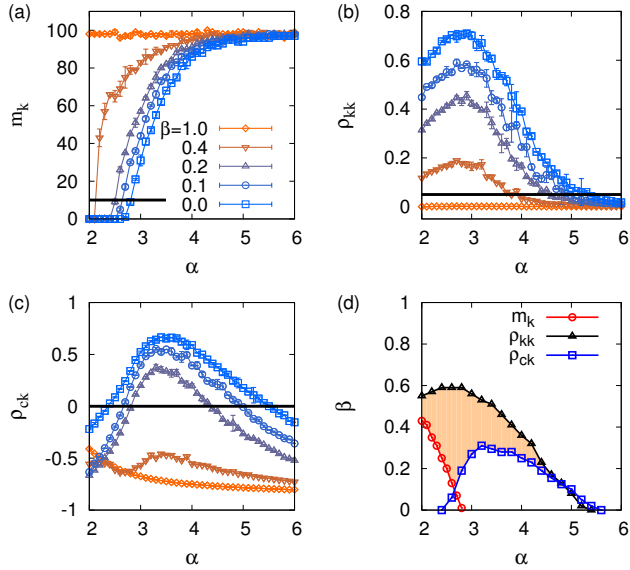


FIG. 2. (Color online) Effects of α and β on topological properties: (a) m_k , (b) ρ_{kk} , and (c) ρ_{ck} , with corresponding threshold values (black lines) for stylized facts. (d) Assuming that the topological stylized facts are reproduced when $m_k > 10$, $\rho_{kk} > 0.05$, and $\rho_{ck} < 0$, we derive the parameter region (shaded area) surrounded by three curves from (a–c). Here the results have been averaged over 10–40 networks generated using $N = 3 \cdot 10^4$, $\langle k \rangle = 100$, $g_0 = 3$, and $g_{\max} = 10^3$.

nities. Therefore the negative ρ_{ck} is observed, while the global clustering, i.e., the average of c for all nodes in the network, is still high (not shown).

B. Intensity-related properties

Next we numerically obtain the intensity-related quantities, i.e., $\frac{m_s}{\langle s \rangle}$, ρ_{sk} , ρ_{ow} , and Δf_c , by varying α and γ while keeping $\beta = 0.5$, for which it is expected to show the topological stylized facts for $\alpha < 3.5$. Here the mode of $P(s)$ has been normalized by the average strength $\langle s \rangle$ as $\langle s \rangle$ also depends on the parameters. We find in Fig. 3(a) that $\frac{m_s}{\langle s \rangle}$ overall increases with α but decreases with γ . In order to distinguish the region of overall decreasing $P(s)$ for small α and large γ from that of overall peaked $P(s)$ for large α and small γ , we take a threshold value for $\frac{m_s}{\langle s \rangle}$ as 0.1 to imitate the threshold value of m_k . In Fig. 3(b), we find that $\rho_{sk} \geq 0$ for the entire range of the parameter space, and that ρ_{sk} overall increases with α but decreases with γ . We take 0.1 as the threshold value for ρ_{sk} to separate the correlated region for large α and small γ from the uncorrelated region for small α and large γ . As for ρ_{ow} , it has a negative value for the almost entire range of the parameter space, except when both α and γ are small, as shown in Fig. 3(c). Finally, Δf_c shows the non-monotonic behavior according to α in Fig. 3(d). For small γ , there exists an intermediate re-

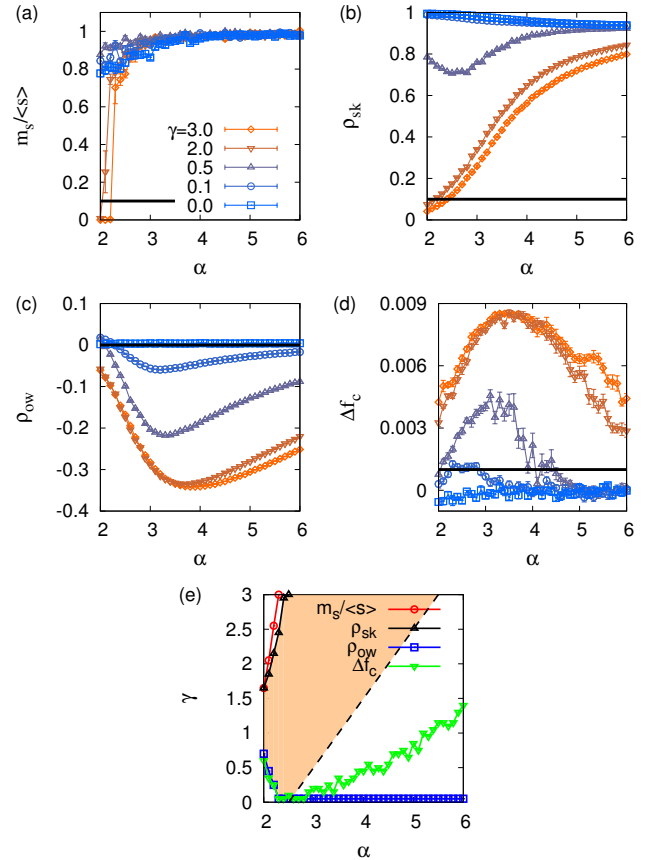


FIG. 3. (Color online) Effects of α and γ on intensity-related properties for a given $\beta = 0.5$: (a) $\frac{m_s}{\langle s \rangle}$, (b) ρ_{sk} , (c) ρ_{ow} , and (d) Δf_c , with corresponding threshold values (black lines) for stylized facts. (e) Assuming that the intensity-related stylized facts are reproduced when $\frac{m_s}{\langle s \rangle} > 0.1$, $\rho_{sk} > 0.1$, $\rho_{ow} > 0$, and $\Delta f_c > 0.001$, we barely find the parameter region surrounded by four curves from (a–d). If the condition for positive ρ_{ow} is relaxed, then we have the wide range of parameter space (shaded area) for the stylized facts, i.e., for large α and γ . Here the results have been averaged over 10 networks generated using $N = 3 \cdot 10^4$, $\langle k \rangle = 100$, $g_0 = 3$, $g_{\max} = 10^3$, and $w_0 = 1$. The dashed line, $\gamma = \alpha + \beta - 3$, indicates the criterion for the decreasing behavior of $P(w)$ obtained from Eq. (37).

gion of α showing $\Delta f_c > 0.001$, with the threshold value of 0.001. As γ increases, this region becomes wider.

In summary, we barely find a region in the parameter space of (α, γ) for a fixed $\beta = 0.5$, where all intensity-related stylized facts are reproduced, i.e., relatively large $\frac{m_s}{\langle s \rangle}$, ρ_{sk} , and Δf_c , as well as positive ρ_{ow} . However, if the condition for positive ρ_{ow} is relaxed, we have the wide range of the parameter space for the stylized facts to be reproduced, i.e., for large α and γ . This region is depicted as shaded in Fig. 3(e). We note that despite the negative ρ_{ow} , $o(w)$ shows the increasing and then decreasing behavior for some range of parameter values, which will be discussed in the next Section.

In order to understand these results, we remind that

the larger value of γ leads to more hierarchical weights as larger communities contain even weaker links on average. We first consider the limiting case with small γ and large α , and then the case with large γ and not too small α .

In the limiting case when $\gamma = 0$, all weights are drawn from the uniform distribution with the range of $[0, 1]$, independent of the community size. It implies no intensity-topology correlations, irrespective of topological structure of the network. Therefore, we have $s(k) \approx \frac{k}{2}$ for sufficiently large networks, which leads to $\frac{m_s}{\langle s \rangle} \approx \frac{m_k}{\langle k \rangle} \sim \mathcal{O}(1)$ and $\rho_{sk} \approx 1$ for a wide range of α . Since the neighborhood overlap of a link is statistically independent of its weight, we have $\rho_{ow} \approx 0$. Finally, for the link percolation the ascending link removal is just the same as the descending link removal, thus $\Delta f_c \approx 0$ is expected. These expected behaviors are numerically confirmed as shown in Fig. 3(a–d). These tendencies are also observed for very large α , independent of γ , as the network consists of triangles only and then the weights are fully uncorrelated with topological structure.

On the other hand, the large value of γ enhances the hierarchical structure of weights. If α decreases from infinity, the network is dominated by larger communities, i.e., by the larger number of links with very small weights. Then as more nodes have smaller strength, both m_s and $\langle s \rangle$ decrease to 0, while $\frac{m_s}{\langle s \rangle}$ decreases to 0 as $P(s)$ becomes more skewed due to the nodes belonging only to large communities. Since the majority of links are weak, the higher degree does not necessarily mean the larger strength, leading to the smaller ρ_{sk} for decreasing α . Further, since the nodes in the large communities tend to have more common neighbors and weaker links between them, we find negative correlations between neighborhood overlap and weight of links, except for small α and γ , as shown in Fig. 3(c). Finally, the Granovetterian community structure appears to exist in the networks, as evidenced by $\Delta f_c > 0$ for sufficiently large γ and for the wide range of α in Fig. 3(d). This is because by construction weak links in larger communities connect smaller communities containing strong links. Moreover, as the larger γ enhances the hierarchical structure of weights, we find the stronger weight-topology correlation, hence a larger Δf_c .

We have also studied the effects of β and γ for a fixed value of $\alpha = 2.5$ to find the range of the parameter space (β, γ) for the intensity-related stylized facts. See Appendix C and Fig. 6 for the details.

To conclude the Section, we note that our simple model with a few assumptions about communities could reproduce almost all stylized facts, except for the increasing neighborhood overlap as a function of link weight, for the wide range of the parameter space (α, β, γ) .

V. ANALYSIS OF LOCAL NETWORK QUANTITIES

Thanks to the random nature of assigning communities to nodes in our model, we can obtain analytical solutions of the local network quantities to some extent, which is important for the rigorous understanding of the consequences of the model. Since the network consists of many random communities whose sizes are randomly drawn from a given distribution, the network can be understood in terms of an aggregate of different layers, where each layer consists of communities with the same size that are overlapping by construction. Based on this concept of layers, we first get analytical results for topological quantities, regardless of the rule of assigning weights to links: $P(k)$, $c(k)$, and $k_{\text{nn}}(k)$. We then consider intensity-related quantities: $P(w)$, $P(s)$, $s(k)$, and $o(w)$. According to the discussion in Appendix B, we assume that communities do not overlap over more than one node.

A. Topological quantities

The number of communities of size g is simply given by $n_g \equiv CP(g) = CAg^{-\alpha}$. The average number of links in the layer g is equal to $L_g \equiv n_g \frac{g(g-1)}{2} p(g)$ and the total number of links in the network is obtained by $L = \sum_g L_g$. Then C is determined from the relation $L = \frac{N\langle k \rangle}{2}$. The degree distribution in the layer g can be naively approximated by

$$P_g(k) = \binom{N-1}{k} p_g^k (1-p_g)^{N-1-k} \approx e^{-\langle K \rangle_g} \frac{\langle K \rangle_g^k}{k!} \quad (14)$$

with $p_g \equiv \frac{2L_g}{N(N-1)}$ and $\langle K \rangle_g \equiv \frac{2L_g}{N}$ respectively denoting the link density and the average degree in the layer g . As we assume that links in different layers are rarely overlapping, the degree distribution for the whole network is obtained as a Poisson distribution:

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}, \quad (15)$$

where we have used the relation $\langle k \rangle = \sum_g \langle K \rangle_g$. $P(k)$ is found to have a peak around at $\langle k \rangle$, i.e., $m_k \approx \langle k \rangle$. The numerical $P(k)$ in Fig. 4(a) also has the peak around at $\langle k \rangle$ but it is broader than the above Poisson distribution possibly due to the correlation between links imposed by the community structure. This effect due to the correlation can be explained by separating nodes chosen for communities from those not chosen in each layer. The number of nodes chosen for communities of size g is approximated as $N_g \approx gn_g$, leading to $\langle k \rangle_g \equiv \frac{2L_g}{N_g} \approx (g-1)p(g)$ denoting the average degree only for nodes chosen for communities in the layer g . Based on this idea, we derive the analytical result for the standard deviation σ for the degrees in Appendix D. For example, when $\alpha = 2.7$ and $\beta = 0.55$, we obtain $\sigma \approx 31.6$

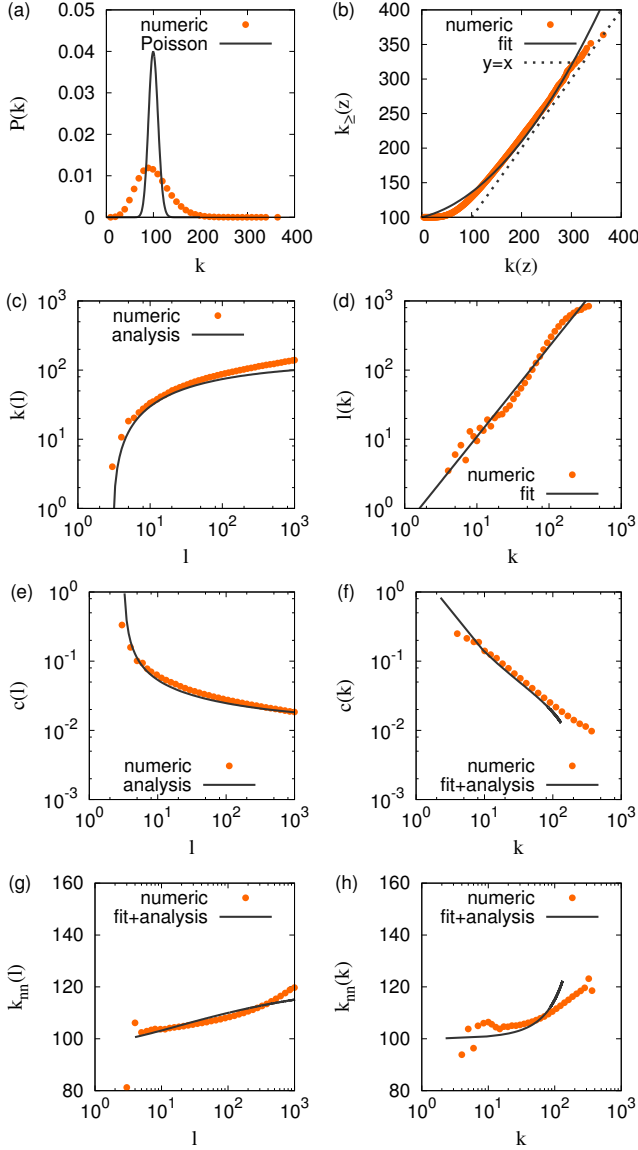


FIG. 4. (Color online) Simulation results of topological quantities for $N = 3 \cdot 10^4$, $\langle k \rangle = 100$, $g_0 = 3$, $g_{\max} = 10^3$, $\alpha = 2.7$, and $\beta = 0.55$ (circles) are compared to the analytical results (solid lines). The simulation results were averaged over 50 generated networks.

using Eq. (D5), which is comparable with the simulation result of ≈ 34.3 . Note that the standard deviation for the Poisson distribution with $\langle k \rangle = 100$ is 10. Thus, the correlation between links imposed by the communities is important in understanding the broader $P(k)$ than the Poissonian case.

Now we focus on the egocentric network that consists of a node and its neighbors, namely, the ego and its alters. The degree of the node is determined by the layers in which the node was chosen to create links. We denote by l the largest community size or the *outermost layer* involving the node. Since the layers involving the node or

the ego are not necessarily consecutive, we take a mean-field approach: The indexes of layers involving the ego are assumed to be consecutive from g_0 to l . Since the probability of an ego being chosen to have links in the layer g is $\frac{N_g}{N}$, one obtains the expected degree of nodes with outermost layer l as

$$\begin{aligned} k(l) &\equiv \sum_{g=g_0}^l \frac{N_g}{N} \langle k \rangle_g \approx \frac{g_0^\beta CA}{N} \int_{g_0}^l (g^{2-\alpha-\beta} - g^{1-\alpha-\beta}) dg \\ &= \frac{g_0^\beta CA}{N} \left[H_{1,0}^{3,1}(g_0, l) - H_{1,0}^{2,1}(g_0, l) \right]. \end{aligned} \quad (16)$$

Here we have defined for convenience

$$\begin{aligned} H_{m,n}^{u,v}(x_0, x_1) &\equiv \int_{x_0}^{x_1} x^{u-v\alpha-m\beta-n\gamma-1} dx \\ &= \begin{cases} \frac{x_1^{u-v\alpha-m\beta-n\gamma} - x_0^{u-v\alpha-m\beta-n\gamma}}{u-v\alpha-m\beta-n\gamma} & \text{if } u-v\alpha-m\beta-n\gamma \neq 0, \\ \ln \frac{x_1}{x_0} & \text{otherwise.} \end{cases} \end{aligned} \quad (17)$$

Figure 4(c) shows that this analytical result is comparable with the simulation results. If $\alpha + \beta \neq 2, 3$, then $k(l)$ can be explicitly written as follows:

$$k(l) \approx a_0 + a_1 l^{3-\alpha-\beta} + a_2 l^{2-\alpha-\beta}, \quad (19)$$

where

$$a_0 \equiv \frac{g_0^\beta CA}{N} \left(-\frac{g_0^{3-\alpha-\beta}}{3-\alpha-\beta} + \frac{g_0^{2-\alpha-\beta}}{2-\alpha-\beta} \right), \quad (20)$$

$$a_1 \equiv \frac{g_0^\beta CA}{N(3-\alpha-\beta)}, \quad a_2 \equiv -\frac{g_0^\beta CA}{N(2-\alpha-\beta)}. \quad (21)$$

Later we will need to use the expected outermost layer of nodes with degree k , denoted by $l(k)$. As the calculation of $l(k)$ is not trivial, we fit the simulation result in Fig. 4(d) with a simple scaling function as follows:

$$l(k) \approx \tilde{a} k^\mu \quad (22)$$

with $\tilde{a} \approx 0.542$ and $\mu \approx 1.30$. We remark that the inverse function of $k(l)$ does not necessarily match with $l(k)$, implying that in general $k[l(k)] \neq k$.

Next, we calculate the expected local clustering coefficient in two forms: $c(k)$ and $c(l)$. Since the probability of linking two alters in the same layer is $p(g)$, we obtain

$$\begin{aligned} c(l) &\equiv \frac{1}{k(l)[k(l)-1]} \sum_{g=g_0}^l \frac{N_g}{N} \langle k \rangle_g (\langle k \rangle_g - 1) p(g) \\ &\approx \frac{g_0^{3\beta} CA}{Nk(l)[k(l)-1]} \left[H_{3,0}^{4,1}(g_0, l) - 2H_{3,0}^{3,1}(g_0, l) + H_{3,0}^{2,1}(g_0, l) \right] \\ &\quad - \frac{g_0^{2\beta} CA}{Nk(l)[k(l)-1]} \left[H_{2,0}^{3,1}(g_0, l) - H_{2,0}^{2,1}(g_0, l) \right]. \end{aligned} \quad (23)$$

Figure 4(e) shows that this analytical result is comparable with the simulation results. Then, we numerically get $c(k)$ by plugging $l(k)$ in Eq. (22) into Eq. (24), i.e.,

$$c(k) \equiv c[l = l(k)], \quad (25)$$

which is plotted against $k[l(k)]$ in Fig. 4(f). We find our approximated solution to be comparable with the simulation results.

Then, we study the assortative mixing by calculating $k_{\text{nn}}(k)$ and $k_{\text{nn}}(l)$. Let us consider the above egocentric network, where each alter, say j , has been connected to the ego in some layer, say z . If z is large, z is most likely to be the outermost layer of j . Then the degree of j is $k(z)$ that is obtained using Eq. (16). However, for small z , the outermost layer of j is most likely to be larger than z . Hence the degree of j can be estimated as the average degree of nodes whose outermost layer is equal to or larger than z . If z is very small, the expected degree of j would be $\langle k \rangle$. In general, the expected degree of the alter having a link to the ego in the layer z can be obtained as follows:

$$k_{\geq}(z) \equiv \frac{\sum_{k=k(z)}^{\infty} kP(k)}{\sum_{k=k(z)}^{\infty} P(k)}. \quad (26)$$

In Fig. 4(b), we plot $k_{\geq}(z)$, calculated using the numerical $P(k)$, against $k(z)$. As expected, $k_{\geq}(z)$ has a value of $\langle k \rangle$ for small z , while it approaches $k(z)$ for large z . As the calculation of $k_{\geq}(z)$ is not trivial, we make an approximation in a quadratic form as

$$k_{\geq}(z) \approx \langle k \rangle + b_1 k(z) + b_2 k(z)^2 \quad (27)$$

with coefficients $b_1 \approx 0.178$ and $b_2 \approx 0.00185$, see the solid line in Fig. 4(b). More complicated functional form may give a better result, but it makes further calculations much more difficult.

Since the ego has on average $\langle k \rangle_g$ neighbors in the layer g whose expected degree is $k_{\geq}(g)$, using Eq. (27) we obtain

$$\begin{aligned} k_{\text{nn}}(l) &\equiv \frac{1}{k(l)} \sum_{g=g_0}^l \frac{N_g}{N} \langle k \rangle_g k_{\geq}(g) \\ &\approx \frac{g_0^{\beta} C A}{N k(l)} \{ c_0 [H_{1,0}^{3,1}(g_0, l) - H_{1,0}^{2,1}(g_0, l)] - c_1 H_{2,0}^{4,2}(g_0, l) \\ &+ (c_1 - c_2) H_{2,0}^{5,2}(g_0, l) + c_2 H_{2,0}^{6,2}(g_0, l) - c_3 H_{3,0}^{6,3}(g_0, l) \\ &+ (c_3 - c_4) H_{3,0}^{7,3}(g_0, l) + (c_4 - c_5) H_{3,0}^{8,3}(g_0, l) + c_5 H_{3,0}^{9,3}(g_0, l) \}, \end{aligned} \quad (28)$$

where

$$\begin{aligned} c_0 &\equiv \langle k \rangle + a_0 b_1 + a_0^2 b_2, \quad c_1 \equiv a_2 (b_1 + 2a_0 b_2), \\ c_2 &\equiv a_1 (b_1 + 2a_0 b_2), \quad c_3 \equiv a_2^2 b_2, \\ c_4 &\equiv 2a_1 a_2 b_2, \quad c_5 \equiv a_1^2 b_2. \end{aligned} \quad (30)$$

Here we have assumed that $\alpha + \beta \neq 2, 3$. Figure 4(g) shows that Eq. (29) is comparable with the simulation results. If $\alpha + \beta < 3$, as $b_1 \gg b_2$, one may expect that for very large l

$$k_{\text{nn}}(l) \sim l^{3-\alpha-\beta}, \quad (31)$$

which hints at the role of exponent values such that the higher assortativity is obtained with the smaller values of

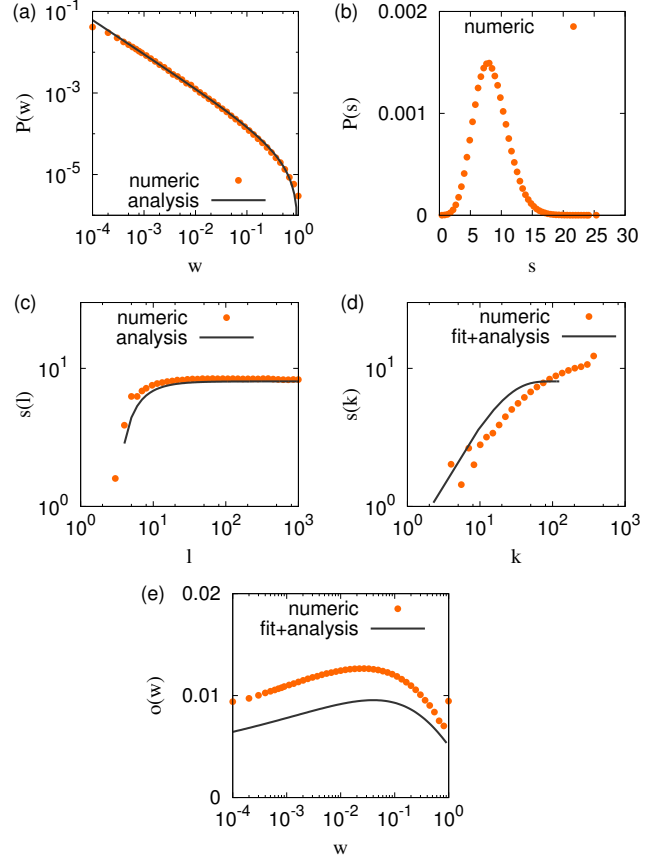


FIG. 5. (Color online) Simulation results of intensity-related quantities for the same parameter values as in Fig. 4 with $\gamma = 1.5$ and $w_0 = 1$ (circles) are compared to the analytical results (solid lines).

α and β , i.e., the more heterogeneous size of communities and the more links for larger communities. This is consistent with the results in Fig. 2(b). Then, we numerically get $k_{\text{nn}}(k)$ by assuming that

$$k_{\text{nn}}(k) \equiv k_{\text{nn}}[l = l(k)]. \quad (32)$$

In Fig. 4(h), we plot $k_{\text{nn}}(k)$ against $k[l(k)]$ to find our approximated solution to be comparable with the simulation results.

B. Intensity-related quantities

Here we calculate local network quantities related to the intensity. The link weight distribution can be obtained as follows:

$$P(w) \equiv \frac{1}{L} \sum_{g=g_0}^{g_{\text{max}}} P_g(w) L_g \approx \frac{1}{L} \int_{g_0}^{g(w)} \frac{n_g g(g-1) P(g)}{2w(g)} dg, \quad (33)$$

where $g(w)$ is the inverse of $w(g)$ in Eq. (12), i.e.,

$$g(w) = g_0 \left(\frac{w}{w_0} \right)^{-\frac{1}{\gamma}}. \quad (34)$$

Then one obtains

$$P(w) \approx \frac{g_0^{\beta-\gamma} CA}{2Lw_0} \left\{ H_{1,-1}^{3,1}[g_0, g(w)] - H_{1,-1}^{2,1}[g_0, g(w)] \right\}. \quad (35)$$

This analytical result compares favorably with the simulation results in Fig. 5(a). For relatively large w , we can get a simple scaling form from Eq. (35) as follows:

$$P(w) \sim w^{-\frac{3-\alpha-\beta+\gamma}{\gamma}}. \quad (36)$$

Thus, in order for $P(w)$ to decrease for large w , the following condition must be satisfied:

$$\gamma > \alpha + \beta - 3. \quad (37)$$

This condition narrows down the parameter region for stylized facts, as depicted in Fig. 3(e).

Next, we obtain the expected strength in two forms: $s(k)$ and $s(l)$. We obtain $s(l)$ as

$$s(l) \equiv \sum_{g=g_0}^l \frac{N_g}{N} \langle k \rangle_g \langle w(g) \rangle, \quad (38)$$

where $\langle w(g) \rangle \equiv \int_0^\infty w P_g(w) dw = \frac{w(g)}{2}$. One then obtains

$$s(l) \approx \frac{g_0^{\beta+\gamma} w_0 CA}{2N} \left[H_{1,1}^{3,1}(g_0, l) - H_{1,1}^{2,1}(g_0, l) \right]. \quad (39)$$

This analytical result is compared with the simulation results in Fig. 5(c). In most cases of our simulations, we use parameter values satisfying that $\alpha + \beta + \gamma > 3$. It implies that $s(l)$ is approaching a constant from below, rather than increasing indefinitely. For the calculation of $s(k)$, we assume that

$$s(k) \equiv s[l = l(k)]. \quad (40)$$

In Fig. 5(d), we plot $s(k)$ against $k[l(k)]$ to find our approximated solution to be comparable with the simulation results, but with systematic discrepancy.

Then, one can get the strength distribution $P(s)$ from the identity $P(s)ds = P(k)dk$ using the above $s(k)$. The peaked $P(k)$ combined with the increasing $s(k)$ leads to the peaked $P(s)$, as evidenced by the simulation results in Fig. 5(b). The peak turns out to be around at the average strength $\langle s \rangle \approx \langle k \rangle \langle w \rangle \approx 8.4$.

Finally, we discuss the behavior of the neighborhood overlap $o(w)$. Once a link with weight w is given, the link must have been created in a layer g with $g_0 \leq g \leq g(w)$. The probability $q_{w,g}$ that the link of weight w is created in the layer g reads

$$q_{w,g} \equiv \frac{L_g/w(g)}{\sum_{g'=g_0}^{g(w)} L_{g'}/w(g')}, \quad (41)$$

where $\frac{1}{w(g)}$ comes from $P_g(w)$. If the link ij is created in the layer g , nodes i and j have $(g-2)p(g)^2$ common

neighbors on average, leading to the expected number of common neighbors as

$$\begin{aligned} \langle e \rangle_w &\equiv \sum_{g=g_0}^{g(w)} q_{w,g} (g-2)p(g)^2. \\ &\approx \frac{g_0^{2\beta} \{ H_{3,-1}^{4,1}[g_0, g(w)] - 3H_{3,-1}^{3,1}[g_0, g(w)] + 2H_{3,-1}^{2,1}[g_0, g(w)] \}}{H_{1,-1}^{3,1}[g_0, g(w)] - H_{1,-1}^{2,1}[g_0, g(w)]}. \end{aligned} \quad (42)$$

Next, we take an approximation of both k_i and k_j in Eq. (8) as $k_{\geq}[g(w)]$ in Eq. (27) to finally obtain

$$o(w) \approx \frac{\langle e \rangle_w}{2k_{\geq}[g(w)] - 2 - \langle e \rangle_w}. \quad (44)$$

For our choice of parameter values, this analytical result of $o(w)$ turns out to be increasing and then decreasing with w , but with systematic discrepancy with the simulation results as shown in Fig. 5(e). The increasing behavior of $o(w)$ for small w can be understood by considering a simple scaling form for large $g(w)$. Since $\alpha + \beta > 3$ and $3 - \alpha - 3\beta + \gamma > 0$ for our choice of exponent values, we get

$$o(w) \sim \frac{\langle e \rangle_w}{k_{\geq}[g(w)]} \sim w^{\frac{2\beta-1}{\gamma}}. \quad (45)$$

Thus, the increasing behavior of $o(w)$ for small w is realized for $\beta > \frac{1}{2}$, and enhanced by the smaller γ . This is because the relatively large β reduces the possibility of having common neighbors in large communities, which leads to smaller neighborhood overlaps for weaker links. However, this argument does not apply to the very strong links as the value of $p(g)$ for small g is still high for large β . This may account for the decreasing behavior of $o(w)$ for very large w that has been empirically observed in some datasets [9, 69].

VI. CONCLUSIONS

Despite a number of empirical findings for social networks, the multi-channel weighted social (MWS) network has never been comprehensively identified, prompting us to ask a series of important questions: What does the MWS network look like? Then how can it be parsimoniously modeled or reproduced? As for the first question, we expect that the structure of the MWS network is reflected in the empirical findings from partial datasets of the network. For this, we have found several commonly observed features or stylized facts in various datasets of social networks, such as broad distributions of local network quantities, existence of communities, assortative mixing, and intensity-topology correlations. These stylized facts are listed in Table I. Among them, the overall decreasing degree and strength distributions are not fully consistent with common sense. In particular, we expect the degree and strength distributions to be peaked [22, 23].

As for the parsimonious modeling of the MWS network, we have devised the community-based static model

for reproducing the stylized facts with the expected peaked behavior of degree and strength distributions. For our model, we have randomly assigned a number of communities to a given set of isolated nodes using three assumptions, such that (i) the size g of each community is drawn from a power-law distribution with the exponent α , (ii) the link density for the community of size g is given as a decreasing function of g , controlled by the power-law exponent β , and (iii) the characteristic weight of links created in the community of size g is given as a decreasing function of g , controlled by the power-law exponent γ . With these few assumptions about communities, realistic social networks have successfully been reproduced such that they show almost all stylized facts for a wide range of parameter space of (α, β, γ) . Note that our assumptions are reasonable, yet they could be deduced from more fundamental mechanisms for the creation, aging, and severance of links.

We could obtain some analytical results for local network quantities that compare favorably with the numerical results. It is because our model can be interpreted as the aggregate of random uniform hypergraphs with various degrees. Such analytic results indeed provide deeper understanding of the consequences of the model. More importantly, thanks to the simplicity and explicitness of our model, we expect it to serve as a general reference system for further research in the direction of investigating the structure and dynamics of the MWS network, as well as the dynamical processes taking place in the MWS network. In addition, once we better understand how the sampling relates the MWS network to the partial observations, we can better translate the conclusions drawn from partial datasets into those for the MWS network.

We remark that we have ignored the known correlations such as those due to the geographical constraint [28] and/or demographic information [30]. These correlations can be incorporated for more realistic modeling of the MWS network. For example, since the geographical distance between a pair of nodes is found to be negatively correlated with the link probability between them [28], geographically close nodes can be chosen when assigning communities to nodes, as done in Ref. [39].

Finally, we briefly discuss the rank curve analysis as one of the stylized facts for social networks. How individuals distribute their limited resources like time to their neighbors is also indicative in characterizing the social networks at the individual level. The rank curve of a node is defined by the weights of links involving the node in a descending order. The layered structure in rank curves has been claimed [66, 82], while the exponential or power-law functional forms have been successfully used for fitting empirical rank curves [83, 84]. In fact, there is a chance to obtain the layered or stepwise rank curves from our model by assuming that all links in the same layer have the same weight, e.g., by using $P_g(w) = \delta[w - w(g)]$ in Eq. (A1). However, this functional form for $P_g(w)$ requires too strict conditions for the parameter values, as discussed in Appendix A. As the functional form of rank

curves is yet inconclusive, we leave this issue for a future work.

ACKNOWLEDGMENTS

H.-H.J. acknowledges financial support by Basic Science Research Program through the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (2015R1D1A1A01058958) and the framework of international cooperation program managed by the National Research Foundation of Korea (NRF-2016K2A9A2A08003695). Y.M. appreciates hospitality at Aalto University and acknowledges support from CREST, JST. J.T. thanks for financial support of Aalto ASci internship programme. J.K. acknowledges support from EU Grant No. FP7 317532 (MULTIPLEX). K.K. acknowledges support from Academy of Finland's COSDYN project (No. 276439) and EU's Horizon 2020 FET Open RIA 662725 project IBSEN. This project was partly supported by JSPS and NRF under the Japan-Korea Scientific Cooperation Program. Partial support by OTKA, K112713 is also acknowledged. The systematic simulations in this study were assisted by OACIS [85].

Appendix A: Alternative forms of $P_g(w)$

In the main text, we studied the case with the uniform distribution for $P_g(w)$. Here we discuss two alternative forms of $P_g(w)$. The first case is the delta function as

$$P_g(w) = \delta[w - w(g)], \quad (\text{A1})$$

with the same $w(g) = w_0(\frac{g_0}{g})^\gamma$ as in Eq. (12). We then find that the parameter range for stylized facts is strongly limited. For example, we obtain the weight distribution as $P(w) \sim w^{-\frac{2-\alpha-\beta}{\gamma}}$. It implies that for decreasing $P(w)$ one must have $\alpha + \beta < 2$, which is hardly realistic, considering the empirical values of these exponents [11, 35, 36, 70, 71]. As for the second case, we can study the exponential distribution as follows:

$$P_g(w) = \frac{1}{w(g)} e^{-w/w(g)}. \quad (\text{A2})$$

We find the same scaling behavior for large w as in Eq. (36). Thus, the advantage of this form is only marginal compared with our choice of the uniform distribution.

Appendix B: Overlapping communities

In order to consider the case of overlapping communities, we introduce an overlapping probability $q_{gg'm}$ that two communities of size g and g' share m nodes with $1 \leq m \leq \min\{g, g'\}$. Here we assume that $p(g) = 1$

for all g , which provides an upper bound of $q_{gg'm}$. One obtains

$$\begin{aligned} q_{gg'm} &\equiv \frac{\binom{N}{m} \binom{N-m}{g-m} \binom{N-g}{g'-m}}{\binom{N}{g} \binom{N}{g'}} \\ &= \frac{g!g'!}{m!(g-m)!(g'-m)!} \frac{(N-g)!(N-g')!}{N!(N-g-g'+m)!} \\ &\approx \frac{m!}{N^m} \binom{g}{m} \binom{g'}{m}. \end{aligned} \quad (\text{B1})$$

We get the basic quantities as

$$q_{gg1} \approx \frac{g^2}{N}, \quad q_{gg2} \approx \frac{[g(g-1)]^2}{2N^2}, \quad (\text{B2})$$

$$q_{gg'1} \approx \frac{gg'}{N}, \quad q_{gg'2} \approx \frac{g(g-1)g'(g'-1)}{2N^2}. \quad (\text{B3})$$

Provided that $N \gg g$, $q_{gg'2}$ is found to be mostly negligible. It implies that it is very unlikely to find a pair of nodes belonging to multiple communities, whether communities have the same size or not, i.e., whether they are in the same layer or not. The average number of links in the layer g can be obtained as

$$\begin{aligned} L_g &= n_g \frac{g(g-1)}{2} p(g) - \frac{n_g(n_g-1)}{2} \sum_{m=2}^g \frac{m(m-1)}{2} p(g) q_{ggm} \\ &\approx n_g \frac{g(g-1)}{2} p(g) - \frac{n_g(n_g-1)}{2} p(g) \frac{[g(g-1)]^2}{2N^2}. \end{aligned} \quad (\text{B4})$$

For the second line, we have taken only the dominant term in the summation.

We can also calculate the exact number of nodes chosen for communities of size g , denoted by N_g , by considering the effect by overlapping communities. The maximum of N_g is gn_g as n_g communities are assigned with g nodes for each community. N_g can be smaller than gn_g due to two factors: (i) Some nodes may be isolated in the layer g if $p(g) < 1$. (ii) Some nodes may belong to multiple communities. For the former, the probability of a node being isolated is $[1 - p(g)]^{g-1}$, provided that this node could belong to only one community. For the latter, the dominant case is when two communities overlap over one node, whose probability is q_{gg1} . In sum, one obtains

$$N_g \approx gn_g \{1 - [1 - p(g)]^{g-1}\} - \frac{n_g(n_g-1)}{2} \frac{g^2}{N}. \quad (\text{B5})$$

We then get the average degree only for nodes chosen for communities in the layer g as

$$\langle k \rangle_g = \frac{2L_g}{N_g} \approx \frac{(g-1)p(g) - \frac{1}{2N^2}(n_g-1)p(g)g(g-1)^2}{1 - [1 - p(g)]^{g-1} - \frac{1}{2N}(n_g-1)p(g)}. \quad (\text{B6})$$

If $p(g) \ll 1$ and N is sufficiently large, one obtains $\langle k \rangle_g \approx 1$, implying that links are mostly isolated.

Appendix C: Effects of β and γ for a fixed $\alpha = 2.5$

We numerically obtain the intensity-related quantities, i.e., $\frac{m_s}{\langle s \rangle}$, ρ_{sk} , ρ_{ow} , and Δf_c , by varying β and γ while keeping $\alpha = 2.5$, for which it is expected to show the

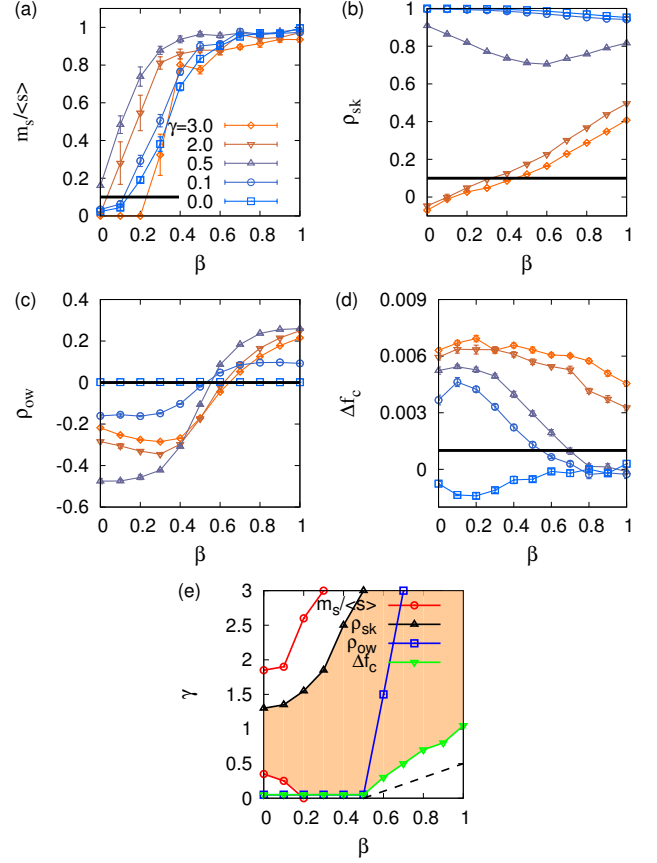


FIG. 6. (Color online) Effects of β and γ on intensity-related properties for a given $\alpha = 2.5$: (a) $\frac{m_s}{\langle s \rangle}$, (b) ρ_{sk} , (c) ρ_{ow} , and (d) Δf_c , with corresponding threshold values (black lines) for stylized facts. (e) Assuming that the intensity-related stylized facts are reproduced when $\frac{m_s}{\langle s \rangle} > 0.1$, $\rho_{sk} > 0.1$, $\rho_{ow} > 0$, and $\Delta f_c > 0.001$, we barely find the parameter region surrounded by four curves from (a-d). If the condition for positive ρ_{ow} is relaxed, then we have the wide range of parameter space (shaded area) for the stylized facts, i.e., for large β and intermediate γ . Here the results have been averaged over 10 networks generated using $N = 3 \cdot 10^4$, $\langle k \rangle = 100$, $g_0 = 3$, $g_{\max} = 10^3$, and $w_0 = 1$. The dashed line, $\gamma = \alpha + \beta - 3$, indicates the criterion for the decreasing behavior of $P(w)$ obtained from Eq. (37).

topological stylized facts for $0.1 < \beta < 0.6$. As shown in Fig. 6, we apparently find a wide region in the parameter space of (β, γ) , where all intensity-related stylized facts are reproduced, i.e., relatively large $\frac{m_s}{\langle s \rangle}$, ρ_{sk} , and Δf_c , as well as positive ρ_{ow} , by using the same threshold values for these quantities as discussed in the main text. However, as the topological stylized facts are reproduced for $0.1 < \beta < 0.6$, the parameter region for the intensity-related stylized facts must be narrow. Thus, if the condition for positive ρ_{ow} is relaxed again, we have the wide range of the parameter space for the stylized facts to be reproduced, i.e., for large β and intermediate γ . This region is depicted as shaded in Fig. 6(e).

Appendix D: Standard deviation for the degrees

In order to take into account the correlation between links imposed by the communities in the degree distribution, we separate nodes chosen for communities from those not chosen in each layer g . Then $P_g(k)$ consists of two parts, one for nodes chosen for communities and the other for those not chosen:

$$P_g(k) = \left(1 - \frac{N_g}{N}\right) \delta_{k,0} + \frac{N_g}{N} \binom{g-1}{k} p(g)^k [1-p(g)]^{g-1-k}, \quad (\text{D1})$$

where $k = 0, \dots, g-1$. Note that isolated nodes are either those not chosen for communities or those chosen but left with no links when $p(g) < 1$. The calculation of $P(k)$ from the above $P_g(k)$ is not trivial. In order to calculate the standard deviation for the degrees, we assume that

$$P_g(k) = \left(1 - \frac{N_g}{N}\right) \delta(k) + \frac{N_g}{N} \delta(k - \langle k \rangle_g). \quad (\text{D2})$$

One obtains the variance for the layer g as

$$\sigma_g^2 = \frac{N_g}{N} \left(1 - \frac{N_g}{N}\right) \langle k \rangle_g^2, \quad (\text{D3})$$

leading to the variance for the entire network as

$$\begin{aligned} \sigma^2 &\equiv \sum_{g=g_0}^{g_{\max}} \sigma_g^2 \approx \int_{g_0}^{g_{\max}} \sigma_g^2 dg \quad (\text{D4}) \\ &= \frac{g_0^{\beta} CA}{N} \left[H_{2,0}^{4,1}(g_0, g_{\max}) - 2H_{2,0}^{3,1}(g_0, g_{\max}) + H_{2,0}^{2,1}(g_0, g_{\max}) \right] \\ &\quad - \frac{g_0^{2\beta} (CA)^2}{N^2} \left[H_{2,0}^{5,2}(g_0, g_{\max}) - 2H_{2,0}^{4,2}(g_0, g_{\max}) + H_{2,0}^{3,2}(g_0, g_{\max}) \right], \quad (\text{D5}) \end{aligned}$$

where $H_{m,n}^{u,v}$ is defined in Eq. (18).

-
- [1] R. Albert and A.-L. Barabási, Review of Modern Physics **74**, 47 (2002), arXiv:cond-mat/0106096.
- [2] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, Science **323**, 892 (2009).
- [3] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Reviews of Modern Physics **87**, 925 (2015), arXiv:1408.2701.
- [4] N. Eagle and A. Pentland, Personal and Ubiquitous Computing **10**, 255 (2006).
- [5] K. Zhao, J. Stehlé, G. Bianconi, and A. Barrat, Physical Review E **83**, 056109+ (2011).
- [6] J. Fournet and A. Barrat, PLoS ONE **9**, e107878 (2014).
- [7] J.-P. Eckmann, E. Moses, and D. Sergi, Proceedings of the National Academy of Sciences **101**, 14333 (2004).
- [8] B. Klimt and Y. Yang, in *Machine Learning: ECML 2004*, Lecture Notes in Computer Science, Vol. 3201, edited by J.-F. Boulicaut, F. Esposito, F. Giannotti, and D. Pedreschi (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004) Chap. 22, pp. 217–226.
- [9] J. P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A. L. Barabási, Proceedings of the National Academy of Sciences **104**, 7332 (2007), arXiv:physics/0610104v1.
- [10] V. D. Blondel, A. Decuyper, and G. Krings, EPJ Data Science **4**, 10 (2015), arXiv:1502.03406.
- [11] D. Hric, R. K. Darst, and S. Fortunato, Physical Review E **90**, 062805 (2014).
- [12] Y.-H. Eom and H.-H. Jo, Scientific Reports **5**, 9752 (2015).
- [13] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, “The anatomy of the facebook social graph,” (2011), arXiv:1111.4503.
- [14] H. Kwak, C. Lee, H. Park, and S. Moon, in *Proceedings of the 19th international conference on World wide web, WWW ’10* (ACM, New York, NY, USA, 2010) pp. 591–600.
- [15] M. Szell and S. Thurner, Social Networks **32**, 313 (2010).
- [16] M. Szell, R. Lambiotte, and S. Thurner, Proceedings of the National Academy of Sciences **107**, 13636 (2010).
- [17] M. P. H. Stumpf, C. Wiuf, and R. M. May, Proceedings of the National Academy of Sciences of the United States of America **102**, 4221 (2005).
- [18] M. P. H. Stumpf and C. Wiuf, Physical Review E **72**, 036118 (2005).
- [19] S. H. Lee, P.-J. Kim, and H. Jeong, Physical Review E **73**, 016102 (2006), arXiv:cond-mat/0505232.
- [20] J. Török, Y. Murase, H.-H. Jo, J. Kertész, and K. Kaski, Physical Review E **94**, 052319 (2016).
- [21] M. O. Jackson, *Social and Economic Networks* (Princeton University Press, 2010).
- [22] Y. Murase, H.-H. Jo, J. Török, J. Kertész, and K. Kaski, PLoS ONE **10**, e0133005 (2015).
- [23] J. Kertész, J. Torok, Y. Murase, H.-H. Jo, and K. Kaski, “Multiplex modeling of the society,” (2016), arXiv:1609.08381.
- [24] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, M. Argollo de Menezes, K. Kaski, A.-L. Barabási, and J. Kertész, New Journal of Physics **9**, 179 (2007).
- [25] S. Fortunato, Physics Reports **486**, 75 (2010), arXiv:0906.0612.
- [26] M. McPherson, L. Smith-Lovin, and J. M. Cook, Annual Review of Sociology **27**, 415 (2001).
- [27] M. E. J. Newman, Physical Review Letters **89**, 208701 (2002).
- [28] J.-P. Onnela, S. Arbesman, M. C. González, A.-L. Barabási, and N. A. Christakis, PLoS ONE **6**, 16939 (2011).
- [29] V. Palchykov, K. Kaski, J. Kertész, A.-L. Barabási, and R. I. M. Dunbar, Scientific Reports **2**, 370 (2012).
- [30] H.-H. Jo, J. Saramäki, R. I. M. Dunbar, and K. Kaski, Scientific Reports **4**, 6988 (2014).
- [31] M. S. Granovetter, American Journal of Sociology **78**, 1360 (1973).
- [32] L. Pappalardo, G. Rossetti, and D. Pedreschi, in *2012 IEEE/ACM International Conference on Advances in*

- Social Networks Analysis and Mining* (IEEE, 2012) pp. 1040–1045.
- [33] J. M. Kumpula, J.-P. Onnela, J. Saramäki, K. Kaski, and J. Kertész, *Physical Review Letters* **99**, 228701 (2007).
- [34] H.-H. Jo, R. K. Pan, and K. Kaski, *PLoS ONE* **6**, e22687 (2011), arXiv:1106.0288.
- [35] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, *Nature* **435**, 814 (2005).
- [36] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, *Nature* **466**, 761 (2010).
- [37] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, *Journal of Complex Networks* **2**, 203 (2014).
- [38] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña Nadal, Z. Wang, and M. Zanin, *Physics Reports* **544**, 1 (2014).
- [39] Y. Murase, J. Török, H.-H. Jo, K. Kaski, and J. Kertész, *Physical Review E* **90**, 052810 (2014).
- [40] R. Toivonen, L. Kovanen, M. Kivelä, J.-P. Onnela, J. Saramäki, and K. Kaski, *Social Networks* **31**, 240 (2009).
- [41] F. Papadopoulos, M. Kitsak, M. A. Serrano, M. Boguñá, and D. Krioukov, *Nature* **489**, 537 (2012).
- [42] J. P. Bagrow and D. Brockmann, *Physical Review X* **3**, 021016 (2013), arXiv:1209.3307.
- [43] P. Erdős and A. Rényi, *Publication of the Mathematical Institute of the Hungarian Academy of Sciences* **5**, 17 (1960).
- [44] A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airoldi, *Found. Trends Mach. Learn.* **2**, 129 (2010).
- [45] E. M. Airoldi, T. B. Costa, and S. H. Chan, in *Advances in Neural Information Processing Systems 26*, edited by C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Curran Associates, Inc., 2013) pp. 692–700.
- [46] D. Garlaschelli, *New Journal of Physics* **11**, 073005 (2009).
- [47] C. Seshadhri, T. G. Kolda, and A. Pinar, *Physical Review E* **85**, 056109 (2012).
- [48] M. E. J. Newman, *Networks: An Introduction*, 1st ed. (Oxford University Press, 2010).
- [49] B. Karrer and M. E. J. Newman, *Physical Review E* **82**, 066118 (2010).
- [50] T. Britton, M. Deijfen, and F. Liljeros, *Journal of Statistical Physics*, *Journal of Statistical Physics* **145**, 1368 (2011).
- [51] J. Palowitch, S. Bhamidi, and A. B. Nobel, “The continuous configuration model: A null for community detection on weighted networks,” (2016), arXiv:1601.05630.
- [52] D. Jin, D. He, Q. Hu, C. Baquero, and B. Yang, *Journal of Statistical Mechanics: Theory and Experiment* **2013**, P09013 (2013).
- [53] C. Stegehuis, R. van der Hofstad, and J. S. H. van Leeuwen, *Physical Review E* **94**, 012302 (2016).
- [54] P. W. Holland, K. B. Laskey, and S. Leinhardt, *Social Networks* **5**, 109 (1983).
- [55] K. Faust and S. Wasserman, *Social Networks* **14**, 5 (1992).
- [56] T. A. B. Snijders and K. Nowicki, *Journal of Classification* **14**, 75 (1997).
- [57] B. Karrer and M. E. J. Newman, *Physical Review E* **83**, 016107 (2011).
- [58] P. Fronczak, A. Fronczak, and M. Bujok, *Physical Review E* **88**, 032810 (2013).
- [59] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, *Journal of Machine Learning Research* **9**, 1981 (2008), arXiv:0705.4485.
- [60] P. K. Gopalan and D. M. Blei, *Proceedings of the National Academy of Sciences* **110**, 14534 (2013).
- [61] C. Aicher, A. Z. Jacobs, and A. Clauset, “Adapting the stochastic block model to edge-weighted networks,” (2013), arXiv:1305.5782.
- [62] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, *Special Section: Advances in Exponential Random Graph (p^*) Models*, *Social Networks* **29**, 173 (2007).
- [63] J. Park and M. E. J. Newman, *Physical Review E* **70**, 066117 (2004), arXiv:cond-mat/0405566.
- [64] J. D. Wilson, M. J. Denny, S. Bhamidi, S. J. Cranmer, and B. A. Desmarais, *Social Networks* **49**, 37 (2017).
- [65] J. Szüle, D. Kondor, L. Dobos, I. Csabai, and G. Vattay, *PLoS ONE* **9**, e111973 (2014).
- [66] P. Mac Carron, K. Kaski, and R. Dunbar, *Social Networks* **47**, 151 (2016), arXiv:1604.02400.
- [67] M. Galesic, H. Olsson, and J. Rieskamp, *Psychological Science* **23**, 1515 (2012).
- [68] M. E. J. Newman, *SIAM Review* **45**, 167 (2003).
- [69] E. Noka and F. Hoxha, *Journal of Multidisciplinary Engineering Science and Technology* **3**, 4505 (2016).
- [70] D. Liu, N. Blenn, and P. Van Mieghem, *Procedia Computer Science* **9**, 1400 (2012).
- [71] P. A. Grabowicz, J. J. Ramasco, E. Moro, J. M. Pujol, and V. M. Eguiluz, *PLoS ONE* **7**, e29358 (2012).
- [72] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, *Proceedings of the National Academy of Sciences* **101**, 3747 (2004).
- [73] P. Csermely, *Weak Links: The Universal Key to the Stability of Networks and Complex Systems* (Springer, 2006).
- [74] T. Takaguchi, M. Nakamura, N. Sato, K. Yano, and N. Masuda, *Physical Review X* **1**, 011008 (2011).
- [75] S. Pajevic and D. Plenz, *Nature Physics* **8**, 429 (2012).
- [76] K. Koroleva and V. Stimac, in *Proceedings of the International Conference on Information Systems, ICIS 2012*, edited by F. G. Joey (Association for Information Systems, Orlando, USA, 2012).
- [77] Q. Ke and Y. Y. Ahn, *Physical Review E* **90**, 032804 (2014).
- [78] D. Stauffer and A. Aharony, *Introduction to percolation theory*, 2nd ed. (Taylor & Francis, 1994).
- [79] Note that the layers in the real social networks do not have to consist of the communities of the same size.
- [80] M. Karoński and T. Łuczak, *Journal of Computational and Applied Mathematics* **142**, 125 (2002).
- [81] E. López, *Journal of Physics A: Mathematical and Theoretical* **46**, 305003 (2013).
- [82] R. I. M. Dunbar, *Journal of Institutional Economics* **7**, 345 (2011).
- [83] C. Song, D. Wang, and A.-L. Barabási, “Connections between human dynamics and network science,” (2013), arXiv:1209.1411.
- [84] J. Saramäki, E. A. Leicht, E. López, S. G. B. Roberts, F. Reed-Tsochas, and R. I. M. Dunbar, *Proceedings of the National Academy of Sciences* **111**, 942 (2014).
- [85] Y. Murase, T. Uchitane, and N. Ito, *Physics Procedia* **57**, 73 (2014).