PAPER • OPEN ACCESS

Ensemble analysis of complex network properties—an MCMC approach

New Journal of Physics

The open access journal at the forefront of physics

PAPER

Ensemble analysis of complex network properties—an MCMC approach

Oskar Pfeffer* , Nora Molkenthin and Frank Hellmann

Potsdam Institute for Climate Impact Research, Potsdam 14473, Germany
* Author to whom any correspondence should be addressed.

E-mail: o.pfeffer@posteo.net

Abstract

What do generic networks that have certain properties look like? We use relative canonical network ensembles as the ensembles that realize a property R while being as indistinguishable as possible from a background network ensemble. This allows us to study the most generic features of the networks giving rise to the property under investigation. To test the approach we apply it to study properties thought to characterize 'small-world networks'. We consider two different defining properties, the 'small-world-ness' of Humphries and Gurney, as well as a geometric variant. Studying them in the context of Erdős-Rényi and Watts–Strogatz ensembles we find that all ensembles studied exhibit phase transitions to systems with large hubs and in some cases cliques. Such features are not present in common examples of small-world networks, indicating that these properties do not robustly capture the notion of small-world networks. We expect the overall approach to have wide applicability for understanding network properties of real world interest, such as optimal ride-sharing designs, the vulnerability of networks to cascades, the performance of communication topologies in coordinating fluctuation response or the ability of social distancing measures to suppress disease spreading.

1. Introduction

Network ensembles are sets of networks together with a probability distribution of their occurrence and have been successfully used to model a wide range of natural, social and technical systems, in which the interaction structure is subject to, or the outcome of, stochasticity [1–6]. Typically those ensembles are generated through a heuristic process, thought to capture some aspect of the microscopic formation process, which underlies the real-world system they are trying to model. The resulting ensemble can then be studied and characterized by means of network measures that quantify certain properties of the networks. Examples for this are Watts–Strogatz (WS) networks, which are characterized by a low average shortest path length and a high clustering coefficient [7], and Barabasi–Albert networks, which are characterized by their power-law degree distribution [8].

Here we want to approach network ensembles from the other side. Rather than trying to model real world networks we ask: what do generic networks that have certain properties look like? From this we hope to gain insights into the effects of the network measure itself. This will be particularly useful for more complex network measures, used to detect practically relevant features, such as a network's propensity for disease spreading or failure cascades. Thus, we will use ensembles *defined* through a particular property captured by a *property function* $R(G)$ on network G and a background ensemble that defines our notion of generic networks in the given context. To this end, we will consider *exponential random graphs*. Exponential random graphs have long been a tool in network science, starting with [9–12], see [13] for a recent review, and are also sometimes known as canonical network ensembles (CNEs) [14–16]. We will consider CNEs relative to a background ensemble as defined in [17, 18]. Given a set of networks $\mathcal{E}$, denote the probability distribution of the background ensemble as $q(G)$ for $G \in \mathcal{E}$. The *relative canonical network ensemble* (RCNE) of R relative to q is given by the probability distribution proportional to $\exp(-\beta R(G))q(G)$. This ensemble minimizes the relative entropy to q. Of all ensembles with a fixed expectation value of R it is the least

distinguishable from q. As is typical in the information theoretic point of view, the relative entropy plays the role of a (generalized) distance between ensembles. For details see appendix A. These ensembles also form the backbone of techniques to sample from the tails of distributions [19], i.e. to evaluate $q(R > R^*)$. This is why we interpret the RCNE to tell us about the structure of R in the ensemble q and study them as ensembles of independent interest.

We emphasize that our aim is not to model empirically observed networks with certain properties. There is no reason to expect empirical networks, that are the outcome of subtle formation processes, to be CNEs unless the formation processes can be shown to maximize entropy in some sense. Instead, we will study the properties themselves, specifically the most generic features that produce them. The background ensemble does not serve as an uninformed prior to be updated by adding new information. Instead the background ensemble encodes what 'generic' means in the context in which the property R is of interest.

Our aim is to understand properties of considerable practical interest. A companion paper studies epidemic thresholds [20], while optimal ride-sharing designs, the vulnerability to failure cascades and the performance of communication topologies for distributed secondary control designed to suppress persistent frequency fluctuations in power grids were studied in [21, 22]. To introduce the approach, this paper focuses on well-known and well-established network measures, that are computationally challenging. Specifically, we analyze the notion of small-world networks.

We consider two measures, the first S, called *small-world-ness* in [23], and the second a combination of Euclidean link length and average shortest path length similar to [24]. We study both of these properties in the context of the *Erdős-Rényi* [25] (ER) ensemble and the WS ensemble of the same link density. The former studies these properties in the context of classical random networks, and the latter in the context of a model for small-world networks. To study these four RCNEs we sample them using the straightforward Metropolis–Hastings (MH)[19, 26, 27] algorithm.

In all cases we find phase transitions as we go from fully generic networks to highly specific ones. At these phase transitions certain features arise, e.g. hubs and cliques start appearing in the ensemble. Surprisingly, we find that neither of these measures give rise to the 'rewired grid' structure produced by the WS model, which is commonly associated with small-world networks. On the contrary, even if we start from WS-networks as a background ensemble, hubs and/or cliques emerge further increasing S. This feature is not typically found in other network ensembles, used for constructing high values of S, such as network optimization models as discussed in [28] or rewiring-based social networks [6].

2. CNE sampling as an analysis tool for network properties

Given the set of simple graphs $\mathcal{E}$, the relative CNE of a network measure R is defined by the probability distribution over $\mathcal{E}$ relative to a background ensemble $q(G)$:

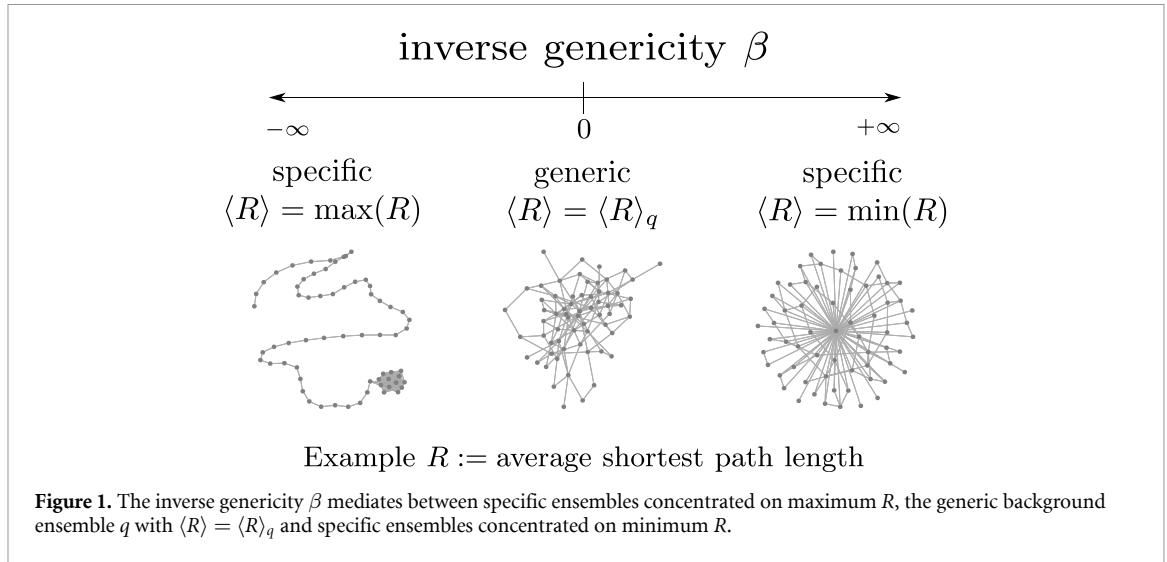$$p^R_{\beta,q}(G) = \frac{1}{Z_{R,q}(\beta)} e^{-\beta R(G)} q(G) \,, \tag{1}$$

with normalization/partition function $Z_{R,q}(\beta) = \sum_{G \in \mathcal{E}} e^{-\beta R(G)} q(G)$. This ensemble is characterized by being the ensemble of minimum relative entropy $D(p||q)$ for a fixed expectation value of R. From a statistical perspective it is the ensemble hardest to distinguish from the ensemble q while having fixed expectation value $\langle R \rangle = R^*$, for a more detailed discussion see appendix A. If we take q to characterize what a generic network should look like in the context of interest, then the RCNE with property R provides us with those networks that are as generic as possible while realizing the property under investigation. Studying these networks therefore reveals the most generic fashion in which high/low R can arise.

The parameter β modulates the trade-off between generic and specific ensembles peaked on networks that are high or low in R, see figure 1. It can range from $-\infty$ to $+\infty$ with the sign depending on whether the expectation value of R is higher or lower than in the generic network ensemble given by $\beta = 0$. At $\beta \to -\infty$ we have an ensemble concentrated on $\max(R)$, at $\beta \to +\infty$ on $\min(R)$. This, and the fact that the interpretation of the relative entropy is purely statistical (rather than thermodynamic), motivates us to refer to β^{-1} as the genericity rather than as a temperature.

Of particular note are phase transitions that occur as we lower the absolute genericity. The structure of the ensemble changes at and beyond the phase transition. This change in structure allows us to identify specific features that contribute to property R but are not generic enough to occur before.

What counts as a generic network is highly dependent on context. A generic social network does not look like a generic power grid. Throughout the rest of this manuscript we compare CNEs with a spatially embedded variant of WS-networks as a background (WSB) to those with an ER background ensemble (ERB).

To study these ensembles we need to sample from them. An important property of RCNEs is that they are potentially well suited for sampling using MH algorithms. Whenever the background distribution q or the

inverse genericity β

$-\infty$ 0 $+\infty$

specific   generic   specific
$\langle R \rangle = \max(R)$ $\langle R \rangle = \langle R \rangle_q$ $\langle R \rangle = \min(R)$

Example $R := $ average shortest path length

Figure 1. The inverse genericity β mediates between specific ensembles concentrated on maximum R, the generic background ensemble q with $\langle R \rangle = \langle R \rangle_q$ and specific ensembles concentrated on minimum R.

ratio between any two networks in the distribution $q(G)/q(G')$ is known, it is possible to include it directly in the acceptance probability, however this is usually not the case. Then the central challenge to use MH on our relative ensemble is finding a background process g that generates proposed steps compatible with the background distribution q in the sense that the proposed steps are in detailed balance with q:

$$g(G'|G)q(G) = g(G|G')q(G'),\tag{2}$$

see appendix B for details, and a proposal for how to generate such processes for ensembles generated by more general network formation processes. For our background ensembles this can be achieved by appropriate rewiring of edges respecting the constraints of the background ensembles. Starting from a system in state G_t the algorithm proposes rewirings in accordance with q that are then accepted with probability:

$$P_\beta(G_t \to G_{t+1}) = \min\left(1, \frac{e^{-\beta R(G_{t+1})}}{e^{-\beta R(G_t)}}\right)$$
$$= \min\left(1, e^{-\beta \Delta R}\right).\tag{3}$$

This algorithm satisfies the detailed balance condition for the RCNE (note that the ratios of q in the RCNE distribution cancels the gs from the MH acceptance probability) and if the process g is ergodic, the Markov chain defined by it is ergodic as well and strongly connected. In the limit of infinite steps T the time average of an observable O for this chain converges to the expectation value of O in the relative CNE:

$$\lim_{T \to \infty} \frac{1}{T}\sum_{t=1}^{T} O(G_t) = \sum_G p_{\beta,q}^R(G)O(G).\tag{4}$$

Unfortunately, there are no guarantees for finite time samples and we have to resort to heuristics to understand whether convergence has occurred. To do so we typically also run several chains in parallel from random initial conditions. More details on our sampling approach are provided in the next section.

3. Small-world properties

We analyze two different properties that have been argued to be related to small-world networks. We look at the features that give rise to networks that are high in these properties and thus whether they generically characterize small-world networks. To do this we compare the resulting network structures to the ensemble of WS-networks. In the first instance we consider the measure

$$S = (C/L)(C_{\mathrm{ER}}/L_{\mathrm{ER}})^{-1},\tag{5}$$

introduced in [23], where C is the global clustering coefficient, defined as the number of closed triplets divided by the number of open triplets, L the average shortest path length, C_{ER} and L_{ER} the average clustering coefficient [25] and the expected average shortest path length [29] of an ER network of the same order and size.

Figure 2. The 'small-world' region is similar in both network models. Clustering coefficient *C*, average shortest path length *L* and *S* vs fraction of rewired edges *p* for (a): the standard WS ensemble and (b): the spatially embedded WS ensemble for networks with $N = 256$ vertices and average degree $\bar{k} = 4$. The values are averaged over 5000 networks.
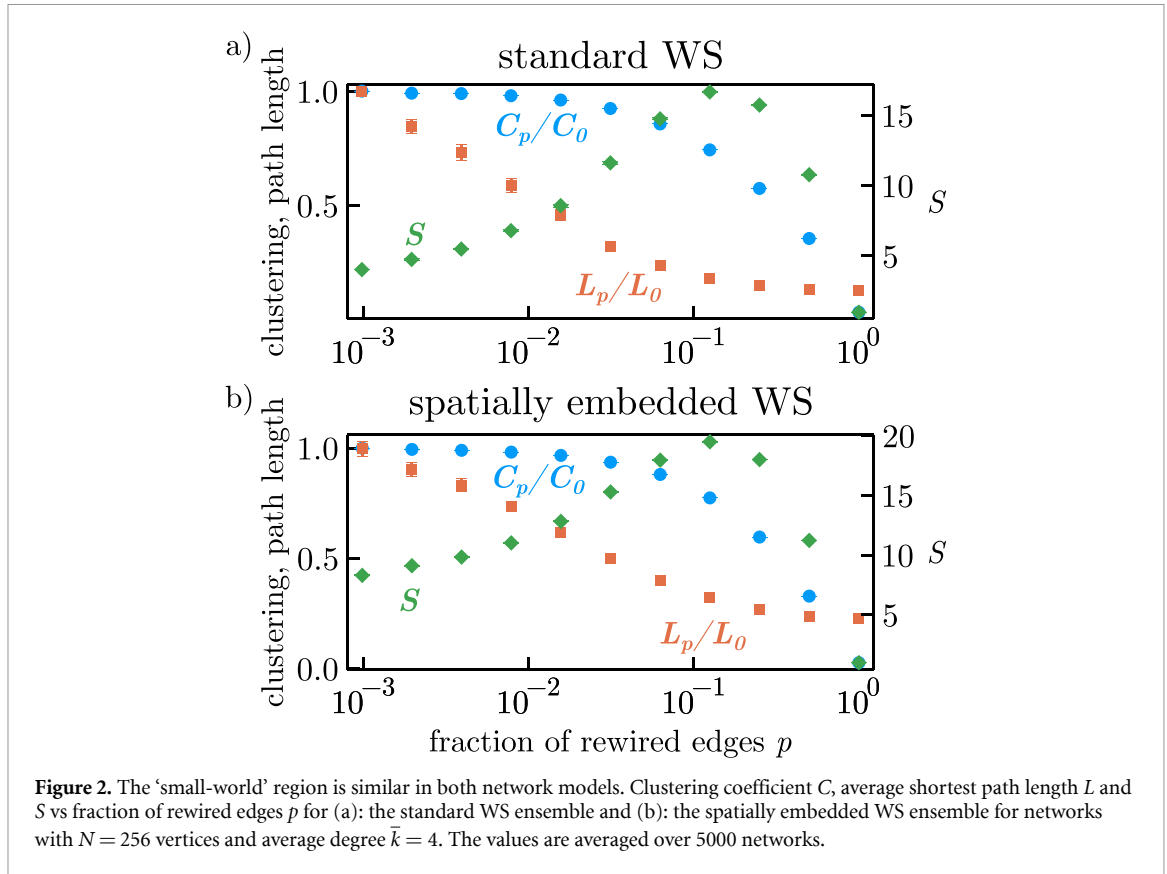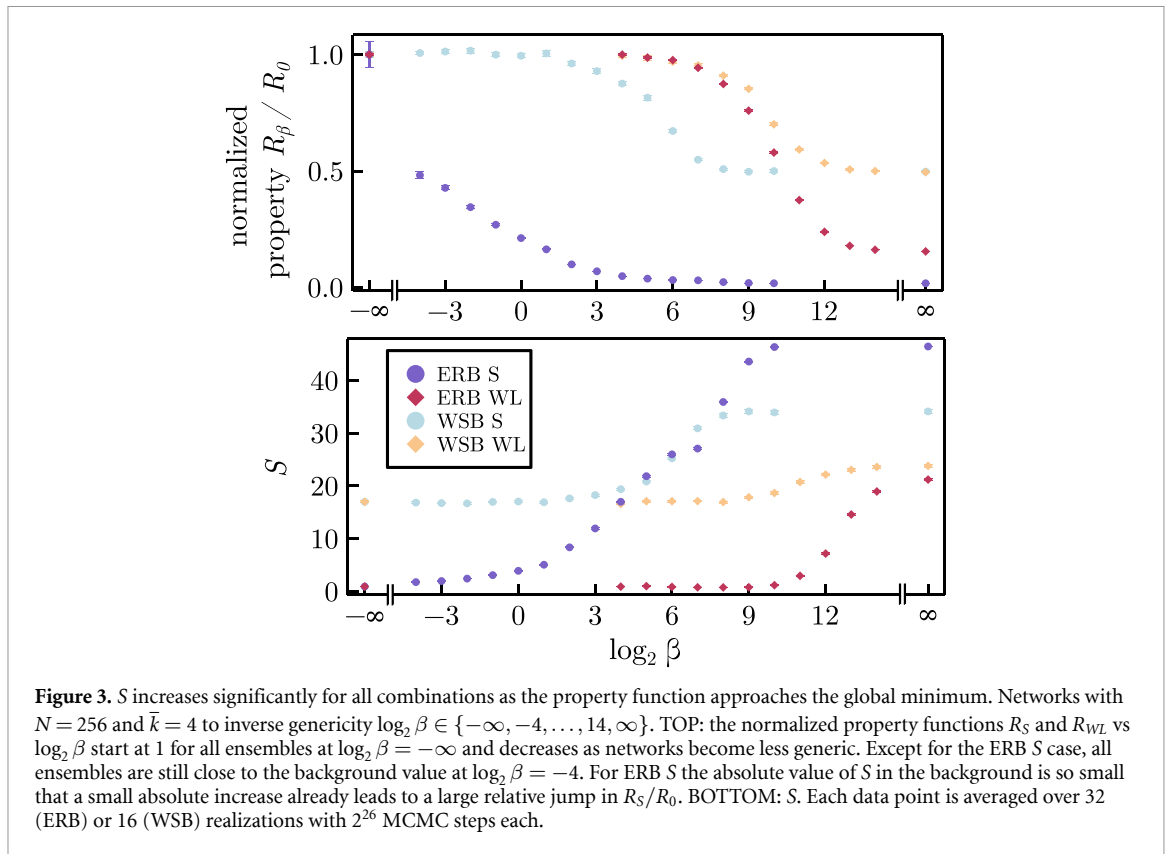
Finally, the form of the property we will consider is:

$$R_S = \frac{L}{C}. \tag{6}$$

That is, proportional to the inverse of *S*. Thus, small values of R_S indicate high *S*, and we are interested in positive β. The second property

$$R_{WL} = WL \tag{7}$$

is given in terms of the average shortest path length *L* and the wiring length *W* in a geometrically embedded network. *W* is the sum of the Euclidean lengths of all edges. The networks of this ensemble are embedded in a 2D square plane with periodic boundaries, where the location of the vertices are randomly drawn. The introduction of *W* was inspired by [24], where it was argued that small-world networks might arise as a secondary feature from a trade-off between maximal connectivity and minimal edge lengths. Again small R_{WL} is expected to yield small-world networks and we consider positive β.

Both properties are considered relative to the $G(N, M)$ model of the ER-network ensemble with *N* vertices and $M = \bar{k}N/2$ edges, where we consider only connected networks, and a spatially embedded variant of WS-networks of the same order and link density. Rather than being arranged in a ring, the vertices are randomly placed in a 2D square plane with periodic boundaries, and connected by a Euclidean minimum spanning tree, where the $N + 1$ geometrically shortest edges not already in the tree are added on top, resulting in a graph with $M = 2N$ edges. We call the edges that are part of this initial network *on-lattice edges*, analogous to the edges composing the ring structure of the WS-network. A fraction *p* of the edges are rewired randomly, leading to $M_1 = (1 - p)M$ on-lattice edges and $M_2 = pM$ off-lattice edges. Figure 2 outlines the conceptual similarity of this approach with the standard WS model. Both ensembles have a high *L* and *C*, when the fraction of rewired edges *p* is low. *L* decreases rapidly as *p* is raised, while *C* decreases only at higher values of *p*, leading to a high *S* region in both models, peaked at $p \sim 0.1$ for networks with $N = 256$ vertices and average degree $\bar{k} = 4$. Thus, we set $p = 0.1$ in our WSB model throughout the manuscript. The two property functions and the two background ensembles lead us to four distinct ensembles that we analyze. We call the ensembles characterized by the property function R_S ERB-*S* and WSB-*S*, where ERB and WSB are the corresponding background ensembles. We refer to them collectively as the *S*-ensembles. Analogously, the

Figure 3. *S* increases significantly for all combinations as the property function approaches the global minimum. Networks with $N = 256$ and $\bar{k} = 4$ to inverse genericity $\log_2 \beta \in \{-\infty, -4, \ldots, 14, \infty\}$. TOP: the normalized property functions R_S and R_{WL} vs $\log_2 \beta$ start at 1 for all ensembles at $\log_2 \beta = -\infty$ and decreases as networks become less generic. Except for the ERB *S* case, all ensembles are still close to the background value at $\log_2 \beta = -4$. For ERB *S* the absolute value of *S* in the background is so small that a small absolute increase already leads to a large relative jump in R_S/R_0. BOTTOM: *S*. Each data point is averaged over 32 (ERB) or 16 (WSB) realizations with 2^{26} MCMC steps each.
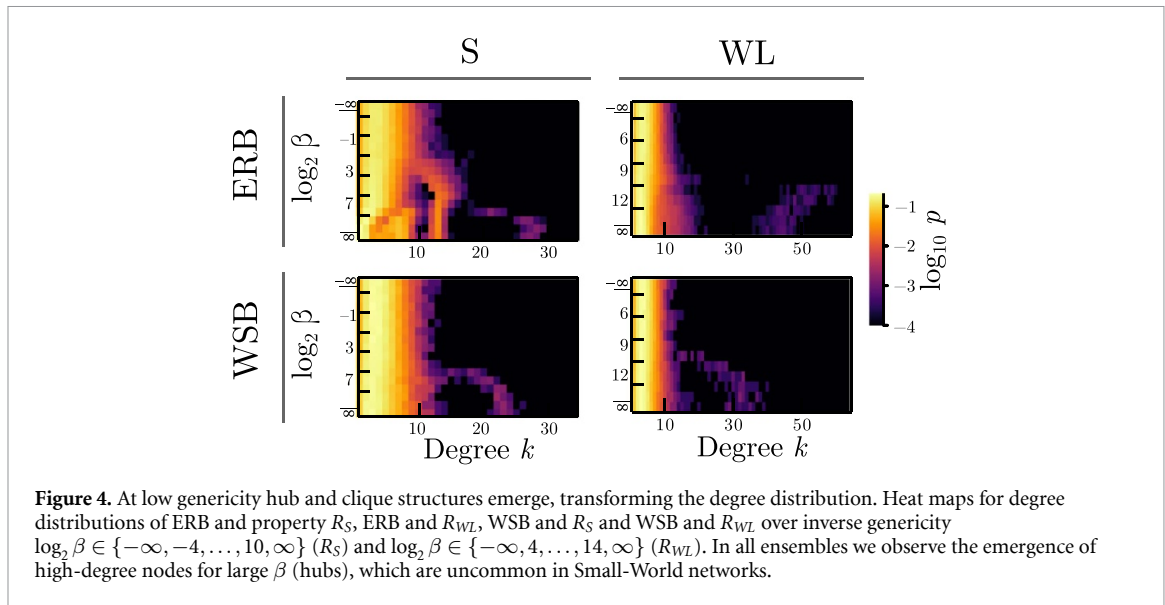
ensembles characterized by R_{WL} are called ERB-*WL* and WSB-*WL* and we collectively refer to them as the *WL*-ensembles.

It is noteworthy, that the two background ensembles ERB and WSB are just a uniform distribution in a restricted set of graphs. However, our sampling method is valid for more complex or even unknown background ensembles, if, for example, the proposal distribution $g(G'|G)$ of the sampler is in detailed balance with the background ensemble (see appendix B).

To sample from the ensembles we implemented the MH algorithm in the following way. We generate the Monte Carlo step proposal in two distinct ways respecting the constraints of the background ensembles. For the ERB ensemble we rewire an edge randomly, i.e. we replace an existing edge by a previously missing one. For the WSB ensemble we rewire consistently on-lattice and off-lattice edges, i.e. an existing edge of the corresponding category is deleted at random and another previously missing edge of the same category is added at random, keeping the number of on/off-lattice edges constant. The proposal is then accepted with the transition probability given in equation (4). Proposals of disconnected graphs are always rejected since *L* is infinite. To ensure convergence at low genericity, we use an exponential schedule $\beta^{-1}(t) = \left(\beta_{\text{start}}^{-1} \alpha^t + \beta_{\text{end}}^{-1} \right)$, similar to the Simulated Annealing approach [30], where *t* is the step parameter, $\alpha = 0.993$ [0.992] (*S* and *WL* respectively) is an algorithmic parameter. $\beta_{\text{start}}^{-1} = 0.05$ [0.005] (*S* and *WL* respectively) and β_{end}^{-1} are the start and final genericities. We generated ensembles of networks of orders $N = (16, 32, 64, 128, 256)$ and average degree $\bar{k} = 4$. Due to the higher computation times for larger networks, the ensembles for smaller networks contain a higher number of realizations, for ERB we have $(512, 512, 128, 32, 16)$ for WSB $(512, 512, 256, 64, 32)$. All the realizations appeared to converge, allowing qualitative evaluation. Throughout the manuscript, genericity was decreased over 2^{10} equally long periods, each containing 2^{16} MCMC steps for a total of 2^{26} MCMC steps. Each network is sampled after the last step of the MCMC chain at a genericity $\beta^{-1} \approx \beta_{\text{end}}^{-1}$. Note that while the properties we consider here are conceptually simple, the presence of the average shortest path length, which needs to be recomputed for every proposed step, renders them computationally expensive.

As shown in figure 3, *S* increases for all network ensembles as they become less generic. For the *S*-ensembles, for which it is mathematically guaranteed that the expectation value of *S* increases for decreasing genericity, this is an important sanity check on our sampling. In the *WL*-ensembles this arises as a secondary effect as Euclidean and network distances are reduced, showing that generic R_{WL} networks do indeed reach high values of *S*, as anticipated in [24]. With the WS-background ensemble the possible networks and thus the final values of *S* are constrained, but *S* still increases with increasing β.

Figure 4. At low genericity hub and clique structures emerge, transforming the degree distribution. Heat maps for degree distributions of ERB and property R_S, ERB and R_{WL}, WSB and R_S and WSB and R_{WL} over inverse genericity $\log_2 \beta \in \{-\infty, -4, \ldots, 10, \infty\}$ (R_S) and $\log_2 \beta \in \{-\infty, 4, \ldots, 14, \infty\}$ (R_{WL}). In all ensembles we observe the emergence of high-degree nodes for large β (hubs), which are uncommon in Small-World networks.

To gain a first overview of what the various ensembles at various genericities look like we can now look at the degree distribution. Figure 4 shows the degree distribution heat maps of the four ensembles for decreasing genericity. All ensembles show Poisson distributed degrees at low β, as both background ensembles are (in case of WSB slightly shifted) Poisson distributed. As the genericity decreases additional structures emerge due to the property functions, causing additional peaks in the degree distributions.

The branching of the degree distribution heat maps happens suddenly in all four cases after $\log_2 \beta \approx (7, 10, 5, 9)$ respectively in the four ensembles. We take this as an indication that a phase transition may be happening as the ensembles become less and less generic, taking the ensemble from near the background ensemble to an ensemble dominated by the property function. In the following we will further investigate these ensembles as well as the transition between them.

4. Phase transition with genericity

To further investigate the phase transition we take a look at example networks in figure 5 as well as the developments of the maximum degree and the clique number, the size of the largest clique, with genericity in figure 6.
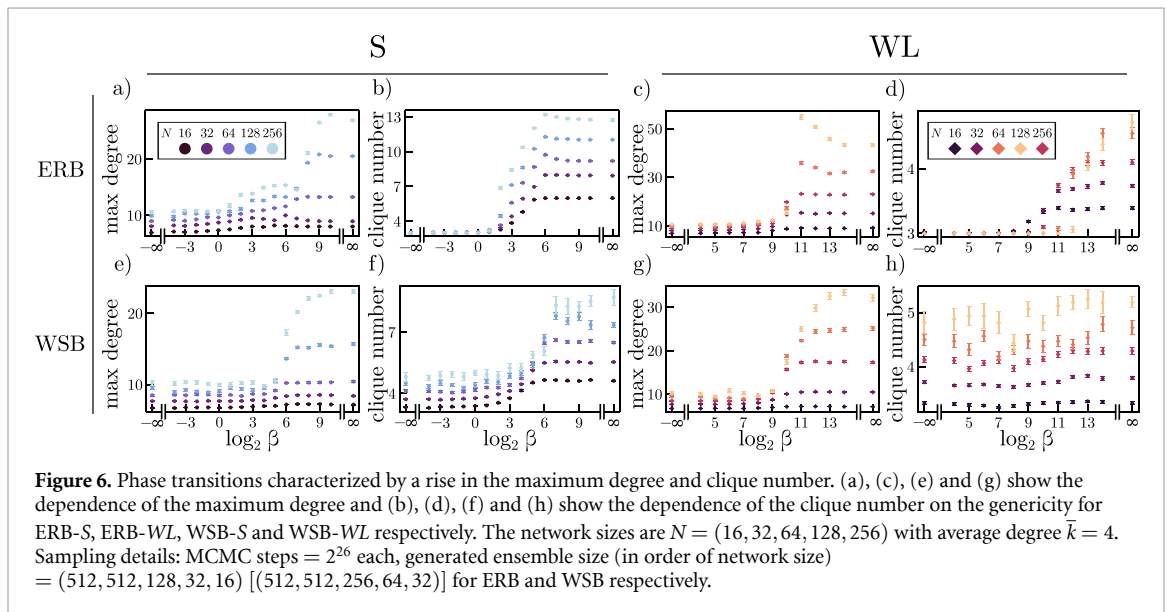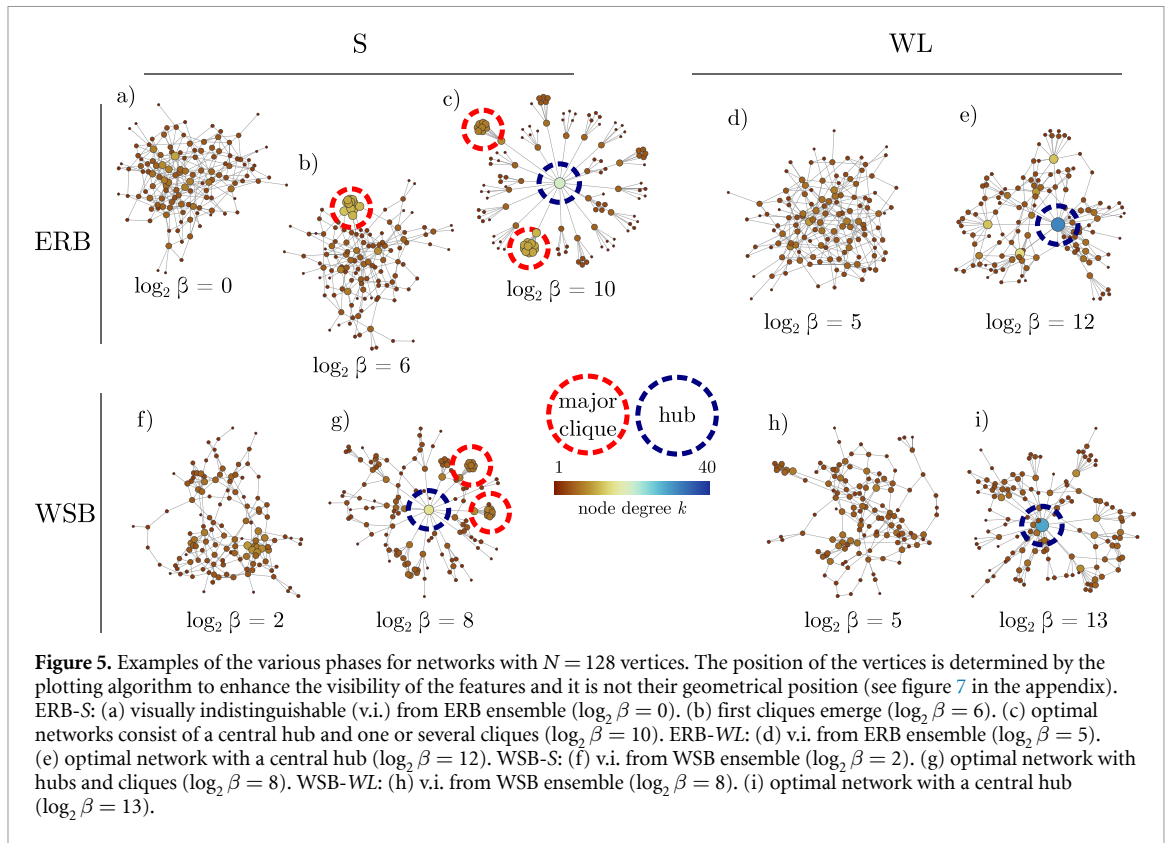
In figures 5(a) and (d) we see examples of networks before the phase transition, which are essentially indistinguishable from the ER-background ensemble. Similarly (f) and (h) show examples indistinguishable from the WS-background ensemble.

In figures 5(c) and (g) we see examples of low genericity networks from the *S*-ensembles. Both are notably different from their respective background ensembles, showing a star-like structure with a central hub and several larger cliques at the leaves of the network. The existence of hubs drastically decreases the average shortest path length L in the network, while cliques contribute to a large clustering coefficient in the network. Note that in the WSB-*S*-ensemble there are still some remaining interconnections between peripheral nodes, which are not present in the unconstrained ERB-*S*-ensemble. This difference probably explains that for $N = 256$ S only rises to ≈ 35 in the WSB-*S*-ensemble rather ≈ 45 in the ERB-*S*-ensemble, as these links leave fewer links to maximize the clustering coefficient by forming cliques. Sub-figure (b) seems to show a network from an intermediate ensemble, in which cliques have formed but hubs are not yet obviously visible.

Sub-figures (e) and (i) of figure 5 show low genericity networks from the *WL*-ensembles. Again we observe the formation of central hubs reducing the average shortest path length. Instead of the formation of cliques, short-range links appear. As a result the ERB-*WL* and WSB-*WL* ensembles appear quite similar. Both differ from the WS-background ensemble by the additional hub-structure.

To better understand the transitions between these ensembles we now take a closer look at the behavior of the maximum degree (as a proxy for the existence and size of hubs) and the clique number (as a proxy for the existence and size of cliques) with genericity for a range of network sizes from $N = 2^4$ to $N = 2^8$. The results are shown in figure 6.

The maximum degree is the maximum number of connections a single node has in the network. In all four ensembles it rises sharply across all network sizes as a certain genericity is reached. In case of the

Figure 5. Examples of the various phases for networks with $N = 128$ vertices. The position of the vertices is determined by the plotting algorithm to enhance the visibility of the features and it is not their geometrical position (see figure 7 in the appendix). ERB-*S*: (a) visually indistinguishable (v.i.) from ERB ensemble ($\log_2 \beta = 0$). (b) first cliques emerge ($\log_2 \beta = 6$). (c) optimal networks consist of a central hub and one or several cliques ($\log_2 \beta = 10$). ERB-*WL*: (d) v.i. from ERB ensemble ($\log_2 \beta = 5$). (e) optimal network with a central hub ($\log_2 \beta = 12$). WSB-*S*: (f) v.i. from WSB ensemble ($\log_2 \beta = 2$). (g) optimal network with hubs and cliques ($\log_2 \beta = 8$). WSB-*WL*: (h) v.i. from WSB ensemble ($\log_2 \beta = 8$). (i) optimal network with a central hub ($\log_2 \beta = 13$).

Figure 6. Phase transitions characterized by a rise in the maximum degree and clique number. (a), (c), (e) and (g) show the dependence of the maximum degree and (b), (d), (f) and (h) show the dependence of the clique number on the genericity for ERB-*S*, ERB-*WL*, WSB-*S* and WSB-*WL* respectively. The network sizes are $N = (16, 32, 64, 128, 256)$ with average degree $\bar{k} = 4$. Sampling details: MCMC steps $= 2^{26}$ each, generated ensemble size (in order of network size) $= (512, 512, 128, 32, 16) \left[(512, 512, 256, 64, 32) \right]$ for ERB and WSB respectively.

ERB-*S*-ensemble one might spot two such shifts, one at $\log_2 \beta = 1$ and one at $\log_2 \beta = 7$, where the first shift indicates the formation of cliques (compare figure 5(b)) and the second the formation of hubs.

The clique number is the size of the largest clique in the network, where a clique is a fully connected subgraph of the network. In figure 6(b) we observe that for the ERB-*S*-ensemble the clique number reveals the first phase transition, which already showed a slight shift in the maximum degree in figure 6(a). In the WSB-*S*-ensemble the clique number rise coincides with the rise of the maximum degree, but happens more gradually as seen in figure 6(f). This corresponds to the observation that there is only one genericity, at which a shift happens in the degree distribution heat map in figure 4(c), as well as no examples of networks with cliques but no hubs as in figure 5(b). We conclude that in the WSB-*S* ensemble hubs and cliques appear at the same genericity. In the *WL*-ensembles we observe no rise of the clique size beyond the clique size of the WS-background ensemble. Since WS networks have a higher degree of cliquishness than ER networks, this

still amounts to an increase in clique sizes in figure 6(d) All transitions observed become steeper and more pronounced with network size, indicating that we are looking at genuine phase transitions in the ensembles.

These results show that various discrete features suddenly emerge at certain genericities. As detailed above the transitions become qualitatively visible in the degree distribution (figure 4), clearly appear in their graphical representations (figure 5) and can be quantified in the maximum degree and clique number (figure 6). We conclude that these phase transitions and the emergence of hubs and cliques are a driving element in the increase of the property S.

5. Discussion and conclusion

We introduced the use of RCNE's as a means to study the features of generic networks with a given property. These ensembles are amenable to MH and MCMC methods, providing a simple and straightforward (if potentially computationally expensive) way of sampling from non-trivial network ensembles defined through network measures of practical interest.

To challenge the method we studied two properties traditionally expected to characterize small-world networks. We found that neither one of them results in WS-like networks and will even move initially WS background networks away from this ensemble. Instead we find that the most generic networks with high S all contain cliques, boosting the clustering coefficient, and hub structures, minimizing the average shortest path length. Neither of these substructures are commonly found in the WS ensemble. This indicates that high S is not the defining feature of small-world networks and even spatial embedding may not fully explain their widespread occurrence.

The transition from highly generic to very specific ensembles in the four cases studied is characterized by well defined phase transitions. These are visible in the maximum degree and clique number in the networks.

It is somewhat surprising that new things are still to be learned on properties thought to characterize small-world networks. The fact that our perspective on RCNE's provide novel insights is a promising sign for the study of properties of greater practical interest. In a companion paper we are considering epidemic thresholds [20]. Further areas of application are as diverse as the optimal design of ride-pooling services [21], the vulnerability to cascading power grid failures [22], as well as the performance of different topologies for distributed secondary frequency control. More generally, this method is of great interest wherever one aims to understand and design topologies that fulfill certain functions, rather than describe empirical networks. Besides the phase transitions explored here we more typically expect to observe correlations between the properties in question and more general network measures within and across samples that have been taken at different β. An example of this for the cases above is shown in appendix C, figure 9. Besides helping to gain a better understanding of the property in question this provides a way to find correlates and predictors for the ensemble defining characteristics that are valid far beyond the values typically observable in the background ensemble.

Data availability statement

Acknowledgments

Appendix A. Relative Entropy

The minimization of the relative entropy has a statistical interpretation. Given a distribution q, the asymptotic probability to obtain a sample that looks like p goes as the exponential of the negative relative entropy $D(p||q)$. This result of Chernoff *et al* [31] is known as Stein's Lemma (for a modern account phrased in terms of relative entropy see e.g. [32] Theorem 4.12) and forms the mathematical basis for the interpretation of the relative entropy as a measure of distinguishability of probability distributions. Our ensembles thus have a statistical interpretation as being the ensembles that are hardest to distinguish from the generic ensemble q. In particular we do not presuppose that real network formation processes maximize entropy subject to some constraints, and do not interpret the resulting ensembles as modeling real networks that have the property R.

For completeness, we recall here the standard argument that the relative entropy, or the Kullback–Leibler divergence, is minimized by the exponential ensemble. We are looking for the probability distribution p^* such that

$$p^* = \underset{\substack{p \\ \langle R \rangle = R^*}}{\arg\min}\, D(p||q)$$

$$= \underset{\substack{p \\ \langle R \rangle = R^*}}{\arg\min} \sum_i p_i \ln\left(\frac{p_i}{q_i}\right). \tag{A1}$$

First, note that this formula diverges to positive infinity if p has support outside the support of q. We thus only consider p whose support is contained in that of q. Then, we introduce the Lagrange multipliers β_n and β_R for the normalization constraint $0 = 1 - \sum_i p_i$ and the expectation value constraint $0 = R^* - \sum_i p_i R_i$ respectively. This allows us to write the free optimization over the components p_i:

$$p^*(\beta_n, \beta_R) = \underset{\{p_i\}}{\arg\min} \sum_i p_i \ln\left(\frac{p_i}{q_i}\right)$$

$$+ \beta_n \left(\sum_i p_i - 1\right) + \beta_R \left(\sum_i p_i R_i - R^*\right), \tag{A2}$$

and the constraints fix the values of the Lagrange multipliers:

$$R^* = \sum_i R_i\, p_i^*(\beta_n, \beta_R)$$

$$1 = \sum_i p_i^*(\beta_n, \beta_R). \tag{A3}$$

to obtain a minimum of (A1).
Now the variation in the direction p_j produces the following condition:

$$0 = \frac{\partial}{\partial p_j}\left[\sum_i p_i\left(\ln(p_i) - \ln(q_i)\right) + \right.$$

$$\left. + \beta_n\left(\sum_i p_i - 1\right) + \beta_R\left(\sum_i p_i R_i - R^*\right)\right]$$

$$= \ln(p_j) - \ln(q_j) + 1 + \beta_n + \beta_R R_j. \tag{A4}$$

From which we can conclude

$$p_j^* = \exp(\ln(q_j) - 1 - \beta_n - \beta_R R_j)$$

$$= \frac{1}{Z} e^{-\beta_R R_j} q_j \tag{A5}$$

with $Z = e^{1 + \beta_n} = \sum_i e^{-\beta_R R_i} q_i$ fixed by the condition $\sum_i p_i^* = 1$ and β_R determined implicitly by the condition $R^* = \sum_i R_i p_i^*$.
Note that

$$\frac{\partial}{\partial \beta_R}\langle R \rangle = -\frac{1}{Z}\sum_i R_i^2 e^{-\beta_R R_i} q_i$$

$$- \frac{\partial Z}{\partial \beta_R}\frac{1}{Z^2}\sum_i R_i e^{-\beta_R R_i} q_i$$

$$= -\langle R^2 \rangle + \langle R \rangle^2$$

$$= -\mathrm{var}(R) \leqslant 0. \tag{A6}$$

Further, for $\beta_R = -\infty$ we have the distribution peaked completely on the global maxima: $\langle R \rangle = R_{\max}$ and for $\beta_R = +\infty$ we have the minima instead $\langle R \rangle = R_{\min}$. For $\beta_R = 0$ we have exactly the expectation value of R in the generic background ensemble q.

Appendix B. MH on a background

Our goal is to sample a distribution of the type

$$p(G) = \frac{1}{Z} e^{-R(G)} q(G) . \tag{B1}$$

To do so we use a MCMC chain with transition probabilities $P(G|G')$ built, according to MH, from some proposal distribution $g(G|G')$ and acceptance ratio $A(G,G')$. In detailed balance we have:

$$P(G'|G) p(G) = P(G|G') p(G') \tag{B2}$$

$$\frac{p(G)}{p(G')} = \frac{P(G|G')}{P(G'|G)} \tag{B3}$$

$$\frac{p(G)}{p(G')} = \frac{g(G|G')A(G,G')}{g(G'|G)A(G',G)} \tag{B4}$$

$$\frac{A(G,G')}{A(G',G)} = \frac{p(G)g(G'|G)}{p(G')g(G|G')} \tag{B5}$$

which we can solve by the classic formula

$$A(G,G') = \min\left(1, \frac{p(G)g(G'|G)}{p(G')g(G|G')}\right) . \tag{B6}$$

Substituting in our relative canonical distribution this becomes

$$A(G,G') = \min\left(1, e^{R(G')-R(G)} \frac{q(G)g(G'|G)}{q(G')g(G|G')}\right) . \tag{B7}$$

If we chose our proposal distribution in such a way that it is in detailed balance with the background ensemble

$$g(G'|G)q(G) = g(G|G')q(G') \tag{B8}$$

this simplifies to the acceptance probability we used above:

$$A(G,G') = \min\left(1, e^{R(G')-R(G)}\right) . \tag{B9}$$

For background ensembles that are already defined as the equilibrium of some network formation process such a proposal distribution is typically available. This is the case for the relatively simple distributions used in the paper where rewiring processes provide us very easily with the appropriate background. In general it might be more difficult to craft such a proposal process for a given background ensemble. We can outline one strategy for an important special case here. If the background is the result of a random process of fixed length itself, we can do so by keeping the history of the process: $H = (G, G^{N-1}, \dots, G^1)$. Call the probability density of the random process $\bar{q}(H)$. Then we define a walk $\bar{g}(H'|H)$ on the space of histories by discarding a randomly chosen (according to some discarding probability $p^d(n)$) number of steps $N - n$, and rerunning the process from there. Call $H^n = (G^n, G^{n-1}, \dots, G^1)$ the process truncated before step $n+1$ such that $H^N = H$, then:

$$\bar{g}(H'|H) = \sum_{n}^{N-1} \bar{q}(H'|H^n) p^d(n). \tag{B10}$$

The detailed balance condition then reads:

$$\bar{g}(H'|H)\bar{q}(H) = \bar{g}(H|H')\bar{q}(H') \tag{B11}$$

$$\sum_{n}^{N-1} \bar{q}(H'|H^n)\bar{q}(H) p^d(n) = \sum_{n}^{N-1} \bar{q}(H|H'^n)\bar{q}(H') p^d(n). \tag{B12}$$

Looking at this equality term by term we can satisfy it if we have $\overline{q}(H'|H^n)\overline{q}(H) = \overline{q}(H|H'^n)\overline{q}(H')$. If H and H' do not agree on the first n terms then both sides are immediately zero, as $\overline{q}(H'|H^n)$ is trivially zero. If they agree we have $H'^n = H^n$. Using $\overline{q}(H) = \overline{q}(H|H^n)\overline{q}(H^n)$ we then immediately have

$$\overline{q}(H'|H^n)\overline{q}(H) = \overline{q}(H|H'^n)\overline{q}(H') \tag{B13}$$

$$\overline{q}(H'|H^n)\overline{q}(H|H^n)\overline{q}(H^n) = \overline{q}(H|H^n)\overline{q}(H'|H^n)\overline{q}(H^n). \tag{B14}$$

Thus this process satisfies detailed balance for the intended distribution on the space of histories.

Appendix C. Further plots

We show the various phases of the ensembles using geometrical positions for the plots in figure 7, and the convergence of four different chains at selected genericities in figure 8.

Figure 7. Examples of the various phases for networks with $N = 128$ vertices, where the vertices are plotted according to their geometrical position. ERB-*S* networks have no geometrical embedding and are not plotted. Edges connecting vertices over the boundary are plotted in grey, direct edges in black. ERB-*WL*: (a) v.. from the ERB ensemble ($\log_2 \beta = 5$). (b) the network contains most of the shortest edges and includes a big hub ($\log_2 \beta = 12$). WSB: (c)–(f) the minimum spanning tree in addition to the N next shortest edges with only 10% of rewired edges is visible WSB-*S*: (c) v.i. from WSB ($\log_2 \beta = 2$). (d) hub and cliques are formed with few long range connections ($\log_2 \beta = 8$). (e) no apparent feature is visible ($\log_2 \beta = 8$). (f) network contains a hub and otherwise only short edges ($\log_2 \beta = 13$).

Figure 8. Convergence of the MCMC algorithm for networks with $N = 256$ vertices. The plot shows the property function averaged over the several chains vs. the iteration step of the algorithm for genericity values after the phase transition. The property function fluctuates around a specific value in the last steps of the chain in all four plots.

Figure 9. Scatter of property *R* vs variance of the degree distribution (a)–(d) and clique number (e)–(h) for samples from network ensembles with $N = 128$ vertices at different genericities (darker colors are more generic). Plots (a) and (e) are cut off on the left as the background ensemble contains very large R_S.

ORCID iDs

Oskar Pfeffer ● https://orcid.org/0000-0002-9456-7694
Nora Molkenthin ● https://orcid.org/0000-0002-7190-0714
Frank Hellmann ● https://orcid.org/0000-0001-5635-4949

References

[1] Schultz P, Heitzig J and Kurths J 2014 *Eur. Phys. J. Spec. Top.* **223** 2593
[2] Opitz D W and Shavlik J W 1995 *Advances in Neural Information Processing Systems* vol 8 pp 535–41
[3] Snoeijer J H, Vlugt T J, van Hecke M and van Saarloos W 2004 *Phys. Rev. Lett.* **92** 054302
[4] Dorogovtsev S N, Mendes J F F and Samukhin A N 2000 *Phys. Rev. Lett.* **85** 4633
[5] Molkenthin N and Timme M 2016 *Phys. Rev. Lett.* **117** 168301

[6] Molkenthin N, Schröder M and Timme M 2018 *Phys. Rev. Lett.* **121** 138301
[7] Watts D J and Strogatz S H 1998 *Nature* **393** 440
[8] Barabási A-L and Albert R 1999 *Science* **286** 509
[9] Holland P W and Leinhardt S 1981 *J. Am. Stat. Assoc.* **76** 33
[10] Frank O and Strauss D 1986 *J. Am. Stat. Assoc.* **81** 832
[11] Strauss D 1986 *SIAM Rev.* **28** 513
[12] Newman M E 2003 *SIAM Rev.* **45** 167
[13] Cimini G, Squartini T, Saracco F, Garlaschelli D, Gabrielli A and Caldarelli G 2019 *Nat. Rev. Phys.* **1** 58
[14] Park J and Newman M E 2004 *Phys. Rev.* E **70** 066117
[15] Bianconi G 2007 *Europhys. Lett.* **81** 28005
[16] Bianconi G 2009 *Phys. Rev.* E **79** 036114
[17] Butts C T 2019 *J. Math. Sociol.* **43** 40
[18] Krivitsky P N 2012 *Electron. J. Stat.* **6** 1100
[19] Iba Y, Saito N and Kitajima A 2014 *Ann. Inst. Stat. Math.* **66** 611
[20] Ansari S, Anvari M, Pfeffer O, Molkenthin N, Moosavi M R, Hellmann F, Heitzig J and Kurths J 2021 *Eur. Phys. J Spec. Top.* **230** 3273
[21] Schmaus A 2022 A Markov Chain Monte Carlo Approach for Stop Pooling in On-Demand Ride-Pooling Services *Master's Thesis* Technische Universität München
[22] Brandner N 2020 *Linear Flow Networks Vulnerable for Cascades Using Simulated Annnealing Bachelor's thesis* Humboldt University Berlin
[23] Humphries M D and Gurney K 2008 *PLoS One* **3** 1
[24] Mathias N and Gopal V 2001 *Phys. Rev.* E **63** 021117
[25] Erdös P and Rényi A 1959 *Publ. Math. Debrecen* **6** 290
[26] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 *J. Chem. Phys.* **21** 1087
[27] Hastings W D 1970 *Biometrika* **57** 97
[28] Newman M 2018 *Networks* (Oxford: Oxford University Press)
[29] Fronczak A, Fronczak P and Hołyst J A 2004 *Phys. Rev.* E **70** 056110
[30] Kirkpatrick S, Gelatt C D and Vecchi M P 1983 *Science* **220** 671
[31] Chernoff H *et al* 1952 *Ann. Math. Stat.* **23** 493
[32] Jaksic V 2018 Lectures on entropy. part i (arXiv:1806.07249 [math-ph])