

PAPER • OPEN ACCESS

Network processes on clique-networks with high average degree: the limited effect of higher-order structure

To cite this article: Clara Stegehuis and Thomas Peron 2021 *J. Phys. Complex.* **2** 045011

View the [article online](#) for updates and enhancements.

You may also like

- [Connectedness percolation in monodisperse rod systems: clustering effects](#)
Avik P Chatterjee
- [General clique percolation in random networks](#)
Jingfang Fan and Xiaosong Chen
- [Spectral theory of sparse non-Hermitian random matrices](#)
Fernando Lucas Metz, Izaak Neri and Tim Rogers

OPEN ACCESS



PAPER

Network processes on clique-networks with high average degree: the limited effect of higher-order structure

Clara Stegehuis^{1,*} and Thomas Peron² ¹ Department of Electrical Engineering, Mathematics and Computer Science, Twente University, The Netherlands² Institute of Mathematics and Computer Science, University of São Paulo, São Carlos 13566-590, São Paulo, Brazil

* Author to whom any correspondence should be addressed.

E-mail: c.steghuis@utwente.nl and thomas.peron@usp.br**Keywords:** random graphs, configuration model, network processes, percolation, synchronization, transitivity

RECEIVED

15 July 2021

REVISED

26 October 2021

ACCEPTED FOR PUBLICATION

2 November 2021

PUBLISHED

22 November 2021

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

**Abstract**

In this paper, we investigate the effect of local structures on network processes. We investigate a random graph model that incorporates local clique structures, and thus deviates from the locally tree-like behavior of most standard random graph models. For the process of bond percolation, we derive analytical approximations for large percolation probabilities and the critical percolation value. Interestingly, these derivations show that when the average degree of a vertex is large, the influence of the deviations from the locally tree-like structure is small. In our simulations, this insensitivity to local clique structures often already kicks in for networks with average degrees as low as 6. Furthermore, we show that the different behavior of bond percolation on clustered networks compared to tree-like networks that was found in previous works can be almost completely attributed to differences in degree sequences rather than differences in clustering structures. We finally show that these results also extend to completely different types of dynamics, by deriving similar conclusions and simulations for the Kuramoto model on the same types of clustered and non-clustered networks.

1. Introduction

One of the problems that has motivated research in network science to a large extent is the assessment of how structural characteristics of real-world networks determine the performance of dynamical processes that take place on them [26]. Most analytical approaches to this problem use networks constructed via configuration models [5] as the substrate for the dynamics. In such models, one specifies the fraction of vertices with k neighbors p_k . A sequence of vertex degrees $\{k_1, \dots, k_N\}$ is then drawn independently following p_k , and the network is assembled by choosing from this sequence pairs of ‘half-edges’ (or stubs) uniformly at random, which are joined to form complete edges. While this method is able to generate networks with any prescribed degree distribution along with offering great analytic tractability, it has the shortcoming that the generated networks are locally tree-like. That is, the density of cycles vanishes asymptotically as the network size increases. This contrasts markedly with the rich topological structure of real-world networks, which often exhibit short cycles, degree correlations and clustering (i.e., the tendency of groups of three vertices to form triangles). Clustering is common to a variety of systems, but it is specially important in social networks, where the average probability that two neighbors of a vertex are also neighbors themselves (also referred to as clustering coefficient) often reaches values of tens of percent [26]. Other classes of systems known to be highly clustered in this sense comprise biological and information networks [26]. Hence, the inclusion of triangles and other types of subgraphs in random network models appears to be a crucial step to model dynamical process on networks accurately.

A practical method to create analytically tractable random networks with a more realistic clustering structure is to extend the standard configuration in order to explicitly include the generation of motifs that yield clustering. The first model of this kind was proposed independently by Newman [27] and Miller [24]. This

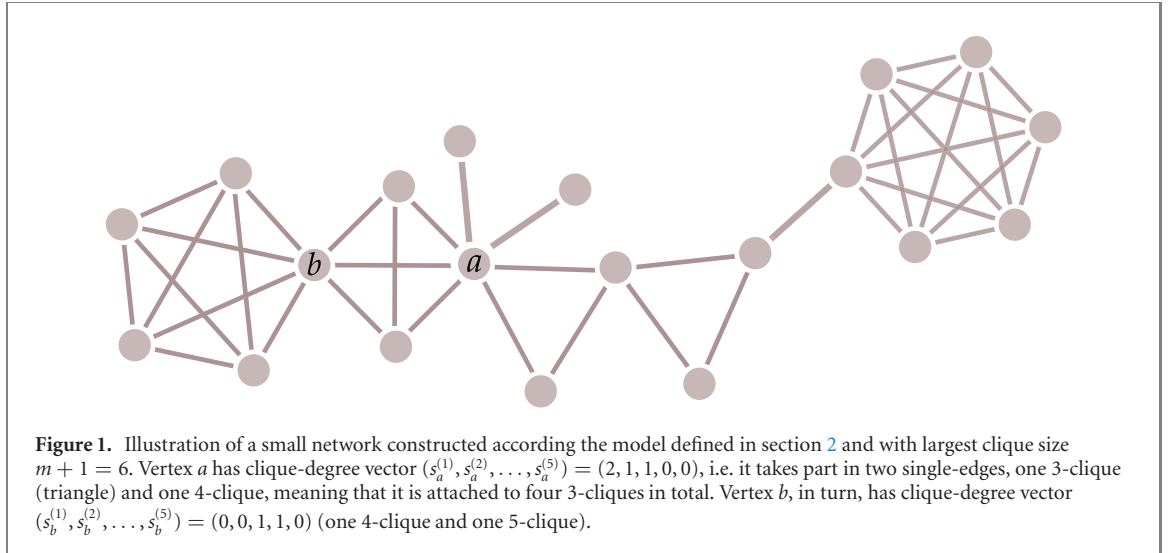
model sets two degree sequences drawn from a joint degree distribution: the first sequence prescribes how many edges each vertex is incident to, exactly as in the standard configuration model; and the second degree sequence defines the number of triangles to which each vertex is attached. As the model then matches these stubs accordingly into edges and triangles, it generates networks with non-vanishing clustering even in the limit of large sizes [24, 27]. This strategy can be adapted to produce networks not only with triangles, but also with distributions of cliques of larger size [9], different types of subgraphs [19], or edge-multiplicities [38].

A number of previous authors have investigated the impact of added clustering on several types of network dynamics by employing such extensions of the standard configuration model. For instance, using a model that created networks with arbitrary distributions of cliques [9], Gleeson *et al* [10] showed that clustered networks exhibit higher bond percolation thresholds in comparison to locally tree-like structures with same degree distributions and correlation properties. Very recently, Mann *et al* [21] studied the percolation properties of the model by Karrer and Newman [19] under different combinations of cycles and cliques as building blocks for the networks. The authors confirmed that the increased clustering created by cliques leads to higher percolation thresholds [21]. On the other hand, the dynamics of networks containing only cycles were shown to approach the result obtained for the configuration model when the length of these cycles increases, as the model then becomes more locally tree-like. A different method to add clique structures to standard configuration models is to use household models, where every vertex of the configuration model is exploded into a clique of a specified size [1]. In this model, clustering was found to increase the percolation threshold [2, 6]. However, when including other clustered subgraphs than cliques, the percolation threshold may either increase or decrease compared to a locally tree-like model [18, 34].

Network processes on configuration models with higher-order clustering find their widest application in mathematical epidemiology, because of the natural importance of modeling of outbreaks in real-world scenarios and the close analogy between disease spreading and percolation processes. Indeed, many results uncovered in the context of percolation have counterparts in disease spreading. For instance, the presence of triangles has been found to increase the epidemic threshold while decreasing the outbreak size [24]. Likewise, networks composed of cycles have been shown to yield epidemic dynamics similar to those of tree-like networks as the length of these cycles increases [32]. Examples of other dynamics investigated with higher-order configuration models include cascade propagation [13, 14], the Ising model [16], and synchronization of coupled oscillators [30].

In this paper we reveal an effect that seems to have remained unnoticed in many previous works; namely, we show that the influence of higher-order subgraphs on network dynamics is negligible when the average degree is large. This phenomenon has also been observed in several real-world networks [23, 25]. In this paper, we show this insignificance of clustering analytically for a popular random graph model that includes clustering in the standard configuration model [27] and show that, in this model, the percolation dynamics of clustered networks for large outbreaks as well as the critical percolation value converge to the one expected for locally tree-like networks. We focus on the most clustered subgraphs possible: cliques of different sizes. While our analytical results are for the large average degree limit, in our simulations, this convergence kicks in for average degrees as small as 6 for several degree distributions. We also show that these conclusions hold for the synchronization transition of phase oscillators modeled by the Kuramoto model [33], indicating that the insensitivity to local network structures may hold for a wide range of network processes. While the effect of added clustering is non-trivial and can either decrease or increase the percolation threshold in any finite network, as was shown in a large body of previous research [2, 6, 10, 18, 21, 34], we show that these increases or decreases are small, and vanish when the average vertex degree becomes large.

Organization of the paper. We first describe the random graph model with subgraphs in section 2. We then focus on the setting where the network is formed by k -cliques of one given size. In section 3, we show that in such networks, size of the largest component under percolation becomes *independent* of the clique structures for large percolation probabilities. We then turn to small percolation probabilities in section 3.1, where we show that the critical percolation threshold also can be approximated by a k -independent value when the average degree of the network is large. We then investigate a setting where different clique sizes are present, in section 4. We show that even in this setting, where it has been reported that the possible introduction of degree correlations can affect the size of the largest component under percolation, when the average degree grows, the giant component only depends on the degree distribution of the network, not on the specific clique sizes. Finally, in section 5, we use analytical approximations as well as simulations to show that for a very different network process, the Kuramoto model, this insensitivity for local clustered network structures also appears for networks of large average degrees.



2. Random graph model with clique subgraphs

As a random graph model, we employ the random graph model with clustering developed in [19, 27]. This random graph model is a general framework that extends the configuration model to create networks with specified densities of arbitrary specified subgraphs. Including clustered subgraphs in the set of specified subgraphs enables to overcome the locally tree-like property of the standard configuration model.

In this manuscript we focus on the most clustered sets of subgraphs, cliques. That is, every vertex has a joint clique degree vector $(s^{(1)}, \dots, s^{(m)})$, where $m + 1$ denotes the largest clique size. Here $s_i^{(1)}$ denotes the edge-degree of vertex i , and $s_i^{(j)}$ denotes the clique-degree of size $j + 1$ of vertex i . The clique-degree of a vertex describes the prescribed involvement of a vertex in non-overlapping maximal cliques of a specified size. Thus, a vertex of clique degree $s_i^{(2)} = 3$ is prescribed to participate in 3 cliques of size 3. Note that this clique-degree encompasses the number of cliques of a given size that are prescribed by the clique-degree vector. However, a vertex of k -clique-degree l may in fact be incident of more cliques of size $k + 1$. For example, a vertex with clique-degree vector $(2, 1, 1)$ (see figure 1) takes part in one 3-clique and one 4-clique, but this 4-clique also contains three 3-cliques incident to the same vertex, which are not counted in the clique-degree vector. Thus, this particular vertex is in fact part of four 3-cliques. We denote the probability that a vertex has clique-degrees $s^{(1)}, \dots, s^{(m)}$ by $q_{s^{(1)}, \dots, s^{(m)}}$. The degree or the total number of connections of vertex i is then described by $\sum_{j=1}^m j s_i^{(j)}$, because every clique of size $j + 1$ adds j connections to the vertex. We denote the degree distribution of a vertex by p_k , so that

$$p_k = \sum_{s^{(1)}, \dots, s^{(m)}=1}^{\infty} q_{s^{(1)}, \dots, s^{(m)}} \mathbb{1}_{s^{(1)} + 2s^{(2)} + \dots + ms^{(m)} = k}, \quad (1)$$

where $\mathbb{1}$ is the indicator function. After sampling a joint clique-degree for every vertex, the network is then formed by selecting j uniformly chosen clique-edges of size j , and pairing the corresponding vertices into a clique. This process is continued for all j until all clique-edges have been paired into a clique. This is an extension of the standard configuration model, where the network is formed by pairing two uniformly chosen half-edges until all half-edges have been paired. Such pairing schemes can form multiple edges or self-loops, creating overlapping cliques, but the probability of such multiple edges or self-loops becomes small in the large-graph limit [19, 24]. The structure of a network constructed via the model described above is illustrated in figure 1.

3. Bond percolation with general cliques

We now investigate the behavior of this network model under bond percolation, where every edge is removed independently with probability $1 - \pi$. We first focus on the case where every vertex is part of only k -cliques. Let q_i denote the probability that a randomly chosen vertex is part of i -cliques. Define the generating functions

$$g(x) = \sum_{i=1}^{\infty} q_i x^i, \quad g_p(x) = \frac{1}{\langle s \rangle} \sum_{i=1}^{\infty} i q_i x^{i-1} = \frac{g'(x)}{\langle s \rangle}, \quad (2)$$

where $\langle s \rangle$ denotes the average number of k -cliques a vertex is part of. Let u denote the probability that a randomly chosen clique-edge is not connected to the giant component. We are interested in the fraction of vertices in the largest component after percolation, S , which can be obtained by [27],

$$u = g_p \left(\sum_{j=0}^{k-1} h(k, j, \pi) u^j \right) \quad (3)$$

$$S = 1 - g \left(\sum_{j=0}^{k-1} h(k, j, \pi) u^j \right),$$

where $h(k, j, \pi)$ is the probability that a given vertex of a k -clique is still connected to j other vertices of the clique after percolation with probability π . These implicit equations are in general difficult to solve [19, 22], so that it is difficult to make general observations on the solution of these equations. Therefore, we here focus on an approximation of S , first for large component sizes (π large), and then for small ones (approximating the critical value where S becomes larger than zero). In these approximations, we will assume that the number of connections of a vertex is large.

When the degree of a typical vertex is large, the probability that a randomly chosen clique-edge is not connected to the giant component becomes small, as it is likely to lead to a large-degree vertex. As u denotes the probability that a randomly chosen clique-edge does not lead to the giant component, we expand (3) with a first-order Taylor expansion around $u = 0$. This yields

$$u = g_p(h(k, 0, \pi) + h(k, 1, \pi)u). \quad (4)$$

For a clique vertex to be disconnected from the rest of the clique, $k - 1$ edges need to be not present, so that $h(k, 0, \pi) = (1 - \pi)^{k-1}$. For a clique vertex to be connected to only one other clique vertex, there are $k - 1$ choices for this other vertex, and then the connections from these two vertices to the other $k - 2$ vertices need to be not present, giving $h(k, 1, \pi) = (k - 1)\pi(1 - \pi)^{2(k-2)}$. Filling in these yields

$$u \approx g_p((1 - \pi)^{k-1} + u\pi(1 - \pi)^{2(k-2)})$$

$$\approx g_p((1 - \pi)^{k-1}) + g'_p((1 - \pi)^{k-1})(k - 1)\pi(1 - \pi)^{2(k-2)}u. \quad (5)$$

This results in

$$u \approx \frac{g_p((1 - \pi)^{k-1})}{1 - g'_p((1 - \pi)^{k-1})(k - 1)\pi(1 - \pi)^{2(k-2)}}. \quad (6)$$

Using a first order Taylor expansion, (3) then yields for S ,

$$S \approx 1 - g(h(k, 0, \pi) + h(k, 1, \pi)u)$$

$$\approx 1 - g(h(k, 0, \pi)) + g'(h(k, 0, \pi))h(k, 1, \pi)u$$

$$= 1 - g((1 - \pi)^{k-1}) - \frac{g_p((1 - \pi)^{k-1})(k - 1)(1 - \pi)^{2(k-2)}\pi g'_p((1 - \pi)^{k-1})}{1 - g'_p((1 - \pi)^{k-1})(k - 1)\pi(1 - \pi)^{2(k-2)}}$$

$$= 1 - g((1 - \pi)^{k-1}) - \frac{\langle s \rangle g_p((1 - \pi)^{k-1})^2 (k - 1)(1 - \pi)^{2(k-2)}\pi}{1 - g'_p((1 - \pi)^{k-1})(k - 1)\pi(1 - \pi)^{2(k-2)}}, \quad (7)$$

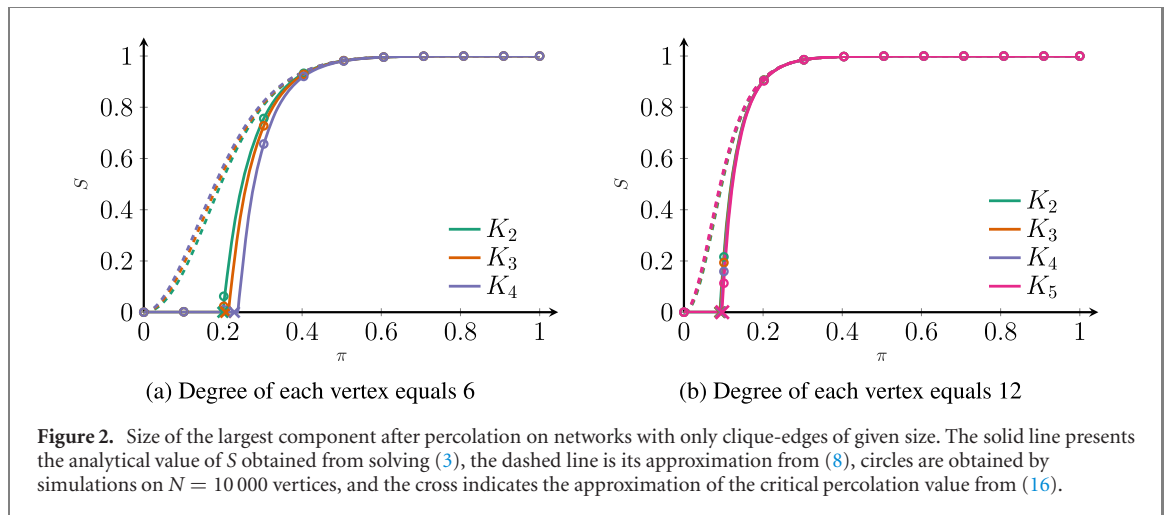
where $\langle s \rangle$ again denotes the average number of cliques a vertex is part of.

Now $g((1 - \pi)^{k-1}) = g_D(1 - \pi)$, where $g_D(x) = \sum_k p_k x^k$ is the generating function of the vertex degrees from (1). This means that for a given degree distribution D , the leading order term of the approximation of the largest component size does not depend on the clique size in which the vertex degrees are split. Furthermore, we show in appendix B that the numerator of the second term also only depends on the degree distribution, not on the clique structure. Furthermore, $g'_p((1 - \pi)^{k-1})$ decreases when the network degrees increase. Thus, large giant components become asymptotically independent of the clique structures in the networks.

Example: regular degrees. We now apply our approximations to several frequently used degree distributions. In regular networks, every vertex is part of sk -cliques. Then, $g(x) = x^s$ and $g_p(x) = x^{s-1}$, so that (7) becomes

$$S = 1 - (1 - \pi)^{s(k-1)} - \frac{(1 - \pi)^{2s(k-1)-2} s(k-1)\pi}{1 - (s-1)(k-1)(1 - \pi)^{(k-1)s-2}\pi}. \quad (8)$$

Now $s(k-1)$ is the degree of a vertex. Equation (8) therefore shows that fixing the degree of a vertex, and changing k (by decreasing or increasing s) does not influence the leading term for the giant component size



S . Furthermore, the larger s , so the larger the average degree of a vertex, the more dominant the first term becomes. Thus, the larger the degree of a vertex, the smaller the influence of the clique structure of the network on percolation processes.

In particular, fixing the degree of a vertex at $s(k-1) = d$ and investigating the difference between choosing cliques of size $k = i$ or $k = j$ yields

$$S_{K_i} - S_{K_j} = \frac{d\pi(1-\pi)^{2d-2}}{1 - d\pi(1-\pi)^{d-2} + (i-1)\pi(1-\pi)^{d-2}} - \frac{d\pi(1-\pi)^{2d-2}}{1 - d\pi(1-\pi)^{d-2} + (j-1)\pi(1-\pi)^{d-2}} = O(d\pi^2(1-\pi)^{3d-4}(j-i)). \quad (9)$$

Thus, by making d larger, it is always possible to get $S_{K_i} - S_{K_j}$ arbitrarily small. This indeed shows that when the average degree of a network is large, the influence of the clique structure of the model becomes irrelevant.

Figure 2 shows the behavior of the approximation of (8) for three networks, one consisting of only edges (the standard configuration model), one only of triangle-edges, and the other only of K_4 -edges. We see that the approximation of (8) works well when S is large for all networks. Furthermore, the size of the largest component under percolation differs more between K_3 and K_4 than between K_2 and K_3 under small average degree in figure 2(a), while these differences have washed away in figure 2(b) under higher average degree. In the simulations, double edges, cliques, self-loops that can possibly be created by the configuration model-like construction of the network have been removed. However, these events are sufficiently rare to not affect the limiting degree distribution when the network size tends to infinity [17].

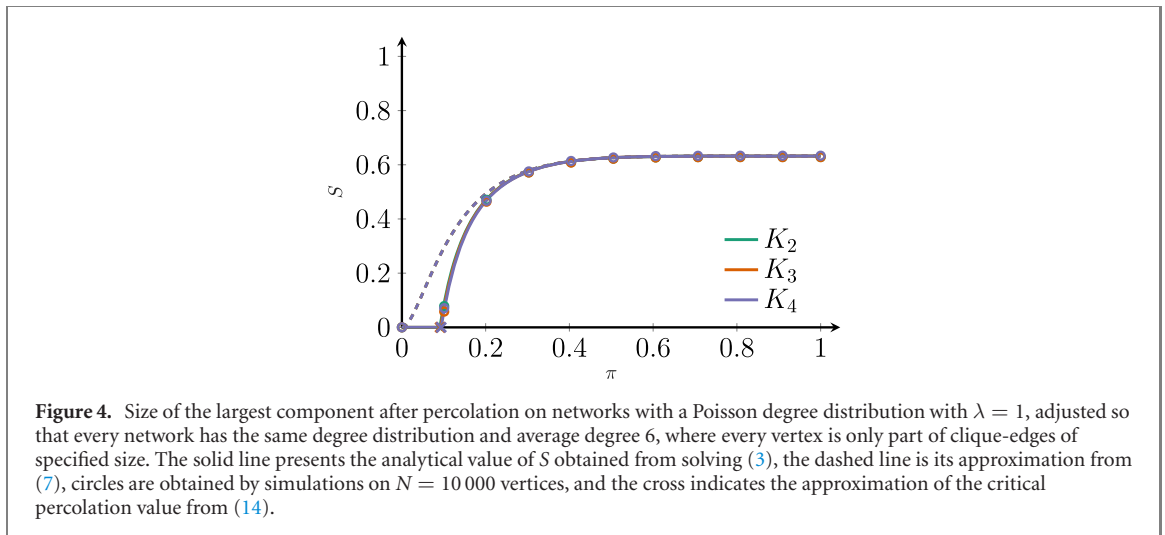
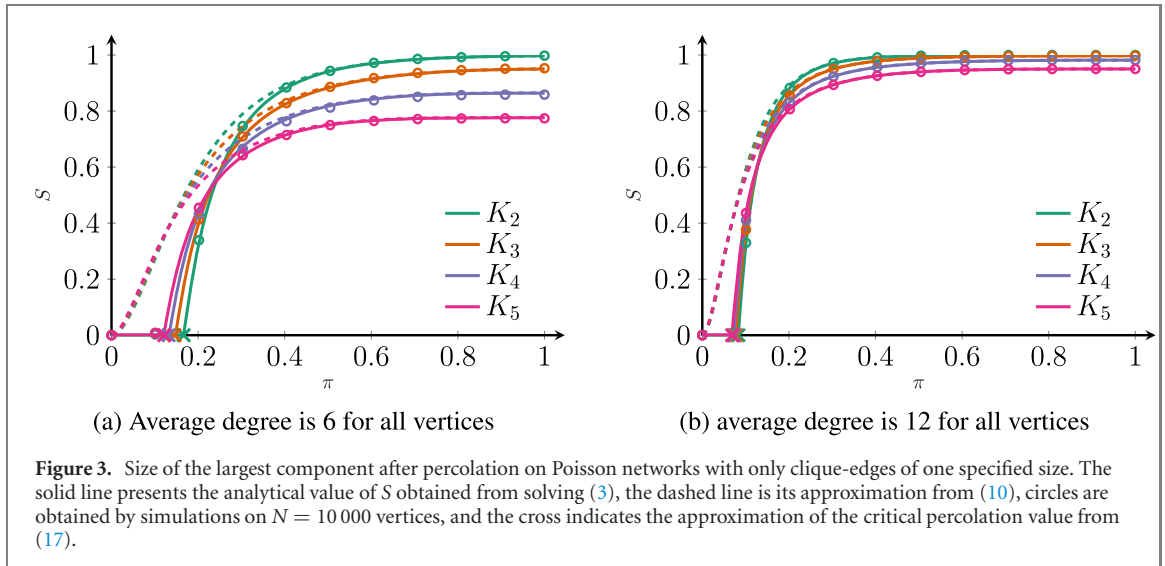
Example: Poisson degrees. Under a Poisson degree distribution where every vertex is part of on average sk -cliques, the generating functions of (3) become $g_p(x) = g(x) = e^{s(x-1)}$. Then, (7) becomes

$$S \approx 1 - e^{s((1-\pi)^{k-1}-1)} \times \left(1 + \frac{s(k-1)\pi(1-\pi)^{2(k-2)}e^{s((1-\pi)^{k-1}-1)}}{1 - s e^{s((1-\pi)^{k-1}-1)}(k-1)\pi(1-\pi)^{2(k-2)}} \right). \quad (10)$$

Figure 3 shows the behavior of the approximation of (8) for four networks, one consisting of only edges, one of only triangle-edges, one of only of K_4 -edges and one of only K_5 -edges. We see that for these Poisson degree distributions, the difference between the large component sizes are well approximated by (10), but that these final sizes still differ quite a bit even for large average degrees. This is caused by the fact that for Poisson clique-degrees, the degree distributions of the different clique sizes are *not the same*.

Indeed, if we focus on the 2-clique case, a vertex can have degree 0, 1, 2, ... when its degree is sampled from a Poisson degree distribution. However, a vertex that is part of triangles, can only have degrees 0, 2, 4, 6, ... when the number of triangles is sampled from a Poisson distribution. In general, a vertex that is only part of k -cliques can only have degrees 0, $k-1$, $2(k-1)$, ... Even when the average values of the Poisson distributions are tuned as $\lambda/(k-1)$ to make sure that, on average, all vertices have the same average number of connections, the degree distributions are not the same. This makes the leading order term in (10) different for different clique sizes. In particular, the probability of having zero connections increases, which makes the final component size smaller when the clique size increases.

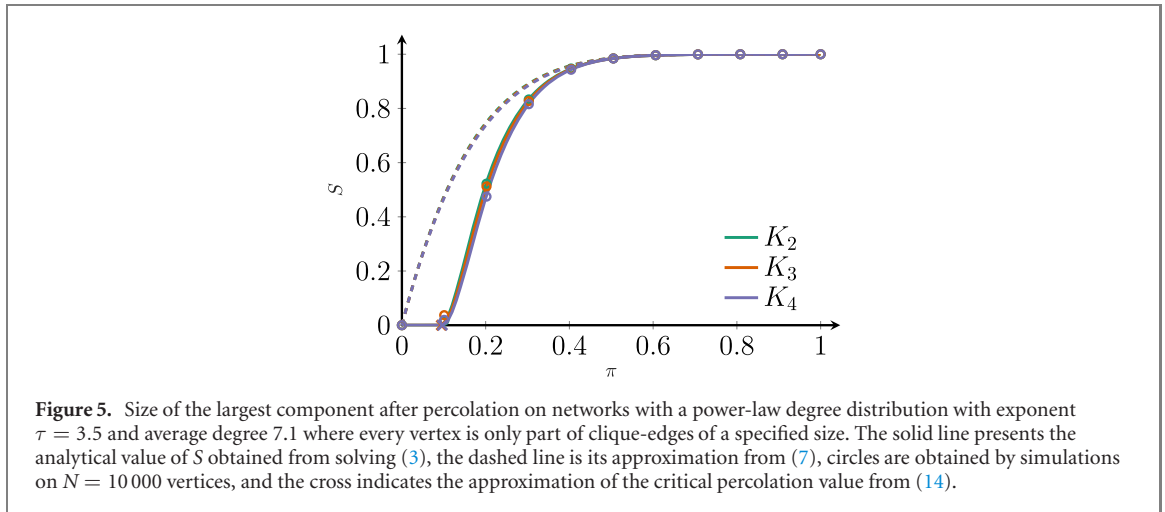
To overcome this problem, we now generate networks with different clique sizes with the same degree distribution. We do this by generating the K_4 network by sampling a Poisson random variable for each vertex, which we multiply by 2. This is the K_4 degree for each vertex. For the K_3 network, we sample a Poisson random



variable with the same mean for each vertex, which we multiply by 3. This is the K_3 degree for each vertex. For the edge-network we again sample a Poisson random variable with the same mean for each vertex, which we multiply by 6. This is the edge-degree for each vertex. Now, in all three networks, vertices can only have degrees $0, 6, 12, \dots$, and the degree distribution across the three networks is the same. Figure 4 shows the results on percolation on these types of networks. We see that in this case, the percolation curves of these Poisson networks of different clique sizes completely overlap, even while the average degree in this setting is only 6. Thus, the difference between networks of different clique structures under Poisson degree distributions reported in figure 3, but also in [10, 19, 27], does in fact not seem to be caused by the clique structure of the network, but by the fact that the degree distributions of the networks are different, changing the leading order term in (7).

For these Poisson networks, figure 4 shows that the percolation threshold and percolation behavior almost perfectly overlaps for cliques of size 2, 3 and 4, in contrast with the results on the regular graph of figure 2(a), where a higher average degree was necessary for convergence. This faster convergence in the Poisson setting can be ascribed to the fact that for the Poisson distribution, the excess degree distribution when arriving at a vertex through an edge is the same as the degree distribution. In general distributions this is not the case. Arriving at a vertex through a clique of size k already uses up $k - 1$ edges. Thus, for general distributions, the larger k , the lower the typical number of new cliques that will be reached from this vertex. This effect is particularly large when k is large and the average degree is small. In Poisson distributions on the other hand this effect vanishes due to the fact that the excess distribution is the same as the original degree distribution, so that the overlap already appears for lower average degrees.

Example: power-law degrees. For networks with power-law degrees, we can follow the same approach as for the Poisson networks. We generate power-law random variables, multiply them by 2 for the K_4 network, by



3 for the triangle-networks, and by 6 for the edge-network to ensure that all networks have the same degree distribution. Using that $g(x) = \text{Li}_\tau(z)/\zeta(z)$ is the generating function of a power-law random variable with exponent τ , we can again find the approximation of the largest component size under percolation for large π from (7). Figure 5 shows that also for power-law random networks, these component sizes of the different networks are similar.

3.1. Approximation of π_c for general clique-degree distributions

We now turn to investigating the similarity of small component sizes. In particular, we approximate the critical percolation value π_c . The critical value π_c is obtained when the average number of neighbors of a vertex reached by following a randomly chosen edge after percolation equals one. Thus,

$$1 = \frac{\langle s^2 \rangle - \langle s \rangle}{\langle s \rangle} \sum_{j=1}^{k-1} j h(k, j, \pi_c), \quad (11)$$

where $\frac{\langle s^2 \rangle - \langle s \rangle}{\langle s \rangle}$ equals the average number of k -cliques connected to a vertex reached from an arbitrary k -clique.

For large average degrees, the critical percolation value is achieved at small π . Therefore, we only keep terms of order π^2 or less. The only terms in the summation above with terms of order π^2 or less are $h(k, 1, \pi)$ and $h(k, 2, \pi)$, as reaching 3 or more other vertices in a clique requires at least 3 edges to be present, giving a contribution of at least π^3 . By filling in $h(k, 1, \pi) = (k-1)\pi(1-\pi)^{2(k-2)}$ and $h(k, 2, \pi) = (k-1)(k-2)(3\pi^2(1-\pi)^{3(k-3)+1} + \pi^3(1-\pi)^{3(k-3)})$, we approximate

$$1 \approx \frac{\langle s^2 \rangle - \langle s \rangle}{\langle s \rangle} \left((k-1)\pi_c(1-\pi_c)^{2(k-2)} + (k-1)(k-2) \left(3\pi_c^2(1-\pi_c)^{3(k-3)+1} + \pi_c^3(1-\pi_c)^{3(k-3)} \right) \right). \quad (12)$$

Keeping only the terms of order π_c^2 or less gives

$$\begin{aligned} 1 &\approx \frac{\langle s^2 \rangle - \langle s \rangle}{\langle s \rangle} (k-1) (\pi_c - 2(k-2)\pi_c^2 + (k-2)3\pi_c^2) \\ &= \frac{\langle s^2 \rangle - \langle s \rangle}{\langle s \rangle} (k-1) (\pi_c + (k-2)\pi_c^2). \end{aligned} \quad (13)$$

This is a quadratic equation that has its positive solution at

$$\pi_c = \frac{-1 + \sqrt{1 + \frac{4(k-2)}{\frac{\langle s^2 \rangle - \langle s \rangle}{\langle s \rangle} (k-1)}}}{2(k-2)}. \quad (14)$$

When the average degree, and therefore also $\frac{\langle s^2 \rangle - \langle s \rangle}{\langle s \rangle} (k-1)$, becomes large, we use a first order Taylor expansion of $\sqrt{1 + 1/x}$ for large x . Then, π_c can be approximated by

$$\pi_c \approx \frac{1}{\frac{\langle s^2 \rangle - \langle s \rangle}{\langle s \rangle} (k-1)}. \quad (15)$$

The term in the denominator describes the average number of vertices reached by coming from a randomly chosen clique-edge, without percolation. When we compare two networks with different clique structures but with the same degree distribution, $\frac{\langle s^2 \rangle}{\langle s \rangle} (k-1)$ is the same for the different networks. Furthermore, this quantity is increasing in the average degree. Thus, in the large average degree-regime π_c converges to a value that is independent of the clique structure of the network.

Regular networks. In networks where every vertex is connected to s k -cliques, we can reduce (14) in the following way. Using that the degree of a vertex $d = s(k-1)$, (14) becomes

$$\pi_c = \frac{-1 + \sqrt{1 + \frac{4(k-2)}{d-(k-1)}}}{2(k-2)} \approx \frac{1}{d - (k-1)}. \quad (16)$$

Thus, when d increases, π_c approaches the same value for all cliques sizes k . Furthermore, the larger k , the larger the difference between π_c when increasing k by one. Figure 2 shows the approximated value of π_c from equation (16) versus the analytical values of the giant component sizes. We see that already for an average degree of 6 this approximation is quite good, and that for larger average degree of 12, indeed the values of π_c for the network of triangles and K_4 cliques almost overlap.

Poisson networks. In Poisson networks where the average vertex is part of sk -cliques with fixed $d = s(k-1)$, $\frac{\langle s^2 \rangle - \langle s \rangle}{\langle s \rangle} = s$. Then (14) becomes

$$\pi_c = \frac{-1 + \sqrt{1 + \frac{4(k-2)}{d}}}{2(k-2)} \approx \frac{1}{d}. \quad (17)$$

Thus, when d gets large, again π_c approaches the same value for all cliques sizes k . Figure 3 shows that this is a good approximation of the critical percolation value π_c , and that for an average degree of 12, these values become very close under different clique sizes.

4. Mixed clique sizes

We now investigate networks where cliques of different sizes are present. By introducing different clique sizes, it is possible to create degree–degree correlations that have often been said to influence the largest component size after percolation [3, 4, 12]. Thus, we now investigate to what extent the introduction of mixed clique sizes influences the size of the largest component.

Under bond percolation of networks where every vertex is part of s_1 cliques of size k_1 , and s_2 cliques of size k_2 with probability q_{s_1, s_2} , the generating-function methodology gives the following results. Let $g(x, y) = \sum_{s_1, s_2 \geq 0} p_{s_1, s_2} x^{s_1} y^{s_2}$ be the generating function of the clique degrees. Furthermore, let

$$g_p(x, y) = \frac{1}{\langle s_1 \rangle} \sum_{s_1, s_2 \geq 0} s_1 q_{s_1, s_2} x^{s_1-1} y^{s_2}, \quad (18)$$

$$g_q(x, y) = \frac{1}{\langle s_2 \rangle} \sum_{s_1, s_2 \geq 0} s_2 q_{s_1, s_2} x^{s_1} y^{s_2-1}, \quad (19)$$

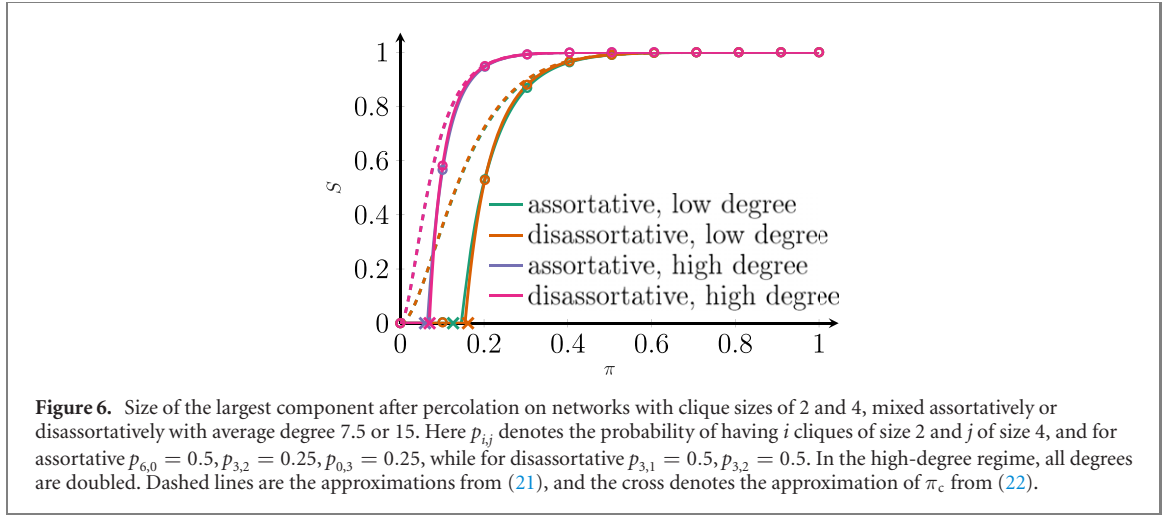
with

$$\langle s_1 \rangle := \sum_{s_1, s_2 \geq 0} s_1 q_{s_1, s_2} \quad \text{and} \quad \langle s_2 \rangle := \sum_{s_1, s_2 \geq 0} s_2 q_{s_1, s_2}$$

being the generating functions of the number of cliques that are reached by following a randomly chosen clique-edge. Let u denote the probability that a randomly chosen k_1 -clique-edge is not connected to the giant component. Similarly, let v denote the probability that following a randomly chosen k_2 -clique edge does not lead to the largest component.

We show in appendix A that u and v can be approximated by

$$\begin{aligned} u &\approx \frac{g_p((1-\pi)^{k_1-1}, (1-\pi)^{k_2-1})}{A(k_1, k_2, \pi)} \\ v &\approx \frac{g_q((1-\pi)^{k_1-1}, (1-\pi)^{k_2-1})}{A(k_1, k_2, \pi)}. \end{aligned} \quad (20)$$



Furthermore, the giant component size S is then approximated as

$$S = 1 - g_D(1 - \pi) - \pi(1 - \pi)^{2(k_1-2)}(k_1 - 1) \frac{g_p((1 - \pi)^{k_1-1}, (1 - \pi)^{k_2-1})^2 \langle s_1 \rangle}{A(k_1, k_2, \pi)} - \pi(1 - \pi)^{2(k_2-2)}(k_2 - 1) \frac{g_q((1 - \pi)^{k_1-1}, (1 - \pi)^{k_2-1})^2 \langle s_2 \rangle}{A(k_1, k_2, \pi)}, \quad (21)$$

where

$$A(k_1, k_2, \pi) = 1 - (k_1 - 1)\pi(1 - \pi)^{2(k_1-1)} \frac{\partial g_p((1 - \pi)^{k_1-1}, (1 - \pi)^{k_2-1})}{\partial x} - (k_2 - 1)\pi(1 - \pi)^{2(k_2-1)} \frac{\partial g_q((1 - \pi)^{k_1-1}, (1 - \pi)^{k_2-1})}{\partial y},$$

and where $g_D(x)$ is the generating function of the total vertex degrees. Thus, the leading order term of the giant component size does not depend on the distribution of the clique degrees k_1 and k_2 , but only on the total vertex degree. Furthermore, the numerators of the second order terms also only depend on the degree distribution, and not on the clique sizes, similarly to the one clique-size case.

It is not difficult to extend this analysis to include more than two different clique sizes, where (21) contains terms for all size biased generating functions of the clique sizes g_{k_i} , instead of only g_p and g_q in (21). Therefore, even in the presence of multiple clique sizes that can generate degree–degree correlations, large component sizes are clique-structure independent for large average degrees.

Example: assortative mixing. In several sources of previous work, degree–degree correlations were found to be important for the behavior of percolation processes (see, e.g., [3, 4, 12]). Furthermore, the clustering assortativity, describing the tendency of high-degree vertices to be more clustered than high-degree vertices or vice versa, has also been ascribed strong importance on the behavior of a network under percolation [21]. However, (21) shows that large component sizes only depend on the degree distribution, so that it is independent of any clique correlations in the large degree limit. Figure 6 shows that indeed the influence of mixed clique sizes on the giant component is small, especially in the large average degree regime.

4.1. Approximation of π_c for mixed clique networks

In appendix C we show that π_c can be approximated by

$$\pi_c = \frac{-E_{k_1, k_1} - E_{k_2, k_2}}{2(E_{k_1, k_1} E_{k_2, k_2} - E_{k_1, k_2} E_{k_2, k_1} - E_{k_1, k_1}(k_1 - 2) - E_{k_2, k_2}(k_2 - 2))} + \frac{\sqrt{(E_{k_1, k_1} + E_{k_2, k_2})^2 - 4(E_{k_1, k_1} E_{k_2, k_2} - E_{k_1, k_2} E_{k_2, k_1} - E_{k_1, k_1}(k_1 - 2) - E_{k_2, k_2}(k_2 - 2))}}{2(E_{k_1, k_1} E_{k_2, k_2} - E_{k_1, k_2} E_{k_2, k_1} - E_{k_1, k_1}(k_1 - 2) - E_{k_2, k_2}(k_2 - 2))}, \quad (22)$$

where

$$E_{k_i, k_j} = \left(\frac{\langle s_i s_j \rangle}{\langle s_j \rangle} - \delta_{k_i, k_j} \right) (k_i - 1). \quad (23)$$

For large average degrees, this value can be approximated by

$$\begin{aligned}\pi_c &\approx \frac{1}{E_{k_1,k_1} + E_{k_2,k_2}} \\ &= \frac{1}{\left(\frac{\langle s_1^2 \rangle}{\langle s_1 \rangle} - 1\right)(k_1 - 1) + \left(\frac{\langle s_2^2 \rangle}{\langle s_2 \rangle} - 1\right)(k_2 - 1)}.\end{aligned}\quad (24)$$

In assortative networks, where cliques of a given size are typically also connected to many cliques of the same size, $\left(\frac{\langle s_1^2 \rangle}{\langle s_1 \rangle} - 1\right)(k_1 - 1)$ and $\left(\frac{\langle s_2^2 \rangle}{\langle s_2 \rangle} - 1\right)(k_2 - 1)$ are large, so that we expect π_c to be small. In disassortative networks, where the different clique sizes are more mixed, $\left(\frac{\langle s_1^2 \rangle}{\langle s_1 \rangle} - 1\right)(k_1 - 1)$ and $\left(\frac{\langle s_2^2 \rangle}{\langle s_2 \rangle} - 1\right)(k_2 - 1)$ are smaller. Thus, the degree–degree correlations that are created by the different clique sizes play a role in the critical percolation value π_c , whereas the giant component size of (21) is asymptotically independent of such degree correlations. However, figure 6 shows that these correlations still vanish in the large-degree regime.

5. Phase oscillators coupled on clique-networks

In this section we illustrate the limited effect of higher-order structure on more complex dynamic network processes than bond percolation. In particular, we focus on the dynamics of coupled oscillators. For this purpose, we employ the paradigmatic Kuramoto model [33] that can describe synchronization phenomena on complex networks. In the Kuramoto model, the oscillator of vertex i is characterized by a phase variable θ_i , and the dynamics on a heterogeneous network is dictated by the following equations [33]:

$$\frac{d\theta_i}{dt} = \omega_i + K \sum_{j=1}^N A_{ij} \sin(\theta_j - \theta_i), \quad i = 1, \dots, N, \quad (25)$$

where ω_i is the natural frequency of oscillation of oscillator i , and A_{ij} is the network adjacency matrix. If there is an edge connecting i and j , $A_{ij} = 1$ (0 otherwise), and the interaction between the vertices is weighted by the coupling K . If K is lower than a certain K_c , the oscillators rotate incoherently, each one at its own rhythm set by the natural frequency ω_i . For $K > K_c$, the incoherent state loses stability: a cluster of oscillators is formed around an average phase value, and these units begin to rotate locked in the same frequency [7, 35]. This transition from asynchrony to a partially synchronized state is measured by the Kuramoto order parameter R given by [7, 35]

$$R e^{i\psi} = \frac{1}{N} \sum_{j=1}^N e^{i\theta_j} \quad (0 \leq R \leq 1), \quad (26)$$

where R quantifies the level of synchrony achieved by the oscillators, and ψ is their average phase. While one can monitor the synchronization transition of a heterogeneous network with equation (26), it is not possible to decouple equation (25) in terms of a global order parameter. Instead, in order to perform a self-consistent analysis and characterize the onset of synchronization analytically, we need to employ heterogeneous degree mean-field approximations [33]. This is equivalent to replacing the terms of the adjacency matrix A_{ij} by their ensemble averages in the configuration model, which in the single-edge version is $\langle A_{ij} \rangle = d_i d_j / N \langle d \rangle$. In the model that generates networks with a single clique type the expression is analogous, namely, $\langle A_{ij}^{(c)} \rangle = (k-1) s_i s_j / N \langle s \rangle$, where s_i is the number of cliques attached to vertex i . Replacing A_{ij} by $\langle A_{ij}^{(c)} \rangle$ in equation (25) we obtain

$$\frac{d\theta_i}{dt} = \omega_i + \frac{K(k-1)s_i}{N \langle s \rangle} \sum_{j=1}^N s_j \sin(\theta_j - \theta_i), \quad i = 1, \dots, N, \quad (27)$$

which motivates the definition of the following order parameter

$$r e^{i\phi} = \frac{1}{N \langle s \rangle} \sum_{j=1}^N s_j e^{i\theta_j}, \quad (28)$$

which in turn allows us to rewrite equation (27) as

$$\frac{d\theta_i}{dt} = \omega_i + Kr(k-1)s_i \sin(\phi - \theta_i). \quad (29)$$

In the limit of $N \rightarrow \infty$, we assume that the assignment of cliques and natural frequencies is well described by distributions q_s and $g(\omega)$; we further assume that the collections of vertices with clique number s and frequency ω form a phase density $\rho(\theta, t|s, \omega)$. Thus, we rewrite equation (28) in the continuum limit as

$$r e^{i\phi} = \frac{1}{\langle s \rangle} \sum_s s q_s \iint d\omega d\theta \rho(\theta, t|s, \omega) g(\omega) e^{i\theta}. \quad (30)$$

By choosing $g(\omega) = (\sqrt{2\pi})^{-1} e^{-\omega^2/2}$, we can set $\phi = 0$ without loss of generality. Substituting the stationary synchronous solution of equation (29)—i.e. $\rho(\theta|s, \omega) = \delta\{\theta - \arcsin[\omega/Kr(k-1)s]\}$, for $|\omega| \leq Kr(k-1)s$ —into equation (30), we arrive at the following implicit equation

$$\begin{aligned} \frac{\langle s \rangle}{K} \sqrt{\frac{8}{\pi}} &= (k-1) \sum_s s^2 q_s e^{-K^2(k-1)^2 s^2 r^2 / 4} \\ &\times \left\{ I_0 \left[\frac{K^2(k-1)^2 s^2 r^2}{4} \right] + I_1 \left[\frac{K^2(k-1)^2 s^2 r^2}{4} \right] \right\}, \end{aligned} \quad (31)$$

where $I_0(\cdot)$ and $I_1(\cdot)$ are the modified Bessel functions of the first kind. Thus, with equation (31) we can find the dependence of the order parameter r on the coupling strength K , and thereby assess the impact of different clique sizes on the onset of synchronization. Letting $r \rightarrow 0^+$, we also obtain the expression for the critical coupling

$$K_c = \frac{1}{(k-1)} \frac{\langle s \rangle}{\langle s^2 \rangle} \sqrt{\frac{8}{\pi}}. \quad (32)$$

Thus, again, substituting $d = (k-1)s$ shows that the critical coupling is independent of the clustering structure. However, an immediate problem we face is the fact that the mean field approximations behind equation (31) are accurate only for sufficiently dense networks, typically when the average degree is at least of order of a few dozen [11, 33]. This limits the analytical verification of the effect of cliques on the dynamics of networks as sparse as the ones considered in the previous sections. Nonetheless, in the appropriate regime in which the mean field approach is valid, equation (31) suggests that the conclusions drawn for bond percolation may be similar for synchronization processes: notice that in equation (31) $(k-1)s$ is the actual degree of a vertex; substituting $d = (k-1)s$ in the implicit equation for r and rewriting it in terms of the new variable, we find that the emergence of a synchronous component depends only on the final degree sequence of the network and not on the sizes of the cliques. Therefore, clustered and unclustered networks are expected to exhibit similar dynamics also in the synchronization of coupled oscillators.

In order to confirm the above result, let us first investigate how the critical point K_c changes according to the clique structure. In figure 7 we compare the predictions of K_c by equation (32) with the corresponding quantities obtained via numerical integration of the system (25) for several average degrees $\langle d \rangle$. We numerically detect the transition point between incoherence and partial synchronization by identifying K_c as the position of the divergent peak of the susceptibility $\chi = N(\langle r^2 \rangle_t - \langle r \rangle_t^2) / \langle r \rangle_t$ [28], where $\langle \cdot \rangle_t$ is a long temporal average. As can be seen in figure 7, the agreement between simulation and theoretical values is satisfactorily good for low $\langle d \rangle$, but it is progressively improved as the networks get denser. Furthermore, figure 7 indicates that the transition to synchrony tends to occur sooner as the clique size, and hence the clustering, increases. The numerical value of K_c for different clique sizes becomes statistically equivalent at high $\langle d \rangle$. Yet, the solutions obtained from equation (32) in figure 7 suggest that clustering always ameliorates the network synchrony, an effect that, as seen in figure 7, asymptotically vanishes as $\langle d \rangle$ increases. This is in apparent contradiction with our analysis of equation (31), in that networks with the same degree distribution, regardless of their clustering structure, ought to have identical dependence $r = r(K)$ and critical couplings K_c . However, similarly to the experiments of figure 3, these networks of different clique sizes do not have the same degree distributions.

To verify whether the differences in the onset of synchronization shown in figure 7 are due to discrepancies in the degree sequences or are, on the other hand, a true effect of the clustered structure, we repeat the methodology of the experiments depicted in figure 4. That is, we simulate the oscillators on networks with different clique sizes but adjusted to have the exact same degree sequence. The result is seen figure 8. Again, as in the percolation experiment in figure 4, the synchronization values match almost perfectly despite the difference in the clustering levels: in the examples in figure 8, the single-edge networks (K_2) have transitivity coefficient [27] equal to $C \approx 0$, while networks with K_4 cliques exhibit $C \approx 0.18$ —a significant structural difference that is not reflected in the dynamics. It is also noteworthy that this phenomenon occurs at a low average degree ($\langle d \rangle = 6$), i.e., the regime depicted in figure 7 with the most prominent discrepancies between clustered and unclustered networks. As discussed for the percolation case in section 3, actually those discrepancies are due the fact that the degree distributions are not identical for different clique sizes; as a consequence, the mismatches in the

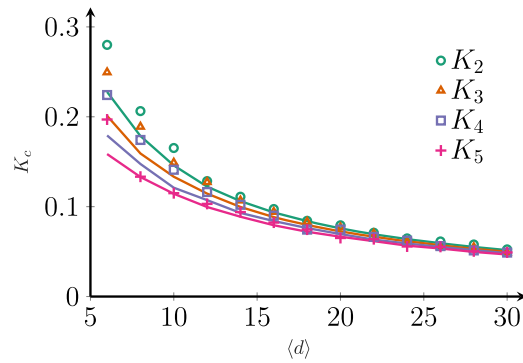


Figure 7. Critical coupling K_c for the onset of synchronization as a function of the average degree $\langle d \rangle$ for the clique-networks with Poisson degree distributions. Solid lines are the solutions of equation (32). Dots correspond to the simulation results obtained by numerically integrating equation (25) using the Heun's method with time step $dt = 0.05$. For each coupling K , the quantities are average over $t \in [500, 1000]$. In all numerical experiments we have $N = 10^4$ oscillators, whose frequencies are distributed according to $g(\omega) = (\sqrt{2\pi})^{-1}e^{-\omega^2/2}$. Different symbols and colors refer to networks constructed with different clique sizes: K_2 denotes networks containing only single-edges (configuration model), and K_5 refers to networks built from sequences of cliques with five vertices.

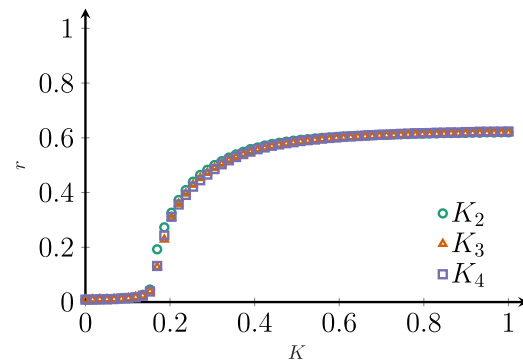


Figure 8. Synchronization diagram showing the evolution of the order parameter r as a function of the coupling K . The networks have Poisson degree distribution adjusted so that every network has the same degree sequence and average degree $\langle d \rangle = 6$, as in figure 4. Dots correspond to the simulation results obtained by numerically integrating equation (25) using the Heun's method with time step $dt = 0.05$. For each coupling K , the order parameters are average over $t \in [500, 1000]$. In all numerical experiments we have $N = 10^4$ oscillators, and the frequencies are distributed according to $g(\omega) = (\sqrt{2\pi})^{-1}e^{-\omega^2/2}$. Different symbols and colors refer to networks constructed with different clique sizes: K_2 denotes networks containing only single-edges (configuration model), and K_5 refers to networks built from sequences of cliques with five vertices.

degree sequence end up generating different solutions for equation (31). The results in figures 4 and 8 therefore show that networks with similar degree distributions and correlations may exhibit equivalent dynamical behavior regardless of their subgraph structure and clustering levels.

We also note that the critical coupling K_c can be estimated via ‘quenched’ mean-field approximations and, for the parameters considered here, expressed in terms of the largest eigenvalue λ_1 of the adjacency matrix as $K_c = \lambda_1^{-1}\sqrt{8/\pi}$ [28, 31]. We can complement the latter expression with the recent results of reference [29] in which the largest eigenvalue for Poisson random networks constructed with K_3 cliques has been estimated as $\lambda_1 = 2\langle s \rangle + 1 + 1/2\langle s \rangle$. In the limit of high average degrees, $\langle s \rangle \rightarrow \infty$, the third term of λ_1 vanishes, and the corresponding result for tree-like Poisson random networks is recovered ($\lambda_1 = \langle d \rangle + 1$). Therefore, also in the quenched mean-field formulation, the value of K_c of clustered networks is expected to asymptotically approach the calculations for tree-like networks, in agreement with the results in figure 7.

6. Conclusion

In this paper, we have investigated the influence of the presence of clustered structures in a random graph model in the form of cliques on two network processes: bond percolation and synchronization. Percolation on such clustered networks has been investigated frequently, but as the equations for the giant component size under percolation are given by several implicit equations that are difficult to analyze mathematically, the factors dominating the behavior of percolation processes on such networks are largely unknown. By approximating the

size of the giant component under large percolation probabilities as well as the critical percolation value where a giant component starts to form, we have found that the degree distribution is the dominant factor in these approximations, especially when the average degree of the network is large. In particular, our approximations are independent of the amount of clustering in the network. This means that introducing clustering by locally inserting cliques or other types of subgraphs in the frequently used locally tree-like random graph models, barely influences the size of the largest component.

We also showed that differences in percolation behavior due to the introduction of cliques in the configuration model that were found in several previous works [10, 19, 21, 27, 36] can be ascribed to the fact that the degree distribution changed in those experiments as well. When keeping the degree distribution fixed while introducing more clustering, this difference becomes small. We also proved that small differences in percolation behavior that remain vanish when the average degree becomes large.

While our approximations show that the dominant factor for the giant component under large percolation probabilities as well as the critical percolation value is the degree distribution, and not the clustering in the network, our simulations show that actually the entire percolation curve seems to become independent of the clustering in the network, once the average degree becomes large. Showing this analytically would be an interesting point for further research.

Furthermore, while we have primarily focused on the process of bond percolation, we also showed that for a different network process of oscillator synchronization, the same independence of higher-order structures is present when the average degree is large. We therefore believe that other processes such as opinion dynamics or contact processes could be independent of the clustering structure of this model as well. Importantly, while we have shown that the size of the giant component and the stationary synchronous states remain unaffected, we note that the transient properties of disease spreading processes may indeed be influenced by the clustering level [15, 37]. Investigating which types of dynamics are independent of the clique structures is therefore an interesting avenue for further research.

In this manuscript, we investigated the percolation properties of a configuration-model type random graph model with enhanced clustering in the form of cliques that do not overlap in their edges. Real-world networks in general however also contain cliques that do have edge-overlaps, or clustered subgraphs that are not necessarily cliques. We believe that our results can be extended to such more realistic settings by using the extension of the random graph model we considered to include arbitrary subgraphs as well [19]. We believe that our Taylor-expansion-based approach still works in this setting with more general subgraphs. In particular, we conjecture that these results also hold for such models with overlapping cliques, where the overlap forms a higher-order locally tree-like structure. Indeed, cliques that overlap at one or more specific edges can then be defined as one subgraph that is included in the model, so that it explicitly includes overlapping cliques. Identifying the exact model conditions under which clustering does not affect the percolation properties would therefore be an interesting line of further research. Furthermore, there is some evidence that clustering properties also barely affect the dynamical processes of real-world networks [11, 23]. It would be interesting to see if such model classifications can also predict for which real-world networks clustering is an important property when investigating network dynamics, and for which ones it is not.

The random graph model we considered connects cliques in a locally-tree like fashion. In other types of models, percolation on locally tree-like networks has been investigated by contracting cliques (or other subgraphs) in a specific way to one or more vertices, to create a reduced tree [6, 17]. Analyzing percolation on this reduced tree then gives the percolation behavior of the clustered network. However, this reduced tree in general has a different degree distribution, and sometimes even a different number of vertices than the original network. As the percolation behavior of locally tree-like networks strongly depends on the degree distribution of the network, this means that this method will still not show the similarity of the process on clustered and non-clustered networks. Furthermore, in the model we investigate, contracting cliques to single vertices is more difficult, as cliques may touch at a single vertex. Still, we believe that the similarity of the two processes on clustered and non-clustered networks has to do with the higher-order locally tree-like structure that is present.

Another interesting line of research following from these results is in higher-order dynamics. We showed that single-edge dynamics on networks where a clique structure is imposed behave similarly as in networks without the clique structure. However, when studying the network model, for example, as a simplicial complex instead, it is possible to impose simplicial dynamics on top of it, where the dynamics involve all clique vertices in the interactions. It would be interesting to see under which conditions on the dynamic process on such a simplicial extension of this model depends on the clique structure, and under which conditions it does not.

Finally, while this work shows that inserting clustering in a locally tree-like model barely affects the behavior of an epidemic process under bond percolation, we believe that in different models, where clustering is introduced by the presence of geometry, or imposed by maximum-entropy based constraints [8, 20], bond percolation can behave very differently under two models of the same degree distribution. Showing general conditions on the network structure under which clustering does or does not affect the size of a giant

component compared to tree-like network models would therefore also be an interesting avenue for further research.

Another interesting question is to identify the conditions on the clique sizes such that clustering does not affect the percolation processes. In this work, we investigated fixed maximal clique sizes, and showed that their effect is small when the average degree becomes large. Thus, an interesting question would be: how large can the clique sizes become compared to the average degree so that these results still hold? Identifying this threshold scaling would be an interesting avenue for further research.

Acknowledgments

CS acknowledges VENI grant 202.001, and TP acknowledges FAPESP (Grant No. 2016/23827-6). This research was carried out using the computational resources of the Center for Mathematical Sciences Applied to Industry (CeMEAI) funded by FAPESP (Grant No. 2013/07375-0).

Data availability statement

No new data were created or analysed in this study.

Appendix A. Computations for the mixed clique sizes

After bond percolation with probability π [27],

$$\begin{aligned} u &= g_p \left(\sum_{j=1}^{k_1-1} h(k_1, j, \pi) u^j, \sum_{j=1}^{k_2-1} h(k_2, j, \pi) v^j \right), \\ v &= g_q \left(\sum_{j=1}^{k_1-1} h(k_1, j, \pi) u^j, \sum_{j=1}^{k_2-1} h(k_2, j, \pi) v^j \right), \end{aligned} \quad (\text{A.1})$$

while $S = 1 - g(\sum_{j=1}^{k_1-1} h(k_1, j, \pi) u^j, \sum_{j=1}^{k_2-1} h(k_2, j, \pi) v^j)$. Again, we expand (A.1) with a first-order Taylor expansion around $u, v = 0$. This yields

$$\begin{aligned} u &= g_p((1 - \pi)^{(k_1-1)} + (k_1 - 1)\pi(1 - \pi)^{2(k_1-2)}u, (1 - \pi)^{(k_2-1)} + (k_2 - 1)\pi(1 - \pi)^{2(k_2-2)}v) \\ v &= g_q((1 - \pi)^{(k_1-1)} + (k_1 - 1)\pi(1 - \pi)^{2(k_1-2)}u, (1 - \pi)^{(k_2-1)} + (k_2 - 1)\pi(1 - \pi)^{2(k_2-2)}v). \end{aligned} \quad (\text{A.2})$$

Taylor expanding $g_p(x, y)$ and $g_q(x, y)$ as well and using that uv is small, we obtain

$$\begin{aligned} u &= g_p((1 - \pi)^{k_1-1}, (1 - \pi)^{k_2-1}) + \frac{\partial g_p((1 - \pi)^{k_1-1}, (1 - \pi)^{k_2-1})}{\partial x} (k_1 - 1)\pi(1 - \pi)^{2(k_1-2)}u \\ &\quad + \frac{\partial g_p((1 - \pi)^{k_1-1}, (1 - \pi)^{k_2-1})}{\partial y} (k_2 - 1)\pi(1 - \pi)^{2(k_2-2)}v \\ v &= g_q((1 - \pi)^{k_1-1}, (1 - \pi)^{k_2-1}) + \frac{\partial g_q((1 - \pi)^{k_1-1}, (1 - \pi)^{k_2-1})}{\partial x} (k_1 - 1)\pi(1 - \pi)^{2(k_1-2)}u \\ &\quad + \frac{\partial g_q((1 - \pi)^{k_1-1}, (1 - \pi)^{k_2-1})}{\partial y} (k_2 - 1)\pi(1 - \pi)^{2(k_2-2)}v. \end{aligned} \quad (\text{A.3})$$

This is a linear system of equations with as solution

$$\begin{aligned} u &= \frac{g_p((1 - \pi)^{k_1-1}, (1 - \pi)^{k_2-1})}{A(k_1, k_2, \pi)} \\ &\quad - \frac{g_p((1 - \pi)^{k_1-1}, (1 - \pi)^{k_2-1})(k_2 - 1)\pi(1 - \pi)^{2(k_2-1)} \frac{\partial g_q((1 - \pi)^{k_1-1}, (1 - \pi)^{k_2-1})}{\partial y}}{A(k_1, k_2, \pi)} \\ &\quad + \frac{g_q((1 - \pi)^{k_1-1}, (1 - \pi)^{k_2-1})(k_2 - 1)\pi(1 - \pi)^{2(k_2-1)} \frac{\partial g_q((1 - \pi)^{k_1-1}, (1 - \pi)^{k_2-1})}{\partial y}}{A(k_1, k_2, \pi)} \\ v &= \frac{g_q((1 - \pi)^{k_1-1}, (1 - \pi)^{k_2-1})}{A(k_1, k_2, \pi)} \end{aligned}$$

$$\begin{aligned}
& + \frac{g_p((1-\pi)^{k_1-1}, (1-\pi)^{k_2-1})(k_1-1)\pi(1-\pi)^{2(k_1-1)} \frac{\partial g_p((1-\pi)^{k_1-1}, (1-\pi)^{k_2-1})}{\partial x}}{A(k_1, k_2, \pi)} \\
& - \frac{g_q((1-\pi)^{k_1-1}, (1-\pi)^{k_2-1})(k_1-1)\pi(1-\pi)^{2(k_1-1)} \frac{\partial g_q((1-\pi)^{k_1-1}, (1-\pi)^{k_2-1})}{\partial x}}{A(k_1, k_2, \pi)}, \tag{A.4}
\end{aligned}$$

where

$$\begin{aligned}
A(k_1, k_2, \pi) = & 1 - (k_1-1)\pi(1-\pi)^{2(k_1-1)} \frac{\partial g_p((1-\pi)^{k_1-1}, (1-\pi)^{k_2-1})}{\partial x} \\
& - (k_2-1)\pi(1-\pi)^{2(k_2-1)} \frac{\partial g_q((1-\pi)^{k_1-1}, (1-\pi)^{k_2-1})}{\partial y}. \tag{A.5}
\end{aligned}$$

This can be approximated by

$$\begin{aligned}
u & \approx \frac{g_p((1-\pi)^{k_1-1}, (1-\pi)^{k_2-1})}{A(k_1, k_2, \pi)} \\
v & \approx \frac{g_q((1-\pi)^{k_1-1}, (1-\pi)^{k_2-1})}{A(k_1, k_2, \pi)}. \tag{A.6}
\end{aligned}$$

This gives for the final component size

$$\begin{aligned}
S = & 1 - g((1-\pi)^{k_1-1}, (1-\pi)^{k_2-1}) - \pi(1-\pi)^{2(k_1-2)}(k_1-1) \\
& \times \frac{g_p((1-\pi)^{k_1-1}, (1-\pi)^{k_2-1}) \frac{\partial g((1-\pi)^{k_1-1}, (1-\pi)^{k_2-1})}{\partial x}}{A(k_1, k_2, \pi)} \\
& - \pi(1-\pi)^{2(k_2-2)}(k_2-1) \frac{g_q((1-\pi)^{k_1-1}, (1-\pi)^{k_2-1}) \frac{\partial g((1-\pi)^{k_1-1}, (1-\pi)^{k_2-1})}{\partial y}}{A(k_1, k_2, \pi)}. \tag{A.7}
\end{aligned}$$

This can be written as

$$\begin{aligned}
S = & 1 - g_D(1-\pi) - \pi(1-\pi)^{2(k_1-2)}(k_1-1) \frac{g_p((1-\pi)^{k_1-1}, (1-\pi)^{k_2-1})^2 \langle s_1 \rangle}{A(k_1, k_2, \pi)} \\
& - \pi(1-\pi)^{2(k_2-2)}(k_2-1) \frac{g_q((1-\pi)^{k_1-1}, (1-\pi)^{k_2-1})^2 \langle s_2 \rangle}{A(k_1, k_2, \pi)}, \tag{A.8}
\end{aligned}$$

where $g_D(x)$ is the generating function of the total vertex degrees.

Appendix B. Equality of numerator second term

We now show that the numerator of the second approximating term in (7) only depends on the degree distribution of the random graph, but not on its clique structures.

$$\begin{aligned}
g_p((1-\pi)^{k-1}) & = \frac{1}{\langle s \rangle} \sum_i i p_i (1-\pi)^{(k-1)(i-1)} \\
& = \frac{1}{\langle d \rangle} \sum_i i(k-1) p_i (1-\pi)^{i(k-1)-(k-2)} \\
& = g_{D^*-1}(1-\pi)(1-\pi)^{-(k-2)}, \tag{B.1}
\end{aligned}$$

where D^* is the size-biased degree distribution. Thus,

$$\langle s \rangle g_p((1-\pi)^{k-1})^2 (k-1)(1-\pi)^{2(k-2)} \pi = \langle d \rangle \pi g_{D^*-1}(1-\pi), \tag{B.2}$$

which is independent of the clique size k , and only depends on the degree distribution.

Appendix C. Approximating π_c for mixed clique sizes

The average number of k_i vertices reached from a k_j -clique vertex for $i, j \in \{1, 2\}$, equals

$$M_{k_i, k_j} = \frac{\langle s_i s_j \rangle - \delta_{k_i, k_j}}{\langle s_j \rangle} \sum_{j=1}^{k_i-1} j h(k_i, j, \pi), \tag{C.1}$$

where δ_{k_i, k_j} is the Kronecker delta. Thus, the matrix M is a branching matrix that describes the average number of vertices of type k_i attached to a randomly chosen clique-edge of type k_j . The average number of vertices at generation j of the offspring distribution can be expressed in terms of M^j . Therefore, if the largest eigenvalue of M becomes larger than one, a giant component forms [19].

Again, we approximate the solution by a second-order polynomial in π assuming again that for large average degrees, the critical percolation value is small. Therefore, similarly to the analysis in section 3.1, we only keep the terms $h(k_i, 1, \pi)$ and $h(k_i, 2, \pi)$. Then, the condition on the largest eigenvalue of M becomes [19]

$$E_{k_1, k_1}(\pi + \pi^2(k_1 - 2)) + E_{k_2, k_2}(\pi + \pi^2(k_2 - 2)) = (\pi + \pi^2(k_1 - 2))(\pi + \pi^2(k_2 - 2)) \\ \times (E_{k_1, k_1}E_{k_2, k_2} - E_{k_1, k_2}E_{k_2, k_1}) + 1, \quad (C.2)$$

where

$$E_{k_i, k_j} = \left(\frac{\langle s_i s_j \rangle}{\langle s_j \rangle} - \delta_{k_i, k_j} \right) (k_i - 1). \quad (C.3)$$

Keeping only second order terms in π yields

$$\pi^2 (E_{k_1, k_1}E_{k_2, k_2} - E_{k_1, k_2}E_{k_2, k_1} - E_{k_1, k_1}(k_1 - 2) - E_{k_2, k_2}(k_2 - 2)) - \pi (E_{k_1, k_1} + E_{k_2, k_2}) = 0. \quad (C.4)$$

This equation has its positive solution as

$$\pi_c = \frac{-E_{k_1, k_1} - E_{k_2, k_2}}{2 (E_{k_1, k_1}E_{k_2, k_2} - E_{k_1, k_2}E_{k_2, k_1} - E_{k_1, k_1}(k_1 - 2) - E_{k_2, k_2}(k_2 - 2))} \\ + \frac{\sqrt{(E_{k_1, k_1} + E_{k_2, k_2})^2 - 4 (E_{k_1, k_1}E_{k_2, k_2} - E_{k_1, k_2}E_{k_2, k_1} - E_{k_1, k_1}(k_1 - 2) - E_{k_2, k_2}(k_2 - 2))}}{2 (E_{k_1, k_1}E_{k_2, k_2} - E_{k_1, k_2}E_{k_2, k_1} - E_{k_1, k_1}(k_1 - 2) - E_{k_2, k_2}(k_2 - 2))}. \quad (C.5)$$

ORCID iDs

Clara Stegehuis  <https://orcid.org/0000-0003-3951-5653>

Thomas Peron  <https://orcid.org/0000-0003-2945-0956>

References

- [1] Ball F, Sirl D and Trapman P 2009 Threshold behaviour and final outcome of an epidemic on a random network with household structure *Adv. Appl. Probab.* **41** 765–96
- [2] Ball F, Sirl D and Trapman P 2010 Analysis of a stochastic SIR epidemic on a random network incorporating household structure *Math. Biosci.* **224** 53–73
- [3] Balogh S G, Palla G and Kryven I 2020 Networks with degree–degree correlations are special cases of the edge-coloured random graph *J. Complex Netw.* **8** cnaa045
- [4] Boguñá M and Pastor-Satorras R 2002 Epidemic spreading in correlated complex networks *Phys. Rev. E* **66** 047104
- [5] Bollobás B 1980 A probabilistic proof of an asymptotic formula for the number of labelled regular graphs *Eur. J. Comb.* **1** 311–6
- [6] Coupechoux E and Lelarge M 2014 How clustering affects epidemics in random networks *Adv. Appl. Probab.* **46** 985–1008
- [7] da Fonseca J D and Abud C V 2018 The Kuramoto model revisited *J. Stat. Mech.* **103204**
- [8] Foster D, Foster J, Paczuski M and Grassberger P 2010 Communities, clustering phase transitions, and hysteresis: pitfalls in constructing network ensembles *Phys. Rev. E* **81** 046115
- [9] Gleeson J P 2009 Bond percolation on a class of clustered random networks *Phys. Rev. E* **80** 036107
- [10] Gleeson J P, Melnik S and Hackett A 2010 How clustering affects the bond percolation threshold in complex networks *Phys. Rev. E* **81** 066114
- [11] Gleeson J P, Melnik S, Ward J A, Porter M A and Mucha P J 2012 Accuracy of mean-field theory for dynamics on real-world networks *Phys. Rev. E* **85** 026106
- [12] Goltsev A V, Dorogovtsev S N and Mendes J F F 2008 Percolation on correlated networks *Phys. Rev. E* **78** 051105
- [13] Hackett A and Gleeson J P 2013 Cascades on clique-based graphs *Phys. Rev. E* **87** 062801
- [14] Hackett A, Melnik S and Gleeson J P 2011 Cascades on a class of clustered random networks *Phys. Rev. E* **83** 056107
- [15] Hébert-Dufresne L and Althouse B M 2015 Complex dynamics of synergistic coinfections on realistically clustered networks *Proc. Natl Acad. Sci. USA* **112** 10551–6
- [16] Herrero C P 2015 Ising model in clustered scale-free networks *Phys. Rev. E* **91** 052812
- [17] van der Hofstad R 2017 *Random Graphs and Complex Networks* vol 1 (Cambridge: Cambridge University Press)
- [18] van der Hofstad R, van Leeuwen J S H and Stegehuis C 2017 Hierarchical configuration model *Internet Math.* **1** 24166
- [19] Karrer B and Newman M E J 2010 Random graphs containing arbitrary distributions of subgraphs *Phys. Rev. E* **82** 066118
- [20] Aguirre López F and Coolen A C C 2021 Transitions in random graphs of fixed degrees with many short cycles *J. Phys. Complex* **2** 035010
- [21] Mann P, Smith V A, Mitchell J B and Dobson S 2021 Random graphs with arbitrary clustering and their applications *Phys. Rev. E* **103** 012309

- [22] Mann P, Smith V A, Mitchell J B O, Jefferson C and Dobson S 2021 Exact formula for bond percolation on cliques *Phys. Rev. E* **104** 024304
- [23] Melnik S, Hackett A, Porter M A, Mucha P J and Gleeson J P 2011 The unreasonable effectiveness of tree-based theory for networks with clustering *Phys. Rev. E* **83** 036112
- [24] Miller J C 2009 Percolation and epidemics in random clustered networks *Phys. Rev. E* **80** 020901
- [25] Miller J C 2009 Spread of infectious disease through clustered populations *J. R. Soc. Interface* **6** 1121–34
- [26] Newman M 2018 *Networks* (Oxford: Oxford University Press)
- [27] Newman M E J 2009 Random graphs with clustering *Phys. Rev. Lett.* **103** 058701
- [28] Peron T, de Resende B M F, Mata A S, Rodrigues F A and Moreno Y 2019 Onset of synchronization of Kuramoto oscillators in scale-free networks *Phys. Rev. E* **100** 042302
- [29] Peron T K D, Ji P, Kurths J and Rodrigues F A 2018 Spectra of random networks in the weak clustering regime *Europhys. Lett.* **121** 68001
- [30] Peron T K D, Rodrigues F A and Kurths J 2013 Synchronization in clustered random networks *Phys. Rev. E* **87** 032807
- [31] Restrepo J G, Ott E and Hunt B R 2005 Onset of synchronization in large networks of coupled oscillators *Phys. Rev. E* **71** 036151
- [32] Ritchie M, Berthouze L and Kiss I Z 2016 Beyond clustering: mean-field dynamics on networks with arbitrary subgraph composition *J. Math. Biol.* **72** 255–81
- [33] Rodrigues F A, Peron T K D, Ji P and Kurths J 2016 The Kuramoto model in complex networks *Phys. Rep.* **610** 1–98
- [34] Stegehuis C, van der Hofstad R and van Leeuwen J S H 2016 Epidemic spreading on complex networks with community structures *Sci. Rep.* **6** 29748
- [35] Strogatz S H 2000 From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators *Physica D* **143** 1–20
- [36] Trapman P 2007 On analytical approaches to epidemics on networks *Theor. Popul. Biol.* **71** 160–73
- [37] Volz E M, Miller J C, Galvani A P and Ancel Meyers L 2011 Effects of heterogeneous and clustered contact patterns on infectious disease dynamics *PLoS Comput. Biol.* **7** e1002042
- [38] Zlatić V, Garlaschelli D and Caldarelli G 2012 Networks with arbitrary edge multiplicities *Europhys. Lett.* **97** 28005