

# Local clustering in scale-free networks with hidden variables

Remco van der Hofstad, A.J.E.M. Janssen, Johan S.H. van Leeuwen, and Clara Stegehuis

*Eindhoven University of Technology*

(Dated: January 16, 2017)

We investigate the presence of triangles in a class of correlated random graphs in which hidden variables determine the pairwise connections between vertices. The class rules out self-loops and multiple edges. We focus on the regime where the hidden variables follow a power law with exponent  $\tau \in (2, 3)$ , so that the degrees have infinite variance. The natural cutoff  $h_c$  characterizes the largest degrees in the hidden variable models, and a structural cutoff  $h_s$  introduces negative degree correlations (disassortative mixing) due to the infinite-variance degrees. We show that local clustering decreases with the hidden variable (or degree). We also determine how the average clustering coefficient  $C$  scales with the network size  $N$ , as a function of  $h_s$  and  $h_c$ . For scale-free networks with exponent  $2 < \tau < 3$  and the default choices  $h_s \sim N^{1/2}$  and  $h_c \sim N^{1/(\tau-1)}$  this gives  $C \sim N^{2-\tau} \ln N$  for the universality class at hand. We characterize the extremely slow decay of  $C$  when  $\tau \approx 2$  and show that for  $\tau = 2.1$ , say, clustering only starts to vanish for networks as large as  $N = 10^9$ .

PACS numbers: 89.75.-k Complex systems, 64.60.aq Networks

## I. INTRODUCTION

Random graphs serve as models for large networked systems that arise in nature or in our technosphere. The sheer complexity of many such networks prevents a detailed microscopic modeling, which is why random graphs only use partial descriptions of networks, such as degree distributions. Statistical analysis of network data suggests that many networks possess a power-law degree distribution [1–4], where the probability  $P(k)$  that a node has  $k$  neighbors scales as  $k^{-\tau}$  for some characteristic exponent  $\tau > 0$ . The power-law distribution leads to the likely presence of high-degree nodes. Many real-world networks are reported to have an exponent  $\tau$  between 2 and 3 [5–7]. Since  $\tau < 3$  for these networks, the second moment of the degree distribution diverges in the infinite-size network limit.

Hidden variable models present a class of popular null models for scale-free networks [8, 9]. In these models vertices are characterized by hidden variables that influence the creation of edges between pairs of vertices. The models can be seen as enlarged ensembles of random graphs that can match in expectation any given degree distribution  $P(k)$ . All topological properties, including correlations and clustering, then become functions of the distribution of the hidden variables and the probability of connecting vertices [10, 11]. The independence between edges makes hidden-variable models analytically tractable. One complication though, is that these hidden-variable models introduce disassortative degree-degree correlations when  $\tau < 3$ : high-degree vertices tend to be connected to low-degree vertices. This negative correlation can have a strong influence on topological network properties, including clustering, defined as the presence of triangles in the network [8, 12].

In [9] it was shown that hidden-variable models with a nonrestrictive cutoff scheme can generate nearly size-independent levels of clustering, and can thus generate networks with high levels of clustering, particularly for  $\tau$

close to 2. In the configuration model, without banning large-degree vertices by installing a cutoff, the long tail of the power law makes that pairs of high-degree vertices would quite likely share more than one edge. But hidden-variable models allow at most one edge between pairs of vertices, so that large-degree vertices must inevitably connect to small-degree vertices due to lack of available large-degree vertices. This phenomenon, in turn, is related to the difference in scaling between the so-called structural cutoff and natural cutoff. The structural cutoff is defined as the largest possible upper bound on the degrees required to guarantee single edges, while the natural cutoff characterizes the maximal degree in a sample of  $N$  vertices. For scale-free networks with  $\tau \in (2, 3]$  the structural cutoff scales as  $N^{1/2}$  while the natural cutoff scales as  $N^{1/(\tau-1)}$  (see Section II), which gives rise to structural negative correlations and possibly other finite-size effects.

Clustering can be measured in various ways. The local clustering coefficient of a vertex is defined as the number of existing edges among the set of its neighbors divided by the total number of possible connections between them. This can be interpreted as the probability that two randomly chosen neighbors of a vertex are neighbors themselves. One could calculate a global clustering coefficient as the total number of triangles (closed triples of three vertices) divided by the number of connected triples. However, when  $\tau \in (2, 3)$ , this version of the clustering coefficient has the undesired property that it always tends to zero when the number of vertices tends to infinity, even if the network is highly clustered [13]. Therefore, we employ as a metric a different global clustering coefficient, the average clustering coefficient  $C$ , defined as the average (over vertices of degree  $\geq 2$ ) of the local clustering coefficient of single vertices. For a node  $i$  the local clustering coefficient is given by  $c_i = 2T_i/k_i(k_i - 1)$  with  $k_i$  the degree of node  $i$  and  $T_i$  the number of triangles that vertex  $i$  is part of.

In the absence of high-degree nodes, the average clus-

tering coefficient is given by [2]

$$C = \frac{1}{N} \sum_{i=1}^N c_i = \frac{\langle k(k-1) \rangle^2}{N \langle k \rangle^3}, \quad (1)$$

which shows that clustering vanishes very fast in the large network limit  $N \rightarrow \infty$  in support of the tree-like approximations of complex networks. However, for scale-free distributions with  $\tau < 3$ , the natural cutoff that scales as  $N^{1/(\tau-1)}$  together with (1) gives  $C \sim N^{(\tau-3)/(\tau-1)}$ . The diverging  $C$  for  $\tau < 7/3$  is caused by the many edges between the high-degree vertices, and can be judged as anomalous or nonphysical behavior if one wants  $C$  to be smaller than 1 and interpret it as a probability or proportion. If a structural cutoff of order  $N^{1/2}$  is imposed, hence banning the largest-degree nodes, formula (1) predicts the correct (in the sense that it matches simulations) scaling  $N^{2-\tau}$  [9, 14].

In a power-law setting, the infinite variance is essential for describing scale-free network behavior, which makes the banning of large-degree vertices unnatural. In this paper we investigate average clustering for an ensemble of scale-free random graphs that allows for an interplay between structural correlations and large-degree nodes. The clustering coefficient in this ensemble turns out to depend on the size of the network, the structural cutoff that arises when conditioning on simplicity and the natural cutoff that accounts for large degrees.

## II. HIDDEN VARIABLES AND CUTOFFS

Given  $N$  nodes, hidden-variable models are defined as follows: (i) associate to each node a hidden variable  $h$  drawn from a given probability distribution function  $\rho(h)$  and (ii) join each pair of vertices independently according to a given probability  $p(h, h')$  with  $h$  and  $h'$  the hidden variables associated to the two nodes. The probability  $p(h, h')$  can be any function of the hidden variables, as long as  $p(h, h') \in [0, 1]$ . Many networks can be embedded in this hidden-variable framework, but particular attention goes to the case in which the hidden variables have themselves the structure of the degrees of a real-world network. One could interpret the hidden-variable model as yielding soft constraints on the degrees, rather than hard constraints often used in the configuration model [1–4, 15, 16]. Chung and Lu [17] introduced this model in the form

$$p(h, h') \sim \frac{hh'}{N \langle h \rangle}, \quad (2)$$

so when allowing self-loops and multiple edges, the expected degree of a node equals its hidden variable. For (2) to make sense we need that the maximal value of the product  $hh'$  never exceeds  $N \langle h \rangle$  and this can be guaranteed by the assumption that the hidden degree  $h$  is smaller than the structural cutoff  $h_s = \sqrt{N \langle h \rangle}$ . While this restricts  $p(h, h')$  within the interval  $[0, 1]$ , the structural

cutoff strongly violates the reality of scale-free networks. Regarding the hidden variables as the desired degrees in the CM, the structural cutoff conflicts with the fact that the natural cutoff for the degree scales as  $N^{1/(\tau-1)}$ .

In [8, 10, 11] more general hidden-variable models were introduced, constructed to preserve conditional independence between edges, while making sure there is only one edge between every vertex pair and that the natural extreme values of power-law degrees are not neglected. Within that large spectrum of models, we focus on the subset of models for which

$$P(k) \sim \rho(k), \quad (3)$$

so that the degrees and the hidden variables in the network are similar. The class of models considered in this paper starts from the ansatz  $p(h, h') \approx hh'/N \langle h \rangle$ , but like [8, 10, 11, 18, 19] *adapts* this setting to incorporate extreme values and to rule out self-loops.

### A. Class of random graphs

Within the wide class of hidden-variable models [8, 10, 11] we consider probabilities of the form

$$p(h, h') = r(u) = uf(u) \quad \text{with} \quad u = \frac{hh'}{h_s^2} \quad (4)$$

with functions  $f : [0, \infty) \rightarrow (0, 1]$  that belong to the F-class spanned by the properties

F1  $f(0) = 1$ ,  $f(u)$  decreases to 0 as  $u \rightarrow \infty$ .

F2  $r(u) = uf(u)$  increases to 1 as  $u \rightarrow \infty$ .

F3  $f$  is continuous and there are  $0 = u_0 < u_1 < \dots < u_K < \infty$  such that  $f$  is twice differentiable on each of the intervals  $[u_{k-1}, u_k]$  and on  $[u_K, \infty)$ , where

$$f'(u_k) = \frac{1}{2}f'(u_k + 0) + \frac{1}{2}f'(u_k - 0) \quad (5)$$

for  $k = 1, \dots, K$  and  $f'(0) = f'(+0)$ .

F4  $-uf'(u)/f(u)$  is increasing in  $u \geq 0$ .

The class of hidden-variable models considered in this paper is completely specified by all functions  $f$  that satisfy F1-F4. Here are important classical members of the F-class:

(i) (*maximally dense graph*) The Chung-Lu setting

$$r(u) = \min\{u, 1\}. \quad (6)$$

This is the default choice in [11] and leads within the F-class to the densest random graphs.

(ii) (*Poisson graph*) A simple exponential form gives

$$r(u) = 1 - e^{-u}. \quad (7)$$

Here we take  $u$  to define the intensities of Poisson processes of edges, and ignore multiple edges, so that (7) gives the probability that there is an edge between two vertices. Variants of this form are covered in e.g. [11, 19–21].

- (iii) (*maximally random graph*) The next function was considered in [8, 9, 22]:

$$r(u) = \frac{u}{1+u}. \quad (8)$$

This connection probability ensures that the entropy of the ensemble is maximal [9]. This random graph is also known in the literature as the generalized random graph [18, 23].

The conditions F1-F4 will prove to be the minimally required conditions for the results that we present for the clustering coefficient. The F-class is constructed so that it remains amenable to analysis; the technique developed in [9] to characterize the average clustering coefficient despite the presence of correlations can be applied to our class. Notice that the technical condition F3 allows to consider piecewise smooth functions with jumps in their derivatives, such as  $\min(1, 1/u)$  that comes with (6). It can be shown that F4 is slightly stronger than the condition of concavity of  $r(u)$ . It appears in Section III that F4 is necessary and sufficient for monotonicity in a general sense of the local clustering coefficient  $c(h)$ .

### B. Cutoffs and correlation

The hidden-variable model by definition excludes self-loops and avoids multiple connections after imposing the structural cutoff  $h_s \sim N^{1/2}$ . Since the natural cutoff is of the order  $h_c \sim N^{1/(\tau-1)}$ , for  $\tau \geq 3$  the structural cutoff dominates and correlations are avoided. For  $\tau < 3$ , however, the structural cutoff is smaller than the natural cutoff predicted by extreme value theory. All actual cutoffs larger than  $h_s$  will then result in a network with a structure that can only be analyzed by considering non-trivial degree-degree correlations.

The structural cutoff  $h_s$  now marks the point as of which correlations imposed by the network structure arise. All pairs of vertices with hidden variables smaller than this cutoff are connected with probability close to  $u = hh'/h_s^2$  and do not show degree-degree correlation. The extent to which the network now shows correlation is determined by the gap between the natural cutoff  $h_c$  and the structural cutoff  $h_s$ . A fully uncorrelated network arises when  $h_c < h_s$ , while correlation will be present when  $h_c > h_s$ . Let  $\langle h \rangle$  denote the average value of the random variable  $h$  with density  $\rho(h) = Ch^{-\tau}$  on  $[h_{\min}, N]$ , so that

$$\langle h \rangle = \frac{\int_{h_{\min}}^N h^{1-\tau} dh}{\int_{h_{\min}}^N h^{-\tau} dh} = \frac{\tau - 1}{\tau - 2} \frac{h_{\min}^{2-\tau} - N^{2-\tau}}{h_{\min}^{1-\tau} - N^{1-\tau}}. \quad (9)$$

With the default choices

$$h_s = \sqrt{N\langle h \rangle}, \quad (10)$$

$$h_c = (N\langle h \rangle)^{1/(\tau-1)} \quad (11)$$

in mind, the regime in terms of cutoffs we are interested in is, just as in [9],

$$h_s \leq h_c \ll h_s^2, \quad (12)$$

where we regard these cutoffs as indexed by  $N$  and consider what happens as  $N \rightarrow \infty$ , with emphasis on the asymptotic regime  $h_s \ll h_c$  for  $N$  large.

In Appendix H we show that  $h_c$  as given in (11) is an accurate approximation of  $\mathbb{E}[\max(\underline{h}_1, \dots, \underline{h}_N)]$ , where the  $\underline{h}_i$  are i.i.d. with  $\rho(h)$  as density.

### III. UNIVERSAL PROPERTIES

For the class of hidden-variable models described in Section II A, we will characterize the large-network asymptotics of the local clustering coefficient  $c(h)$  and average clustering coefficient  $C$ . The first result in this direction for  $C$  was obtained for a class of uncorrelated random scale-free networks with a cutoff of  $N^{1/2}$  [14] for which  $C$  turned out to scale as  $N^{2-\tau}$ , a decreasing function of the network size for  $\tau > 2$ . In [9] the more general setup discussed in Section II was used, with the specific choice of  $r(u) = u/(1+u)$ . After involved calculations with Lerch's transcendent, [9] revealed the scaling relation

$$C \sim h_s^{-2(\tau-2)} \ln(h_c/h_s). \quad (13)$$

For the default choices (10) and (11) this predicts  $C \sim N^{2-\tau} \ln N$  (ignoring the constant).

We adopt the hidden variables formalism developed in [10] that leads, among other things, to explicit expressions for the local clustering coefficient  $c(h)$  of a node with hidden variable  $h$  and for the average clustering coefficient  $C$ .

The clustering coefficient of a vertex with hidden variable  $h$  can be interpreted as the probability that two randomly chosen edges from  $h$  are neighbors. The clustering of a vertex of degree one or zero is defined as zero. Then, if vertex  $h$  has degree at least two,

$$c(h) = \int_{h_{\min}}^{h_c} \int_{h_{\min}}^{h_c} p(h'|h)p(h', h'')p(h''|h)dh'dh'', \quad (14)$$

with  $p(h'|h)$  the conditional probability that a randomly chosen edge from an  $h$  vertex is connected to an  $h'$  vertex given by

$$p(h'|h) = \frac{\rho(h')p(h, h')}{\int_{h''} \rho(h'')p(h, h'')dh''}. \quad (15)$$

The degree of a vertex conditioned on its hidden variable  $h$  is distributed as a Poisson random variable with

parameter  $h$  [10] and [23, Chapter 6]. Therefore, the probability that a vertex with hidden variable  $h$  has degree at least two is given by

$$\mathbb{P}(k \geq 2 | h) = \sum_{k=2}^{\infty} \frac{h^k e^{-h}}{k!} = 1 - e^{-h} - he^{-h}. \quad (16)$$

Therefore, for  $\rho(h) \sim h^{-\tau}$  [10, Eq. (29)]

$$c(h) = (1 - e^{-h} - he^{-h}) \cdot \frac{\int_{h_{\min}}^{h_c} \int_{h_{\min}}^{h_c} \rho(h')p(h, h')\rho(h'')p(h, h'')p(h', h'')dh'dh''}{\left[\int_{h_{\min}}^{h_c} \rho(h')p(h, h')dh'\right]^2}, \quad (17)$$

and hence

$$C = \int_{h_{\min}}^{h_c} \rho(h)c(h)dh. \quad (18)$$

The Poisson distribution is sharply peaked around  $k = h$ , which for large  $k$  yields

$$P(k) \sim \rho(k) \quad \text{and} \quad \bar{c}(k) \sim c(k), \quad (19)$$

where  $\bar{c}(k)$  denotes the average clustering coefficient over all vertices of degree  $k$ , so that the hidden variables become hidden degrees.

We make the change of variables

$$a = \frac{1}{h_s}, \quad b = \frac{h_c}{h_s} \quad (20)$$

and assume henceforth, in line with (12), that

$$0 < ah_{\min} \leq ah_{\min}b \leq 1 \leq b < \infty, \quad 2 < \tau < 3. \quad (21)$$

This gives  $c(h) = (1 - e^{-h} - he^{-h})c_{ab}(h)$  with

$$c_{ab}(h) = \frac{\int_{ah_{\min}}^b \int_{ah_{\min}}^b (xy)^{-\tau} r(ahx)r(ahy)r(xy)dx dy}{\left[\int_{ah_{\min}}^b x^{-\tau} r(ahx)dx\right]^2}. \quad (22)$$

Note that within the domain of integration  $[ah_{\min}, b]$  in Eq. (22) the arguments  $ahx$  and  $ahy$  do not exceed a maximum value  $O(ab)$  as long as  $h < h_s^2/h_c$ , which tends to zero under assumption (12). Therefore, since  $r(u) \approx u$ ,  $c_{ab}(h) \approx c_{ab}(0)$  for  $h < h_s^2/h_c$ . When choosing  $h_s$  as in (10), this means that  $c_{ab}(h) \approx c_{ab}(0)$  for  $h \leq N\langle h \rangle/h_c$ . In Proposition 1 below we will prove that  $h \mapsto c_{ab}(h)$  is a bounded monotonically decreasing function for the class of models at hand. Furthermore, the density  $\rho(h) \sim h^{-\tau}$  with  $\tau \geq 2$ , decays sufficiently rapidly for the integral in (18) for  $C$  to have converged already before  $c_{ab}(h)$  starts to drop significantly below its value  $c_{ab}(0)$  at  $h = 0$ . Thus,  $C$  can be approximated with

$$C_{ab}(\tau) = c_{ab}(0) \int_{h_{\min}}^N \rho(h)(1 - (1+h)e^{-h})dh \\ := c_{ab}(0)A(\tau), \quad (23)$$

where we have conveniently extended the integration range to the  $\tau$ -independent interval  $[h_{\min}, N]$  at the expense of a negligible additional error.

**Proposition 1.** *Assume that  $f$  satisfies F1 - F3. Then  $c_{ab}(h)$  is decreasing in  $h \geq 0$  for all  $a, b$  with  $0 < a < b$  if and only if  $f$  satisfies F4.*

Proposition 1 shows that for large enough  $h$  local clustering decreases with the hidden variable. For the default choices with (6), (10), (11), local clustering  $c(h)$  is plotted in Fig. 1, which shows both exact formulas and extensive simulations. Because our model starts from a single-edge constraint, Proposition 1 also provides support for the asserted disassortative mixing observed in many technological and biological networks [8, 12]. The proof of Proposition 1 can be found in Appendix B.

**Proposition 2.** *Assume that  $f$  is positive, and satisfies F3. Then  $c_{ab}(0)$  is decreasing in  $\tau > 0$  for all  $a, b$  with  $0 < a \leq b < \infty$  if and only if  $f$  satisfies F2.*

We shall prove Proposition 2 in Appendix A. Proposition 2 gives evidence for the fact that clustering increases as  $\tau$  decreases, as confirmed in Fig. 3. More precisely, Proposition 2 shows the monotonicity of  $c_{ab}(0)$ , which is one of the factors of  $C_{ab}(\tau)$  in (23). The issue of monotonicity of  $C_{ab}(\tau)$  is more delicate, since  $a$  and  $b$  are function of  $\tau$  themselves. In Appendix G we present several other monotonicity properties of the remaining building blocks that together give  $C_{ab}(\tau)$ . It follows that for  $\tau > 2$ ,  $C_{ab}(\tau)$  is bounded from above by an envelope function of  $\tau$  that is very close to  $C_{ab}(\tau)$  and that is decreasing in  $\tau$ . Figure 2 provides empirical evidence for the monotonicity of  $C_{ab}(\tau)$  in  $\tau$ . This monotonicity seems to conflict observations made in [9], where the clustering coefficient of a hidden-variable model first increases in  $\tau$  when  $\tau$  is close to 2, and then starts decreasing. The difference is caused by the choice of the structural cutoff. Where we take  $h_s = \sqrt{N\langle h \rangle}$  with  $\langle h \rangle$  as in (9), in [9]  $h_s = \sqrt{N(\tau - 1)/(\tau - 2)}$  was used. Thus, in [9], the structural cutoff includes the infinite system size limit of  $\langle h \rangle$ , where we use the size-dependent version of  $\langle h \rangle$ .

Figure 3 suggests that  $C$  falls off with  $N$  according to a function  $N^\delta$  where  $\delta$  depends on  $\tau$ . In Proposition 3 below, we will show that for the F-class of hidden-variable models and the standard cutoff levels,  $C$  decays as  $N^{\tau-2} \ln N$ . On a log-scale, moreover, the clustering coefficient of different hidden-variable models in the F-class only differs by a constant, which is confirmed in Fig. 3 and substantiated in Proposition 4. Then, we will focus in Section V on  $\tau \approx 2$ , for which Fig. 3 suggests that the clustering remains nearly constant as a function of  $N$ , and characterize how large a network should be for  $C$  to start showing decay. This again will depend on  $\tau$ .

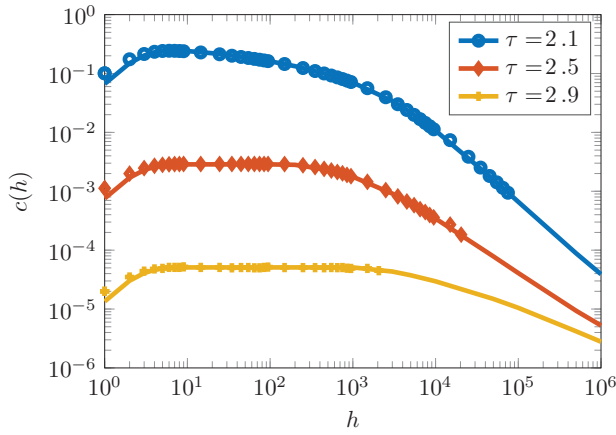


Figure 1.  $c(h)$  for  $\tau = 2.1, 2.5, 2.9$  and networks of size  $N = 10^6$ , using  $h_{\min} = 1$ . The markers indicate the average of  $10^5$  simulations, and the solid lines follow from (17).

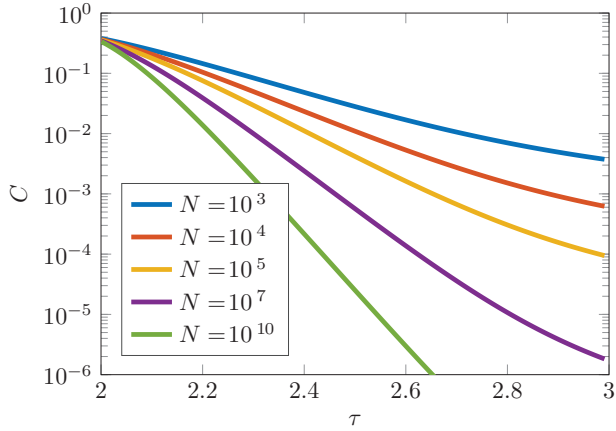


Figure 2.  $C_{ab}^{\max}(\tau)$  as a function of  $\tau$ , with  $r$  as in (6), using  $h_{\min} = 1$ .

#### IV. UNIVERSAL BOUNDS

We next compute the clustering coefficient  $C_{ab}^{\max}(\tau) = C_{ab}(\tau)$  for the maximally dense graph with  $f(u) = f_{\max}(u) = \min(1, 1/u)$ ,  $u \geq 0$  and  $a, b$  satisfying (21). In this case we have  $c_{ab}(h) = c_{ab}(0)$  for  $h \leq 1/(ab) = h_s^2/h_c = N\langle h \rangle/h_c$ . It is easy to see that  $f_{\max}$  is the maximal element in the F-class in the sense that  $f(u) \leq f_{\max}(u)$  for all  $u \geq 0$  and all  $f \in F$ . For a general  $f \in F$  we shall also bound  $C_{ab}^f(\tau)$  in terms of  $C_{ab}^{\max}(\tau)$ . This yields a scaling relation similar to (13), but then for the whole F-class. We start from an explicit representation for  $C_{ab}^{\max}(\tau)$ :

**Proposition 3.**

$$C_{ab}^{\max}(\tau) = \frac{A(\tau)(\tau-2)^2}{((ah_{\min})^{2-\tau} - b^{2-\tau})^2} \times I_{ab}^{\max}(\tau), \quad (24)$$

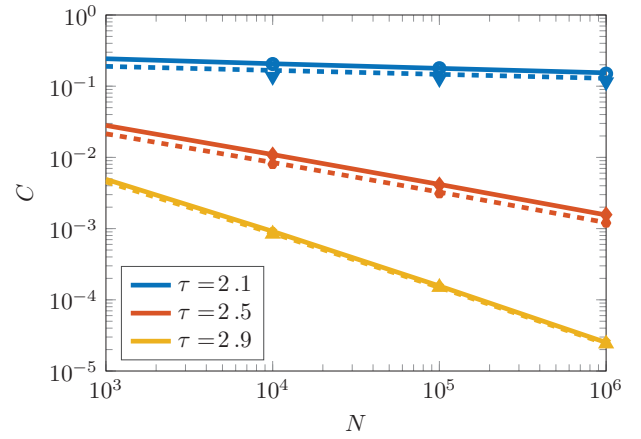


Figure 3.  $C_{ab}(\tau)$  for  $\tau = 2.1, 2.5, 2.9$ , choices (6) (solid line) and (8) (dashed line) and networks of size  $N = 10^k$  for  $k = 4, 5, 6$ , using  $h_{\min} = 1$ . The markers indicate the average of  $10^5$  simulations, and the solid lines follow from (24) and (F2).

with  $A(\tau)$  given in (23) and

$$I_{ab}^{\max}(\tau) = \frac{\ln(b^2)}{(\tau-2)(3-\tau)} - \frac{1-b^{2(2-\tau)}}{(\tau-2)^2} + \frac{1-2(ah_{\min}b)^{3-\tau} + (ah_{\min})^{2(3-\tau)}}{(3-\tau)^2}. \quad (25)$$

In Appendix F we give the counterpart of (24) for the maximally random graph (8) studied in [9], and show that on a log-scale the leading asymptotics differs only by a constant, so that the decay exponent describing how the clustering decays with network size is the same. This can also be seen in Fig. 3. In fact, for all functions  $f$  in the F-class we show below that the decay exponent is universal, and that the difference in constants can be bounded.

When  $\tau$  is away from 2 and 3, and  $b$  is large and  $a$  is small, we can ignore the  $b^{2-\tau}$  in the front factor of (24) and the second term in (25). Furthermore,  $ab = O(N^{\frac{2-\tau}{\tau-1}})$ , so that we may ignore this factor in the third term of (25). In this case we get the approximation

$$C_{ab}^{\max}(\tau) \approx A(\tau) \frac{\tau-2}{3-\tau} (ah_{\min})^{2(\tau-2)} \ln(b^2). \quad (26)$$

Using the default choices for  $a$  and  $b$  from (10), (11) and (20) then shows that  $C_{ab}^{\max} \sim N^{2-\tau} \ln(N)$  (ignoring the constant). Proposition 3 can be used to find upper and lower bounds for  $C_{ab}^f(\tau)$  with general  $f \in F$ . Since  $f(u) \leq f_{\max}(u)$ ,  $u \geq 0$ , it follows from (23) that

$$C_{ab}^f(\tau) \leq C_{ab}^{\max}(\tau). \quad (27)$$

**Proposition 4.** For all  $u_0 \geq 1$ ,

$$C_{ab}^f(\tau) \geq u_0 f(u_0) C_{a_0 b_0}^{\max}(\tau), \quad (28)$$

with  $a_0 = a/\sqrt{u_0}$  and  $b_0 = b/\sqrt{u_0}$ .

In particular, the choice  $u_0 = 1$  yields

$$C_{ab}^f(\tau) \geq f(1)C_{ab}^{\max}(\tau), \quad (29)$$

which together with (26) and (27) gives the large-network behavior of  $C_{ab}^f(\tau)$ , when  $\tau \in (2, 3)$  and away from 2 and 3, and  $a$  is small and  $b$  is large (up to a multiplicative constant that is less interesting). In particular, this shows that  $C_{ab}^f \sim N^{2-\tau} \ln(N)$  (again ignoring the constant). The proofs of Propositions 3 and 4 can be found in Appendix C and D, respectively.

## V. PERSISTENT CLUSTERING

In [9] it was observed that for values of the exponent  $\tau \approx 2$ , clustering remains nearly constant up to extremely large network sizes, which makes the convergence to the thermodynamic limit extremely slow (as also observed in [24, 25]). Here we now use the explicit results for the maximally dense graph to characterize this rate of convergence as a function of the network size  $N$ . For convenience, we assume in this section that  $h_{\min} = 1$ .

In view of the lower and upper bounds obtained in Section IV for  $C_{ab}^f(\tau)$  with general  $f \in \mathbb{F}$  in terms of  $C_{ab}^{\max}(\tau) = C_{ab}(\tau)$ , it suffices to consider  $C_{ab}^{\max}(\tau)$  for  $\tau$  close to 2. In Appendix E we show that when  $\tau$  is close to 2 and  $|\ln(ab)/\ln(b^2)|$  is small,  $C_{ab}^{\max}(\tau)$  can be approximated by

$$C_{ab}^{\max}(\tau) \approx \frac{A(\tau)(1 - \frac{1}{3}(\tau - 2)\ln(b^2))}{2(1 - \frac{1}{2}(\tau - 2)\ln(ab) + \frac{1}{6}(\tau - 2)^2 \ln^2 b)^2}. \quad (30)$$

It is the term  $-\frac{1}{3}(\tau - 2)\ln(b^2)$  in the numerator and the term  $\frac{1}{6}(\tau - 2)^2 \ln^2 b$  in the denominator (the term  $\frac{1}{2}(\tau - 2)\ln(ab)$  being less important) of the right-hand side of (30) that are the main influencers for when  $C_{ab}^{\max}(\tau)$  starts to decay. The decay is certainly absent as long as the numerator  $1 - \frac{1}{3}(\tau - 2)\ln(b^2)$  is away from zero.

We then apply this reasoning to the canonical choices  $h_s = \sqrt{N\langle h \rangle}$  and  $h_c = (N\langle h \rangle)^{1/(\tau-1)}$ , for which

$$b = (N\langle h \rangle)^{\frac{3-\tau}{2(\tau-1)}}, \quad ab = (N\langle h \rangle)^{-\frac{\tau-2}{\tau-1}}, \quad (31)$$

ensuring  $|\ln(ab)/\ln(b^2)| = (\tau - 2)/(\tau - 3)$  to be small indeed. Then, choosing a threshold  $t \in (0, 3)$  and solving  $N$  from

$$(\tau - 2)\ln(b^2) = t, \quad (32)$$

we get

$$N\langle h \rangle = \exp\left(\frac{\tau - 1}{(\tau - 2)(3 - \tau)}t\right). \quad (33)$$

In Table I we consider the case that  $t = 2$  and use that  $\langle h \rangle$  can accurately be bounded above by  $\ln N$  when  $h_{\min} = 1$

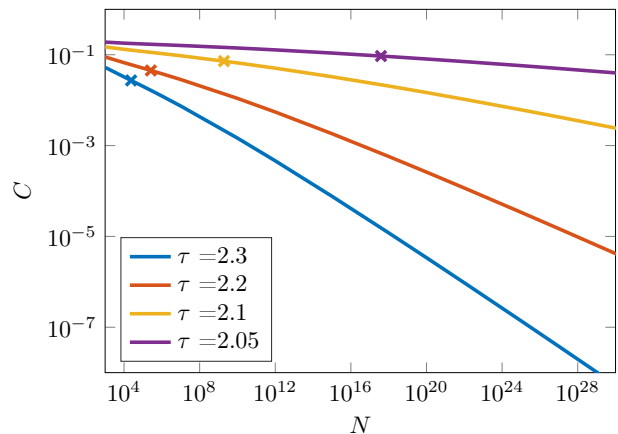


Figure 4.  $C_{ab}^{\max}(\tau)$  as a function of  $N$  for  $\tau$  close to 2, with  $r(u)$  as in (6) and  $h_{\min} = 1$ . The marks indicate the value of  $N_{\tau,2}$  as calculated in Table I.

$\tau$	$N_{\tau,2}$
2.3	$2.37 \cdot 10^4$
2.2	$2.62 \cdot 10^5$
2.1	$1.93 \cdot 10^9$
2.05	$3.92 \cdot 10^{17}$

Table I. Solution  $N_{\tau,t}$  to  $(\tau - 2)\ln(b^2) = 2$ .

and  $\tau$  is close to 2, and we let  $N_{\tau,2}$  be such that  $N \ln N$  equals the right-hand side of (33). For  $\tau = 2.1$ , the value of  $N$  where the clustering starts to decay is much larger than the typical size of real-world network data sets. This supports the observation that clustering is persistent for  $\tau$  close to 2. Figure 4 shows the decay of the clustering coefficient for  $\tau \approx 2$ , together with  $N_{\tau,2}$ , the value where the clustering coefficient is expected to decay. We see that for  $\tau$  very close to 2, clustering indeed barely decays for  $N$  smaller than  $N_{\tau,2}$ . On the other hand, we see that for  $\tau = 2.2$  and  $\tau = 2.3$  the decay already starts before  $N_{\tau,2}$ , so there the approximation is less accurate.

## VI. OUTLOOK

For hidden-variable models with scale-free degree distributions and connection probabilities in the F-class, we have shown that the local clustering coefficient  $c(h)$  decays with the hidden variable  $h$  and that the average clustering coefficient  $C(\tau)$  roughly decreases with the tail exponent  $\tau$  according to some function that depends on the structural and natural cutoffs. For the typical cutoff choices  $\sqrt{N}$  and  $N^{1/(\tau-1)}$  this showed that  $C$  decays as  $N^{2-\tau} \ln N$ , confirming an earlier result in [9] and suggesting universal behavior for the entire F-class introduced in this paper. By analyzing the special case of maximally dense graphs, a member of the F-class, we estimated the constant  $C(\tau)/N^{2-\tau} \ln N$  and the extremely slow decay that occurs when  $\tau \downarrow 2$ .

The hidden-variable model is a widely adopted null model, not only because of its ability to generate inhomogeneous random graphs, but also because of recent work that uses the hidden-variable model for constructing geometric versions of random graphs [26]. Here, the common thread is to equip every vertex not only with a weight, but also with a randomly chosen position in some space. Two vertices then form an edge independently with a probability that is proportional to the product of their weights and inversely proportional to some power of their Euclidian distance, that gives rise to a class of connection probabilities that generalizes the F-class with a geometric feature. It would be interesting to use the methods developed in this paper to investigate the clustering in relation to cutoffs and tail exponent in these graphs with an underlying geometry.

Another possible thread is to compare clustering in the hidden-variable model with clustering in other null models, like the configuration model (CM). For any given real-world network, the CM preserves the degree distribution  $P(k)$ , and makes connections between vertices in the most random way possible [1–4, 15, 16]. Given the random nature of the edge assignment, the CM has in principle no degree correlations. But in case of scale-free networks with diverging second moment, this random assignment leads to uncorrelated networks with non-negligible fractions of self-loops (a vertex joined to itself) and multiple connections (two vertices connected by more than one edge). This could be avoided by forbidding self-loops and multiple edges, for instance by ge-

nerating a sample from the CM and then erasing all the self-loops and multiple edges. This comes at the cost, however, of introducing non-trivial degree correlations among vertices.

In future work we want to investigate the clustering in such erased configuration models. Hidden-variable models are *soft* models, where unlike *hard* models such as the (erased) CM, graph constraints are satisfied only on average. Soft models are probabilistically more tractable because of the weak dependence structures, which makes it easier to show model properties. Transferring results for soft models to hard models is not straightforward, and the question is whether the important clustering properties are asymptotically invariant to the soft and hard constraints in the large-network limit. If this is the case, then that would extend the universality class from hidden-variable models in the F-class to a larger class of random network models with single-edge constraints.

## ACKNOWLEDGEMENTS

This work is supported by NWO TOP grant 613.001.451 and by the NWO Gravitation Networks grant 024.002.003. The work of RvdH is further supported by the NWO VICI grant 639.033.806. The work of JvL is further supported by an NWO TOP-GO grant and by an ERC Starting Grant.

- 
- [1] A. Clauset, C. R. Shalizi, and M. Newman, *SIAM Rev.* **51**, 661 (2009).
- [2] M. Newman, *Networks: An Introduction* (Oxford Univ Pr, 2010).
- [3] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. E* **65**, 035108 (2002).
- [4] H. Ebel, L. Mielsch, and S. Bornholdt, *Phys. Rev. E* **66**, 035103 (2002).
- [5] R. Albert, H. Jeong, and A.-L. Barabási, *Nature* **401**, 130 (1999).
- [6] C. Faloutsos, P. Faloutsos, and M. Faloutsos, *Computer Communications Rev.* **29**, 251 (1999).
- [7] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, *Nature* **407**, 651 (2000).
- [8] J. Park and M. E. Newman, *Physical Review E* **70**, 066117 (2004).
- [9] P. Colomer-de Simon and M. Boguñá, *Phys. Rev. E* **86**, 026120 (2012).
- [10] M. Boguñá and R. Pastor-Satorras, *Phys. Rev. E* **68**, 036112 (2003).
- [11] B. Bollobás, S. Janson, and O. Riordan, *Random Structures & Algorithms* **31**, 3 (2007).
- [12] S. Maslov and K. Sneppen, *Science* **296**, 910 (2002).
- [13] L. Ostroumova Prokhorenkova and E. Samosvat, “Global clustering coefficient in scale-free networks,” in *Algorithms and Models for the Web Graph: 11th International Workshop, WAW 2014, Beijing, China, December 17-18, 2014, Proceedings*, edited by A. Bonato, F. C. Graham, and P. Prałat (Springer International Publishing, Cham, 2014) pp. 47–58.
- [14] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, *Physical Review E* **71**, 027103 (2005).
- [15] S. Dhara, R. v. d. Hofstad, J. S. H. van Leeuwen, and S. Sen, arXiv:1605.02868 (Preprint (2016)).
- [16] S. Dhara, R. v. d. Hofstad, J. S. H. van Leeuwen, and S. Sen, arXiv:1612.00650 (Preprint (2016)).
- [17] F. Chung and L. Lu, *Adv. in Appl. Math.* **26**, 257 (2001).
- [18] T. Britton, M. Deijfen, and A. Martin-Löf, *Journal of Statistical Physics* **124**, 1377 (2006).
- [19] I. Norros and H. Reittu, *Adv. in Appl. Probab.* **38**, 59 (2006).
- [20] S. Bhamidi, R. v. d. Hofstad, and J. S. H. van Leeuwen, *Electron. J. of Probab.* **15**, 1682 (2010).
- [21] S. Bhamidi, R. v. d. Hofstad, and J. S. H. v. Leeuwen, *Ann. Probab.* **40**, 2299 (2012).
- [22] T. Squartini and D. Garlaschelli, *New Journal of Physics* **13**, 083001 (2011).
- [23] R. v. d. Hofstad, *Random Graphs and Complex Networks* (To be published by Cambridge Univ. Pr., 2016) see <http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf>.
- [24] M. Boguñá, C. Castellano, and R. Pastor-Satorras, *Physical Review E* **79**, 036110 (2009).
- [25] A. J. E. M. Janssen and J. S. H. van Leeuwen, *EPL (Europhysics Letters)* **112**, 68001 (2016).

[26] D. Krioukov, Physical Review Letters **116**, 208302 (2016).

### Appendix A: Proof of Proposition 2

We consider

$$c_{ab}(0) = D_{ab}(\tau) = \frac{\int_a^b \int_a^b f(xy)(xy)^{2-\tau} dx dy}{\left(\int_a^b x^{1-\tau} dx\right)^2}, \quad (\text{A1})$$

where we have written the lower integration limit  $ah_{\min}$  in (24) as  $a$  for notational convenience. We fix  $N$ , and study the dependence of  $D_{ab}(\tau)$  on  $\tau$ . We assume here that  $a$  and  $b$  are fixed, and do not depend on  $\tau$ . Assume that  $f$  is positive and satisfies F2 and F3. We have for  $D'_{ab}(\tau) = \frac{d}{d\tau} D_{ab}(\tau)$ ,

$$D'_{ab}(\tau) = \frac{-\int_a^b \int_a^b f(xy) \ln(xy)(xy)^{2-\tau} dx dy}{\left(\int_a^b x^{1-\tau} dx\right)^2} + \frac{2\int_a^b \int_a^b f(xy)(xy)^{2-\tau} dx dy \int_a^b x^{1-\tau} \ln x dx}{\left(\int_a^b x^{1-\tau} dx\right)^3}. \quad (\text{A2})$$

Observe that  $D'_{ab}(\tau) \leq 0$  if and only if

$$\frac{\int_a^b \int_a^b f(xy) \ln(xy)(xy)^{2-\tau} dx dy}{\int_a^b \int_a^b f(xy)(xy)^{2-\tau} dx dy} \geq 2 \frac{\int_a^b x^{1-\tau} \ln x dx}{\int_a^b x^{1-\tau} dx}. \quad (\text{A3})$$

Symmetry of  $f(xy)/(xy)^{\tau-2}$  and  $\ln(xy) = \ln x + \ln y$  gives

$$\begin{aligned} & \int_a^b \int_a^b f(xy) \ln(xy)(xy)^{2-\tau} dx dy \\ &= 2 \int_a^b \ln x \left( \int_a^b f(xy)(xy)^{2-\tau} dy \right) dx. \end{aligned} \quad (\text{A4})$$

Letting

$$W(x) = \frac{\int_a^b f(xy)(xy)^{2-\tau} dy}{\int_a^b \int_a^b f(vy)(vy)^{2-\tau} dv dy}, \quad a \leq x \leq b, \quad (\text{A5})$$

$$V(x) = \frac{x^{1-\tau}}{\int_a^b v^{1-\tau} dv}, \quad a \leq x \leq b, \quad (\text{A6})$$

we thus need to show that

$$\int_a^b \ln x W(x) dx \geq \int_a^b \ln x V(x) dx. \quad (\text{A7})$$

Observe that

$$x^{\tau-1} \int_a^b f(xy)(xy)^{2-\tau} dy = \int_a^b (xyf(xy)) y^{1-\tau} dy, \quad (\text{A8})$$

which increases in  $x > 0$  when  $f$  satisfies F2. Therefore,  $W(x)/V(x)$  increases in  $x > 0$  when  $f$  satisfies F2. Furthermore,  $\ln x$  increases in  $x > 0$ , so the inequality in (A7) follows from the following lemma:

**Lemma A.1.** *Let  $0 < a < b$  and assume that  $p(x)$  and  $q(x)$  are two positive, continuous probability distribution functions (pdf's) on  $[a, b]$  such that  $p(x)/q(x)$  is increasing in  $x \in [a, b]$ . Let  $g(x)$  be an increasing function of  $x \in [a, b]$ . Then*

$$g_p = \int_a^b g(x)p(x)dx \geq \int_a^b g(x)q(x)dx = g_q. \quad (\text{A9})$$

*Proof.* For any  $R \in \mathbb{R}$ ,

$$g_p - g_q = \int_a^b (g(x) - g_q)(p(x) - Rq(x))dx, \quad (\text{A10})$$

since  $p$  and  $q$  are pdf's. Let  $x_q$  be a point in  $[a, b]$  such that  $g(x) \leq g_q$  when  $x \leq x_q$  and  $g(x) \geq g_q$  when  $x \geq x_q$ . Choose  $R = p(x_q)/q(x_q)$ , so that by monotonicity of  $g$  and  $p/q$ , there are the logical representations

$$\begin{aligned} a \leq x \leq x_q &\Rightarrow (g(x) - g_q) \leq 0 \wedge p(x) - Rq(x) \leq 0, \\ x_q \leq x \leq b &\Rightarrow (g(x) - g_q) \geq 0 \wedge p(x) - Rq(x) \geq 0. \end{aligned}$$

Hence, the integrand in (A10) is everywhere nonnegative, so that  $g_p - g_q \geq 0$  as required.  $\square$

**Remark A.1.** *The following observation will prove useful later: (i) The inequality in (A9) is strict when both  $g(x)$  and  $p(x)/q(x)$  are strictly increasing. (ii) When  $g(x)$  is (strictly) decreasing and  $p(x)/q(x)$  is (strictly) increasing, there is  $\leq$  ( $<$ ) rather than  $\geq$  ( $>$ ) in (A9).*

Now that we have shown F2 to be a sufficient condition for  $D_{ab}(\tau)$  to be increasing, we next show that F2 is also a necessary condition. Suppose we have two points  $u_1$  and  $u_2$  with  $0 < u_1 < u_2$  such that  $u_1 f(u_1) > u_2 f(u_2)$ . Since  $f$  is continuous and piecewise continuous differentiable, there is a  $u_0 \in (u_1, u_2)$  and  $\varepsilon > 0$  such that  $uf(u)$  is strictly decreasing in  $u \in [u_0 - \varepsilon, u_0 + \varepsilon]$ . In (A8), take  $a = \sqrt{u_0 - \varepsilon}$  and  $b = \sqrt{u_0 + \varepsilon}$  so that  $xy \in [u_0 - \varepsilon, u_0 + \varepsilon]$  when  $xy \in [a, b]$ . Therefore,  $W(x)/V(x)$  is strictly decreasing in  $x \in [a, b]$ . By the version of Lemma A.1 with  $g(x) = \ln x$  strictly increasing and  $p(x)/q(x) = W(x)/V(x)$  strictly decreasing, we see that  $g_p - g_q < 0$ . Therefore, we have (A7) with  $<$  instead of  $\geq$ , and so  $D'_{ab}(\tau)$  is positive for all  $\tau$  with this particular choice of  $a$  and  $b$ . This completes the proof of Proposition 2(i).  $\square$

### Appendix B: Proof of Proposition 1

We consider for a fixed  $a, b, \tau$  and  $h > 0$ ,

$$c_{ab}(h) = \frac{\int_a^b \int_a^b (xy)^{2-\tau} f(ahx)f(ahy)f(xy) dx dy}{\left(\int_a^b x^{1-\tau} f(ahx) dx\right)^2}. \quad (\text{B1})$$



Observe that  $\frac{d}{dh}c_{ab}(h) \leq 0$  if and only if

$$\frac{\frac{d}{dh} \left[ \int_a^b \int_a^b (xy)^{2-\tau} f(ahx)f(ahy)f(xy) dx dy \right]}{\int_a^b \int_a^b (xy)^{2-\tau} f(ahx)f(ahy)f(xy) dx dy} \quad (\text{B2})$$

$$\leq 2 \frac{\frac{d}{dh} \left[ \int_a^b f(ahx)x^{1-\tau} dx \right]}{\int_a^b f(ahx)x^{1-\tau} dx}. \quad (\text{B3})$$

Using

$$\frac{d}{dh} [f(ahx)f(ahy)] = axf'(ahx)f(ahy) + ayf(ahx)f'(ahy) \quad (\text{B4})$$

and the symmetry of the function  $f(xy)/(xy)^{\tau-2}$  gives

$$\begin{aligned} \frac{d}{dh} \left[ \int_a^b \int_a^b f(ahx)f(ahy)f(xy)(xy)^{2-\tau} dx dy \right] \\ = 2 \int_a^b \int_a^b axf'(ahx)f(ahy)f(xy)(xy)^{2-\tau} dx dy. \end{aligned} \quad (\text{B5})$$

Also,

$$\frac{d}{dh} \left[ \int_a^b f(ahx)x^{1-\tau} dx \right] = \int_a^b axf'(ahx)x^{1-\tau} dx. \quad (\text{B6})$$

So we write the left-hand side of (B2) as

$$2 \int_a^b \frac{axf'(ahx)}{f(ahx)} T(x) dx, \quad (\text{B7})$$

and the right-hand side of (B2) as

$$2 \int_a^b \frac{axf'(ahx)}{f(ahx)} U(x) dx, \quad (\text{B8})$$

where the pdf's  $T(x)$  and  $U(x)$  on  $[a, b]$  are defined as

$$T(x) = \frac{f(ahx) \int_a^b f(ahy)f(xy)(xy)^{2-\tau} dy}{\int_a^b \int_a^b f(ahv)f(ahy)f(vy)(vy)^{2-\tau} dv dy} \quad (\text{B9})$$

and

$$U(x) = \frac{f(ahx)x^{1-\tau}}{\int_a^b f(ahv)v^{1-\tau} dv}. \quad (\text{B10})$$

The inequality in (B2) thus becomes

$$\int_a^b \frac{-ahxf'(ahx)}{f(ahx)} T(x) dx \geq \int_a^b \frac{-ahxf'(ahx)}{f(ahx)} U(x) dx, \quad (\text{B11})$$

where we have multiplied by  $h > 0$ . Assume that  $f$  satisfies F2. Then

$$\begin{aligned} x^{\tau-1} \int_a^b f(ahy)f(xy)(xy)^{2-\tau} dy \\ = \int_a^b xyf(ahy)f(xy)y^{1-\tau} dy \end{aligned} \quad (\text{B12})$$

is increasing in  $x > 0$ . Therefore, see (B9) and (B10),  $T(x)/U(x)$  is increasing in  $x > 0$ . Hence, from Lemma A.1 we get (B11) when  $g(x) = -ahxf'(ahx)/f(ahx)$  is increasing in  $x > 0$ , i.e. when  $f$  satisfies F4.

We have now shown that when  $f$  satisfies F1-F3, the condition F4 is sufficient for  $c_{ab}(h)$  to be decreasing in  $h > 0$ . For the result in the converse direction we argue as follows. The function  $uf(u)$  is continuous, piecewise smooth, increasing and not constant, and so there is a  $u_0 > 0$ ,  $\varepsilon > 0$  such that  $uf(u)$  is strictly increasing in  $u \in [u_0 - \varepsilon, u_0 + \varepsilon]$ . Let  $z(v) = -vf'(v)/f(v)$ , and assume there are  $0 < v_1 < v_2$  such that  $z(v_1) > z(v_2)$ . We may assume that  $z$  is continuous at  $v = v_1, v_2$ . Indeed, when  $z$  is discontinuous at  $v_1$  say,  $z(v_1) = \frac{1}{2}(z(v_1 + 0) + z(v_2 - 0))$  and so at least one of  $z(v_1 - 0) = \lim_{v \uparrow v_1} z(v)$  and  $z(v_1 + 0) = \lim_{v \downarrow v_1} z(v)$  is larger than  $z(v_2)$ . Since  $z$  has only finitely many discontinuities, it suffices to decrease or increase  $v_1$  somewhat, to a point of continuity, while maintaining  $z(v_1) > z(v_2)$ . We have to consider two cases.

A. Assume that  $z(v)$  is continuous on  $[v_1, v_2]$ . We can then basically argue as in the proof of the only-if part of Proposition 2. Thus, there is a  $v_0 > 0$ ,  $\delta > 0$  such that  $z(v)$  is strictly decreasing in  $v \in [v_0 - \delta, v_0 + \delta]$ . We choose  $a, b$  such that  $xy \in [u_0 - \varepsilon, u_0 + \varepsilon]$  when  $x, y \in [a, b]$ . This is satisfied when  $(u_0 - \varepsilon)^{1/2} \leq a < b \leq (u_0 + \varepsilon)^{1/2}$ , and it guarantees that  $T(x)/U(x)$  is strictly increasing in  $x \in [a, b]$ . Next, we choose  $h$  such that  $ahx \in [v_0 - \delta, v_0 + \delta]$  when  $x \in [a, b]$ , so that  $z(ahx)$  is strictly decreasing in  $x \in [a, b]$ . For this, we need to take  $h$  such that  $a^2h \geq v_0 - \delta$  and  $abh \leq v_0 + \delta$ . This can be done indeed when  $a/b \geq (v_0 - \delta)/(v_0 + \delta)$ . Choosing  $a$  and  $b$  with  $a < b$ ,  $a, b \in [(u_0 - \varepsilon)^{1/2}, (u_0 + \varepsilon)^{1/2}]$  such that this latter condition is satisfied, we can apply the version of Lemma A.1 with strictly decreasing  $g(x) = z(ahx)$  and strictly increasing  $p(x)/q(x) = T(x)/U(x)$ . Thus we get in (B11) strict inequality  $<$  for these  $a, b$  and  $h$ , and this means that  $c'_{ab}(h) < 0$ . This proves Proposition 1 for this case.

B. Assume that  $z(v)$  has discontinuities on  $[v_1, v_2]$ , say at  $c_1 < c_2 < \dots < c_j$  with  $v_1 < c_1$  and  $v_2 > c_j$ . In the case that there is an interval  $[v_0 - \delta, v_0 + \delta]$  contained in one of  $(v_1, c_1), (c_1, c_2), \dots, (c_j, v_2)$  where  $z$  is strictly decreasing, we are in the position of case A, and then we are done. Otherwise, we have by F3 that  $z'(v) \geq 0$  for all  $v \in [v_1, v_2]$ ,  $v \neq c_1, \dots, c_j$ . Then we must have  $z(v_0 - 0) > z(v_0 + 0)$  for at least one  $v_0 = c_1, \dots, c_j$ , for else we would have  $z(v_1) \leq z(c_1 - 0) \leq z(c_1 + 0) \leq \dots \leq z(c_j - 0) \leq z(c_j + 0) \leq z(v_2)$ .

We want to find  $a, b$  such that

$$\int_a^b z(ahx)T(x) dx < \int_a^b z(ahx)U(x) dx \quad (\text{B13})$$

for the case that  $z(v)$  has a downward jump at  $v = v_0 > 0$  while being increasing to the left and to the right of  $v_0$ . Set

$$\Delta = z(v_0 - 0) - z(v_0 + 0), \quad M = \frac{1}{2}(z(v_0 - 0) + z(v_0 + 0)), \quad (\text{B14})$$

and observe that  $M \geq \frac{1}{2}\Delta > 0$  since  $z(v) \geq 0$  for all  $v$ . We can find  $\delta > 0$  such that

$$z(v_0 - 0) \geq z(v) \geq z(v_0 - 0) - \frac{1}{8}\Delta, \quad v_0 - \delta \leq v < v_0 \quad (\text{B15})$$

$$z(v_0 + 0) \leq z(v) \leq z(v_0 + 0) + \frac{1}{8}\Delta, \quad v_0 < v \leq v_0 + \delta. \quad (\text{B16})$$

Next, let

$$l(v) = f(v)v^{1-\tau}, \quad v > 0, \quad (\text{B17})$$

and observe that  $l(v)$  is positive and continuous at  $v = v_0$ . Hence, we can choose  $\delta > 0$  such that, in addition to (B15) and (B16),

$$\left| \frac{l(v)}{l(v_0)} - 1 \right| \leq \lambda, \quad v \in [v_0 - \delta, v_0 + \delta], \quad (\text{B18})$$

where  $\lambda$  is any number between 0 and  $\frac{5}{16}\Delta/(2M + \frac{7}{16}\Delta)$ . As in case A of the proof, we choose  $a, b$  and  $h$  such that

$$xy \in [u_0 - \varepsilon, u_0 + \varepsilon], \quad (\text{B19})$$

when  $x, y \in [a, b]$  and

$$ahx \in [v_0 - \delta, v_0 + \delta], \quad (\text{B20})$$

when  $x \in [a, b]$ . Thus, we let  $(u_0 - \varepsilon)^{1/2} \leq a < b \leq (u_0 + \varepsilon)^{1/2}$  such that  $1 > a/b \geq (v_0 - \delta)/(v_0 + \delta)$ . Below, we shall transform the two integrals by the substitution  $v = ah_0x$  for a special choice of  $h = h_0$  to an integral over an interval  $[w_1, w_2]$  having  $v_0$  as midpoint. This  $h_0$  is given by

$$h_0 = \frac{2v_0}{a^2 + ab} \in \left[ \frac{v_0 - \delta}{a^2}, \frac{v_0 + \delta}{ab} \right]. \quad (\text{B21})$$

Indeed, this  $h_0$  satisfies (B20) since

$$\begin{aligned} \frac{2v_0}{a^2 + ab} \leq \frac{v_0 + \delta}{ab} &\iff 2bv_0 \leq (b+a)(v_0 + \delta) \\ &\iff \left(1 - \frac{a}{b}\right)v_0 \leq \left(1 + \frac{a}{b}\right)\delta \\ &\iff \frac{a}{b} \geq \frac{v_0 - \delta}{v_0 + \delta}, \end{aligned} \quad (\text{B22})$$

and

$$\begin{aligned} \frac{2v_0}{a^2 + ab} \geq \frac{v_0 - \delta}{ab} &\iff 2av_0 \leq (b+a)(v_0 - \delta) \\ &\iff \left(\frac{a}{b} - 1\right)v_0 \geq -\left(1 + \frac{a}{b}\right)\delta \\ &\iff \frac{a}{b} \geq \frac{v_0 - \delta}{v_0 + \delta}. \end{aligned} \quad (\text{B23})$$

In the integrals in the inequality in (B13) with  $h = h_0$ , we substitute  $ah_0x = v$ , and the inequality to be proved becomes

$$z_t := \int_{w_1}^{w_2} z(v)t(v)dv < \int_{w_1}^{w_2} z(v)u(v)dv =: z_u. \quad (\text{B24})$$

Here

$$w_1 = a^2h_0, \quad w_2 = abh_0 \quad (\text{B25})$$

so that  $v_0 = \frac{1}{2}(a^2 + ab)h_0$  is the midpoint of the integration interval  $[w_1, w_2] \subset [v_0 - \delta, v_0 + \delta]$ , and  $t(v)$  and  $u(v)$  are the pdf's

$$t(v) = \frac{1}{ah_0}T\left(\frac{v}{ah_0}\right), \quad u(v) = \frac{1}{ah_0}U\left(\frac{v}{ah_0}\right) \quad (\text{B26})$$

for which  $t(v)/u(v)$  is strictly increasing in  $v \in [w_1, w_2]$ . We shall show that  $z_u \in (M - \frac{3}{8}\Delta, M + \frac{3}{8}\Delta)$ , and so, by (B15) and (B16),

$$z(v) - z_u > 0, \quad w_1 \leq v < v_0; \quad z(v) - z_u < 0, \quad v_0 < v \leq w_2. \quad (\text{B27})$$

With  $R = t(v_0)/u(v_0)$ , this implies that

$$z_t - z_u = \int_{w_1}^{w_2} (z(v) - z_u)(t(v) - Ru(v))dv < 0, \quad (\text{B28})$$

since the integrand is negative for all  $v \neq v_0$ .

To show that  $z_u \in (M - \frac{3}{8}\Delta, M + \frac{3}{8}\Delta)$ , we note that the pdf  $u(v)$  is built from the function  $l(v)$  in (B17) via (A10) and (B26). In terms of this  $l(v)$  we can write  $z_u$  as

$$z_u = \frac{\int_{w_1}^{w_2} z(v)l(v)dv}{\int_{w_1}^{w_2} l(v)dv}. \quad (\text{B29})$$

Now, by (B18),

$$(w_2 - w_1)l(v_0)(1 - \lambda) \leq \int_{w_1}^{w_2} l(v)dv \leq (w_2 - w_1)(1 + \lambda)l(v_0). \quad (\text{B30})$$

Also, by (B15), (B16) and (B18) and the fact that  $v_0 = \frac{1}{2}(w_1 + w_2)$ ,

$$\begin{aligned} \int_{w_1}^{w_2} z(v)l(v)dv &= \int_w^{v_0} z(v)l(v)dv + \int_{v_0}^z z(v)l(v)dv \\ &\leq z(v_0 - 0) \int_w^{v_0} l(v)dv \\ &\quad + z(v_0 + \frac{1}{8}\Delta) \int_{v_0}^z l(v)dv \\ &\leq \frac{1}{2}(w_2 - w_1)(1 + \lambda)l(v_0) \\ &\quad \cdot [z(v_0 - 0) + z(v_0 + 0) + \frac{1}{8}\Delta] \\ &= (w_2 - w_1)(1 + \lambda)l(v_0) \left[ M + \frac{1}{16}\Delta \right], \end{aligned} \quad (\text{B31})$$

and in a similar fashion

$$\int_{w_1}^{w_2} z(v)l(v)dv \geq (w_2 - w_1)(1 - \lambda)l(v_0) \left[ M - \frac{1}{16}\Delta \right]. \quad (\text{B32})$$

From (B30), (B31) and (B32) we then get

$$\frac{1 - \lambda}{1 + \lambda} \left( M - \frac{1}{16}\Delta \right) \leq z_u \leq \frac{1 + \lambda}{1 - \lambda} \left( M + \frac{1}{16}\Delta \right). \quad (\text{B33})$$

Now

$$\frac{1+\lambda}{1-\lambda}(M+\frac{1}{16}\Delta) < M+\frac{3}{8}\Delta \iff \lambda < \frac{\frac{5}{16}\Delta}{2M+\frac{7}{16}\Delta},$$

$$\frac{1-\lambda}{1+\lambda}(M-\frac{1}{16}\Delta) < M+\frac{3}{8}\Delta \iff \lambda < \frac{\frac{5}{16}\Delta}{2M-\frac{7}{16}\Delta}.$$

Then it follows from the choice of  $\lambda$  in (B18) that  $z_u \in (M-\frac{3}{8}\Delta, M+\frac{3}{8}\Delta)$  for such  $\lambda$ .

### Appendix C: Proof of Proposition 3

Taking the limit  $h \downarrow 0$  in (22), with  $r(u) = u \min(1, 1/u)$ , we have

$$c_{ab}^{\max}(0) = \frac{\int_a^b \int_a^b (xy)^{2-\tau} \min(1, (xy)^{-1}) dx dy}{\left[ \int_a^b x^{1-\tau} dx \right]^2}, \quad (C1)$$

where we have written  $a$  instead of  $ah_{\min}$  for ease of notation. The denominator in (C1) is evaluated as

$$\left( \int_a^b x^{1-\tau} dx \right)^2 = \frac{1}{(\tau-2)^2} (a^{2-\tau} - b^{2-\tau})^2. \quad (C2)$$

For the numerator in (C1) we compute

$$\begin{aligned} & \int_a^b \int_a^b \min(1, (xy)^{-1}) (xy)^{2-\tau} dx dy \\ &= \int_a^{1/b} \int_a^b (xy)^{2-\tau} dy dx \\ & \quad + \int_{1/b}^b \left( \int_a^{1/x} (xy)^{2-\tau} dy + \int_{1/x}^b (xy)^{1-\tau} dy \right) dx \\ &= \int_a^{1/b} x^{2-\tau} dx \int_a^b y^{2-\tau} dy + \int_{1/b}^b x^{2-\tau} \int_a^{1/x} y^{2-\tau} dy dx \\ & \quad + \int_{1/b}^b x^{1-\tau} \int_{1/x}^b y^{1-\tau} dy dx \\ &= \frac{(b^{\tau-3} - a^{3-\tau})(b^{3-\tau} - a^{3-\tau})}{(3-\tau)^2} \\ & \quad + \frac{1}{3-\tau} \int_{1/b}^b x^{2-\tau} (x^{\tau-3} - a^{3-\tau}) dx \\ & \quad + \frac{1}{2-\tau} \int_{1/b}^b x^{1-\tau} (b^{2-\tau} - x^{\tau-2}) dx \\ &= \frac{(b^{\tau-3} - a^{3-\tau})(b^{3-\tau} - a^{3-\tau})}{(3-\tau)^2} \\ & \quad + \frac{1}{3-\tau} \left( \ln(b^2) - \frac{a^{3-\tau}(b^{3-\tau} - b^{\tau-3})}{3-\tau} \right) \\ & \quad + \frac{1}{2-\tau} \left( \frac{b^{2-\tau}(b^{2-\tau} - b^{\tau-2})}{2-\tau} - \ln(b^2) \right) \\ &= \frac{\ln(b^2)}{(3-\tau)(\tau-2)} - \frac{1-b^{2(2-\tau)}}{(2-\tau)^2} + \frac{1-2(ab)^{3-\tau}+a^{2(3-\tau)}}{(3-\tau)^2}. \end{aligned} \quad (C3)$$

The last member of (C3) equals  $I_{ab}^{\max}(\tau)$  in (25), and the result follows from (23), (C1), (C2) and (C3).  $\square$

### Appendix D: Proof of Proposition 4

Take  $u_0 \geq 1$  and note that

$$f(u) \geq u_0 f(u_0) \min(u_0^{-1}, u^{-1}), \quad u \geq 0, \quad (D1)$$

since, for  $f \in \mathbb{F}$ ,

$$f(u) \geq f(u_0), \quad 0 \leq u \leq u_0; \quad uf(u) \geq u_0 f(u_0), \quad u \geq u_0. \quad (D2)$$

Now for any  $c > 0$ ,

$$\begin{aligned} & \int_a^b \int_a^b \min(c, (xy)^{-1}) (xy)^{2-\tau} dx dy \\ &= c^{\tau-2} \int_{a\sqrt{c}}^{b\sqrt{c}} \int_{a\sqrt{c}}^{b\sqrt{c}} \min(1, (xy)^{-1}) (xy)^{2-\tau} dx dy \\ &= c^{\tau-2} I_{\max; a\sqrt{c}, b\sqrt{c}}(\tau). \end{aligned} \quad (D3)$$

Also,

$$\frac{(\tau-2)^2}{(a^{2-\tau} - b^{2-\tau})^2} = c^{\tau-2} \frac{(\tau-2)^2}{((a\sqrt{c})^{2-\tau} - (b\sqrt{c})^{2-\tau})^2}. \quad (D4)$$

The result then follows from combining (D1), (D3) and (D4).  $\square$

### Appendix E: Derivation of Equation (30)

We shall derive (30) assuming (21) and that  $|\ln(ab)/\ln(b^2)|$  is small. In the present case, where  $a$  and  $b$  are given through (10), (11) and (20) with  $h_{\min} = 1$ , this indeed holds since  $|\ln(ab)/\ln(b^2)| = (\tau-2)/(3-\tau)$ . With  $s = \tau - 2$  we consider

$$\begin{aligned} C_{ab}^{\max}(\tau) &= \frac{s^2}{(a^{-s} - b^{-s})^2} \left[ \frac{\ln(b^2)}{s(1-s)} - \frac{1-b^{-2s}}{s^2} \right. \\ & \quad \left. + \frac{1-2(ab)^{1-s}+a^{2(1-s)}}{(1-s)^2} \right], \end{aligned} \quad (E1)$$

where we have written  $a$  instead of  $ah_{\min}$  for notational convenience. We develop, assuming  $s \ln(b^2)$  of order unity and less,

$$\begin{aligned} & \frac{\ln(b^2)}{s(1-s)} - \frac{1-b^{-2s}}{s^2} + \frac{1-2(ab)^{1-s}+a^{2(1-s)}}{(1-s)^2} \\ &= \frac{1}{2} \ln^2 b^2 \left( 1 - \frac{1}{3} s \ln(b^2) + \dots + O((\ln^2 b^2)^{-1}) \right). \end{aligned} \quad (E2)$$

Also,

$$\begin{aligned} \frac{a^{-s} - b^{-s}}{s} &= \ln(b/a) \left[ 1 - \frac{1}{2} s \ln(ab) \right. \\ & \quad \left. + \frac{1}{6} s^2 (\ln^2 b + \ln b \ln a + \ln^2 a) - \dots \right]. \end{aligned} \quad (E3)$$

Note that

$$\ln(b/a) = \ln(b^2) - \ln(ab) = \ln(b^2) \left(1 - \frac{\ln(ab)}{\ln(b^2)}\right), \quad (\text{E4})$$

$$\ln^2 b + \ln b \ln a + \ln^2 a = \ln^2 b \left(1 - \frac{\ln(ab)}{\ln b} + \left(\frac{\ln(ab)}{\ln b}\right)^2\right). \quad (\text{E5})$$

Thus we get

$$\begin{aligned} \frac{a^{-s} - b^{-s}}{s} &= \ln(b^2) \left(1 - \frac{\ln(ab)}{\ln(b^2)}\right) \left(1 - \frac{1}{2}s \ln(ab)\right) \\ &\quad + \frac{1}{6}s^2 \ln^2 b \left(1 + O\left(\frac{\ln(ab)}{\ln b}\right)\right) \\ &\approx \ln(b^2) \left[1 - \frac{1}{2}s \ln(ab) + \frac{1}{6}s^2 \ln^2 b\right], \quad (\text{E6}) \end{aligned}$$

where we have used the assumption that  $|\ln(ab)/\ln b|$  is small.

When we insert (E2) and (E6) into (E1) and divide through  $\ln^2(b^2)$ , we arrive at (30).

### Appendix F: Maximally random graph

Define Lerch's transcendent

$$\Phi(z, s, v) = \sum_{k=0}^{\infty} \frac{z^k}{(k+v)^s}. \quad (\text{F1})$$

In [9] it was shown that for the maximally random graph (8)

$$\begin{aligned} C_{ab}(\tau) &= \frac{(\tau-2)^2}{(a^{2-\tau} - b^{2-\tau})^2} \left\{ \frac{\pi \ln(b^2)}{\sin \pi(\tau-2)} - \frac{\pi^2 \cos \pi(\tau-2)}{(\sin \pi(\tau-2))^2} \right. \\ &\quad + b^{-2(\tau-2)} \Phi(-b^{-2}, 2, \tau-2) \\ &\quad + a^{2(3-\tau)} \Phi(-a^2, 2, 3-\tau) \\ &\quad \left. - 2(ab)^{3-\tau} \Phi(-ab, 2, 3-\tau) \right\}. \quad (\text{F2}) \end{aligned}$$

(The expression is slightly simplified compared to [9, Eq. (5)].) Comparing (F2) and (24) shows that the front factor is identical, and that the terms in between brackets differ. Table II compares the two dominant terms in (F2) and (24) and shows that these terms are of comparable magnitude for  $\tau-2$  small.

### Appendix G: Monotonicity properties for $C_{ab}(\tau)$

In this appendix we show that  $C_{ab}(\tau)$  is bounded from above by a closely related function that decreases in  $\tau$ . Notice that Proposition 2 assumes  $a$  and  $b$  fixed. We have from (22) and (10), (11) that

$$a = a(\tau) = \left(\frac{1}{N\langle h \rangle}\right)^{1/2}, \quad b = b(\tau) = (N\langle h \rangle)^{\frac{3-\tau}{2(\tau-1)}} \quad (\text{G1})$$

$s$	$\frac{\pi}{\sin \pi s}$	$\frac{1}{s(1-s)}$	$\frac{\pi^2 \cos \pi s}{(\sin \pi s)^2}$	$\frac{1}{s^2} - \frac{1}{(1-s)^2}$
0.1	10.1664	11.1111	98.2972	98.7654
0.2	5.3448	6.2500	23.1111	23.4375
0.3	3.8832	4.7619	8.8635	9.0703
0.4	3.3033	4.1666	3.3719	3.4722
0.5	3.1416	4.0000	0.0000	0.0000

Table II. Dominant terms in (F2) and (24) for several values of  $s = \tau - 2$ .

where  $\langle h \rangle$  is given in (9). Below we shall use that  $\langle h \rangle$  decreases in  $\tau \in (2, 3]$ ; this is clear intuitively and can be proved rigorously by using Lemma A.1. We have

$$C_{ab}(\tau) = A(\tau)G(\tau, a(\tau), b(\tau)), \quad (\text{G2})$$

where

$$A(\tau) = \int_{h_{\min}}^N \rho(h)(1 - (1+h)e^{-h})dh, \quad (\text{G3})$$

with density  $\rho(h) = Ch^{-\tau}$  on  $[h_{\min}, N]$  and

$$G(\tau, a, b) = \frac{\int_{ah_{\min}}^b \int_{ah_{\min}}^b (xy)^{2-\tau} f(xy) dx dy}{\left(\int_{ah_{\min}}^b x^{1-\tau} dx\right)^2}. \quad (\text{G4})$$

Proposition 2 says that

- (i)  $G$  decreases in  $\tau$  (when  $a$  and  $b$  are fixed).

With the method of the proof of Proposition 2 in Appendix A, we will show that

- (ii)  $A$  decreases in  $\tau$ ,
- (iii)  $G$  increases in  $a$  and in  $b$ .

Showing that  $G(\tau, a(\tau), b(\tau))$  decreases in  $\tau$  is complicated by the facts that  $a(\tau)$  increases in  $\tau$ , see (iii), and that the dependence of  $a(\tau)$  and  $b(\tau)$  on  $\tau$  is rather involved. Let  $m$  and  $M$  be the minimum and maximum, respectively, of  $\langle h \rangle$  when  $\tau \in [2, 3]$  ( $m = \langle h \rangle|_{h=3} \approx 2h_{\min}$ ,  $M = \langle h \rangle|_{\tau \downarrow 2} \approx h_{\min} \ln(N/h_{\min})$  from (9) and the monotonicity of  $\langle h \rangle$ ). Letting

$$\bar{a} := (Nm)^{-1/2}, \quad \bar{b}(\tau) = (NM)^{\frac{3-\tau}{2(\tau-1)}}, \quad (\text{G5})$$

we have  $a(\tau) \leq \bar{a}$ ,  $b(\tau) \leq \bar{b}(\tau)$ , and so by (iii)

$$G(\tau, a(\tau), b(\tau)) \leq G(\tau, \bar{a}, \bar{b}(\tau)). \quad (\text{G6})$$

The right-hand side of (G6) decreases in  $\tau$  by (i) and (iii) and the fact that  $\bar{b}(\tau)$  decreases in  $\tau$ . Therefore,  $C_{ab}(\tau)$  in (G2) is bounded above by a closely related function that does decrease in  $\tau$ .

We have shown that  $G(\tau, a, b)$  decreases in  $\tau \in [2, 3]$ . We shall show now that

- (ii)  $A$  decreases in  $\tau$  and increases in  $l$ ,

(iii)  $G$  increases in  $a$  and in  $b$ .

*Proof that  $A$  decreases in  $\tau$ .* Since  $\frac{d}{d\tau}h^{-\tau} = -h^{-\tau} \ln h$ , we have

$$\begin{aligned} \frac{\partial A}{\partial \tau} &\leq 0 \\ &\iff \int_{h_{\min}}^N -h^{-\tau}(1 - (1+h)e^{-h}) \ln h dh \int_{h_{\min}}^N h^{-\tau} dh \\ &\quad - \int_{h_{\min}}^N h^{-\tau}(1 - (1+h)e^{-h}) dh \int_{h_{\min}}^N -hh^{-\tau} \ln h dh \leq 0 \\ &\iff \frac{\int_{h_{\min}}^N h^{-\tau}(1 - (1+h)e^{-h}) \ln h dh}{\int_{h_{\min}}^N h^{-\tau}(1 - (1+h)e^{-h}) dh} \\ &\quad \geq \frac{\int_{h_{\min}}^N h^{-\tau} \ln h dh}{\int_{h_{\min}}^N h^{-\tau} dh}. \end{aligned} \quad (\text{G7})$$

Consider on  $[h_{\min}, N]$  the pdf's

$$p(h) = \frac{h^{-\tau}(1 - (1+h)e^{-h})}{\int_{h_{\min}}^N h_1^{-\tau}(1 - (1+h_1)e^{-h_1}) dh_1}, \quad (\text{G8})$$

$$q(h) = \frac{h^{-\tau}}{\int_{h_{\min}}^N h_1^{-\tau} dh_1} = \rho(h). \quad (\text{G9})$$

Clearly  $p(h)/q(h) = C(1 - (1+h)e^{-h})$  with  $C$  independent of  $h \in [h_{\min}, N]$ . Hence,  $p(h)/q(h)$  is increasing in  $[h_{\min}, N]$ . Also,  $\ln(h) = g(h)$  is increasing in  $[h_{\min}, N]$ . Hence, by Lemma A.1,

$$\int_{h_{\min}}^N g(h)p(h)dh \geq \int_{h_{\min}}^N g(h)q(h)dh, \quad (\text{G10})$$

and this is the last inequality in (G7).  $\square$

*Proof that  $G$  increases in  $b$ .* Again, for notational simplicity, we write  $a$  and  $b$  instead of  $ah_{\min}$  and  $bh_{\min}$  respectively. Let  $\tau$  and  $a$  be fixed, and set

$$p(x, y) = (xy)^{2-\tau} f(xy), \quad P(b, y) = \int_a^b p(x, y) dx. \quad (\text{G11})$$

We have

$$\begin{aligned} \frac{d}{db} \left[ \int_a^b \int_a^b (xy)^{2-\tau} f(xy) dx dy \right] &= \frac{d}{db} \left[ \int_a^b P(b, y) dy \right] \\ &= P(b, b) + \int_a^b \frac{\partial P}{\partial b}(b, y) dy \\ &= \int_a^b p(x, b) dx + \int_a^b p(b, y) dy = 2 \int_a^b p(x, b) dx \\ &= 2 \int_a^b (xb)^{2-\tau} f(xb) dx \end{aligned} \quad (\text{G12})$$

because of symmetry of  $p(x, y)$ . Also,

$$\frac{d}{db} \left( \int_a^b x^{1-\tau} dx \right)^2 = 2b^{1-\tau} \int_a^b x^{1-\tau} dx. \quad (\text{G13})$$

Therefore,

$$\begin{aligned} \frac{\partial G}{\partial b} &\geq 0 \\ &\iff 2 \int_a^b (xb)^{2-\tau} f(xb) dx \left( \int_a^b x^{1-\tau} dx \right)^2 \\ &\quad - 2 \int_a^b \int_a^b (xy)^{2-\tau} f(xy) dx dy b^{1-\tau} b^{1-\tau} \int_a^b x^{1-\tau} dx \geq 0 \\ &\iff \frac{\int_a^b (xb)^{2-\tau} f(xb) dx}{\int_a^b \int_a^b (xy)^{2-\tau} f(xy) dx dy} \geq \frac{b^{1-\tau}}{\int_a^b x^{1-\tau} dx} \\ &\iff W(b) \geq V(b), \end{aligned} \quad (\text{G14})$$

where  $W(x)$  and  $V(x)$  are the pdf's as defined in (A5) and (A6). Since  $W(x)/V(x)$  increases in  $x \in [a, b]$ , we get

$$\begin{aligned} 1 &= \int_a^b W(x) dx = \int_a^b \frac{W(x)}{V(x)} V(x) dx \\ &\leq \frac{W(b)}{V(b)} \int_a^b V(x) dx = \frac{W(b)}{V(b)}, \end{aligned} \quad (\text{G15})$$

and this gives the last inequality in (G14).  $\square$

*Proof that  $G$  increases in  $a$ .* This proof is very similar to the proof that  $G$  increases in  $b$ . Let  $\tau$  and  $b$  be fixed. We now have

$$\frac{d}{da} \left[ \int_a^b \int_a^b (xy)^{2-\tau} f(xy) dx dy \right] = -2 \int_a^b (xa)^{2-\tau} f(xa) dx, \quad (\text{G16})$$

and

$$\frac{d}{da} \left( \int_a^b x^{1-\tau} dx \right)^2 = -2a^{1-\tau} \int_a^b x^{1-\tau} dx. \quad (\text{G17})$$

Then, as in (G14),

$$\begin{aligned} \frac{\partial G}{\partial a} \geq 0 &\iff \frac{\int_a^b (xa)^{2-\tau} f(xa) dx}{\int_a^b \int_a^b (xy)^{2-\tau} f(xy) dx dy} \leq \frac{a^{1-\tau}}{\int_a^b x^{1-\tau} dx} \\ &\iff W(a) \leq V(a), \end{aligned} \quad (\text{G18})$$

and the inequality follows again from increasingness of  $W(x)/V(x)$ .  $\square$

## Appendix H: Choice of the natural cutoff

We consider random variables  $h_i$ ,  $i = 1, \dots, N$ , that are i.i.d. with density  $p(h) = Ch^{-\tau}$ ,  $h_{\min} \leq h < \infty$ . We have

$$\begin{aligned} \mathbb{E}[\max_i h_i] &= h_{\min} \Gamma \left( \frac{\tau - 2}{\tau - 1} \right) \frac{\Gamma(N + 1)}{\Gamma \left( N + \frac{\tau - 2}{\tau - 1} \right)} \\ &\approx h_{\min} \Gamma \left( \frac{\tau - 2}{\tau - 1} \right) N^{\frac{1}{\tau - 1}}. \end{aligned} \quad (\text{H1})$$

The first identity in (H1) is exact, and follows from

$$\begin{aligned}\mathbb{E}[\max_i h_i] &= \int_{h_{\min}}^{\infty} h d [\mathbb{P}^N(\underline{h} \leq h)] \\ &= \int_{h_{\min}}^{\infty} h d [(1 - \mathbb{P}(\underline{h} > h))^N] \\ &= \int_{h_{\min}}^{\infty} h d \left[ \left( 1 - \left( \frac{h}{h_{\min}} \right)^{1-\tau} \right)^N \right], \quad (\text{H2})\end{aligned}$$

using the substitution  $t = (h/h_{\min})^{1-\tau} \in (0, 1]$  and the expression of the  $B$ -function in terms of the  $\Gamma$  function. The approximate identity in (H1) follows from  $\Gamma(n+a)/\Gamma(n+b) \approx n^{a-b}$ , which is quite accurate when  $a, b \in [0, 1]$  and  $n$  large.

Next, there is for  $2 \leq \tau \leq 3$  the inequality

$$\left( \frac{\tau-1}{\tau-2} \right)^{\frac{1}{\tau-1}} \leq \Gamma \left( \frac{\tau-2}{\tau-1} \right) \leq \frac{4}{3} \left( \frac{\tau-1}{\tau-2} \right)^{\frac{1}{\tau-1}}. \quad (\text{H3})$$

This inequality follows from

$$u^u \leq \Gamma(1+u) = u\Gamma(u) \leq \frac{4}{3}u^u, \quad (\text{H4})$$

with  $u = (\tau-2)/(\tau-1) \in (0, \frac{1}{2}]$  that can be shown by considering the concave function  $\ln(\Gamma(1+u)) - u \ln u$  which vanishes at  $u = 0, 1$  and is positive at  $u = \frac{1}{2}$  (the upper bound in (H4) follows from a numerical inspection of this function).

Then from (H3) we get, using  $\langle h \rangle = h_{\min}(\tau-1)/(\tau-2)$ ,

$$\begin{aligned}h_{\min}^{\frac{\tau-2}{\tau-1}} (N\langle h \rangle)^{\frac{1}{\tau-1}} &\leq h_{\min} \Gamma \left( \frac{\tau-2}{\tau-1} \right) N^{\frac{1}{\tau-1}} \\ &\leq \frac{4}{3} h_{\min}^{\frac{\tau-2}{\tau-1}} (N\langle h \rangle)^{\frac{1}{\tau-1}}, \quad (\text{H5})\end{aligned}$$

showing that the order of magnitude of  $\mathbb{E}[\max_i h_i]$  is  $(N\langle h \rangle)^{1/(\tau-1)}$ . This motivates our choice of  $h_c$  in (11).