Estimating the resolution limit of the map equation in community detection

Tatsuro Kawamoto[1] and Martin Rosvall[2]

[1]*Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, 4259-G5-22, Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa 226-8502, Japan*
[2]*Integrated Science Lab, Department of Physics, Umeå University, SE-901 87 Umeå, Sweden*

A community detection algorithm is considered to have a resolution limit if the scale of the smallest modules that can be resolved depends on the size of the analyzed subnetwork. The resolution limit is known to prevent some community detection algorithms from accurately identifying the modular structure of a network. In fact, any global objective function for measuring the quality of a two-level assignment of nodes into modules must have some sort of resolution limit or an external resolution parameter. However, it is yet unknown how the resolution limit affects the so-called map equation, which is known to be an efficient objective function for community detection. We derive an analytical estimate and conclude that the resolution limit of the map equation is set by the total number of links between modules instead of the total number of links in the full network as for modularity. This mechanism makes the resolution limit much less restrictive for the map equation than for modularity; in practice, it is orders of magnitudes smaller. Furthermore, we argue that the effect of the resolution limit often results from shoehorning multilevel modular structures into two-level descriptions. As we show, the hierarchical map equation effectively eliminates the resolution limit for networks with nested multilevel modular structures.

I. INTRODUCTION

The ability to detect the community structure of networks plays an important role in the analysis of complex systems. Therefore, researchers have developed a suite of community detection algorithms based on different principles or heuristics [1–7]. While graph partitioning methods require the number of modules as input, community detection methods, such as *modularity* [1] and the *map equation* [3], intrinsically identify the number of modules [7]. The widely used method of modularity has been studied extensively [4,8–11], but less is known about the inner workings of the flow-based and information-theoretic map equation [12,13], despite its strong performance on benchmark networks with densely connected components when companioned with its search algorithm Infomap [14,15]. Interestingly, the resolution limit [16], which causes small modules to aggregate in larger modules in modularity maximization and can lead to poor performance in resolving actual communities of real networks, seems to have an unnoticeable effect on the map equation [17]. Since any global objective function for two-level community detection must have a resolution limit [18], or an external resolution parameter [19,20], it is important to understand how the map equation succeeds at suppressing the effect of the resolution limit. In this paper, we analytically derive the resolution limit of the map equation and show why the map equation can resolve a much wider range of module sizes than modularity can. Although the community structure does not necessarily mean the densely connected components, because the resolution limit is about a detectability of densely connected components, we focus on such a case.

The resolution limit is the consequential downside of methods that intrinsically identify a resolution scale in a network to determine the number of modules. We conceptually illustrate this fact with the two-level map equation applied to the global road network. The map equation framework seeks an optimal modular description of a random walker on the network. This maximum compression is achieved by balancing the description length of movements within and between modules. With many small modules, representing city neighborhoods, for example, the within-module description length is short at the cost of a very long between-module description length. Contrarily, with few large modules, representing continents, for example, the between-module description length is short at the cost of a very long within-module description length. Consequently, the optimal two-level description length is achieved by identifying modules of intermediate sizes, such as neighborhoods aggregated into cities. If, however, only a subnetwork was analyzed, such as the network of a single city, modules would likely correspond to neighborhoods. In contrast, with a so-called resolution limit-free method, and given an external and fixed resolution parameter, the modules identified in the subnetwork could also be identified in the full network [19,20]. However, circumventing the resolution limit does, of course, not in itself imply good performance in resolving actual communities of real networks [17]. Moreover, for real networks the network itself must be used to set the resolution parameter, and the method is no longer resolution limit-free [18]. This example makes clear that no two-level community detection method is resolution limit-free in practice and that the resolution limit can arise because a two-level method is applied to a multilevel structure with nested modules. We argue that this case should be considered unproblematic and show later that the natural solution is to use a multilevel community detection method [21].

However, it is more problematic when a method aggregates small modules in a plain modular structure. In this case, a mechanistic understanding of how a method performs is critical for successful application [16,22–24]. For example, in a network with L links, Fortunato and Barthélemy [16] showed that modularity may fail to detect a module of size less than about $\sqrt{L}$ links. This limit is a result of the intrinsic

scale of the method [11,12]. Similarly, for a stochastic block model with model selection based on the minimum description principle, the corresponding block size was found to be about $\sqrt{N}$ nodes [25]. It has been experimentally verified that the map equation has an upper limit of detectability; Infomap splits nonclique structures with large diameters, such as strings and lattices [12], but the resolution limit at which modules cannot be fully resolved by the map equation is still unknown. Unlike modularity and stochastic block models, which build on configuration and generative models, respectively, the map equation operates by compressing a modular description of flow on the network. The different machinery makes it especially interesting to better understand the effect of the resolution limit on the map equation.

II. THE MAP EQUATION

To derive the limit and show how it depends on the network structure, we begin by reviewing the machinery of the map equation, which takes advantage of the fundamental duality in information theory between finding regularities in data and compressing the data [26]. For community detection, the regularities naturally correspond to the modules in a network.

The purest form of a modular network consists of isolated cliques of nodes. With a random walker as a proxy for dynamics on a network, its movements to any node from any other node in the clique occur with equal probability. That is, node visits are independent and identically distributed. Consequently, the map equation's underlying code structure is based precisely on the assumption of independent and identically distributed node visits and module entries and exits such that it can efficiently compress the description length of the random walker's trajectory in a modular network.

Specifically, given a partition M of nodes i assigned to modules $\iota = 1, 2, \ldots, m$, the map equation measures the per-step average description length $L(\mathsf{M})$ of dynamics on a network. The movements are encoded as follows: *m module codebooks*, one for each module, map node visits within modules and exits from modules to codewords for describing movements within modules; and one *index codebook* maps entries into modules to codewords for describing movements between modules. The length of the codewords are optimally derived from the rates of the corresponding movements they describe. However, the explicit description with codewords is not necessary for taking advantage of the duality between finding regularities in data and compressing the data. Instead, only the description length is required. Therefore, for a given partition of the network, the map equation simply measures the average codelength of each codebook and weights them by how often they are used. In any case, the modular partition that provides the most efficient compression of the random walker's movements also best captures the community structure with respect to the dynamics on the network.

According to Shannon's source coding theorem [26], the average minimum description length of each codebook directly from the associated probability distribution X of the corresponding node-visit and module-transition rates in terms of the Shannon entropy $H(X) = -\sum_i P(x_i) \log_2 P(x_i)$. Then, the complete average description length $L(\mathsf{M})$ is simply the sum of the average description length of all codebooks

weighted by their rate of use. That is,

$$L(\mathsf{M}) = q_\curvearrowright H(\mathcal{Q}) + \sum_{i=1}^{m} p_{i\circlearrowleft} H(\mathcal{P}_i). \tag{1}$$

The first term is the average description length of the index codebook. Its rate of use $q_\curvearrowright = \sum_{i=1}^{m} q_{i\curvearrowright}$ is the sum of the entering rates $q_{i\curvearrowright}$ into each module i, and $H(\mathcal{Q}) = -\sum_{i=1}^{m} (q_{i\curvearrowright}/q_\curvearrowright) \log_2(q_{i\curvearrowright}/q_\curvearrowright)$ is the average description length of the index codebook given by the Shannon entropy of the entering rates into the modules. The second term is the average description length of the module codebooks. The rate of use of module codebook i, $p_{i\circlearrowleft} = q_{i\curvearrowright} + \sum_{\alpha \in i} p_\alpha$, is the sum of the exiting rate $q_{i\curvearrowright}$ and the visiting rates p_α of all nodes α in module i, and $H(\mathcal{P}^i) = -(q_{i\curvearrowright}/p_{i\circlearrowleft}) \log_2(q_{i\curvearrowright}/p_{i\circlearrowleft}) - \sum_{\alpha \in i}(p_\alpha/p_{i\circlearrowleft}) \log_2(p_\alpha/p_{i\circlearrowleft})$ is the average description length of module codebook i. All rates are evaluated at stationarity, such that for undirected networks the rate of entering $q_{i\curvearrowright}$ and exiting $q_{i\curvearrowright}$ a given module i are the same. Using this equality and expansion of Eq. (1) gives

$$L(\mathsf{M}) = q_\curvearrowright \log_2 q_\curvearrowright - 2\sum_{i=1}^{m} q_{i\curvearrowright} \log_2 q_{i\curvearrowright}$$
$$+ \sum_{i=1}^{m} p_{i\circlearrowleft} \log_2 p_{i\circlearrowleft} - \sum_{\alpha} p_\alpha \log_2 p_\alpha. \tag{2}$$

Since the sum in the last term runs over all nodes, it does not depend on the choice of partition.

For undirected, unweighted networks, the exit probability $q_{i\curvearrowright}$ from module i and the rate of use $p_{i\circlearrowleft}$ of module codebook i can be expressed in terms of the number of links as

$$p_\alpha = \frac{k_\alpha}{K}, \tag{3}$$

$$q_{i\curvearrowright} = \sum_{\beta \notin i} \sum_{\alpha \in i} T_{\beta\alpha} p_\alpha = \frac{l_i^{\text{out}}}{K}, \tag{4}$$

$$p_{i\circlearrowleft} = q_{i\curvearrowright} + \sum_{\alpha \in i} p_\alpha = 2\frac{l_i + l_i^{\text{out}}}{K}, \tag{5}$$

where k_α is the degree of node α, $T_{\beta\alpha}$ is the transition probability that the random walker moves from node α to node β, l_i^{out} is the number of links which connect the nodes in module i with nodes in other modules, l_i is the number of links within module i, and $K = 2L = \sum_i (2l_i + l_i^{\text{out}})$ is the total degree of the network. Substituting these expressions into Eq. (2) gives

$$L(\mathsf{M}) = \frac{1}{K}\left[2C \log_2 2C - 2\sum_{i=1}^{m} l_i^{\text{out}} \log_2 l_i^{\text{out}} + K + 2C \right.$$
$$\left. + 2\sum_{i=1}^{m} \left(l_i + l_i^{\text{out}}\right) \log_2 \left(l_i + l_i^{\text{out}}\right) - \sum_{\alpha} k_\alpha \log_2 k_\alpha \right],$$
$$\tag{6}$$

where k_α is the degree of the node α and C is the *cut size* [6,7], i.e., $2C = \sum_{i=1}^{m} l_i^{\text{out}}$.

III. THE RESOLUTION LIMIT OF THE MAP EQUATION

The resolution limit of the map equation depends on which partitions are favored when the map equation is minimized. In general, to minimize Eq. (6), we update the partition $\mathsf{M} = \mathsf{A}$ to a nearby partition $\mathsf{M} = \mathsf{B}$ if $\Delta L(\mathsf{M}) = L(\mathsf{B}) - L(\mathsf{A}) < 0$. From Eq. (6), we readily see that there is no resolution limit caused by the total degree K, because it never changes the sign of $\Delta L(\mathsf{M})$. Instead, the resolution limit must depend on the cut size, which controls the only global term. In other words, as long as the update conserves the cut size C, there is no restriction from the global structure under an arbitrary update. Note that the resolution limit of the stochastic block model with model selection based on the minimum description principle depends on the number of nodes in the network [25]. Therefore, the dependence on the cut size is not a result of the information-theoretic nature of the map equation per se, but rather its machinery to describe flow trajectories on networks.

To identify which partition updates will be accepted, we denote an element in the two sums over modules in Eq. (6) by

$$\mathcal{L}_i = -l_i^{\text{out}} \log_2 l_i^{\text{out}} + \left(l_i + l_i^{\text{out}}\right) \log_2 \left(l_i + l_i^{\text{out}}\right), \quad (7)$$

and denote the total change over all modules before and after the update by

$$R = \sum_{i'} \mathcal{L}_{i'}(\mathsf{B}) - \sum_{i} \mathcal{L}_i(\mathsf{A}). \quad (8)$$

We now consider a local update where the cut size is decreased by a small δ, such that $C \gg \delta > 0$. Then

$$\Delta L(\mathsf{M}) = \frac{1}{K} [2(C - \delta) \log_2 (2(C - \delta)) - 2C \log_2 (2C)$$
$$+ 2R - 2\delta] \quad (9)$$

$$\simeq \frac{2}{K} [-\delta(2 + \log_2(e\,C)) + R], \quad (10)$$

where e is the basis of natural logarithm. Therefore, any local update should be accepted if

$$R \lesssim \delta(2 + \log_2(e\,C)). \quad (11)$$

For updates that increase the cut size, i.e., $\delta < 0$, on the other hand, no local update will be accepted as long as the cut size C is sufficiently large.

To identify the resolution limit, we parallel the analysis of Ref. [16] and pinpoint a partition at the point where the map equation can resolve small modules. As shown in Fig. 1, we use a partition A with two modules $\mathcal{M}_1$ and $\mathcal{M}_2$ connected with l_{int} links between them, and l_{13} and l_{23} links, respectively, with the rest of the network $\mathcal{M}_3$. Note that $\mathcal{M}_3$ may consist of many modules, so that Fig. 1 represents a completely general situation. In partition B, the two modules $\mathcal{M}_1$ and $\mathcal{M}_2$ are merged into a single module $\mathcal{M}_{12}$. We then consider an update from partition A to partition B. From Eq. (8) we have

$$R = - l_3^{\text{out}} \log_2 l_3^{\text{out}} + l_2^{\text{out}} \log_2 l_2^{\text{out}} + l_1^{\text{out}} \log_2 l_1^{\text{out}}$$
$$+ \left(l_1 + l_1^{\text{out}} + l_2 + l_2^{\text{out}} - l_{\text{int}}\right) \log_2 \left(l_1 + l_1^{\text{out}} + l_2 + l_2^{\text{out}} - l_{\text{int}}\right)$$
$$- \left(l_2 + l_2^{\text{out}}\right) \log_2 \left(l_2 + l_2^{\text{out}}\right) - \left(l_1 + l_1^{\text{out}}\right) \log_2 \left(l_1 + l_1^{\text{out}}\right), \quad (12)$$
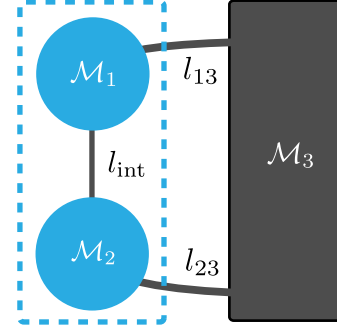

FIG. 1. (Color online) A schematic picture for the greedy update of two modules $\mathcal{M}_1$ and $\mathcal{M}_2$. Note that $\mathcal{M}_3$ may consist of many modules.

where $l_3^{\text{out}} = l_{13} + l_{23}$. We now consider the extreme case in which $l_{\text{int}} = 1$, i.e., $\delta = 1$, and set the sizes of two modules equal, i.e., $l_1 = l_2 = l_c$, because it maximizes R in $0 < l_2 \leqslant l_1$ for a fixed l_1. We also set $l_{13} = l_{23} = h$. Then, using the assumption that $l_c + h \gg 1$, we have

$$R \simeq 1 + 2[l_c + (1 + h) \log_2(1 + h) - h \log_2 h]$$
$$- \log_2 [e(l_c + h)]. \quad (13)$$

Assuming that $h = 1$ and using Eq. (11), we obtain the inequality for the resolution limit

$$\frac{4^{l_c}}{l_c + 1} \lesssim C, \quad (14)$$

where we have dropped a small constant factor $8/e^2 \simeq 1.08$ on the left-hand side, in order to highlight the basic scaling. Assuming that $h \gg 1$, $(1 + h) \log_2(1 + h) - h \log_2 h \simeq \log_2 [e(1 + h)]$, we instead have

$$\frac{(1 + h)^2 4^{l_c}}{2(l_c + h)} \lesssim C. \quad (15)$$

Accordingly, the map equation fails to detect a module with less than l_c links whenever the cut size C satisfies the above conditions, provided that the module is adjacent to modules of equal or smaller size. Note that Eqs. (14) and (15) apply only when evaluated close to the global minimum of the map equation. Otherwise, they are only practical restrictions during an optimization process. Furthermore, with $l_1 = l_2 = l$ and $l_{13} = l_{23} = h$ in Fig. 1, the following two examples are worth mentioning. First, if $l_{\text{int}} = 0$, $\Delta L(\mathsf{M}) = 4l/K > 0$, and disconnected modules never merge with other modules, as they should not. Second, if $\mathcal{M}_3$ is a single module, $\Delta L(h = 1) > 0$ for $l \geqslant 2$ and $\partial \Delta L/\partial h > 0$, and $\Delta L(h \to \infty) = 4(l - 1)/K$ for any l, such that the map equation can detect modules of arbitrary size with $l \geqslant 2$.

To illustrate that the resolution limit of the map equation is much smaller in practice and less restrictive than it is for modularity, we consider a simple modular network shown in the inset of Fig. 2. The network consists of a ring of m modules, each forming a clique with n nodes. For this network, $C = m$. The resolution limit of the map equation is in practice many orders of magnitudes smaller than the resolution limit of modularity. For clique size 6, for example,
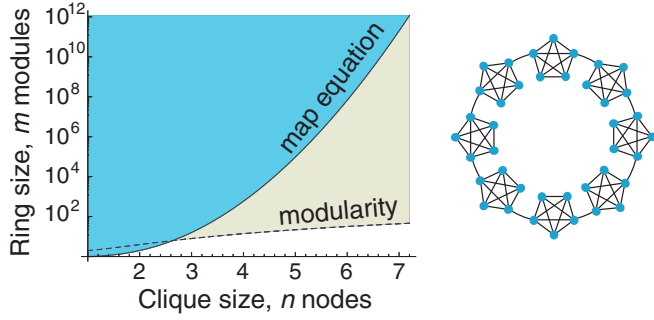
FIG. 2. (Color online) Detectable region of m modules, each forming a clique with n nodes. In the inset network, $n = 5$ and $m = 8$. The solid line is the resolution limits of the map equation according to Eq. (14), while the dashed line is the resolution limit of modularity, $m = n(n - 1) + 2$ [16].

FIG. 3. (Color online) Detected sizes of modules for each Sierpinski triangle of different size l with the two-level method (circular points) and the multilevel method (cross points). For example, the two-level method detects modules of 3 links and 12 links when the hierarchy of the Sierpinski triangle is three. The Sierpinski triangles up to three levels are illustrated at the top. The boundary of the shaded region shows the resolution limit of the module size obtained with Eq. (14).

the map equation can resolve modules in a ring network that is millions of times larger. If this intrinsic scale of resolution set by using one node codeword per step is not appropriate for the problem at hand and a multilevel solution is not an option, the scale can be modified by the Markov time, which effectively works as an intrinsic resolution parameter [11,12]. More importantly, the resolution limit does not depend on the number of links in the network, as for modularity, but on the cut size. This feature makes the resolution limit less restrictive with important performance implications.

IV. THE RESOLUTION LIMIT OF THE HIERARCHICAL MAP EQUATION

We now turn to the hierarchical map equation [21], the multilevel generalization of the two-level method described above. For example, the three-level map equation can consider movements within and between supermodules, modules in supermodules, and nodes in modules. In the general case, this hierarchy of nested modules can be extended locally and independently between branches as long as it reduces the minimal description length. Because modification of a partition at a certain level of a branch only influences the description length of movements within and between affected modules, $\Delta L(\mathsf{M})$ for the hierarchical map equation turns out to be analogous to that of the two-level map equation (see Sec. 1 of the Appendix for details). However, and importantly, the resolution limit now depends on the structure of the associated supermodule rather than on the structure of the full network. As a result, in the absence of a nested multilevel modular structure for the hierarchical map equation to capitalize on, the resolution limit of the two-level method remains. However, with a sufficiently pronounced nested multilevel modular structure, the hierarchical map equation will resolve all modules (see Secs. 2 and 3 of the Appendix for details). A similar effect is true also for methods based on generative models. For example, the typical detectable block size decreases from $\sim\sqrt{N}$ to $\sim\ln N$ for a hierarchical generalization of the stochastic block model mentioned above [27]. In analogy with the two-level analysis, since the hierarchical map equation determines the number of levels
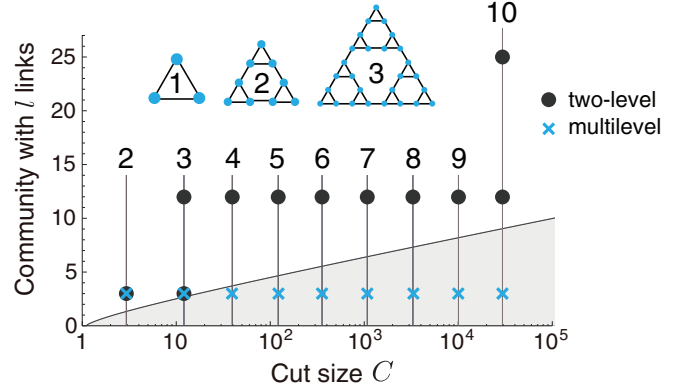
of each branch intrinsically based on the network structure rather than with an external parameter, it is not resolution limit-free in the sense that analysis of the full network or a subnetwork necessarily would give the same result [20]. Nevertheless, as long as the network has a pronounced nested multilevel modular structure, modules at the finest level will be resolved.

To demonstrate the performance of the hierarchical map equation, we use Sierpinski triangles, as shown in the inset of Fig. 3. Using the code distributed at [28], we identify modules in the Sierpinski triangles of different sizes with the two-level and multilevel methods. For the multilevel method, we focus on the results at the finest level. While the result of the two-level method is harmed by the resolution limit, the multilevel method detects the triangles at the lowest level for any network size. Also, cliques in a ring of cliques, as illustrated in Fig. 2, are resolved for any network size (see Sec. 3 of the Appendix for analytical derivation). Furthermore, as we see in the next section, the relaxed resolution limit can be observed in real networks as well. Therefore, we conclude that the hierarchical map equation effectively eliminates the resolution limit for networks with nested multilevel modular structures.

V. EFFECTS OF THE RESOLUTION LIMIT: MODULE SIZE DISTRIBUTIONS OF REAL NETWORKS

In this section, we show that the relaxed resolution limit can be observed in analysis of real networks. Figure 4 shows the module size distributions, which we obtained by running Infomap [28] on the data distributed at [29,30]. The size of each network and the total number of detected modules are listed in Table I. Here size refers to the number of nodes in a module instead of the number of internal links, but the effect of the resolution limit is nevertheless clear. The multilevel method detects many more smaller modules [31].
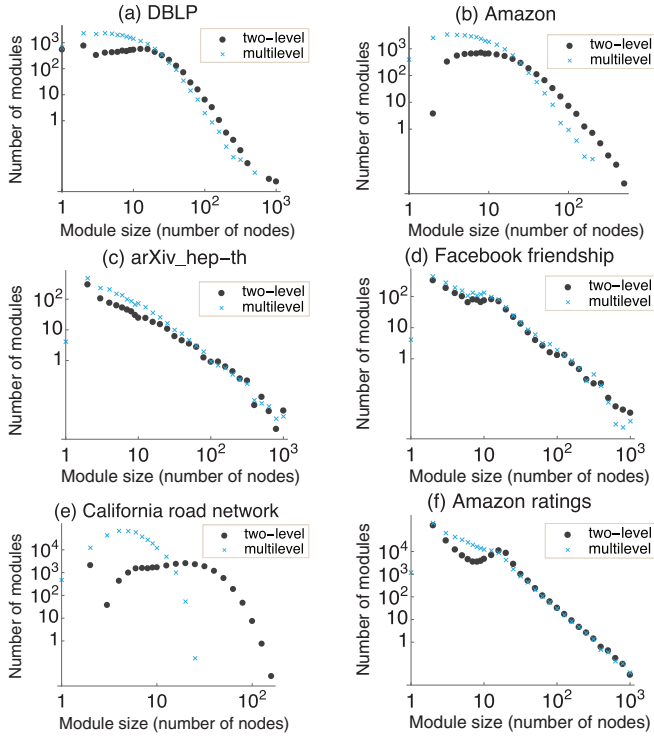
FIG. 4. (Color online) Module size distributions of (a) the collaboration network of authors of scientific papers from DBLP computer science bibliography [29,33], (b) the copurchasing network of products in Amazon.com [29,33], (c) the network of citations in arXiv's High Energy Physics-Theory (hep-th) section [34], (d) the friendship network on Facebook [35], (e) the road network of California [36] in the USA, and (f) the bipartite rating network of products in Amazon.com [37] in the log-log scale with partial-logarithmic binning. The circular points represent the result of the two-level method and the cross points represent the result in the finest level of the multilevel method. Note that the size of a module here does not indicate the number of internal links, but the number of nodes within the module.

Note that the Amazon rating network is a bipartite network; the random walker in a bipartite network has periodic stationary states by nature, but we assume the nonperiodic solution with visit rates given by Eq. (3) and all derived results apply. For some of these networks, such as the arXiv citation network and the Facebook friendship network, the effect of the resolution limit on the two-level method looks small, because these networks are dense compared to the other networks

and the depth of the hierarchy in the multilevel method is shallow.

We can estimate the theoretical resolution limit by estimating the cut size in Eq. (14). The cut size is bounded below by the number of modules detected by the multilevel method and above by the number of links in the network. For the DBLP network, these numbers are 29 252 and 1 049 866, respectively, and the left-hand side of Eq. (14) in the main text falls between these values for $l_c \approx 11$ [32].

As Fig. 4 shows, because real networks have modules of varying strength, the resolution limit does not force a clear separation between detected and undetected module sizes. For real networks, the theoretical resolution limit is instead the point at which we can expect deviations between a two-level method and a multilevel method.

VI. CONCLUSION

In summary, we have revealed the inner workings of the map equation and estimated its resolution limit [Eq. (14)]. While the number of links in the network determines the resolution limit of the configuration-model-based modularity, the number of links between modules instead determines the resolution limit of the flow-based map equation. This less restrictive dependence contributes to the performance difference between the methods. Even if the resolution limit in practice is many orders of magnitudes smaller than it is for modularity, for sufficiently large networks the map equation will eventually be affected, as any global two-level objective function is in practice. We argue that the natural solution is to use the hierarchical map equation, and exemplify both with synthetic and real networks. We conclude that multilevel methods or Markov time sweeping should always be the first choice for simplifying large networks, but that better tools are needed for efficiently working with such structures. Finally, we emphasize that coping with the resolution limit does not in itself imply good performance in resolving actual communities of real networks. It is still an open question as to what structures can be detected in conventional networks and when possibly higher-order information is necessary.

TABLE I. Community sizes in real networks obtained with the two-level and the multilevel map equation.

	No. of nodes	No. of links	No. of communities (two-level)	No. of communities (multilevel)
DBLP	338 029	1 049 866	16 450	29 252
Amazon (copurchase)	334 863	925 872	15 685	34 802
Facebook	63 731	1 269 502	2 268	2 819
arXiv_hep-th	48 239	352 807	1 332	2 247
California	1 965 206	2 766 607	82 322	344 485
Amazon (rating)	3 376 972	5 838 041	350 419	480 810

APPENDIX: THE HIERARCHICAL MAP EQUATION

1. General argument on partitioning

Here we explain the details of the hierarchical map equation [21]. Again, we restrict ourselves to undirected, unweighted networks. Analogously to the two-level method, the quality function of the multilevel method is defined by

$$L(\mathsf{M}) = q_\curvearrowright H(\mathcal{Q}) + \sum_{i=1}^m q_{i\circlearrowleft} H(\mathcal{Q}_i) + \sum_{i=1}^m \sum_{j=1}^{m_i} q_{ij\circlearrowleft} H(\mathcal{Q}_{ij})$$
$$+ \cdots + \sum_{ij,\ldots,k} p_{ij,\ldots,k\circlearrowleft} H(\mathcal{P}_{ij,\ldots,k}), \tag{A1}$$

with

$$H(\mathcal{Q}) = - \sum_{i=1}^m \frac{q_{i\curvearrowright}}{q_\curvearrowright} \log_2 \frac{q_{i\curvearrowright}}{q_\curvearrowright}, \tag{A2}$$

$$\sum_{i=1}^m H(\mathcal{Q}_i) = - \sum_{i=1}^m \frac{q_{i\curvearrowright}}{q_{i\circlearrowleft}} \log_2 \frac{q_{i\curvearrowright}}{q_{i\circlearrowleft}} - \sum_{i=1}^m \sum_{j=1}^{m_i} \frac{q_{ij\curvearrowright}}{q_{i\circlearrowleft}} \log_2 \frac{q_{ij\curvearrowright}}{q_{i\circlearrowleft}}, \tag{A3}$$

$$\sum_{ij,\ldots,k} H(\mathcal{P}_{ij,\ldots,k}) = - \sum_{ij,\ldots,k} \frac{q_{ij,\ldots,k\curvearrowright}}{p_{ij,\ldots,k\circlearrowleft}} \log_2 \frac{q_{ij,\ldots,k\curvearrowright}}{p_{ij,\ldots,k\circlearrowleft}} - \sum_{ij,\ldots,k} \sum_{\alpha \in ij,\ldots,k} \frac{p_\alpha}{p_{ij,\ldots,k\circlearrowleft}} \log_2 \frac{p_\alpha}{p_{ij,\ldots,k\circlearrowleft}}, \tag{A4}$$

where $q_{i_1 i_2,\ldots,i_x\curvearrowright}$ is the probability that a random walker exits from the module i_x in the xth level and $q_{i_1 i_2,\ldots,i_x\circlearrowleft}$ is the probability that the random walker stays within the module i_x and exits from it. As we mentioned in the main text, the probabilities of entering and exiting a module are equal for undirected networks. Thus, we replaced the entering probability with the exiting probability in order to simplify the notation. As before, p_α is the probability that the random walker visits node α. Following what we did for the two-level method, we can write (A1) as

$$L(\mathsf{M}) = q_\curvearrowright \log_2 q_\curvearrowright + \sum_{i_1} q_{i_1\circlearrowleft} \log_2 q_{i_1\circlearrowleft} + \sum_{i_1 i_2} q_{i_1 i_2\circlearrowleft} \log_2 q_{i_1 i_2\circlearrowleft} + \cdots + \sum_{i_1 i_2,\ldots,i_k} q_{i_1 i_2,\ldots,i_k\circlearrowleft} \log_2 q_{i_1 i_2,\ldots,i_k\circlearrowleft} - \sum_\alpha p_\alpha \log_2 p_\alpha$$
$$- 2 \left(\sum_{i_1} q_{i_1\curvearrowright} \log_2 q_{i_1\curvearrowright} + \sum_{i_1 i_2} q_{i_1 i_2\curvearrowright} \log_2 q_{i_1 i_2\curvearrowright} + \cdots + \sum_{i_1 i_2,\ldots,i_k} q_{i_1 i_2,\ldots,i_k\curvearrowright} \log_2 q_{i_1 i_2,\ldots,i_k\curvearrowright} \right). \tag{A5}$$

We then consider the difference of the map equation at an update. The modification of the partition at a certain level does not affect the partitions in higher and lower levels. It only alters the description length of movements between the modules of the modified level, as well as the description length of movements within the modified module (it can also be regarded as movements between submodules) and exiting from it. Therefore, the difference of the map equation from an update in the xth level is

$$\Delta L(\mathsf{M}) = q^{\mathsf{B}}_{i_1 i_2,\ldots,i_{x-1}\circlearrowleft} \log_2 q^{\mathsf{B}}_{i_1 i_2,\ldots,i_{x-1}\circlearrowleft} - q^{\mathsf{A}}_{i_1 i_2,\ldots,i_{x-1}\circlearrowleft} \log_2 q^{\mathsf{A}}_{i_1 i_2,\ldots,i_{x-1}\circlearrowleft}$$
$$- 2 \left(\sum_{i'_x} q^{\mathsf{B}}_{i_1 i_2,\ldots,i'_x\curvearrowright} \log_2 q^{\mathsf{B}}_{i_1 i_2,\ldots,i'_x\curvearrowright} - \sum_{i_x} q^{\mathsf{A}}_{i_1 i_2,\ldots,i_x\curvearrowright} \log_2 q^{\mathsf{A}}_{i_1 i_2,\ldots,i_x\curvearrowright} \right)$$
$$+ \left(\sum_{i'_x} q^{\mathsf{B}}_{i_1 i_2,\ldots,i'_x\circlearrowleft} \log_2 q^{\mathsf{B}}_{i_1 i_2,\ldots,i'_x\circlearrowleft} - \sum_{i_x} q^{\mathsf{A}}_{i_1 i_2,\ldots,i_x\circlearrowleft} \log_2 q^{\mathsf{A}}_{i_1 i_2,\ldots,i_x\circlearrowleft} \right), \tag{A6}$$

which is analogous to that of the two-level method,

$$\Delta L_{\text{two-level}}(\mathsf{M}) = q^{\mathsf{B}}_\curvearrowright \log_2 q^{\mathsf{B}}_\curvearrowright - q^{\mathsf{A}}_\curvearrowright \log_2 q^{\mathsf{A}}_\curvearrowright$$
$$- 2 \left(\sum_{i'=1}^{m'} q^{\mathsf{B}}_{i'\curvearrowright} \log_2 q^{\mathsf{B}}_{i'\curvearrowright} - \sum_{i=1}^m q^{\mathsf{A}}_{i\curvearrowright} \log_2 q^{\mathsf{A}}_{i\curvearrowright} \right) + \left(\sum_{i'=1}^{m'} p^{\mathsf{B}}_{i'\circlearrowleft} \log_2 p^{\mathsf{B}}_{i'\circlearrowleft} - \sum_{i=1}^m p^{\mathsf{A}}_{i\circlearrowleft} \log_2 p^{\mathsf{A}}_{i\circlearrowleft} \right). \tag{A7}$$

Instead of $q_\curvearrowright$ in the two-level method in the above equation, we have

$$q_{i_1 i_2,\ldots,i_{x-1}\circlearrowleft} = q_{i_1 i_2,\ldots,i_{x-1}\curvearrowright} + \sum_{i_x} q_{i_1 i_2,\ldots,i_x\curvearrowright} \tag{A8}$$

for the multilevel method; hence, other than the extra term $q_{i_1 i_2,\ldots,i_{x-1}\curvearrowright}$ in $q_{i_1 i_2,\ldots,i_{x-1}\circlearrowleft}$, the mathematical structure of the hierarchical map equation is analogous to that of the two-level method. That is, the difference between these methods is the effective network size in each update process, as exemplified in Fig. 5 for the Sierpinski triangle. The above result can also be interpreted as follows. The resolution limit of a subgraph H in a graph G remains the same in a subgraph S, where $H \subset S \subset G$, as long as the subgraph S strictly includes the supermodule
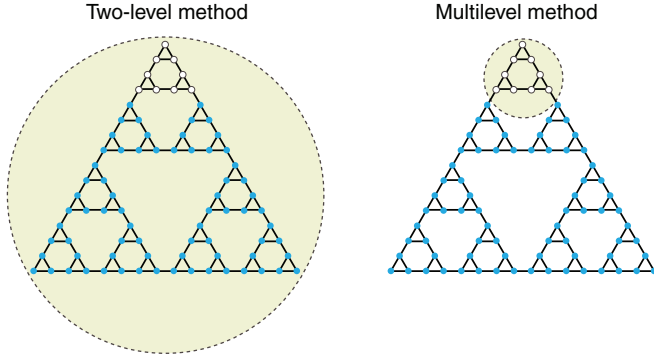
FIG. 5. (Color online) Effective network sizes (the region encircled with the dashed line) for the evaluation of partitioning of the nine nodes at the top of the Sierpinski triangle (the nodes of open circles) in the two-level method and in the finest level of the multilevel method. While the effective network size is the size of the whole network in the two-level method, in the finest level of the multilevel method, it is of the supermodule plus the links lying between the supermodule and the rest of the network.

containing the subgraph H, provided that the structure of the supermodule is conserved in the graph G and the subgraph S. This property is somewhat similar to a mathematical definition of a "resolution limit-free method" in Ref. [20]. Note, however, that the hierarchical map equation is not a resolution limit-free method in their sense, because the hierarchical structure may change for a particular choice of the subgraph S.

As we did for the two-level method, we can write $\Delta L(\mathsf{M})$ of the multilevel method in terms of the number of links, by using the explicit form of the stationary distribution of the random walker. The elements of the objective function reads

$$q_{i_1 i_2, \ldots, i_x \curvearrowright} = \frac{l^{\text{out}}_{i_1 i_2, \ldots, i_x}}{K}, \tag{A9}$$

$$q_{i_1 i_2, \ldots, i_x \circlearrowleft} = \frac{2}{K}\left(l_{i_1 i_2, \ldots, i_x} + l^{\text{out}}_{i_1 i_2, \ldots, i_x}\right). \tag{A10}$$

Substituting them into Eq. (A6), after some algebra, we obtain

$$\frac{K}{2}\Delta L(\mathsf{M}) = \left(l^{\mathsf{B}}_{i_1 i_2, \ldots, i_{x-1}} + l^{\mathsf{B},\text{out}}_{i_1 i_2, \ldots, i_{x-1}}\right) \log_2 \left(l^{\mathsf{B}}_{i_1 i_2, \ldots, i_{x-1}} + l^{\mathsf{B},\text{out}}_{i_1 i_2, \ldots, i_{x-1}}\right) - \left(l^{\mathsf{A}}_{i_1 i_2, \ldots, i_{x-1}} + l^{\mathsf{A},\text{out}}_{i_1 i_2, \ldots, i_{x-1}}\right) \log_2 \left(l^{\mathsf{A}}_{i_1 i_2, \ldots, i_{x-1}} + l^{\mathsf{A},\text{out}}_{i_1 i_2, \ldots, i_{x-1}}\right)$$

$$- 2\left(l^{\mathsf{B}}_{i_1 i_2, \ldots, i_{x-1}} - l^{\mathsf{A}}_{i_1 i_2, \ldots, i_{x-1}}\right) + R_{i_1 i_2, \ldots, i_{x-1}}, \tag{A11}$$

where

$$R_{i_1 i_2, \ldots, i_{x-1}} = \sum_{i'_x} \mathcal{L}_{i_1 i_2, \ldots, i'_x}(\mathsf{B}) - \sum_{i_x} \mathcal{L}_{i_1 i_2, \ldots, i_x}(\mathsf{A}), \tag{A12}$$

$$\mathcal{L}_{i_1 i_2, \ldots, i_x}(\mathsf{M}) = -l^{\mathsf{M},\text{out}}_{i_1 i_2, \ldots, i_x} \log_2 l^{\mathsf{M},\text{out}}_{i_1 i_2, \ldots, i_x} + \left(l^{\mathsf{M}}_{i_1 i_2, \ldots, i_x} + l^{\mathsf{M},\text{out}}_{i_1 i_2, \ldots, i_x}\right) \log_2 \left(l^{\mathsf{M}}_{i_1 i_2, \ldots, i_x} + l^{\mathsf{M},\text{out}}_{i_1 i_2, \ldots, i_x}\right). \tag{A13}$$

Denoting $C_{i_1 i_2, \ldots, i_x} \equiv l_{i_1 i_2, \ldots, i_{x-1}} + l^{\text{out}}_{i_1 i_2, \ldots, i_{x-1}}$ and letting δ be the difference of $C_{i_1 i_2, \ldots, i_x}$ under an update, we have

$$\frac{K}{2}\Delta L(\mathsf{M}) \simeq -\delta\left(2 + \log_2\left(e\, C_{i_1 i_2, \ldots, i_x}\right)\right) + R_{i_1 i_2, \ldots, i_{x-1}}, \tag{A14}$$

where we assumed $\delta \ll l_{i_1 i_2, \ldots, x-1}$. Equation (A14) corresponds to Eq. (10) in the main text. With this correspondence, the same argument holds for the multilevel method as for the two-level method. Since Eq. (A14) depends only on the structure of a subnetwork instead of the full network, the multilevel method has higher resolution than the two-level method, provided that the network has a nested multilevel modular structure.

2. Generation of supermodules

We showed that the existence of the nested structure of modules in the multilevel method of the map equation enables us to resolve smaller modules than the two-level method. When the multilevel method generates only two levels, however, the multilevel method is equivalent to the two-level method, and the resolution limit would be kept the same. As we did for the update of partition in the two-level method in the main text, we can evaluate the generation of a higher level by comparing the partitions with and without a supermoudule. For simplicity, we compare a two-level partition and a three-level partition.

The objective functions of the map equation with two-level structure $L(\mathsf{M}_2)$ and three-level structure $L(\mathsf{M}_3)$ read

$$L(\mathsf{M}_2) = q^{(2)}_{\curvearrowright} \log_2 q^{(2)}_{\curvearrowright} - 2\sum_{i'=1}^{m'} q^{(2)}_{i'\curvearrowright} \log_2 q^{(2)}_{i'\curvearrowright}$$

$$+ \sum_{i'=1}^{m'} p^{(2)}_{i'\circlearrowleft} \log_2 p^{(2)}_{i'\circlearrowleft} - \sum_{\alpha} p_{\alpha} \log_2 p_{\alpha}, \tag{A15}$$

$$L(\mathsf{M}_3) = q^{(3)}_{\curvearrowright} \log_2 q^{(3)}_{\curvearrowright} - 2\sum_{i=1}^{m} q^{(3)}_{i\curvearrowright} \log_2 q^{(3)}_{i\curvearrowright} + \sum_{i=1}^{m} p^{(3)}_{i\circlearrowleft} \log_2 p^{(3)}_{i\circlearrowleft}$$

$$- 2\sum_{i=1}^{m}\sum_{j=1}^{m_i} q^{(3)}_{ij\curvearrowright} \log_2 q^{(3)}_{ij\curvearrowright} + \sum_{i=1}^{m}\sum_{j=1}^{m_i} p^{(3)}_{ij\circlearrowleft} \log_2 p^{(3)}_{ij\circlearrowleft}$$

$$- \sum_{\alpha} p_{\alpha} \log_2 p_{\alpha}. \tag{A16}$$

We consider the difference of the description length given by the map equation when we introduce a supermodule s in addition to the two-level partition, forming a three-level partition. In the following, we denote the label of a module in the three-level structure by i and the label of a submodule in module s by j [see Fig. 6(a)]. Based on Eqs. (A15) and (A16),
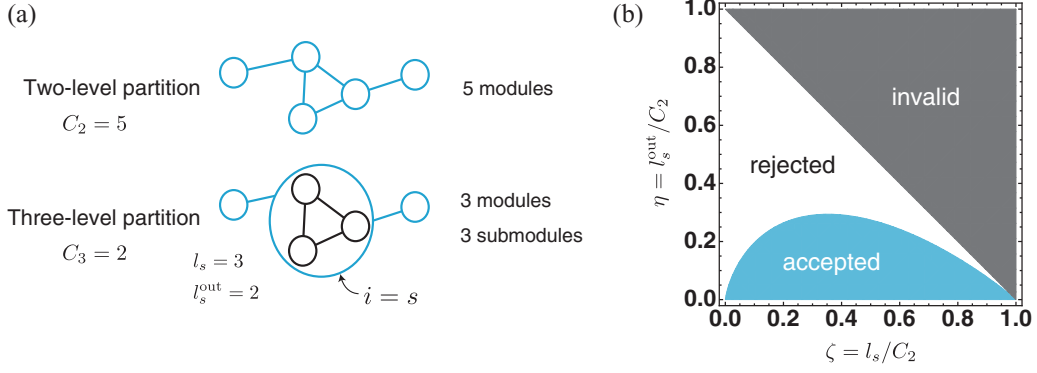
FIG. 6. (Color online) (a) An example of a hierarchical community structure in a network. (b) Phase diagram of the accepted and rejected regions for the generation of a higher level. The shaded region (A28) of the upper-right corner is invalid.

we have

$$
L(\mathsf{M}_3) - L(\mathsf{M}_2) = q_\frown^{(3)} \log_2 q_\frown^{(3)} - 2q_{s\frown}^{(3)} \log_2 q_{s\frown}^{(3)} + p_{s\circlearrowleft}^{(3)} \log_2 p_{s\circlearrowleft}^{(3)} - 2\sum_{j=1}^{m_s} q_{sj\frown}^{(3)} \log_2 q_{sj\frown}^{(3)} + \sum_{j=1}^{m_s} p_{sj\circlearrowleft}^{(3)} \log_2 p_{sj\circlearrowleft}^{(3)}
$$

$$
- \left(q_\frown^{(2)} \log_2 q_\frown^{(2)} - 2\sum_{j=1}^{m_s} q_{j\frown}^{(2)} \log_2 q_{j\frown}^{(2)} + \sum_{j=1}^{m_s} p_{j\circlearrowleft}^{(2)} \log_2 p_{j\circlearrowleft}^{(2)} \right)
$$

$$
= q_\frown^{(3)} \log_2 q_\frown^{(3)} - q_\frown^{(2)} \log_2 q_\frown^{(2)} - 2q_{s\frown}^{(3)} \log_2 q_{s\frown}^{(3)} + p_{s\circlearrowleft}^{(3)} \log_2 p_{s\circlearrowleft}^{(3)}, \tag{A17}
$$

where

$$
q_\frown^{(2)} = \sum_{i=1}^{m} q_{i\frown} - q_{s\frown} + \sum_{j=1}^{m_s} q_{sj\frown}, \quad q_\frown^{(3)} = \sum_{i=1}^{m} q_{i\frown}. \tag{A18}
$$

Note here that $q_{sj\frown}^{(3)} = q_{j\frown}^{(2)}$ and $p_{sj\circlearrowleft}^{(3)} = p_{j\circlearrowleft}^{(2)}$. When Eq. (A17) is negative, the construction of an additional level is accepted, i.e., the three-level description gives a shorter codelength.

For undirected networks, Eq. (A17) can be written in terms of the number of links as follows:

$$
L(\mathsf{M}_3) - L(\mathsf{M}_2)
$$

$$
= \frac{\sum_i l_i^{\text{out}}}{K} \log_2 \frac{\sum_i l_i^{\text{out}}}{K} - \left(\frac{\sum_i l_i^{\text{out}} + 2l_s}{K} \right) \log_2 \left(\frac{\sum_i l_i^{\text{out}} + 2l_s}{K} \right) - 2\frac{l_s^{\text{out}}}{K} \log_2 \frac{l_s^{\text{out}}}{K} + \left(2\frac{l_s + l_s^{\text{out}}}{K} \right) \log_2 \left(2\frac{l_s + l_s^{\text{out}}}{K} \right)
$$

$$
= \frac{2}{K} \left[C_3 \log_2 C_3 - (C_3 + l_s) \log_2 (C_3 + l_s) - l_s^{\text{out}} \log_2 l_s^{\text{out}} + (l_s^{\text{out}} + l_s) \log_2 (l_s^{\text{out}} + l_s) + l_s^{\text{out}} \right], \tag{A19}
$$

where l_s, l_i^{out}, and $C_3 = \sum_i l_i^{\text{out}}/2$ are variables of the coarsest level in the three-level partition, which are the number of links within the module s, the number of links between nodes within module i and nodes outside the module, and the cut size of the network, respectively.

As in the main text, we introduce

$$
\mathcal{L}(x,y) = -x \log_2 x + (x + y) \log_2 (x + y) \quad (x > 0 \text{ and } y > 0), \tag{A20}
$$

which has the following properties:

$$
\frac{\partial \mathcal{L}(x,y)}{\partial x} = \log_2 \left(1 + \frac{y}{x} \right) > 0, \tag{A21}
$$

$$
\mathcal{L}(x - y, y) = -\mathcal{L}(x, -y) \quad \text{for } x \geqslant y, \tag{A22}
$$

$$
\mathcal{L}(ax, ay) = ay \log_2 a + a\mathcal{L}(x,y) \quad \text{for } a > 0. \tag{A23}
$$

Using $\mathcal{L}(x,y)$, we can recast Eq. (A19) as

$$
\frac{K}{2}[L(\mathsf{M}_3) - L(\mathsf{M}_2)] = \mathcal{L}(l_s^{\text{out}}, l_s) - \mathcal{L}(C_3, l_s) + l_s^{\text{out}}. \tag{A24}
$$

Therefore, the transition to the three-level structure occurs when the following condition is satisfied:

$$
l_s^{\text{out}} < \mathcal{L}(C_3, l_s) - \mathcal{L}(l_s^{\text{out}}, l_s). \tag{A25}
$$

The right-hand-side of (A25) is always greater than or equal to zero, since $\mathcal{L}(x,y)$ has monotonicity (A21) and $C_3 \geqslant l_s^{\text{out}}$. If the partition at the coarsest level is a bisection, it would give $C_3 = l_s^{\text{out}}$, which never satisfies Eq. (A25).

In terms of the cut size in the two-level partition $C_2 = C_3 + l_s$, Eq. (A25) can also be written as

$$
l_s^{\text{out}} < \mathcal{L}(C_2 - l_s, l_s) - \mathcal{L}(l_s^{\text{out}}, l_s), \tag{A26}
$$

or using (A22) and (A23), we have

$$0 > \eta + \mathcal{L}(1, -\zeta) + \mathcal{L}(\eta,\zeta), \tag{A27}$$

where $\zeta = l_s/C_2$ and $\eta = l_s^{\text{out}}/C_2$. The phase diagram of the generation of a higher level determined by Eq. (A27) is shown in Fig. 6(b). Note that the cut size C_2 is bounded below by $l_s + l_s^{\text{out}} \leqslant C_2$; i.e.,

$$\zeta + \eta \leqslant 1. \tag{A28}$$

The boundary in (A27) never intersects with the boundary of the invalid region (A28). This observation can be confirmed from Eq. (A25), which never satisfies the inequality at the boundary of the invalid region $l_s^{\text{out}} = C_3 = C_2 - l_s$.

Notice that the analysis here is not the whole story of the optimization of the hierarchical partitioning. The path for optimization can be very complex, because the generation of the deeper hierarchy and the update of partition in each level occur simultaneously. Roughly speaking, while dense networks tend to have large modules with shallow hierarchies, the hierarchies of sparse networks tend to be deep, i.e., the multilevel method has very high resolution. This behavior can be confirmed in a synthetic graph (Sec. 3 of this appendix), as

well as in real networks (Sec. V). As we exemplify in the next section, however, we can use the result here in order to check whether an obtained partition is truly the optimal solution, or can be improved at least by adding another level.

3. Multilevel solution of a ring of cliques

In the previous section, we observed in Fig. 6(b) that the generation of higher levels will not be accepted when the network is only weakly modular. In this section, we exemplify with a ring of cliques that a very modular network indeed has a nested multilevel structure and that the modules of the finest levels are always resolved.

We set the number of cliques equal to m and refer to the number of links within a module as l, as in the main text. We also denote the number of modules in each level as m_g ($g = 1,2,,\ldots,d-1$). In the case of a ring of cliques, the number of links connected to nodes outside of a module is two at any level. Moreover, due to the symmetry of the graph, the sizes of modules for the same level must be equal. With the d-level map equation, we have the following objective function:

$$KL(\mathsf{M}) = 2m_1 \log_2(2m_1) + m_1(2 + 2m_2) \log_2(2 + 2m_2) + m_1 m_2(2 + 2m_3) \log_2(2 + 2m_3) + \cdots$$
$$+ \prod_{g=1}^{d-1} m_g[2 + (2l + 2)] \log_2 (2 + (2l + 2)) - 2\left(2m_1 + 2m_1 m_2 + \cdots + 2\prod_{g=1}^{d-1} m_g\right) - \sum_\alpha k_a \log_2 k_\alpha, \tag{A29}$$

$$\frac{K}{2}L(\mathsf{M}) = m_1 \log_2 m_1 + \sum_{k=2}^{d-1}\prod_{g=1}^{k-1} m_g(1 + m_k) \log_2(1 + m_k) + \prod_{g=1}^{d-1} m_g[(l + 1) + (l + 2)\log_2(l + 2)] - \frac{1}{2}\sum_\alpha k_a \log_2 k_\alpha. \tag{A30}$$

Also, since the number of finest modules m is fixed, we have the constraint

$$m = \prod_{g=1}^{d-1} m_g. \tag{A31}$$

Hence, the optimal solution is obtained by minimizing (A30) subject to (A31). Although the number of modules of each level m_g is an integer, if we approximate it as a continuous variable, the present problem becomes equivalent to solving the following Lagrange multiplier:
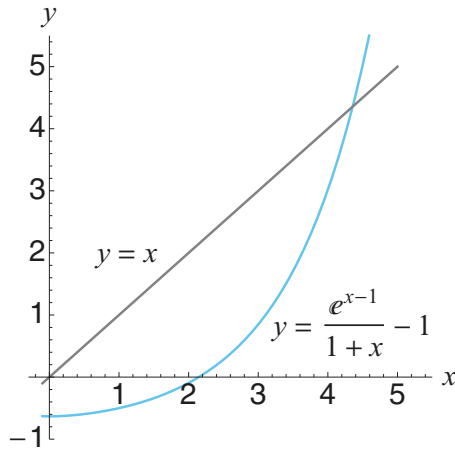
$$\delta\left\{m_1 \log_2 m_1 + \sum_{k=2}^{d-1}\prod_{g=1}^{k-1} m_g(1 + m_k) \log_2(1 + m_k) + \prod_{g=1}^{d-1} m_g[(l + 1) + (l + 2)\log_2(l + 2)] - \lambda\left(\prod_{g=1}^{d-1} m_g - m\right)\right\}$$
$$= \delta\left[m_1 \log_2 m_1 + \sum_{k=2}^{d-1}\prod_{g=1}^{k-1} m_g(1 + m_k) \log_2(1 + m_k) + \prod_{g=1}^{d-1} m_g \Xi(l,\lambda) + \lambda m\right] = 0, \tag{A32}$$

where we set

$$\Xi(l,\lambda) \equiv (l + 1) + (l + 2)\log_2(l + 2) - \lambda. \tag{A33}$$

After some straightforward calculations, we obtain the value of m_g as a function of m_{g+1} as follows:

$$m_1 = \frac{e^{m_2 - 1}}{1 + m_2},$$

$$m_g = \frac{e^{m_{g+1} - 1}}{1 + m_{g+1}} - 1 \quad \text{for } 2 \leqslant g \leqslant d - 2,$$

$$m_{d-1} = \frac{2^\Xi}{e} - 1. \tag{A34}$$

FIG. 7. (Color online) The plot $y = f(x)$ in Eq. (A35).

The value of λ in $\Xi(l,\lambda)$ is determined by the constraint of m, the total number of modules. Hence, the number of modules at the deepest level in a supermodule m_{d-1} is determined by

m, d, and l. In the actual multilevel method, the number of hierarchal levels d is adjusted so that the average codelength is minimized.

We now estimate the values of m_{d-1}. Defining

$$f(x) = \frac{e^{x-1}}{1+x} - 1, \qquad (A35)$$

we have $f(3) = 0.85$, $f(4) = 3.02$, and $f(5) = 8.10$ [see Fig. 7 for the shape of $f(x)$]. First, we readily see that $m_{d-1} \leqslant 3$ is not the choice for a large ring, because the number of modules in a higher level decreases in such cases, which restricts the value of m in (A31). If $m_{d-1} \geqslant 5$ and we assume that there are more than two levels, m_1 would be very large. However, this is unlikely the optimal, because in the phase diagram of Fig. 6(b), when $l_s^{out} = 2$ and m_1 is large enough, there must be a supermodule of a higher level which makes the total description length shorter. Therefore, the number of modules inside of a supermodule tends to be close to four in the ring of cliques. Accordingly, every clique is always detectable with the multilevel method no matter how large the size of the ring is.

[1] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, 026113 (2004).

[2] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Nature (London) **435**, 814 (2005).

[3] M. Rosvall and C. Bergstrom, Proc. Natl. Acad. Sci. U.S.A. **105**, 1118 (2008).

[4] J.-C. Delvenne, S. N. Yaliraki, and M. Barahona, Proc. Natl. Acad. Sci. U.S.A. **107**, 12755 (2010).

[5] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, Science **328**, 876 (2010).

[6] U. Luxburg, Stat. Comput. **17**, 395 (2007).

[7] S. Fortunato, Phys. Rep. **486**, 75 (2010).

[8] M. E. J. Newman, Phys. Rev. E **74**, 036104 (2006).

[9] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **103**, 8577 (2006).

[10] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, J. Stat. Mech. (2008) P10008.

[11] M. T. Schaub, J.-C. Delvenne, S. N. Yaliraki, and M. Barahona, PloS One **7**, e32210 (2012).

[12] M. T. Schaub, R. Lambiotte, and M. Barahona, Phys. Rev. E **86**, 026112 (2012).

[13] R. Lambiotte and M. Rosvall, Phys. Rev. E **85**, 056107 (2012).

[14] A. Lancichinetti and S. Fortunato, Phys. Rev. E **80**, 056117 (2009).

[15] R. Aldecoa and I. Marín, Sci. Rep. **3**, 2216 (2013).

[16] S. Fortunato and M. Barthélemy, Proc. Natl. Acad. Sci. U.S.A. **104**, 36 (2007).

[17] A. Lancichinetti and S. Fortunato, Phys. Rev. E **84**, 066122 (2011).

[18] V. A. Traag, G. Krings, and P. Van Dooren, Sci. Rep. **3**, 2930 (2013).

[19] P. Ronhovde and Z. Nussinov, Phys. Rev. E **81**, 046114 (2010).

[20] V. A. Traag, P. Van Dooren, and Y. Nesterov, Phys. Rev. E **84**, 016114 (2011).

[21] M. Rosvall and C. T. Bergstrom, PloS One **6**, e18209 (2011).

[22] J. M. Kumpula, J. Saramäki, K. Kaski, and J. Kertész, Eur. Phys. J. B **56**, 41 (2007).

[23] B. H. Good, Y.-A. de Montjoye, and A. Clauset, Phys. Rev. E **81**, 046106 (2010).

[24] J. W. Berry, B. Hendrickson, R. A. LaViolette, and C. A. Phillips, Phys. Rev. E **83**, 056119 (2011).

[25] T. P. Peixoto, Phys. Rev. Lett. **110**, 148701 (2013).

[26] C. E. Shannon, A Mathematical Theory of Communication **27**, 379 (1948).

[27] T. P. Peixoto, Phys. Rev. X **4**, 011047 (2014).

[28] http://www.mapequation.org/.

[29] http://snap.stanford.edu/data/.

[30] http://konect.uni-koblenz.de/.

[31] The multilevel solution is not constructed as a simple decomposition of the two-level solution, but by an attempt to find a global optimal modular description. Therefore, the multilevel method does not trivially detect small modules more than the two-level method does.

[32] Note that the number of nodes n within a module is bounded by the number of links l within the module by $n \leqslant l + 1$. The equality holds for a module of a tree. Therefore, the results of the two-level method and the multilevel method should deviate from each other by about $n \simeq l$.

[33] J. Yang and J. Leskovec, in *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics* (ACM, New York, 2012), MDS '12, pp. 3:1–3:8.

[34] J. Leskovec, J. Kleinberg, and C. Faloutsos, ACM Trans. Knowl. Discovery Data **1**, 1 (2007).

[35] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, in *Proceedings of the Workshop on Online Social Networks* (ACM, New York, NY, 2009), pp. 37–42.

[36] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, in *Proceedings of the International Conference on World Wide Web* (ACM, New York, NY, 2008), pp. 695–704.

[37] N. Jindal and B. Liu, in *Proceedings of the International Conference on Web Search and Web Data Mining* (ACM, New York, 2008), WSDM '08, pp. 219–230.