Higher-order clustering in networks

Hao Yin, Austin R. Benson, and Jure Leskovec

Phys. Rev. E **97**, 052306 — Published 18 May 2018

DOI: 10.1103/PhysRevE.97.052306

Higher-order clustering in networks

Hao Yin[*]
Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA, 94305, USA

Austin R. Benson[†]
Department of Computer Science, Cornell University, Ithaca, NY, 14850, USA

Jure Leskovec[‡]
Computer Science Department, Stanford University, Stanford, CA, 94305, USA

(Dated: April 30, 2018)

A fundamental property of complex networks is the tendency for edges to cluster. The extent of the clustering is typically quantified by the clustering coefficient, which is the probability that a length-2 path is closed, i.e., induces a triangle in the network. However, higher-order cliques beyond triangles are crucial to understanding complex networks, and the clustering behavior with respect to such higher-order network structures is not well understood. Here we introduce higher-order clustering coefficients that measure the closure probability of higher-order network cliques and provide a more comprehensive view of how the edges of complex networks cluster. Our higher-order clustering coefficients are a natural generalization of the traditional clustering coefficient. We derive several properties about higher-order clustering coefficients and analyze them under common random graph models. Finally, we use higher-order clustering coefficients to gain new insights into the structure of real-world networks from several domains.

I. INTRODUCTION

Networks are a fundamental tool for understanding and modeling complex physical, social, informational, and biological systems [1]. Although such networks are typically sparse, a recurring trait of networks throughout all of these domains is the tendency of edges to appear in small clusters or cliques [2, 3]. In many cases, such clustering can be explained by local evolutionary processes. For example, in social networks, clusters appear due to the formation of triangles where two individuals who share a common friend are more likely to become friends themselves, a process known as *triadic closure* [2, 4]. Similar triadic closures occur in other networks: in citation networks, two references appearing in the same publication are more likely to be on the same topic and hence more likely to cite each other [5] and in co-authorship networks, scientists with a mutual collaborator are more likely to collaborate in the future [6]. In other cases, local clustering arises from highly connected functional units operating within a larger system, e.g., metabolic networks are organized by densely connected modules [7].

The *clustering coefficient* quantifies the extent to which edges of a network cluster in terms of triangles. The clustering coefficient is defined as the fraction of length-2 paths, or *wedges*, that are closed with a triangle [3, 8] (Fig. 1, row C_2). In other words, the clustering coefficient measures the probability of triadic closure in the network.

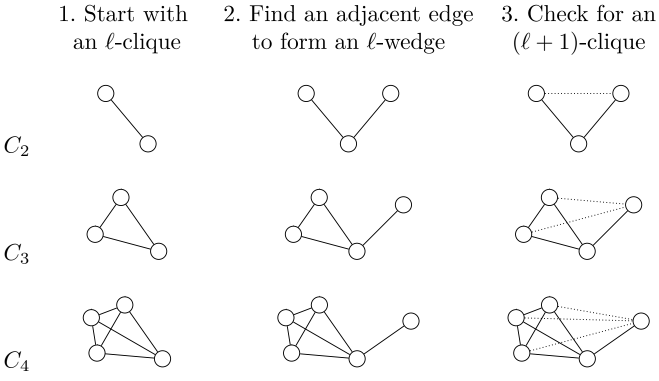

1. Start with an ℓ-clique
2. Find an adjacent edge to form an ℓ-wedge
3. Check for an $(\ell+1)$-clique
C_2
C_3
C_4

FIG. 1. Overview of higher-order clustering coefficients as clique expansion probabilities. The ℓth-order clustering coefficient C_ℓ measures the probability that an ℓ-clique and an adjacent edge, i.e., an ℓ-wedge, is closed, meaning that the $\ell-1$ possible edges between the ℓ-clique and the outside node in the adjacent edge exist to form an $(\ell+1)$-clique.

The clustering coefficient is an important statistic for data modeling in network science [9–11], as well as a useful feature in machine learning pipelines for, e.g., role discovery [12] and anomaly detection [13]. The statistic has also been identified as an important covariate in sociological studies [14].

However, the clustering coefficient is inherently restrictive as it measures the closure probability of just one simple structure—the triangle. Moreover, higher-order structures such as larger cliques are crucial to the structure and function of complex networks [15–17]. For example, 4-cliques reveal community structure in word association and protein-protein interaction networks [18] and cliques of sizes 5–7 are more frequent than triangles

[*] yinh@stanford.edu
[†] arb@cs.cornell.edu
[‡] jure@cs.stanford.edu

in many real-world networks with respect to certain null models [19]. However, the extent of clustering of such higher-order structures has not been well understood nor quantified.

Here, we provide a framework to quantify higher-order clustering in networks by measuring the normalized frequency at which higher-order cliques are closed, which we call *higher-order clustering coefficients*. We derive our higher-order clustering coefficients by extending a novel interpretation of the classical clustering coefficient as a form of clique expansion (Fig. 1). We then derive several properties about higher-order clustering coefficients and analyze them under the $G_{n,p}$ and small-world null models.

Using our theoretical analysis as a guide, we analyze the higher-order clustering behavior of real-world networks from a variety of domains. Conventional wisdom in network science posits that practically all real-world networks exhibit clustering; however, we find that the clustering property only holds up to a certain order. More specifically, once we control for the clustering as measured by the classical clustering coefficient, networks from some domains do not show significant higher-order clustering in terms of higher-order clique closure. Moreover, by examining how the clustering changes with the order of the clustering, we find that each domain of networks has its own higher-order clustering pattern. Since the traditional clustering coefficient only provides one measurement, it does not show such trends by itself. In addition to the theoretical properties and empirical findings exhibited in this paper, our related work also theoretically connects higher-order clustering and community detection [20].

II. DERIVATION OF HIGHER-ORDER CLUSTERING COEFFICIENTS

In this section, we derive our higher-order clustering coefficients and some of their basic properties. We first present an alternative interpretation of the classical clustering coefficient and then show how this novel interpretation seamlessly generalizes to arrive at our definition of higher-order clustering coefficients. We then provide some probabilistic interpretations of higher-order clustering coefficients that will be useful for our subsequent analysis. Throughout this paper, we confine our discussion to homogeneous networks with only one type of node and leave the development of higher-order clustering on bipartite and multilayer networks for further work.

A. Alternative interpretation of the classical clustering coefficient

Here we give an alternative interpretation of the clustering coefficient that will later allow us to generalize it and quantify clustering of higher-order network structures (this interpretation is summarized in Fig. 1). Our interpretation is based on a notion of clique expansion. First, we consider a 2-clique K in a graph G (that is, a single edge K; see Fig. 1, row C_2, column 1). Next, we *expand* the clique K by considering any edge e adjacent to K, i.e., e and K share exactly one node (Fig. 1, row C_2, column 2). This expanded subgraph forms a wedge, i.e., a length-2 path. The classical global clustering coefficient C of G (sometimes called the transitivity of G [21]) is then defined as the fraction of wedges that are *closed*, meaning that the 2-clique and adjacent edge induce a $(2+1)$-clique, or a triangle (Fig. 1, row C_2, column 3) [8, 22]. The novelty of our interpretation of the clustering coefficient is considering it as a form of clique expansion, rather than as the closure of a length-2 path, which is key to our generalizations in the next section.

Formally, the classical global clustering coefficient is

$$C = \frac{6|K_3|}{|W|}, \tag{1}$$

where K_3 is the set of 3-cliques (triangles), W is the set of wedges, and the coefficient 6 comes from the fact that each 3-clique closes 6 wedges—the 6 ordered pairs of edges in the triangle.

We can also reinterpret the local clustering coefficient [3] in this way. In this case, each wedge again consists of a 2-clique and adjacent edge (Fig. 1, row C_2, column 2), and we call the unique node in the intersection of the 2-clique and adjacent edge the *center* of the wedge. The *local clustering clustering coefficient* of a node u is the fraction of wedges centered at u that are closed:

$$C(u) = \frac{2|K_3(u)|}{|W(u)|}, \tag{2}$$

where $K_3(u)$ is the set of 3-cliques containing u and $W(u)$ is the set of wedges with center u (if $|W(u)| = 0$, we say that $C(u)$ is undefined). The *average clustering coefficient* $\bar{C}$ is the mean of the local clustering coefficients,

$$\bar{C} = \frac{1}{|\widetilde{V}|}\sum_{u\in\widetilde{V}} C(u), \tag{3}$$

where $\widetilde{V}$ is the set of nodes in the network where the local clustering coefficient is defined.

B. Generalizing to higher-order clustering coefficients

Our alternative interpretation of the clustering coefficient, described above as a form of clique expansion, leads to a natural generalization to higher-order cliques. Instead of expanding 2-cliques to 3-cliques, we expand ℓ-cliques to $(\ell+1)$-cliques (Fig. 1, rows C_3 and C_4). Formally, we define an ℓ-wedge to consist of an ℓ-clique and an adjacent edge for $\ell \geq 2$. Then we define the global ℓth-order clustering coefficient C_ℓ as the fraction of ℓ-wedges that are closed, meaning that they induce an $(\ell+1)$-clique in the network. We can write this as

$$C_\ell = \frac{(\ell^2 + \ell)|K_{\ell+1}|}{|W_\ell|}, \tag{4}$$

where $K_{\ell+1}$ is the set of $(\ell+1)$-cliques, and W_ℓ is the set of ℓ-wedges. The coefficient $\ell^2 + \ell$ comes from the fact that each $(\ell+1)$-clique closes that many wedges: each $(\ell+1)$-clique contains $\ell+1$ ℓ-cliques, and each ℓ-clique contains ℓ nodes which may serve as the center of an ℓ-wedge. Note that the classical definition of the global clustering coefficient given in Eq. 1 is equivalent to the definition in Eq. 4 when $\ell = 2$.

We also define higher-order local clustering coefficients:

$$C_\ell(u) = \frac{\ell |K_{\ell+1}(u)|}{|W_\ell(u)|}, \tag{5}$$

where $K_{\ell+1}(u)$ is the set of $(\ell+1)$-cliques containing node u, $W_\ell(u)$ is the set of ℓ-wedges with center u (where the center is the unique node in the intersection of the ℓ-clique and adjacent edge comprising the wedge; see Fig. 1), and the coefficient ℓ comes from the fact that each $(\ell+1)$-clique containing u closes that many ℓ-wedges in $W_\ell(u)$. The ℓth-order clustering coefficient of a node is defined for any node that is the center of at least one ℓ-wedge, and the average ℓth-order clustering coefficient is the mean of the local clustering coefficients:

$$\bar{C}_\ell = \frac{1}{|\widetilde{V}_\ell|} \sum_{u \in \widetilde{V}_\ell} C_\ell(u), \tag{6}$$

where $\widetilde{V}_\ell$ is the set of nodes that are the centers of at least one ℓ-wedge.

To understand how to compute higher-order clustering coefficients, we substitute the following useful identity

$$|W_\ell(u)| = |K_\ell(u)| \cdot (d_u - \ell + 1), \tag{7}$$

where d_u is the degree of node u, into Eq. 5 to get

$$C_\ell(u) = \frac{\ell \cdot |K_{\ell+1}(u)|}{(d_u - \ell + 1) \cdot |K_\ell(u)|}. \tag{8}$$

From Eq. 8, it is easy to see that we can compute all local ℓth-order clustering coefficients by enumerating all $(\ell+1)$-cliques and ℓ-cliques in the graph. The computational complexity of the algorithm is thus bounded by the time to enumerate $(\ell+1)$-cliques and ℓ-cliques. Using the Chiba and Nishizeki algorithm [23], the complexity is $O(\ell a^{\ell-2} m)$, where m is the number of edges and a is the arboricity of the graph, that is, the minimum number of edge-disjoint spanning forests to compose the graph [24]. (Arboricity is a specific measure of network density useful in the design of fast algorithms for globally sparse graphs; a dense graph with many edges would have large arboricity.) The arboricity a may be as large as $\sqrt{m}$, so this algorithm is only guaranteed to take polynomial time if ℓ is a constant. In general, determining if there exists a single clique with at least ℓ nodes is NP-complete [25].

For the global clustering coefficient, note that

$$|W_\ell| = \sum_{u \in V} |W_\ell(u)|. \tag{9}$$

Thus, it suffices to enumerate ℓ-cliques (to compute $|W_\ell|$ using Eq. 7) and to count the total number of ℓ-cliques. In practice, we use the Chiba and Nishizeki to enumerate cliques and simultaneously compute C_ℓ and $C_\ell(u)$ for all nodes u. This suffices for our clustering analysis with $\ell = 2, 3, 4$ on networks with over a hundred million edges in Section IV.

C. Probabilistic interpretations of higher-order clustering coefficients

To facilitate understanding of higher-order clustering coefficients and to aid our analysis in Section III, we present a few probabilistic interpretations of the quantities. First, we can interpret $C_\ell(u)$ as the probability that a wedge w chosen uniformly at random from all wedges centered at u is closed:

$$C_\ell(u) = \mathbb{P}\left[w \in K_{\ell+1}(u)\right]. \tag{10}$$

The variant of this interpretation for the classical clustering case of $\ell = 2$ has been useful for graph algorithm development [26].

For the next probabilistic interpretation, it is useful to analyze the structure of the 1-hop neighborhood graph $N_1(u)$ of a given node u (not containing node u). The vertex set of $N_1(u)$ is the set of all nodes adjacent to u, and the edge set consists of all edges between neighbors of u, i.e., $\{(v, w) \mid (u, v), (u, w), (v, w) \in E\}$, where E is the edge set of the graph.

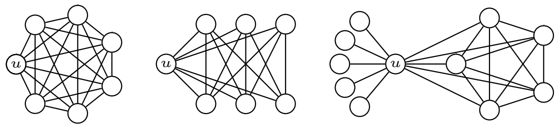
Any ℓ-clique in G containing node u corresponds to a unique $(\ell-1)$-clique in $N_1(u)$, and specifically for $\ell = 2$, any edge (u, v) corresponds to a node v in $N_1(u)$. Therefore, each ℓ-wedge centered at u corresponds to an $(\ell-1)$-clique K and one of the $d_u - \ell + 1$ nodes outside K (i.e., in $N_1(u) \backslash K$). Thus, Eq. 8 can be re-written as

$$\frac{\ell \cdot |K_\ell(N_1(u))|}{(d_u - \ell + 1) \cdot |K_{\ell-1}(N_1(u))|}, \tag{11}$$

where $K_k(N_1(u))$ denotes the number of k-cliques in $N_1(u)$.

If we uniformly at random select an $(\ell-1)$-clique K from $N_1(u)$ and then also uniformly at random select a node v from $N_1(u)$ outside of this clique, then $C_\ell(u)$ is the probability that these ℓ nodes form an ℓ-clique:

$$C_\ell(u) = \mathbb{P}\left[K \cup \{v\} \in K_\ell(N_1(u))\right]. \tag{12}$$



$C_2(u)$	1	$\frac{d}{2(d-1)} \approx \frac{1}{2}$	$\frac{d-2}{4d-4} \approx \frac{1}{4}$
$C_3(u)$	1	0	$\frac{d-4}{2d-4} \approx \frac{1}{2}$
$C_4(u)$	1	0	$\frac{d-6}{2d-6} \approx \frac{1}{2}$

FIG. 2. Example 1-hop neighborhoods of a node u with degree d with different higher-order clustering. Left: For cliques, $C_\ell(u) = 1$ for any ℓ. Middle: If u's neighbors form a complete bipartite graph, $C_2(u)$ is constant while $C_\ell(u) = 0$, $\ell \geq 3$. Right: If half of u's neighbors form a star and half form a clique with u, then $C_\ell(u) \approx \sqrt{C_2(u)}$, which is the upper bound in Proposition 1.

Moreover, if we condition on observing an ℓ-clique from this sampling procedure, then the ℓ-clique itself is selected uniformly at random from all ℓ-cliques in $N_1(u)$. Therefore, $C_{\ell-1}(u) \cdot C_\ell(u)$ is the probability that an $(\ell-1)$-clique and two nodes selected uniformly at random from $N_1(u)$ form an $(\ell+1)$-clique. Applying this recursively gives

$$\prod_{j=2}^{\ell} C_j(u) = \frac{|K_\ell(N_1(u))|}{\binom{d_u}{\ell}}. \tag{13}$$

In other words, the product of the higher-order local clustering coefficients of node u up to order ℓ is the ℓ-clique density amongst u's neighbors.

III. THEORETICAL ANALYSIS AND HIGHER-ORDER CLUSTERING IN RANDOM GRAPH MODELS

We now provide some theoretical analysis of our higher-order clustering coefficients. We first give some extremal bounds on the values that higher-order clustering coefficients can take given the value of the traditional (second-order) clustering coefficient. After, we analyze the values of higher-order clustering coefficients in two common random graph models—the $G_{n,p}$ and small-world models. The theory from this section will be a useful guide for interpreting the clustering behavior of real-world networks in Section IV.

A. Extremal bounds

We first analyze the relationships between local higher-order clustering coefficients of different orders. Our technical result is Proposition 1, which provides essentially tight lower and upper bounds for higher-order local clustering coefficients in terms of the traditional local clustering coefficient. The main ideas of the proof are illustrated in Fig. 2.

Proposition 1. *For any fixed $\ell \geq 3$,*

$$0 \leq C_\ell(u) \leq \sqrt{C_2(u)}. \tag{14}$$

Moreover,

1. *There exists a finite graph G with a node u such that the lower bound is tight and $C_2(u)$ is within ϵ of any prescribed value in $[0, \frac{\ell-2}{\ell-1}]$.*
2. *There exists a finite graph G with a node u such that $C_\ell(u)$ is within ϵ of the upper bound for any prescribed value of $C_2(u) \in [0, 1]$.*

Proof. Clearly, $0 \leq C_\ell(u)$ if the local clustering coefficient is well defined. This bound is tight when $N_1(u)$ is $(\ell-1)$-partite, as in the middle column of Fig. 2. In the $(\ell-1)$-partite case, $C_2(u) = \frac{\ell-2}{\ell-1}$. By removing edges from this extremal case in a sufficiently large graph, we can make $C_2(u)$ arbitrarily close to any value in $[0, \frac{\ell-2}{\ell-1}]$.

To derive the upper bound, consider the 1-hop neighborhood $N_1(u)$, and let

$$\delta_\ell(N_1(u)) = \frac{|K_\ell(N_1(u))|}{\binom{d_u}{\ell}} \tag{15}$$

denote the ℓ-clique density of $N_1(u)$. The Kruskal-Katona theorem [27, 28] implies that

$$\delta_\ell(N_1(u)) \leq [\delta_{\ell-1}(N_1(u))]^{\ell/(\ell-1)}$$
$$\delta_{\ell-1}(N_1(u)) \leq [\delta_2(N_1(u))]^{(\ell-1)/2}.$$

Combining this with Eq. 8 gives

$$C_\ell(u) \leq [\delta_{\ell-1}(N_1(u))]^{\frac{1}{\ell-1}} \leq \sqrt{\delta_2(N_1(u))} = \sqrt{C_2(u)},$$

where the last equality uses the fact that $C_2(u)$ is the edge density of $N_1(u)$.

The upper bound becomes tight when $N_1(u)$ consists of a clique and isolated nodes (Fig. 2, right) and the neighborhood is sufficiently large. Specifically, let $N_1(u)$ consist of a clique of size c and b isolated nodes. When $\ell = 2$,

$$C_\ell(u) = \frac{\binom{c}{2}}{\binom{c+b}{2}} = \frac{(c-1)c}{(c+b-1)(c+b)} \to \left(\frac{c}{c+b}\right)^2$$

and by Eq. 11, when $3 \leq \ell \leq c$,

$$C_\ell(u) = \frac{\ell \cdot \binom{c}{\ell}}{(c+b-\ell+1) \cdot \binom{c}{\ell-1}} = \frac{c-\ell+1}{c+b-\ell+1} \to \frac{c}{c+b}.$$

By adjusting the ratio $c/(b+c)$ in $N_1(u)$, we can construct a family of graphs such that $C_2(u)$ takes any value in the interval $[0, 1]$ as $d_u \to \infty$ and $C_\ell(u) \to \sqrt{C_2(u)}$ as $d_u \to \infty$. □

The second part of the result requires the neighborhoods to be sufficiently large in order to reach the upper bound. However, we will see later that in some real-world data, there are nodes u for which $C_3(u)$ is close to the upper bound $\sqrt{C_2(u)}$ for several values of $C_2(u)$.

Next, we analyze higher-order clustering coefficients in two common random graph models: the Erdős-Rényi model with edge probability p (i.e., the $G_{n,p}$ model [29]) and the small-world model [3].

B. Analysis for the $G_{n,p}$ model

Now, we analyze higher-order clustering coefficients in classical Erdős-Rényi random graph model, where each edge exists independently with probability p (i.e., the $G_{n,p}$ model [29]). We implicitly assume that ℓ is small in the following analysis so that there should be at least one ℓ-wedge in the graph (with high probability and n large, there is no clique of size greater than $(2+\epsilon)\log n/\log(1/p)$ for any $\epsilon > 0$ [30]). Therefore, the global and local clustering coefficients are well-defined.

In the $G_{n,p}$ model, we first observe that any ℓ-wedge is closed if and only if the $\ell - 1$ possible edges between the ℓ-clique and the outside node in the adjacent edge exist to form an $(\ell+1)$-clique. Each of the $\ell - 1$ edges exist independently with probability p in the $G_{n,p}$ model, which means that the higher-order clustering coefficients should scale as $p^{\ell-1}$. We formalize this in the following proposition.

Proposition 2. *Let G be a random graph drawn from the $G_{n,p}$ model. For constant ℓ,*
1. $\mathbb{E}_G\left[C_\ell\right] = p^{\ell-1}$
2. $\mathbb{E}_G\left[C_\ell(u) \mid W_\ell(u) > 0\right] = p^{\ell-1}$ *for any node u*
3. $\mathbb{E}_G\left[\bar{C}_\ell\right] = p^{\ell-1}$

Proof. We prove the first part by conditioning on the set of ℓ-wedges, W_ℓ:

$$\begin{aligned}\mathbb{E}[C_\ell] &= \mathbb{E}_G\left[\mathbb{E}_{W_\ell}\left[C_\ell \mid W_\ell\right]\right] \\ &= \mathbb{E}_G\left[\mathbb{E}_{W_\ell}\left[\tfrac{1}{|W_\ell|}\textstyle\sum_{w\in W_\ell}\mathbb{P}\left[w \text{ is closed}\right]\right]\right] \\ &= \mathbb{E}_G\left[\mathbb{E}_{W_\ell}\left[\tfrac{1}{|W_\ell|}\textstyle\sum_{w\in W_\ell}p^{\ell-1}\right]\right] \\ &= \mathbb{E}_G\left[p^{\ell-1}\right] \\ &= p^{\ell-1}.\end{aligned}$$

As noted above, the second equality is well defined (with high probability) for small ℓ. The third equality comes from the fact that any ℓ-wedge is closed if and only if the $\ell - 1$ possible edges between the ℓ-clique and the outside node in the adjacent edge exist to form an $(\ell+1)$-clique.

The proof of the second part is essentially the same, except we condition over the set of possible cases where $W_\ell(u) > 0$.

Recall that $\widetilde{V}$ is the set of nodes at the center of at least one ℓ-wedge. To prove the third part, we take the conditional expectation over $\widetilde{V}$ and use our result from the second part. □

The above results say that the global, local, and average ℓth order clustering coefficients decrease exponentially in ℓ. It turns out that if we also condition on the second-order clustering coefficient having some fixed value, then the higher-order clustering coefficients still decay exponentially in ℓ for the $G_{n,p}$ model. This will be useful for interpreting the distribution of local clustering coefficients on real-world networks.

Proposition 3. *Let G be a random graph drawn from the $G_{n,p}$ model. Then for constant ℓ,*

$$\begin{aligned}&\mathbb{E}_G\left[C_\ell(u) \mid C_2(u), W_\ell(u) > 0\right] \\ &= \left[C_2(u) - (1 - C_2(u)) \cdot O(1/d_u^2)\right]^{\ell-1} \approx (C_2(u))^{\ell-1}.\end{aligned}$$

Proof. Similar to the proof of Proposition 3, we look at the conditional expectation over $W_\ell(u) > 0$:

$$\begin{aligned}&\mathbb{E}_G\left[C_\ell(u) \mid C_2(u), W_\ell(u) > 0\right] \\ &= \mathbb{E}_G\left[\mathbb{E}_{W_\ell(u)>0}\left[C_\ell(u) \mid C_2(u),\ W_\ell(u)\right]\right] \\ &= \mathbb{E}_G\left[\mathbb{E}_{W_\ell(u)>0}\left[\tfrac{1}{|W_\ell(u)|}\textstyle\sum_{w\in W_\ell(u)}\mathbb{P}\left[w \text{ closed} \mid C_2(u)\right]\right]\right].\end{aligned}$$

Now, note that $N_1(u)$ has $m = C_2(u)\cdot\binom{d_u}{2}$ edges. Knowing that $w \in W_\ell(u)$ accounts for $\binom{\ell-1}{2}$ of these edges. By symmetry, the other $q = m - \binom{\ell-1}{2}$ edges appear in any of the remaining $r = \binom{d_u}{2} - \binom{\ell-1}{2}$ pairs of nodes uniformly at random. There are $\binom{r}{q}$ ways to place these edges, of which $\binom{r-\ell+1}{q-\ell+1}$ would close the wedge w. Thus,

$$\begin{aligned}&\mathbb{P}\left[w \text{ is closed} \mid C_2(u)\right] \\ &= \frac{\binom{r-\ell+1}{q-\ell+1}}{\binom{r}{q}} = \frac{(r-\ell+1)!q!}{(q-\ell+1)!r!} = \frac{(q-\ell+2)(q-\ell+3)\cdots q}{(r-\ell+2)(r-\ell+3)\cdots r}.\end{aligned}$$

Now, for any small nonnegative integer k,

$$\begin{aligned}\frac{q-k}{r-k} &= \frac{C_2(u)\cdot\binom{d_u}{2} - \binom{\ell-1}{2} - k}{\binom{d_u}{2} - \binom{\ell-1}{2} - k} \\ &= C_2(u) - (1 - C_2(u))\left[\frac{\binom{\ell-1}{2} + k}{\binom{d_u}{2} - \binom{\ell-1}{2} - k}\right] \\ &= C_2(u) - (1 - C_2(u)) \cdot O(1/d_u^2).\end{aligned}$$

(Recall that ℓ is constant by assumption, so the big-O notation is appropriate). The above expression approaches $(C_2(u))^{\ell-1}$ when $C_2(u) \to 1$ as well as when $d_u \to \infty$. □

Proposition 3 says that even if the second-order local clustering coefficient is large, the ℓth-order clustering coefficient will still decay exponentially in ℓ, at least in the limit as d_u grows large. By examining higher-order clique closures, this allows us to distinguish between nodes u whose neighborhoods are "dense but random" ($C_2(u)$ is large but $C_\ell(u) \approx (C_2(u))^{\ell-1}$) or "dense and structured" ($C_2(u)$ is large *and* $C_\ell(u) > (C_2(u))^{\ell-1}$). Only the latter case exhibits higher-order clustering. We use this in our analysis of real-world networks in Section IV.

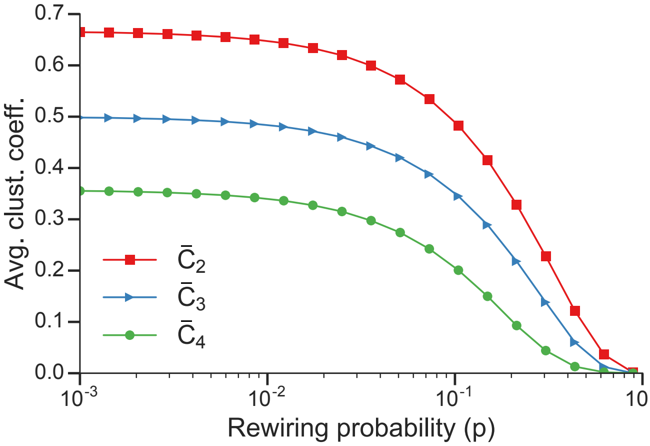


FIG. 3. Average higher-order clustering coefficient $\bar{C}_\ell$ as a function of rewiring probability p in small-world networks for $\ell = 2, 3, 4$ ($n = 20,000$, $k = 5$). Proposition 4 shows that the ℓth-order clustering coefficient when $p = 0$ predicts that the clustering should decrease modestly as ℓ increases.

C. Analysis for the small-world model

We also study higher-order clustering in the small-world random graph model [3]. The model begins with a ring network where each node connects to its $2k$ nearest neighbors. Then, for each node u and each of the k edges (u, v) with v following u clockwise in the ring, the edge is rewired to (u, w) with probability p, where w is chosen uniformly at random.

With no rewiring ($p = 0$) and $k \ll n$, it is known that $\bar{C}_2 \approx 3/4$ [3]. As p increases, the average clustering coefficient $\bar{C}_2$ slightly decreases until a phase transition near $p = 0.1$, where $\bar{C}_2$ decays to 0 [3] (also see Fig. 3). Here, we generalize these results for higher-order clustering coefficients.

Proposition 4. *In the small-world model without rewiring ($p = 0$),*

$$\bar{C}_\ell \to (\ell + 1)/(2\ell)$$

for any constant $\ell \geq 2$ as $k \to \infty$ and $n \to \infty$ while $2k < n$.

Proof. Applying Eq. 8, it suffices to show that

$$|K_\ell(u)| = \frac{\ell}{(\ell-1)!} \cdot k^{\ell-1} + O(k^{\ell-2}) \tag{16}$$

as

$$C_\ell(u) = \frac{\ell \cdot \frac{(\ell+1)k^\ell}{\ell!}}{(2k - \ell + 1) \cdot \frac{\ell k^{\ell-1}}{(\ell-1)!}},$$

which approaches $\frac{\ell+1}{2\ell}$ as $k \to \infty$.

Now we give a derivation of Eq. 16. We first label the $2k$ neighbors of u as $1, 2, \ldots, 2k$ by their clockwise ordering in the ring. Since $2k < n$, these nodes are unique. Next, define the *span* of any ℓ-clique containing u as the difference between the largest and smallest label of the $\ell - 1$ nodes in the clique other than u. The span s of any ℓ-clique satisfies $s \leq k - 1$ since any node is directly connected with a node of label difference no greater than $k - 1$. Also, $s \geq \ell - 2$ since there are $\ell - 1$ nodes in an ℓ-clique other than u. For each span s, we can find $2k-1-s$ pairs of (i, j) such that $1 \leq i, j \leq 2k$ and $j - i = s$. Finally, for every such pair (i, j), there are $\binom{s-1}{\ell-3}$ choices of $\ell - 3$ nodes between i and j which will form an ℓ-clique together with nodes u, i, and j. Therefore,

$$\begin{aligned} |K_\ell(u)| &= \textstyle\sum_{s=\ell-2}^{k-1} (2k - 1 - s) \cdot \binom{s-1}{\ell-3} \\ &= \textstyle\sum_{s=\ell-2}^{k-1} (2k - 1 - s) \cdot \frac{(s-1)(s-2)\cdots(s-\ell+3)}{(\ell-3)!} \\ &= \textstyle\sum_{t=1}^{k-\ell+2} (2k + 2 - t - \ell) \cdot \frac{t(t+1)\cdots(t+\ell-4)}{(\ell-3)!}. \end{aligned}$$

If we ignore lower-order terms k and note that $t = O(k)$, we get

$$\begin{aligned} |K_\ell(u)| &= \textstyle\sum_{t=1}^{k} \left[\frac{(2k-t)t^{\ell-3}}{(\ell-3)!} + O(k^{\ell-3})\right] \\ &= \tfrac{1}{(\ell-3)!} \textstyle\sum_{t=1}^{k} (2kt^{\ell-3} - t^{\ell-2}) + O(k^{\ell-2}). \\ &= \tfrac{1}{(\ell-3)!} \left[2k \cdot \frac{k^{\ell-2}}{\ell-2} - \frac{k^{\ell-1}}{\ell-1}\right] + O(k^{\ell-2}), \\ &= \tfrac{\ell}{(\ell-1)!} \cdot k^{\ell-1} + O(k^{\ell-2}). \end{aligned}$$

□

Proposition 4 shows that, when $p = 0$, $\bar{C}_\ell$ decreases as ℓ increases. When $p \neq 0$, obtaining a closed-form formula for the expected value of $\bar{C}_\ell$ remains an open problem. Via simulation, we observe that $\bar{C}_\ell$ also decreases as ℓ increases. The intuition is that cliques of larger size are hard to form in these synthetic networks. Furthermore, we observe the same behavior as for $\bar{C}_2$ when adjusting the rewiring probability p (Fig. 3). Regardless of ℓ, the phase transition happens near $p = 0.1$. Essentially, once there is enough rewiring, all local clique structure is lost, and clustering at all orders is lost. This is partly a consequence of Proposition 1, which says that $C_\ell(u) \to 0$ as $C_2(u) \to 0$ for any ℓ.

IV. EXPERIMENTAL RESULTS ON REAL-WORLD NETWORKS

We now analyze the higher-order clustering of real-world networks. We first study how the higher-order global and average clustering coefficients vary as we increase the order ℓ of the clustering coefficient on a collection of 20 networks from several domains. After, we concentrate on a few representative networks and compare the higher-order clustering of real-world networks to null models. We find that only some networks exhibit higher-order clustering once the traditional clustering coefficient is controlled. Finally, we examine the local clustering of real-world networks.

A. Higher-order global and average clustering

We compute and analyze the higher-order clustering for networks from a variety of domains (Table I). We briefly describe the collection of networks and their categorization below:

1. Two synthetic networks—a random instance of an Erdős-Rényi graph with $n = 1,000$ nodes and edge probability $p = 0.2$ and a small-world network with $n = 20,000$ nodes, $k = 10$, and rewiring probability $p = 0.1$;
2. Four neural networks—the complete neural systems of the nematode worms *P. pacificus* and *C. elegans* as well as the neural connections of the Drosophila medulla and mouse retina;
3. Four online social networks—two Facebook friendship networks between students at universities from 2005 (fb-Stanford, fb-Cornell) and two complete online friendship networks (Pokec and Orkut);
4. Four collaboration networks—two co-authorship networks constructed from arxiv submission categories (arxiv-AstroPh and arxiv-HepPh), a co-authorship network constructed from DBLP, and the co-committee membership network of United States congresspersons (congress-committees);
5. Four human communication networks—two email networks (email-Enron-core, email-Eu-core), a Facebook-like messaging network from a college (CollegeMsg), and the edits of user talk pages by other users on Wikipedia (wiki-Talk); and
6. Four technological systems networks—three autonomous systems (oregon2-010526, as-caida-20071105, as-skitter) and a peer-to-peer connection network (p2p-Genutella31).

In all cases, we take the edges as undirected, even if the original network data is directed.

Table I lists the ℓth-order global and average clustering coefficients for $\ell = 2, 3, 4$ as well as the fraction of nodes that are the center of at least one ℓ-wedge (recall that the average clustering coefficient is the mean only over higher-order local clustering coefficients of nodes participating in at least one ℓ-wedge; see Kaiser [45] for a discussion on how this can affect network analyses). We highlight some important trends in the raw clustering coefficients, and in the next section, we focus on higher-order clustering compared to what one gets in a null model.

Propositions 2 and 4 say that we should expect the higher-order global and average clustering coefficients to decrease as we increase the order ℓ for both the Erdős-Rényi and small-world models, and indeed $\bar{C}_2 > \bar{C}_3 > \bar{C}_4$ for these networks. This trend also holds for all of the real-world networks except oregon2-010526, where $\bar{C}_4$ is slightly larger than $\bar{C}_3$ (but $\bar{C}_2$ is still the largest). Thus, when averaging over nodes, higher-order cliques are overall less likely to close in both the synthetic and real-world networks.

The relationship between the higher-ordrer global clustering coefficient C_ℓ and the order ℓ is less uniform over the datasets. For the three co-authorship networks (arxiv-HepPh, arxiv-AstroPh, and DBLP) and the three autonomous systems networks (oregon2-010526, as-caida-20071105, and as-skitter), C_ℓ increases with ℓ, although the base clustering levels are much higher for co-authorship networks. This is not simply due to the presence of cliques—a clique has the same clustering for any order (Fig. 2, left). Instead, due to core-periphery network structure [50, 51], these datasets may have nodes that serve as the center of a star and also participate in a clique (Fig. 2, right; see also Proposition 1). On the other hand, C_ℓ decreases with ℓ for the two email networks and the four neural networks. Finally, the change in C_ℓ need not be monotonic in ℓ. In three of the four online social networks, $C_3 < C_2$ but $C_4 > C_3$.

Overall, the trends in the higher-order clustering coefficients can be different within one of our dataset categories, but tend to be uniform within a particular domain: the change of $\bar{C}_\ell$ and C_ℓ with ℓ is the same for the two email networks within the communication networks, the three co-authorship networks within the collaboration networks, and all four neural networks. These trends hold even if the (classical) second-order clustering coefficients differ substantially in absolute value.

While the raw clustering values are informative, it is also useful to compare the clustering to what one expects from null models. We find in the next section that this reveals additional insights into our data.

B. Comparison against null models

For one real-world network from each dataset category, we also measure the higher-order clustering coefficients with respect to two null models (Table II). First, we compare against the Configuration Model (CM) that samples uniformly from simple graphs with the same degree distribution [46, 47]. In real-world networks, $\bar{C}_2$ is much larger than expected with respect to the CM null model. We find that the same holds for $\bar{C}_3$.

Second, we use a null model that samples graphs preserving both degree distribution and $\bar{C}_2$. Specifically, these are samples from an ensemble of exponential graphs where the Hamiltonian measures the absolute value of the difference between the original network and the sampled network [48]. Such samples are referred to as Maximally Random Clustered Networks (MRCN) and are sampled with a simulated annealing procedure [49]. Comparing $\bar{C}_3$ between the real-world and the null network, we observe different behavior in higher-order clustering across our datasets. Compared to the MRCN null model, *C. elegans* has significantly less than expected higher-order clustering (in terms of $\bar{C}_3$), the Facebook friendship and autonomous system networks have significantly more than expected higher-order clustering, and the co-authorship and email networks have slightly (but not significantly) more than expected higher-order clustering (Table II). Put another way, all real-world networks

TABLE I. Higher-order clustering coefficients on random graph models, neural connections, online social networks, collaboration networks, human communication, and technological systems. Broadly, networks from the same domain have similar higher-order clustering characteristics. Since $\widetilde{V}_\ell$ is the set of nodes at the center of at least one ℓ-wedge (see Eq. 6), $|\widetilde{V}_\ell|/|V|$ is the fraction of nodes at the center of at least one ℓ-wedge (the higher-order average clustering coefficient $\bar{C}_\ell$ is only measured over those nodes participating in at least one ℓ-wedge).

Network	Nodes	Edges	C_2	C_3	C_4	$\bar{C}_2$	$\bar{C}_3$	$\bar{C}_4$	$\lvert\widetilde{V}_2\rvert/\lvert V\rvert$	$\lvert\widetilde{V}_3\rvert/\lvert V\rvert$	$\lvert\widetilde{V}_4\rvert/\lvert V\rvert$
Erdős-Rényi [29]	1,000	99,831	0.200	0.040	0.008	0.200	0.040	0.008	1.000	1.000	1.000
Small-world [3]	20,000	100,000	0.480	0.359	0.229	0.489	0.350	0.205	1.000	1.000	0.999
P. pacificus [31]	50	141	0.349	0.234	0.166	0.471	0.274	0.141	0.700	0.520	0.400
C. elegans [3]	297	2,148	0.181	0.080	0.056	0.308	0.137	0.062	0.949	0.926	0.808
Drosophila-medulla [32]	1,781	8,911	0.069	0.025	0.014	0.339	0.150	0.062	0.775	0.585	0.417
mouse-retina [33]	1,076	90,811	0.400	0.269	0.212	0.593	0.468	0.401	0.996	0.995	0.992
fb-Stanford [34]	11,621	568,330	0.157	0.107	0.116	0.253	0.181	0.157	0.955	0.922	0.877
fb-Cornell [34]	18,660	790,777	0.136	0.106	0.121	0.225	0.169	0.148	0.973	0.951	0.923
Pokec [35]	1,632,803	22,301,964	0.047	0.044	0.046	0.122	0.084	0.061	0.900	0.675	0.508
Orkut [36]	3,072,441	117,185,083	0.041	0.022	0.019	0.170	0.131	0.110	0.978	0.949	0.878
arxiv-HepPh [37]	12,008	118,489	0.659	0.749	0.788	0.698	0.586	0.520	0.875	0.723	0.567
arxiv-AstroPh [37]	18,772	198,050	0.318	0.326	0.359	0.677	0.609	0.561	0.932	0.839	0.740
congress-committees [38]	871	79,886	0.424	0.269	0.218	0.499	0.364	0.320	1.000	1.000	1.000
DBLP [39]	317,080	1,049,866	0.306	0.634	0.821	0.732	0.613	0.517	0.864	0.675	0.489
email-Enron-core [40]	148	1,356	0.383	0.245	0.192	0.496	0.363	0.277	0.966	0.946	0.946
email-Eu-core [20, 37]	1,005	16,064	0.267	0.170	0.135	0.450	0.329	0.264	0.887	0.847	0.784
CollegeMsg [41]	1,899	13,838	0.057	0.018	0.009	0.138	0.039	0.014	0.793	0.579	0.331
wiki-Talk [42]	2,394,385	4,659,565	0.002	0.011	0.010	0.201	0.081	0.051	0.262	0.077	0.027
oregon2-010526 [43]	11,461	32,730	0.037	0.085	0.097	0.494	0.294	0.300	0.711	0.269	0.121
as-caida-20071105 [43]	26,475	53,381	0.007	0.012	0.015	0.333	0.159	0.134	0.625	0.171	0.060
p2p-Gnutella31 [37, 44]	62,586	147,892	0.004	0.003	0.000	0.010	0.001	0.000	0.542	0.067	0.001
as-skitter [43]	1,696,415	11,095,298	0.005	0.007	0.011	0.296	0.126	0.109	0.871	0.633	0.335

TABLE II. Average higher-order clustering coefficients for five networks as well as the clustering with respect to two null models: a Configuration Model (CM) that samples random graphs with the same degree distribution [46, 47], and Maximally Random Clustered Networks (MRCN) that preserve degree distribution as well as $\bar{C}_2$ [48, 49]. For the random networks, we report the mean over 100 samples. An asterisk ($*$) denotes when the value in the original network is at least five standard deviations above the mean and a dagger (†) denotes when the value in the original network is at least five standard deviations below the mean. Although all networks exhibit clustering with respect to CM, only some of the networks exhibit higher-order clustering when controlling for $\bar{C}_2$ with MRCN.

	C. elegans			fb-Stanford			arxiv-AstroPh			email-Enron-core			oregon2-010526		
	original	CM	MRCN	original	CM	MRCN	original	CM	MRCN	original	CM	MRCN	original	CM	MRCN
$\bar{C}_2$	0.31	0.15^*	0.31	0.25	0.03^*	0.25	0.68	0.01^*	0.68	0.50	0.23^*	0.50	0.49	0.25^*	0.49
$\bar{C}_3$	0.14	0.04^*	$0.17^\dagger$	0.18	0.00^*	0.14^*	0.61	0.00^*	0.60	0.36	0.08^*	0.35	0.29	0.10^*	0.14^*

exhibit clustering in the classical sense of triadic closure. However, the higher-order clustering coefficients reveal that the friendship and autonomous systems networks exhibit significant clustering beyond what is given by triadic closure. These results suggest the need for models that directly account for closure in node neighborhoods [52, 53].

Our finding about the lack of higher-order clustering in *C. elegans* agrees with previous results that 4-cliques are under-expressed, while open 3-wedges related to cooperative information propagation are over-expressed [15, 54, 55]. This also provides credence for the "3-layer" model of *C. elegans* [55]. The observed clustering in the friendship network is consistent with prior work showing the relative infrequency of open ℓ-wedges in many Facebook network subgraphs with respect to a null model accounting for triadic closure [56]. Co-authorship networks and email networks are both constructed from "events" that create multiple edges—a paper with k authors induces a k-clique in the co-authorship graph and an email sent from one address to k others induces k edges. This event-driven graph construction creates enough closure structure so that the average third-order clustering coefficient is not much larger than random graphs where the classical second-order clustering coefficient and degree sequence is kept the same.

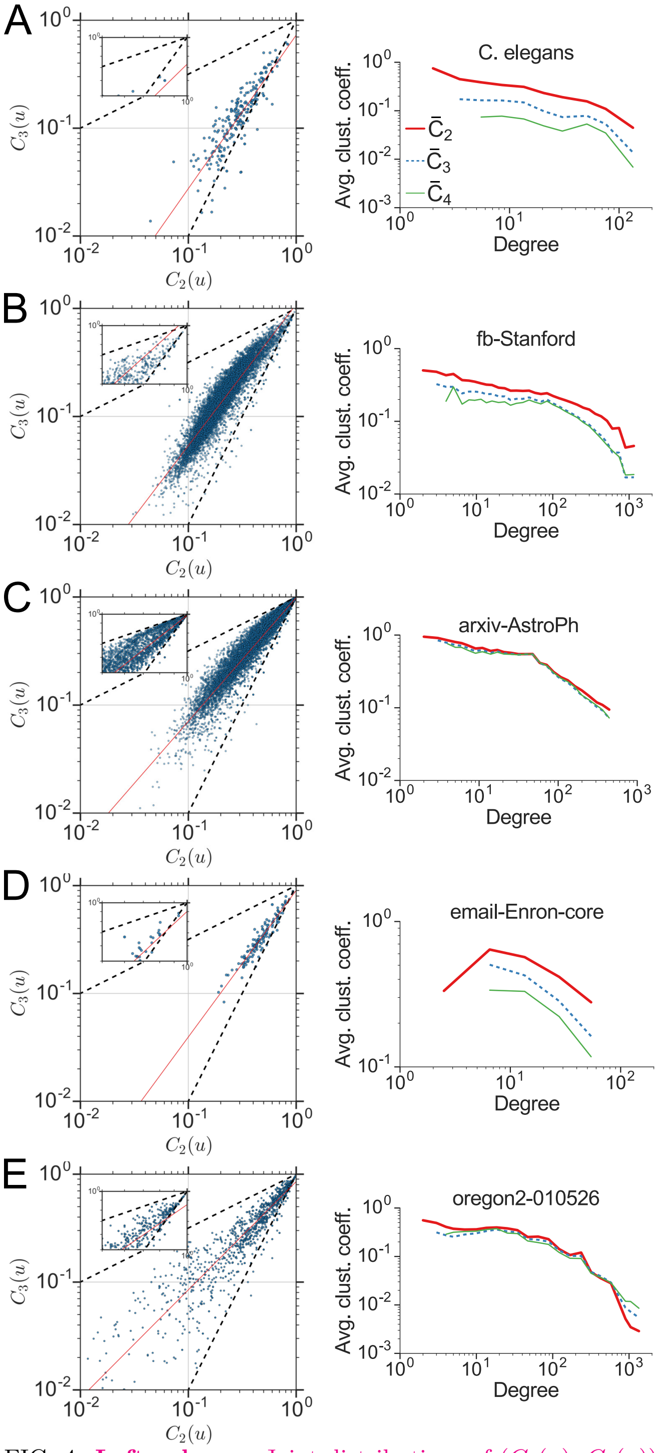


FIG. 4. **Left column:** Joint distributions of $(C_2(u), C_3(u))$ for (A) *C. elegans* (B) friendship, (C) co-authorship, (D) email, and (E) autonomous systems networks. There is a blue dot for each node u with $C_2(u) \neq 0$ and $C_3(u) \neq 0$. The red curve is a linear fit of $\ln C_3(u)$ in terms of $\ln C_2(u)$ and an intercept (see Table III). The upper trend line is the bound in Eq. 14—the largest possible value of $C_3(u)$ given $C_2(u)$. The lower trend line is the expected Erdős-Rényi behavior from Prop. 3. **Left column insets:** The insets are enlarged versions of the figure for the data where both $C_2(u)$ and $C_3(u)$ are inside the interval $[0.5, 1]$. **Right column:** Average second-order (classical), third-order, and fourth-order clustering coefficient as a function of node degree.

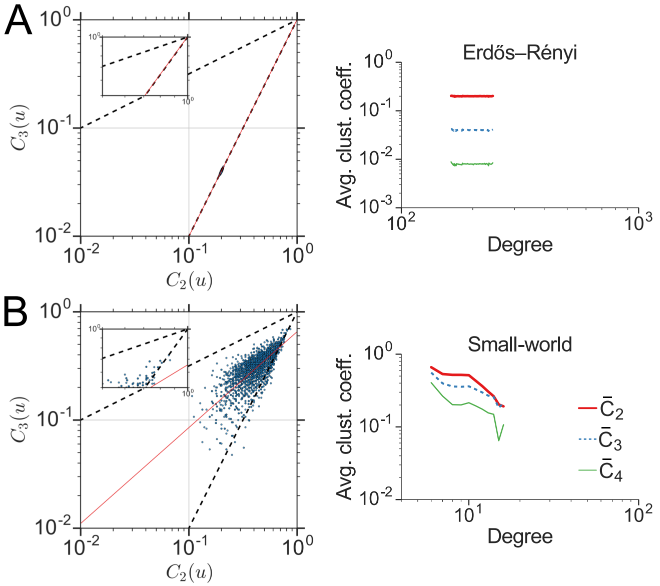


FIG. 5. Analogous plots of Fig. 4 for an instance of (A) Erdős-Rényi and (B) small-world random graphs (see the caption in Fig. 4 for a more complete explanation of the figure). **Left column:** Joint distributions of $(C_2(u), C_3(u))$ with a linear fit of the natural logs of the third-order clustering coefficient in terms of the second-order (classical) clustering coefficient (see Table III). The insets show the domain restricted to $C_2(u)$ and $C_3(u)$ both inside the interval $[0.5, 1]$ **Right column:** Average higher-order clustering coefficients as a function of degree.

We emphasize that simple clique counts are not sufficient to obtain these results. For example, the discrepancy in the third-order average clustering of *C. elegans* and the MRCN null model is not simply due to the presence of 4-cliques. The original neural network has nearly twice as many 4-cliques (2,010) than the samples from the MRCN model (mean 1006.2, standard deviation 73.6), but the third-order clustering coefficient is larger in MRCN. The reason is that clustering coefficients normalize clique counts with respect to opportunities for closure.

Thus far, we have analyzed global and average higher-order clustering, which both summarize the clustering of the entire network. In the next section, we look at more localized properties, namely the distribution of higher-order local clustering coefficients and the higher-order average clustering coefficient as a function of node degree.

C. Higher-order local clustering coefficients and degree dependencies

We now examine more localized clustering properties of our networks. Figure 4 (left column) plots the joint distribution of $C_2(u)$ and $C_3(u)$ for the five networks analyzed in Table II (along with a linear model for $C_3(u)$ in terms of $C_2(u)$; see also Table III), and Fig. 5 (left column) provides the analogous plots for the Erdős-Rényi

TABLE III. Linear regression models of the form $\ln C_3(u) = m \cdot \ln C_2(u) + b$ for the plots in the left column of Figs. 4 and 5. The regression coefficients m and b are listed with the standard error, along with the R^2 values of each model.

network	m	b	R^2
C. elegans	1.424 ± 0.059	-0.318 ± 0.067	0.770
fb-Stanford	1.311 ± 0.005	0.089 ± 0.008	0.879
arxiv-AstroPh	1.146 ± 0.002	-0.014 ± 0.002	0.936
email-Enron-core	1.360 ± 0.037	-0.104 ± 0.030	0.908
oregon2-010526	1.005 ± 0.009	-0.158 ± 0.011	0.886
Erdős-Rényi	2.003 ± 0.021	0.006 ± 0.034	0.899
Small-world	0.886 ± 0.004	-0.424 ± 0.003	0.672

and small-world networks. In these plots, the lower dashed trend line represents the expected Erdős-Rényi behavior, i.e., the expected clustering if the edges in the neighborhood of a node were configured randomly, as formalized in Proposition 3. The upper dashed trend line is the maximum possible value of $C_3(u)$ given $C_2(u)$, as given by Proposition 1.

For many nodes in *C. elegans*, local clustering is nearly random (Fig. 4A, left), i.e., resembles the Erdős-Rényi joint distribution (Fig. 5A, left). In other words, there are many nodes that lie on the lower trend line. The fitted linear model of $C_3(u)$ in terms of $C_2(u)$ further highlights this concept (see also Table III). Overall, this provides further evidence that *C. elegans* lacks higher-order clustering. In the arxiv co-authorship network, there are many nodes u with a large value of $C_2(u)$ that have an even larger value of $C_3(u)$ near the upper bound of Eq. 14 (see the inset of Fig. 4C, left). This implies that some nodes appear in both cliques and also as the center of star-like patterns, as in Fig. 2. On the other hand, only a handful of nodes in the Facebook friendships, Enron email, and Oregon autonomous systems networks are close to the upper bound (insets of Figs. 4B,4D, and 4E, left). However, there are still several nodes in the friendship and autonomous system networks that have a larger third-order clustering coefficient than second-order (classical) clustering coefficient.

Figures 4 and 5 (right column) plot higher-order average clustering as a function of node degree in the real-world and synthetic networks. In the Erdős-Rényi, small-world, *C. elegans*, and Enron email networks, there is a distinct gap between the average higher-order clustering coefficients for nodes of all degrees. Thus, our previous finding that the average clustering coefficient $\bar{C}_\ell$ decreases with ℓ in these networks is independent of degree. In the Facebook friendship network, $C_2(u)$ is larger than $C_3(u)$ and $C_4(u)$ on average for nodes of all degrees, but $C_3(u)$ and $C_4(u)$ are roughly the same for nodes of all degrees, which means that 4-cliques and 5-cliques close at roughly the same rate, independent of degree, albeit at a smaller rate than traditional triadic closure (Fig. 4B, right). In the co-authorship network, nodes u have roughly the same $C_\ell(u)$ for $\ell = 2, 3, 4$, which means that ℓ-cliques close at about the same rate, independent of ℓ (Fig. 4C, right). In the Oregon autonomous systems network, we see that, on average, $C_4(u) > C_3(u) > C_2(u)$ for nodes with large degree (Fig. 4E, right). This explains how the global clustering coefficient increases with the order, but the average clustering does not, as observed in Table I.

V. DISCUSSION

We have proposed higher-order clustering coefficients to study higher-order closure patterns in networks, which generalizes the widely used clustering coefficient that measures triadic closure. Our work compliments other recent developments on the importance of higher-order information in network navigation [17, 57] and on temporal community structure [58]; in contrast, we examine higher-order clique closure and only implicitly consider time as a motivation for closure. Extending our ideas to more network models, such as bipartite and multilayer networks, provides an avenue for future research.

Prior efforts in generalizing clustering coefficients have focused on shortest paths [59], cycle formation [60], and triangle frequency in k-hop neighborhoods [61, 62]. Such approaches fail to capture closure patterns of cliques, suffer from challenging computational issues, and are difficult to theoretically analyze in random graph models more sophisticated than the Erdős-Rényi model. On the other hand, our higher-order clustering coefficients are simple but effective measurements that are analyzable and easily computable (we only rely on clique enumeration, a well-studied algorithmic task). Furthermore, our methodology provides new insights into the clustering behavior of several real-world networks and random graph models, and our theoretical analysis provides intuition for the way in which higher-order clustering coefficients describe local clustering in graphs.

Finally, we focused on higher-order clustering coefficients as a global network measurement and as a node-level measurement. In related work we also show that large higher-order clustering implies the existence of mesoscale clique-dense community structure [20].

The web site associated with this paper, which includes software for computing higher-order clustering coefficients, is `http://snap.stanford.edu/hocc`.

ACKNOWLEDGMENTS

This research has been supported in part by NSF IIS-1149837, ARO MURI, DARPA, ONR, Huawei, and Stanford Data Science Initiative. ARB was supported in part by a Simons Investigator Award and NSF TRIPODS Award #1740822. We thank Will Hamilton and Marinka Žitnik for insightful comments. We thank Mason Porter and Peter Mucha for providing the congress committee membership data.

[1] M. E. J. Newman, SIAM Review **45**, 167 (2003).

[2] A. Rapoport, The Bulletin of Mathematical Biophysics **15**, 523 (1953).

[3] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).

[4] M. S. Granovetter, American Journal of Sociology , 1360 (1973).

[5] Z.-X. Wu and P. Holme, Physical Review E **80**, 037101 (2009).

[6] E. M. Jin, M. Girvan, and M. E. J. Newman, Physical Review E **64**, 046132 (2001).

[7] E. Ravasz and A.-L. Barabási, Physical Review E **67**, 026112 (2003).

[8] A. Barrat and M. Weigt, The European Physical Journal B: Condensed Matter and Complex Systems **13**, 547 (2000).

[9] M. E. J. Newman, Physical Review Letters **103**, 058701 (2009).

[10] C. Seshadhri, T. G. Kolda, and A. Pinar, Physical Review E **85**, 056109 (2012).

[11] P. Robles, S. Moreno, and J. Neville, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2016) pp. 1155–1164.

[12] K. Henderson, B. Gallagher, T. Eliassi-Rad, H. Tong, S. Basu, L. Akoglu, D. Koutra, C. Faloutsos, and L. Li, in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (ACM, 2012) pp. 1231–1239.

[13] T. La Fond, J. Neville, and B. Gallagher, in *Outlier Detection and Description under Data Diversity at the International Conference on Knowledge Discovery and Data Mining* (2014).

[14] P. S. Bearman and J. Moody, American journal of public health **94**, 89 (2004).

[15] A. R. Benson, D. F. Gleich, and J. Leskovec, Science **353**, 163 (2016).

[16] Ö. N. Yaveroğlu, N. Malod-Dognin, D. Davis, Z. Levnajic, V. Janjic, R. Karapandza, A. Stojmirovic, and N. Pržulj, Scientific Reports **4** (2014).

[17] M. Rosvall, A. V. Esquivel, A. Lancichinetti, J. D. West, and R. Lambiotte, Nature Communications **5** (2014).

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

[19] N. Slater, R. Itzchack, and Y. Louzoun, Network Science **2**, 387 (2014).

[20] H. Yin, A. R. Benson, J. Leskovec, and D. F. Gleich, in *Proceedings of the 23rd ACM SIGKDD international conference on Knowledge discovery and data mining* (2017).

[21] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Physics reports **424**, 175 (2006).

[22] R. D. Luce and A. D. Perry, Psychometrika **14**, 95 (1949).

[23] N. Chiba and T. Nishizeki, SIAM Journal on Computing **14**, 210 (1985).

[24] F. Harary, “Graph theory, revised,” (1972).

[25] R. M. Karp, in *Complexity of computer computations* (Springer, 1972) pp. 85–103.

[26] C. Seshadhri, A. Pinar, and T. G. Kolda, in *Proceedings of the 2013 SIAM International Conference on Data Mining* (SIAM, 2013) pp. 10–18.

[27] J. B. Kruskal, Mathematical Optimization Techniques **10**, 251 (1963).

[28] G. Katona, in *Theory of Graphs: Proceedings of the Colloquium held at Tihany, Hungary* (1966) pp. 187–207.

[29] P. Erdös and A. Rényi, Publicationes Mathematicae (Debrecen) **6**, 290 (1959).

[30] B. Bollobás and P. Erdös, in *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 80 (Cambridge University Press, 1976) pp. 419–427.

[31] D. J. Bumbarger, M. Riebesell, C. Rödelsperger, and R. J. Sommer, Cell **152**, 109 (2013).

[32] S.-y. Takemura, A. Bharioke, Z. Lu, A. Nern, S. Vitaladevuni, P. K. Rivlin, W. T. Katz, D. J. Olbris, S. M. Plaza, P. Winston, *et al.*, Nature **500**, 175 (2013).

[33] M. Helmstaedter, K. L. Briggman, S. C. Turaga, V. Jain, H. S. Seung, and W. Denk, Nature **500**, 168 (2013).

[34] A. L. Traud, P. J. Mucha, and M. A. Porter, Physica A: Statistical Mechanics and its Applications **391**, 4165 (2012).

[35] L. Takac and M. Zabovsky, in *International Scientific Conference and International Workshop Present Day Trends of Innovations*, Vol. 1 (2012).

[36] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, in *Proceedings of the 5th ACM/Usenix Internet Measurement Conference (IMC'07)* (San Diego, CA, 2007).

[37] J. Leskovec, J. Kleinberg, and C. Faloutsos, ACM Transactions on Knowledge Discovery from Data (TKDD) **1**, 2 (2007).

[38] M. A. Porter, P. J. Mucha, M. E. J. Newman, and C. M. Warmbrand, Proceedings of the National Academy of Sciences **102**, 7057 (2005).

[39] J. Yang and J. Leskovec, Knowledge and Information Systems **42**, 181 (2015).

[40] B. Klimt and Y. Yang, in *CEAS* (2004).

[41] P. Panzarasa, T. Opsahl, and K. M. Carley, Journal of the Association for Information Science and Technology **60**, 911 (2009).

[42] J. Leskovec, D. P. Huttenlocher, and J. M. Kleinberg, in *Proceedings of the Internatonal Conference on Web and Social Media* (2010).

[43] J. Leskovec, J. Kleinberg, and C. Faloutsos, in *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (ACM, 2005) pp. 177–187.

[44] M. Ripeanu, A. Iamnitchi, and I. Foster, IEEE Internet Computing **6**, 50 (2002).

[45] M. Kaiser, New Journal of Physics **10**, 083042 (2008).

[46] B. Bollobás, European Journal of Combinatorics **1**, 311 (1980).

[47] R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon, arXiv preprint cond-mat/0312028 (2003).

[48] J. Park and M. E. J. Newman, Physical Review E **70**, 066117 (2004).

[49] P. Colomer-de Simón, M. Á. Serrano, M. G. Beiró, J. I. Alvarez-Hamelin, and M. Boguñá, Scientific Reports **3**, 2517 (2013).

[50] S. P. Borgatti and M. G. Everett, Social networks **21**, 375 (2000).

[51] P. Rombach, M. A. Porter, J. H. Fowler, and P. J. Mucha, SIAM Review **59**, 619 (2017).

[52] U. Bhat, P. Krapivsky, R. Lambiotte, and S. Redner, Physical Review E **94**, 062302 (2016).

[53] R. Lambiotte, P. Krapivsky, U. Bhat, and S. Redner, Physical Review Letters **117**, 218301 (2016).

[54] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).

[55] L. R. Varshney, B. L. Chen, E. Paniagua, D. H. Hall, and D. B. Chklovskii, PLOS Computational Biology **7**, e1001066 (2011).

[56] J. Ugander, L. Backstrom, and J. Kleinberg, in *Proceedings of the 22nd international conference on World Wide Web* (ACM, 2013) pp. 1307–1318.

[57] I. Scholtes, arXiv:1702.05499 (2017).

[58] V. Sekara, A. Stopczynski, and S. Lehmann, Proceedings of the National Academy of Sciences **113**, 9977 (2016).

[59] A. Fronczak, J. A. Hołyst, M. Jedynak, and J. Sienkiewicz, Physica A: Statistical Mechanics and its Applications **316**, 688 (2002).

[60] G. Caldarelli, R. Pastor-Satorras, and A. Vespignani, The European Physical Journal B: Condensed Matter and Complex Systems **38**, 183 (2004).

[61] R. F. Andrade, J. G. Miranda, and T. P. Lobão, Physical Review E **73**, 046101 (2006).

[62] B. Jiang and C. Claramunt, Environment and Planning B: Planning and Design **31**, 151 (2004).