Relative Canonical Network Ensembles –
(Mis)characterizing Small-World Networks

Oskar Pfeffer, Nora Molkenthin, Frank Hellmann
Potsdam Institute for Climate Impact Research
TU Berlin
(Dated: January 26, 2021)

What do generic networks that have certain properties look like? We define Relative Canonical Network ensembles as the ensembles that realize a property R while being as indistinguishable as possible from a generic 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 first to the network measure "small-world-ness", thought to characterize small-world networks. We find several phase transitions as we go to less and less generic networks in which cliques and hubs emerge. Such features are not shared by typical small-world networks, showing that high "small-world-ness" does not characterize small-world networks as they are commonly understood. On the other hand we see that for embedded networks, the average shortest path length and total Euclidean link length are better at characterizing small-world networks, with hubs that emerge as a defining feature at low genericity. We expect the overall approach to have wide applicability for understanding network properties of real world interest.

I. 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 networks, which are characterized by low average shortest path length and high clustering coefficients [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? Thus, we will *define* ensembles through a particular property captured by a "property function" $R(G)$ on networks G and a background ensemble that defines our notion of generic networks in the given context. To this end, we will consider slightly generalized 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 (CNE) [14–16]. We will consider CNEs relative to the background ensemble of generic networks. Given some set of networks $\mathcal{E}$ on a finite set of vertices, 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 $\exp(-\beta R(G))q(G)$.

We emphasize that our aim is not to model empiri-

cally observed network ensembles with certain properties. There is no reason to expect empirical networks, that are the outcome of subtle formation processes, to be generic. Instead, we will study the properties themselves, specifically the most generic features that produce them, and whether or not the properties suffice to generically characterize the networks under study. Our aim in this is to understand properties that are of considerable practical interest. Companion papers will consider epidemic thresholds and the vulnerability to failure cascades in power grids. To introduce our approach, this paper will focus on well-known and well-established network measures, that are computationally challenging, instead. Specifically, we will consider the notion of "small-world-ness".

We study two ensembles, the first defined by the *small-world-ness*, as introduced in [17], the second defined by a combination of Euclidean link length and average shortest path length similar to [18]. To study these ensembles we sample them using the straightforward Metropolis-Hastings (MH)[19–21] algorithm.

In both 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 generic networks with high small-world-ness do not resemble small-world networks. Thus, we find that what [17] called small-world-ness does not actually characterize small-world networks generically.

II. RELATIVE CANONICAL NETWORK ENSEMBLE

Exponential random graphs were first introduced in [9–11]. Given the set of simple graphs $\mathcal{E}_N$ on a set of N vertices, they are defined by the probability distribution over $\mathcal{E}_N$, $p_\beta^R(G) = Z_R(\beta)^{-1} \exp\left(-\beta R(G)\right)$. That is,

they are the Gibbs ensemble at temperature $T = 1/k_B\beta$. The use of such network ensembles is sometimes justified by the fact that these are maximum entropy ensembles with a given expectation value for R. However, there is no a priori reason to expect formation processes that lead to real world networks to maximize entropy. For instance, typical formation processes do not resemble exchange with an environment at fixed genericity (in analogy to a heat bath). In fact, it was already noted in [12] that the maximum entropy ensembles do not model real world systems easily and show unexpected structures, interpreted there as an "unfortunate pathology".

Instead, we want to understand the most generic features giving rise to a property R. That is, a feature, that is observed more frequently the more the expectation value of R differs from the value expected for generic networks. As mentioned in the introduction, to define our notion of genericity we specify background ensemble $q(G)$ (for example an Erdős–Rényi ensemble at a fixed number of edges). The *relative canonical network ensemble* of R relative to q is then given by:
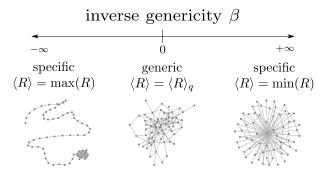
$$p_{\beta,q}^R(G) = \frac{1}{Z_{R,q}(\beta)} e^{-\beta R(G)} q(G) \,, \qquad (1)$$

with normalization/partition function $Z_{R,q}(\beta) = \sum_{G \in \mathcal{E}_N} e^{-\beta R(G)} q(G)$. This ensemble is characterized by being the ensemble of minimum relative entropy $D(p||q)$ for a fixed expectation value of R. From an information-theoretic perspective it is the ensemble hardest to distinguish from the generic ensemble q while having fixed expectation value $\langle R \rangle = R^*$, for a more detailed discussion see Appendix A.

The parameter β moderates the trade-off between the generic ensemble and highly specific ones 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 $\max(R)$, while at $\beta \to +\infty$ it is $\min(R)$. This, and the fact that interpretation of the relative entropy is purely information theoretic (rather than thermodynamic), motivates us to refer to β^{-1} in this context 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.

Throughout the rest of this manuscript we will consider canonical ensembles relative to the Erdős–Rényi ensemble at fixed size N and mean degree k, that is, the equidistribution over all graphs with vertex set $\{1, ..., N\}$ and $kN/2$ edges. Generally what counts as a generic network is highly dependent on context. A generic social network does not look like a generic power grid. In some contexts

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

FIG. 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.

it might also be appropriate to use maximum entropy null-models as generic ensembles[13].

Since exponential random graphs were first introduced, computing capabilities profoundly increased. This means we can now use complex, practically relevant network properties and analyze what features of networks generically give rise to them. This approach may help in the future to gain a better understanding of complex network measures and provide a way to find simpler network measures to act as predictors for the characteristics defining the ensemble.

To study these ensembles we need to sample from them. An important property of RCNEs is that they are well suited for sampling using Metropolis-Hastings (MH) algorithms. To use MH on our relative ensemble, we require a background process that generates proposed steps compatible with the background distribution q. For q_{Nk} this can be provided simply by considering rewiring of edges. Starting from a system in state x the algorithm proposes rewirings that are accepted with probability

$$P_\beta(x \to y) = \min\left(1, \frac{p_\beta(R(x))}{p_\beta(R(y))}\right) = \min\left(1, e^{-\beta \Delta R}\right). \qquad (2)$$

This algorithm satisfies the detailed balance condition and the Markov chain defined by it is strongly connected. In the limit of infinite steps the time average for this ensemble converges to the ensemble average of the ground state which is the relative canonical network ensemble. 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. This further allows us to obtain less correlated samples. More details on our sampling approach are provided in the next section.

For more general background ensembles it might be complicated to find step proposals. If q is explicitly
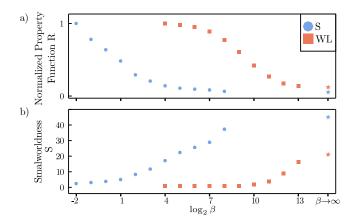
FIG. 2. **Small-world-ness increases significantly as the property function approaches the global minimum.** R_S-ensemble (circles) and R_{WL}-ensemble (squares) networks with $N = 256$ and $\langle k \rangle = 4$ to finite inverse genericity $\beta \in \{2^{-2} \rightarrow 2^{13}\}$. At the start of the range the ensembles are statistically indistinguishable from the background ensemble q. a) shows the small-world-ness S and b) the property R_S and R_{WL} vs β. The stars on the right identify simulations for $\beta \rightarrow \infty$. Each data point is averaged over 32 realizations with 2^{24} MCMC steps each.

FIG. 3. **At low genericity hub and clique structures emerge, transforming the degree distribution.** The degree distribution in a) shows three peaks for the R_S-ensemble while the degree distribution of the R_{WL}-ensemble in c) shows two peaks. b) and d) show a heat-map of degree distributions at various inverse genericities. The degree distributions are an average of 32 realizations with 2^{24} MCMC steps each.

known it can be incorporated into the step acceptance probability. If there is no explicit formula for q, for example because it is implicitly defined by a stochastic growth process, it is necessary to make use of the growth process directly to generate new proposals.

III. SMALL WORLD PROPERTIES

To demonstrate the approach, we analyze two different small-world-ness properties. In particular we look at the features that give rise to them and whether they generically characterize what is commonly known as small-world networks. In the first instance we consider the Small-world-ness measure $S = (C/L) (C_{\mathrm{ER}}/L_{\mathrm{ER}})^{-1}$, introduced in [17], where C is the global clustering coefficient defined as the number of closed triplets divided by the number of all triplets, L the average shortest path length C_{ER} is the average clustering coefficient of an Erdős–Rényi network [22] of the same size and L_{ER} is the expected average shortest path length of an Erdős–Rényi network of the same size. Finally, the form of the property we will consider is:

$$R_S = \frac{L}{C}. \tag{3}$$

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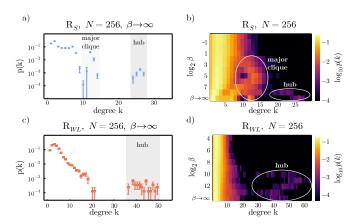
The second property

$$R_{WL} = WL \tag{4}$$

is given in terms of the average shortest path length L and the wiring length W in an embedded network. W is given by the sum of the Euclidean length of all edges. The networks for this ensemble are embedded in a 2D plane. The introduction of W was inspired by [18], 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 ensembles are taken relative to a random Erdős–Rényi network with N vertices and $M = \langle k \rangle N/2$ edges, where $\langle k \rangle$ is the average degree of the network. The positions of the vertices in the embedded networks are initialized randomly on a 2D unit square.

The proposal for each Monte Carlo step is generated by rewiring a single edge, i.e. deleting an existing edge at random and connecting two previously unconnected vertices chosen at random, thus keeping the number of edges constant. The proposal is then accepted with the transition probability given in Eq. (2). Proposals of disconnected graphs are always rejected since L is infinite. To ensure convergence at low genericity, we use an exponential schedule $\beta^{-1}(t) = \left(\beta_{\mathrm{start}}^{-1} \alpha^t + \beta_{\mathrm{end}}^{-1} \right)$, similiar to the Simulated Annealing approach [23], where t is the step parameter, $\alpha = 0.99$ is a simulation parameter. $\beta_{\mathrm{start}}^{-1}$ and $\beta_{\mathrm{end}}^{-1}$ are the start and final genericities. We generated ensembles with $(128, 128, 128, 128, 64, 32, 16)$ networks of size $N = (8, 16, 32, 64, 128, 256, 512)$ correspondingly with average degree $\langle k \rangle = 4$. All the simulations appeared to converge, allowing qualitative evaluation of the simulations. Throughout the manuscript, genericity was
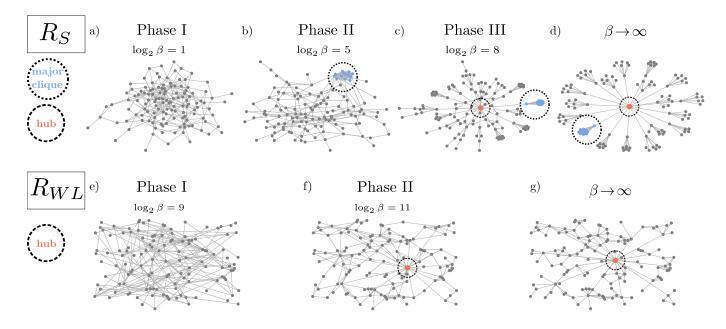
FIG. 4. **Examples of the various phases for** $N = 128$ **networks.** The R_S ensemble starts as a random network without any recognizable structure in a), then a first large clique emerges in b) and finally a central hub that lowers the average shortest path length in c), which is close to the network where the small-world-ness is maximal in d). The R_{WL} ensemble starts from a random phase in e), then a central hub with long range connections emerges in f). The network minimizing R_{WL} resembles a random geometric network with a central hub.

decreased over 2^{11} equally long periods, each containing 2^{13} MCMC steps for a total of 2^{24} MCMC steps. 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. We further note that achieving convergence for the R_S ensemble was considerably harder than for R_{WL}. Thus, they do constitute a real check of the ability of the approach to study complex network properties.

As shown in Fig. 2, the Small-world-ness S increases for both network ensembles as they become less generic. For the R_S-ensemble, 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 R_{WL}-ensemble this arises as a secondary effect as Euclidean and network distances are reduced, showing that generic R_{WL} networks do indeed have high small-world-ness, as anticipated in [18].

As a common and simple network measure, we now look at the degree distribution. Fig. 3 shows the degree distributions of the two ensembles for decreasing genericity. The shift from generic (poisson distributed) to specific networks is evident. The extremal $\beta \to \infty$ case (simulated with the same exponential schedule as above until we observed convergence) is shown explicitly in Fig. 3 a) and c), and we see highly pronounced features in the degree distribution. Example of networks at this state are shown in Fig. 4 d) and g), looking at these allows us to identify the features in the degree distribution as major

cliques and hubs. Note that the degree distribution of the R_S ensemble in particular does not resemble that of the WS-ensemble.

The R_S example network (Fig. 4 d) looks almost star-shaped with a very highly connected central node and a few fully connected branches. This indicates that the two components of the property, namely average shortest path length and clustering are optimized in specialized areas of the network. The star graph, which is the smallest possible sparse graph, is thereby combined with many nodes in fully connected cliques. The R_{WL} network (Fig. 4 g) on the other hand looks like a sparse geometric network with star-shaped shortcuts, making it much closer in spirit to the WS-ensemble and its two-dimensional relatives.

As a result, the nodes in the R_S-networks can be categorized into hub-nodes, clique-nodes, and the rest, where hub-nodes have very high degrees ($k \approx 20 - 30$), clique nodes above-average degrees ($k \approx 10 - 15$) and the rest has low degrees. This can be seen in Fig. 3b) as 3 major peaks.

To understand how these extremal cases come about, we consider the degree distributions over the whole parameter space in (Fig. 3 b and d). Here we see several abrupt transitions. For the R_S ensemble the major clique starts forming at $\beta \approx 2^{-2}$, while the hub only emerges at high inverse genericities of around $\beta \approx 2^8$.

In the embedded networks, nodes fall into two categories: regional nodes and inter-regional hubs. This can be seen in Fig. 3 c), where regional nodes fall into the nor-

mal degree distribution of a (slightly sparser) geometric network and hubs have higher degrees of $k \approx 40 - 50$. This hub emerges at around $\beta \approx 2^{10}$.

IV. GENERICITY PHASE TRANSITION

Fig. 4 shows various examples of networks taken from different genericity phases. We can now study the transition between these phases in more detail.

As seen in Fig.3 b and d, both the R_S and R_{WL} ensembles show a qualitative change in the degree distribution. At high genericity we have essentially random graphs in both cases with the expected Poisson degree distribution. At low genericity both ensembles show multiple peaks. In case of the R_{WL}-ensemble this comes as a sudden appearance of a second peak at $\beta = 2^{10}$. In case of the R_S-ensemble this transition appears to be less clear cut with a structure that resembles branching at $\beta \approx 2^2$ and almost merging again, while another peak appears at $\beta \approx 2^8$.

To better understand these transitions we analyze the mean largest eigenvalues λ_1 of the adjacency matrix, sizes of the largest non-empty k-core and maximum degree as functions of the genericity for network sizes from $N = 2^3$ to $N = 2^9$. The results are shown in Fig. 5. Both ensembles show a phase transition in the largest Eigenvalue between a low λ_1 state and a high λ_1 state. This transition is located at $\beta \approx 2^2$ for the R_S-ensemble and at $\beta \approx 2^{10}$ for the R_{WL}-ensemble.

This transition is mirrored by the maximum k-core in case of the R_S-ensemble (see Fig. 5 c)). This indicates that here the formation of the first dense region in the graph is responsible for the phase transition. This is clearly not the case for the R_{WL}-ensemble, where we find no consistent transition genericity for the largest non-empty k-core, but a shift in its rise depending on the network size as shown in Fig. 5 f).

Instead, the largest Eigenvalue transition in the R_{WL}-ensemble is mirrored by the maximum degree in the network, as displayed in Fig. 5 e). As expected from the degree distributions shown above, the changes of the maximum degree in the R_S-ensemble hint at two transitions, one at $\beta \approx 2^2$, in which the first dense region forms and one at $\beta \approx 2^8$, at which the central hub forms. The phase transition giving birth to the first dense region found at $\beta \approx 2^2$, can be interpreted as similar to [24], where a first order phase transition was analytically found for Strauss's model of clustering [11].

These results show that certain discrete features suddenly emerge at certain genericities. The transitions become qualitatively visible in the degree distribution, clearly appear in their graphical representations and can be quantified in various network measures, where the largest eigenvalue is a good first indicator and are more detailed in the maximum k-core and degree. These phase transitions and the emergence of hubs and cliques are a driving element in the increase of the small-world-ness property.

V. DISCUSSION AND CONCLUSION

Here we introduced the concept of relative canonical network ensembles of arbitrary network properties, as a means to study what the most generic networks with these properties look like. These ensembles are amenable to Metropolis-Hastings 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. Surprisingly we found that generic networks with a high small-world-ness index S in the sense of [17]. Instead we find that as S increases the most generic networks with high S contain first cliques and then hubs, neither of which occur in the WS-ensemble. An alternative property defined as the product of wiring and shortest path length fared better, here also hubs arise for the least generic networks, but the system appears to resemble small world networks more closely. This indicates that at least for some networks, spatial embedding may actually be the defining feature, from which high small-world-ness arises as a secondary effect.

The transition from highly generic to very specific ensembles in both cases is characterized by well defined phase transitions. These are visible in a number of network measures. Notably in both cases we have a rise of the largest eigenvalue of the adjacency matrix. This, however corresponds to the growth of the first dense region in the R_S-ensemble and to the emergence of an inter-regional hub in the R_{WL}-ensemble.

It is somewhat surprising that new things are still to learn on properties thought to characterize small-world networks. The fact that our perspective on relative canonical network ensembles could discover novel features is a promising sign for the study of properties of greater practical interest. In companion papers we are considering epidemic thresholds, and the vulnerability to cascading failures. More generally this method is of great interest wherever we want to understand and design topologies that fulfill certain functions, rather than describe empirical networks.
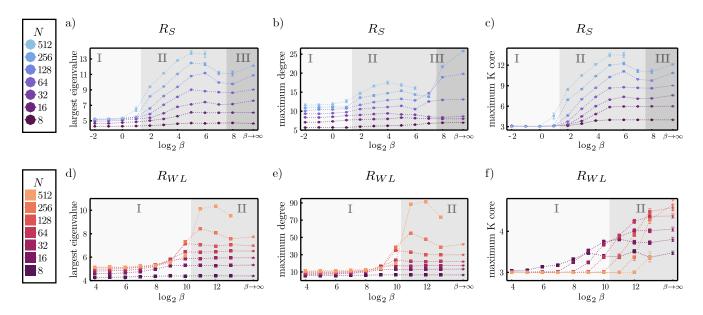
Code and Data availability

FIG. 5. **The phase transition is characterized by a rise in the largest Eigenvalue.** The largest Eigenvalue is plotted in a) and d) for R_S and R_{WL} respectively. b) and e) show the dependence of the largest degree on the genericity and c) and f) show the maximum k-core over genericity for network sizes from $N = \{8, 16, 32, 64, 128, 256, 512\}$ and average degree $\langle k \rangle = 4$. Simulation details: MCMC steps $= 2^{24}$ each, generated ensemble size (in order of network size) $= \{128, 128, 128, 128, 64, 32, 16\}$. The realizations for $R_S, N = 512, \log_2 \beta \geq 7$ did not fully converge and are not plotted.

Appendix A: Relative Entropy

The minimization of the relative entropy has an information theoretic 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 [25] is known as Stein's Lemma (for a modern account phrased in terms of relative entropy see e.g. [26] 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 an information theoretic 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

$$
\begin{aligned}
p^* &= \arg\min_{\substack{p \\ \langle R \rangle = R^*}} D(p||q) \\
&= \arg\min_{\substack{p \\ \langle R \rangle = R^*}} \sum_i p_i \ln\left(\frac{p_i}{q_i}\right)
\end{aligned}
\tag{A1}
$$

First, note that this formula diverges to positive infin-

ity if p has support outside the support of q. We thus only consider p whose support is contained in that of q. Then, by introducing Lagrange multipliers for the expectation value of R as well as for the normalization condition on the distribution p we can rewrite the constrained minimization above as a free minimization:

$$
\begin{aligned}
p^*(\beta_n, \beta_R) &= \arg\min_p \sum_i p_i \ln\left(\frac{p_i}{q_i}\right) + \\
&+ \beta_n \left(\sum_i p_i - 1\right) + \beta_R \left(\sum_i p_i R_i - R^*\right)
\end{aligned}
\tag{A2}
$$

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

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

Now the variation in the direction p_j produces the following condition:

$$
\begin{aligned}
0 &= \frac{\partial}{\partial p_j}\left[\sum_i p_i(\ln(p_i) - \ln(q_i)) + \right. \\
&\left. + \beta_n \left(\sum_i p_i - 1\right) + \beta_R \left(\sum_i p_i R_i - R^*\right)\right] \\
&= \ln(p_j) - \ln(q_j) + 1 + \beta_n + \beta_R R_j
\end{aligned}
\tag{A3}
$$

From which we can conclude

Note that

$$\frac{\partial}{\partial \beta_R} \langle R \rangle = -\frac{1}{Z} \sum_i R_i^2 e^{-\beta_R R_i} q_i$$
$$- \frac{\partial Z}{\partial \beta_R} \frac{1}{Z^2} \sum_i R_i e^{-\beta_R R_i} q_i$$
$$= -\langle R^2 \rangle + \langle R \rangle^2$$
$$= -\mathrm{Var}(R) \leq 0 . \tag{A5}$$

$$p_j^* = \exp(\ln(q_j) - 1 - \beta_n - \beta_R R_j)$$
$$= \frac{1}{Z} e^{-\beta_R R_j} q_j \tag{A4}$$

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 $R^* = \sum_i R_i p_i^*$.

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.

[1] P. Schultz, J. Heitzig, and J. Kurths, A random growth model for power grids and other spatially embedded infrastructure networks, The European Physical Journal Special Topics **223**, 2593 (2014).
[2] D. W. Opitz and J. W. Shavlik, Generating accurate and diverse members of a neural-network ensemble, in *Advances in neural information processing systems* (1996) pp. 535–541.
[3] J. H. Snoeijer, T. J. Vlugt, M. van Hecke, and W. van Saarloos, Force network ensemble: a new approach to static granular matter, Physical review letters **92**, 054302 (2004).
[4] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Structure of growing networks with preferential linking, Physical review letters **85**, 4633 (2000).
[5] N. Molkenthin and M. Timme, Scaling laws in spatial network formation, Physical review letters **117**, 168301 (2016).
[6] N. Molkenthin, M. Schröder, and M. Timme, Adhesion-induced discontinuous transitions and classifying social networks, Physical review letters **121**, 138301 (2018).
[7] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world'networks, nature **393**, 440 (1998).
[8] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, science **286**, 509 (1999).
[9] P. W. Holland and S. Leinhardt, An exponential family of probability distributions for directed graphs, Journal of the american Statistical association **76**, 33 (1981).
[10] O. Frank and D. Strauss, Markov graphs, Journal of the american Statistical association **81**, 832 (1986).
[11] D. Strauss, On a general class of models for interaction, SIAM review **28**, 513 (1986).
[12] M. E. Newman, The structure and function of complex networks, SIAM review **45**, 167 (2003).
[13] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, and G. Caldarelli, The statistical physics of real-world networks, Nature Reviews Physics **1**, 58 (2019).
[14] J. Park and M. E. Newman, Statistical mechanics of networks, Physical Review E **70**, 066117 (2004).
[15] G. Bianconi, The entropy of randomized network ensembles, EPL (Europhysics Letters) **81**, 28005 (2007).
[16] G. Bianconi, Entropy of network ensembles, Physical Review E **79**, 036114 (2009).
[17] M. D. Humphries and K. Gurney, Network 'small-world-ness': A quantitative method for determining canonical network equivalence, PLOS ONE **3**, 1 (2008).
[18] N. Mathias and V. R. Gopal, Small worlds: how and why., Physical review. E, Statistical, nonlinear, and soft matter physics **63 2 Pt 1**, 021117 (2001).
[19] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, Journal of Chemical Physics **21**, 1087 (1953).
[20] W. D. Hastings, Monte carlo sampling methods using markov chains and their applications, Biometrika **57**, 97 (1970).
[21] Y. Iba, N. Saito, and A. Kitajima, Multicanonical mcmc for sampling rare events: an illustrative review, Annals of the Institute of Statistical Mathematics **66**, 611 (2014).
[22] P. Erdős and A. Rényi, On random graphs i, Publ. math. debrecen **6**, 18 (1959).
[23] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Optimization by simulated annealing, Science **220**, 671 (1983).
[24] J. Park and M. Newman, Solution for the properties of a clustered network., Physical review. E, Statistical, nonlinear, and soft matter physics **72 2 Pt 2**, 026136 (2005).
[25] H. Chernoff *et al.*, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, The Annals of Mathematical Statistics **23**, 493 (1952).
[26] V. Jaksic, Lectures on entropy. part i (2018), arXiv:1806.07249 [math-ph].