

# Bootstrap estimators for the tail-index and for the count statistics of graphex processes

**Zacharie Naulet**

*Université Paris-Saclay, Laboratoire de Mathématiques d'Orsay  
91405, Orsay, France  
e-mail: [zacharie.naulet@universite-paris-saclay.fr](mailto:zacharie.naulet@universite-paris-saclay.fr)*

**Daniel M Roy**

*Department of Statistics, University of Toronto; Vector Institute  
University Avenue, Toronto, Ontario, Canada  
e-mail: [daniel.roy@utoronto.ca](mailto:daniel.roy@utoronto.ca)*

**Ekansh Sharma**

*Department of Computer Science, University of Toronto; Vector Institute  
6 King's College Rd, Toronto M5S 3H3, Ontario, Canada  
e-mail: [ekansh@cs.toronto.edu](mailto:ekansh@cs.toronto.edu)*

**Victor Veitch**

*Department of Statistics, University of Chicago  
5747 S Ellis Ave, Chicago, IL 60637  
e-mail: [victorveitch@gmail.com](mailto:victorveitch@gmail.com)*

**Abstract:** Graphex processes resolve some pathologies in traditional random graph models, notably, providing models that are both projective and allow sparsity. Most of the literature on graphex processes study them from a probabilistic point of view. Techniques for inferring the parameter of these processes – the so-called *graphon* – are still marginal; exceptions are a few papers considering parametric families of graphons. Nonparametric estimation remains unconsidered. In this paper, we propose estimators for a selected choice of functionals of the graphon. Our estimators originate from the subsampling theory for graphex processes, hence can be seen as a form of bootstrap procedure.

**MSC2020 subject classifications:** Primary 62F10; secondary 60G55, 60G70.

**Keywords and phrases:** Graphex processes, sparse random graphs, tail-index, estimation, count statistics, bootstrap.

Received April 2019.

## 1. Introduction

Many statistical models for relational data are based on vertex-exchangeable random graphs (also called *graphon* models), see [45] for a review. These models are interpretable and easy to work with, but have the considerable limitation

that, almost surely, they generate dense graphs; that is, the number of edges scales quadratically with the number of vertices as the graph grows. Folk wisdom holds that real-world datasets are sparse and their degree distributions are power-law<sup>1</sup> implying that graphon based approaches are unsatisfactory. Several model classes have been proposed to address this issue. A non exhaustive list includes *edge exchangeable* models [17], *preferential attachment* models [4], *sparse graphon models* [6, 7], and *graphex processes* [13, 52, 8].

Graphex processes have been proposed as a generalization of the graphon models that preserve the key properties, but allow for sparsity and degree distribution with power-law behaviour [13, 52, 14]. Thus, graphex processes are good candidates to model *empirical* graphs. Apart of [13, 51, 53, 29], inference on graphex processes is still marginal in the literature, most of papers studying graphex processes from a *probabilistic* point of view only. Efficient estimation remains challenging for these new models. The present paper intend to contribute to the study of graphex processes from a *statistical* viewpoint, driven by two main motivations described in section 2.

On a high-level, a graphex process  $\{G_t : t \geq 0\}$  is a *graph-valued* stochastic process. They are used as probabilistic models for graph observations, which are seen as the realization of a sample  $G_t$  of the process at a finite but unknown *size*  $t$ . All graphs will be simple, and will always be considered to be finite, except in rare occasions, in which case they will be clearly stated to be infinite. The law of the process  $\{G_t : t \geq 0\}$  is determined by a jointly measurable symmetric function  $W : \mathbb{R}_+^2 \rightarrow [0, 1]$ , called the *graphon* [13, 52]. Our goal is to do nonparametric inference about some functionals of  $W$ . The functionals we consider are the *tail-index* of the process (see section 4) and the *rescaled injective homomorphism densities* (see section 5). The motivations for these functionals are detailed in section 2, and general definitions about graphex processes are recalled in section 3. Along the way, we use our estimators to construct statistical tests for certain counts of motifs. We propose some applications of our results in section 6.

All the estimators we propose are constructed using the subsampling invariance of graphex processes under the  $p$ -sampling mechanism uncovered by [53]. In particular, we use the scale invariance of certain statistics of  $G_t$  to estimate the functionals of interest using  $p$ -samples of  $G_t$ . For this reason, our estimators may be viewed as some form of *bootstrap* estimators [24, 5].

## 2. Motivations

### 2.1. Sparsity and power-law

Since graphex processes intend to improve on graphon processes in term of sparsity and power-law behaviour, these notions need to be clarified. Sparsity, is usually understood as a property of a sequence of graphs, *i.e.* as the growth of the number of edges  $e(G_t)$  in term of the number of vertices  $v(G_t)$ . In many data

---

<sup>1</sup>This second point is subject to long debates, which we discuss thoroughly in section 4.4.

analysis situations we just have a single observation at some particular size. This begs the question: does sparsity have practical relevance in this case? Regarding the power-law behaviour of the degree distribution, there is a long historical debate about whether or not this is desirable to model *empirical* graphs. We come back to this controversy in section 4.4. Anyway, it is fair to ask if it makes sense to account for sparsity and power-law in modeling. The main issue is that we often think about sparsity and power-law as a property of the sample  $G_t$  rather than a property of the *population*. This is misleading since without modeling assumptions there is no reason that sparsity and degree distribution of  $G_t$  inform us about the population.

Regarding graphex processes, both sparsity and power-law behaviour arise as a structural property of  $W$ , *i.e.* the tail-index  $\sigma \in [0, 1]$ , as explained in section 3.4. Consequently,  $\sigma$  a distinguished feature of the population, constituting an interesting summarizing statistic, as it explains the growth of the network and part of its structure. Whence, rather than talking about sparsity and power-law, we shall rather talk about the tail-index. We show in section 4 that sparse graphex samples have a readily identifiable signature which can be uncovered using the subsampling invariance described in section 3.3. This can be leveraged to visually diagnose if a given graph can be reasonably assumed to be sample of a sparse graphex process (see section 6.1).

Also, we note that it is of practical interest to construct graphex models that are close analogues of graphon models that have been proven to be useful in practice; this is the approach of [51, 29], and a number of models of this kind are described in [8]. Then, one wants to combine the efficient procedures already established for the graphon model with some new procedure for estimating the additional parameters governing the sparse behaviour, such as the tail-index.

## 2.2. Nonparametric inference, bootstrap, and hypothesis testing

Apart of the work of [51, 12, 13, 29], algorithms for statistical inference of graphex processes are still marginal in the literature. Moreover, even the most advanced algorithm of [51] assumes a parametric structure for  $W$ . Full nonparametric estimation of  $W$  is to the best of our knowledge still a challenge. One possibility is to adapt the *method of moments* [7] to graphex processes. The main idea is that  $W$  is entirely determined by the collection of all rescaled injective homomorphism densities [7, 5, 21], which can be further estimated using the *count statistics*, *i.e.* the number of occurrence of certain motifs in the graph. We elaborate on this in section 5.

Besides interest for nonparametric estimation of  $W$ , the count statistics also give nice hints on the structure of the graph. They have been used in testing equality of features of networks and finding confidence intervals of the count feature [42, 47], and similarly to determine if the dearth of certain motifs is statistically significant [5].

While in principle count statistics are exactly computable, in practice the problem becomes rapidly intractable for large motifs. In [5], the authors propose to use two different types of bootstrap techniques to make computations

feasible. The idea is to count the motifs in a possibly small subsampled subgraph to estimate the count in the whole graph. Here we propose an alternative using  $p$ -sampling. Though for counting motifs the performance is equivalent (and thus for inference too), there are some advantages when it comes to test some hypotheses, or simply if one is interested in also estimating the variance of the count statistics.

As in [5], we apply our method to construct statistical tests about the count statistics. However, in contrast with [5] our null hypotheses is a graphex process, instead of sparse graphon process. Though it might seem anecdotal<sup>2</sup>, this has a major advantage: bootstrapped samples of sparse graphon processes do not have distinguished distributional properties, while  $p$ -sampling preserves the distribution of graphex processes. We exploit this invariance to construct estimators for the variances of count statistics that are based on recycling the bootstrap samples used to estimate them, while [5] have to rely on the construction of ad-hoc estimators. Our estimators are simply rescaled version of the empirical variance of the bootstrap samples, and thus straightforward to compute.

### 3. Simple graphex processes

#### 3.1. Notations

Most of our notations are taken or inspired from [21]. We denote the vertex and edge sets of a graph  $G$  by  $V(G)$  and  $E(G)$ , and the numbers of vertices and edges by  $v(G) := |V(G)|$  and  $e(G) := |E(G)|$ . Also for every  $j \in \mathbb{N} := \{1, 2, \dots\}$ , we let  $d_j(G)$  be the number of vertices with degree  $j$  in  $G$ . We consider both labeled and unlabeled graphs; the labels will always be the integers  $1, \dots, n$  where  $n$  is the number of vertices in the graph. A labeled graph is thus a graph with vertex set  $\mathbb{N}_n := \{1, \dots, n\}$  for some  $n \geq 1$ ; we let  $\mathcal{L}_n$  denote the set of the  $2^{\binom{n}{2}}$  labeled graphs on  $\mathbb{N}_n$  and let  $\mathcal{L} := \bigcup_{n=1}^{\infty} \mathcal{L}_n$ . An unlabeled graph can be regarded as a labeled graph where we ignore the labels; formally, we define  $\mathcal{U}_n$ , the set of unlabeled graphs or order  $n$ , as the quotient  $\mathcal{L}_n / \cong$  of labeled graphs modulo isomorphisms. We let  $\mathcal{U} := \bigcup_{n=1}^{\infty} \mathcal{U}_n = \mathcal{L} / \cong$ , the set of all unlabeled graphs.

We write  $X_t \sim Y_t$  for  $\lim_t X_t/Y_t = 1$ ,  $X_t = O(Y_t)$  for  $\limsup_t |X_t/Y_t| < \infty$ , and  $X_t = o(Y_t)$  for  $\limsup_t |X_t/Y_t| = 0$ . We also write  $X_t \asymp Y_t$  if there is a constant  $c \in \mathbb{R}$  such that  $\lim_t X_t/Y_t = c$ . If  $X_t$  and  $Y_t$  are random variables, the limits are understood almost-surely. In case where the limits hold in probability, we use  $X_t = O_p(Y_t)$  and  $X_t = o_p(Y_t)$ . Inequalities up to generic constants are denoted by the symbols  $\lesssim$  and  $\gtrsim$ . The symbol  $\mathbb{E}$  denote expectations, and when it is needed to emphasize that  $\{G_t : t \geq 0\}$  is a graphex process with parameter  $W$ , we use  $\mathbb{E}_W$ .

---

<sup>2</sup>In particular these two models are tightly connected, see [8].

### 3.2. Model

We now define the (simple) graphex process. The word simple refers to the fact that we consider undirected graphs with no loops. The model has been introduced in [13] and further studied in [52, 53, 51, 8, 9, 31, 14]. We recall that  $W : \mathbb{R}_+^2 \rightarrow [0, 1]$  is a jointly measurable symmetric function, usually referred to as the *graphon* [52].

A graphex process  $\{G_t : t \geq 0\}$  is a *graph-valued* stochastic process which can be constructed as follows. Sample a unit-rate Poisson random measure  $\mathcal{V}$  on  $\mathbb{R}_+^2$  [see 18, 19], and identify  $\mathcal{V}$  with the collection of points  $(t, x) \in \mathbb{R}_+^2$  such that  $\mathcal{V}$  has a point mass, that is  $\mathcal{V} \equiv \{(t_1, x_1), (t_2, x_2), \dots\}$ . Let  $\tilde{G} \in \mathcal{L}$  be a (eventually infinite) graph with vertex set  $\mathcal{V}$ , such that for each pair of vertices  $v_i = (t_i, x_i)$  and  $v_j = (t_j, x_j)$ , there is an edge between  $v_i$  and  $v_j$  with probability  $W(x_i, x_j)$ , independently for any two  $i \neq j$ . Let  $G \in \mathcal{L}$  be the subgraph of  $\tilde{G}$  consisting only of the vertices with degree at least one. Note that  $G$  is almost-surely a graph with countably infinitely many vertices, and that the set of edges is also countably infinite except if  $W$  is equal to 0 almost-everywhere. For each  $t \geq 0$  let  $G_t \in \mathcal{U}$  be the unlabeled induced subgraph of  $G$  consisting only on the vertices  $(t', x)$  such that  $t' \leq t$ . The process  $\{G_t : t \geq 0\}$  is a graphex process with parameter  $W$ . For any fixed  $t \geq 0$ ,  $G_t \in \mathcal{U}$  is called a sample of size  $t$ . We insist that  $G_t$  is unlabeled. In other words, the labels  $(t_i, x_i)$  are not observed, and the statistician may think of them as latent variables.

If  $\int_{\mathbb{R}_+^2} W(x, y) dx dy < \infty$ , then at each  $t \geq 0$  the sample  $G_t$  has finitely many vertices and edges almost-surely [13, 52].

**Assumption 1.** *The graphon satisfies  $0 < \int_{\mathbb{R}_+^2} W(x, y) dx dy < \infty$ .*

**Remark 1.** *In this paper, we only consider graphex processes with no loops. We note that in the seminal papers [13, 52] graphex processes are defined with the possibility of having loops, and hence the definition here is a bit less general. Indeed, the processes in this paper are a special case of [13, 52] corresponding to let  $W(x, x) = 0$  for all  $x \in \mathbb{R}_+$  in their model.*

### 3.3. Subsampling of graphex processes

Various sampling designs have been proposed in the statistical and computer science literature to derive representative samples of a given network [35, 49, 38, 45]. Many of these sampling designs have been analyzed from the design-based sampling point of view [50, 15]. It is common in the statistical network literature to oppose model-based and design-based sampling. In the context of graphex processes, [53, 9] make the bridge between the two points of view by identifying the sampling design naturally associated with graphex processes.

Of particular interest, [53, 9] defines the *p-sampling* procedure. Given a graph  $G$ , a *p*-sample of  $G$ , written later  $\text{ps}(G, p)$ , is the random graph obtained from  $G$  by sampling the vertices of  $G$  independently with probability  $p$  and returning the

induced subgraph where only non-isolated vertices are conserved. The following lemma is taken from [53] and will be at the core of our inference technique.

**Lemma 1.** *Let  $\{G_t : t \geq 0\}$  be a graphex process and  $p \in [0, 1]$ . Then  $\text{ps}(G_t, p) \stackrel{d}{=} G_{pt}$  for any  $t \geq 0$ .*

From [9] we also know the converse result that if  $\text{ps}(G_t, p) \stackrel{d}{=} G_{pt}$ , then  $\{G_t : t \geq 0\}$  is a graphex process. This result shed light on the sampling design associated with graphex processes: we may also view  $G_t$  as a sample obtained from a *population* graph  $G_{t'}$  with  $t' > t > 0$  and  $G_t = \text{ps}(G_{t'}, t/t')$ .

### 3.4. Graphex marginal and tail-index

Recently, [14] have studied asymptotics for an important class of graphex models (encompassing most known examples). They derive the remarkable result that, for this model class, the sparsity and the degree distribution of the graph are governed by the tail behaviour of the graphex marginal  $\mu : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  defined by

$$\mu(x) := \int_{\mathbb{R}_+} W(x, x') dx'.$$

More specifically, assuming without loss of generality that  $\mu$  is monotone, they show that if there is  $\sigma \in [0, 1]$  and a slowly varying function<sup>3</sup>  $\ell$  such that as  $z \rightarrow 0$ ,

$$\mu^{-1}(z) := \inf\{y > 0 : \mu(y) \leq z\} \sim z^{-\sigma} \ell(z^{-1}), \tag{3.1}$$

then, under mild supplementary assumptions on  $W$  (related to our assumption 2 below), the asymptotic behaviour of  $\mathbf{v}(G_t)$ ,  $\mathbf{e}(G_t)$  and  $\mathbf{d}_j(G_t)$  is determined by  $\sigma$ . They characterize the four regimes listed below.

- *Dense:*  $\sigma = 0$  and  $\lim_{t \rightarrow \infty} \ell(t) < \infty$ . In that case  $\mu^{-1}$  has compact support and  $\mathbf{e}(G_t) \asymp \mathbf{v}(G_t)^2$  almost-surely.
- *Almost-dense:*  $\sigma = 0$  and  $\lim_{t \rightarrow \infty} \ell(t) = \infty$ . In that case  $\mathbf{e}(G_t)/\mathbf{v}(G_t)^2 \rightarrow 0$  almost-surely, but  $\mathbf{e}(G_t)/\mathbf{v}(G_t)^{2-\epsilon} \rightarrow \infty$  almost-surely for every  $\epsilon > 0$ .
- *Sparse with power-law:*  $\sigma \in (0, 1)$ . In that case,  $\mathbf{e}(G_t) \asymp [\mathbf{v}(G_t)/\ell(t)]^{\frac{2}{1+\sigma}}$  almost-surely. Moreover, for every  $j \in \mathbb{N}$ ,  $\mathbf{d}_j(G_t)/\mathbf{v}(G_t) \rightarrow \sigma \Gamma(j - \sigma)/j!$  almost-surely as  $t \rightarrow \infty$ .
- *Very sparse:*  $\sigma = 1$ . In that case,  $\ell$  has to go to zero sufficiently fast and  $\mathbf{e}(G_t)/\mathbf{v}(G_t) \rightarrow \infty$  almost-surely while  $\mathbf{e}(G_t)/\mathbf{v}(G_t)^{1+\epsilon} \rightarrow 0$  almost-surely for all  $\epsilon > 0$ .

We call  $\sigma$  in equation (3.1) the tail-index of  $\mu$ . Here we are mostly interested in the sparse with power-law regime, *i.e.*  $\sigma \in (0, 1)$ . It is the most interesting regime, since the dense regime is equivalent to the classical graphon model, and the very sparse is an extreme case unlikely to be useful in statistical applications.

---

<sup>3</sup>We recall that a function  $\ell : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is called slowly varying if  $\lim_{x \rightarrow \infty} \ell(cx)/\ell(x) = 1$  for all  $c > 0$ .

## 4. Estimating the tail-index of graphex processes

### 4.1. Heuristic and the $p$ -sampling argument

Before introducing our estimator for  $\sigma$  and results on its risk, we review here the heuristic we used to construct the estimator. Since it relies mostly on lemma 1, we shall refer to this heuristic as the  $p$ -sampling argument in the next. The idea is to use the scale-invariance of certain statistics.

From [14], we have  $v(G_t) \sim t^{1+\sigma} \ell(t) \Gamma(1-\sigma)$  almost-surely as  $t \rightarrow \infty$ , under mild conditions. Imagine we could observe  $\{G_t : t \geq 0\}$  at different sizes  $t < t'$ , with  $t/t'$  known. Then, we could asymptotically recover  $\sigma$  because

$$\frac{v(G_{t'})}{v(G_t)} \sim \left(\frac{t'}{t}\right)^{1+\sigma} \frac{\ell(t')}{\ell(t)} \approx \left(\frac{t'}{t}\right)^{1+\sigma}$$

almost-surely as  $t \rightarrow \infty$ , at least if  $t'/t$  remains constant. Of course, we usually have only one sample  $G_t$  with unknown  $t > 0$ . The key insight is that  $p$ -sampling allows us to effectively simulate a sample  $G_{t'}$  with  $t' < t$ ; intuitively, this is because  $\text{ps}(G_t, p) \stackrel{d}{=} G_{pt}$  by lemma 1. Of course, here we don't have access to the marginal law of  $\text{ps}(G_t, p)$ , only to the conditional  $\text{ps}(G_t, p) \mid G_t$ . Hence some cares have to be taken, but the intuition remains useful. In particular, we can asymptotically recover  $\sigma$  from  $v(G_t)$  and  $v(\text{ps}(G_t, p))$  for some  $p \in [0, 1]$ , or even better, from  $v(G_t)$  and

$$N_p(G_t) := \mathbb{E}[v(\text{ps}(G_t, p)) \mid G_t].$$

It is easy to see that for any simple graph  $G \in \mathcal{U}$  with no isolated vertex,

$$N_p(G) = p \sum_{j \geq 1} d_j(G) (1 - (1-p)^j).$$

In fig. 1, we illustrate on a simulation example how to infer  $\sigma$  from  $p \mapsto N_1(G_t)/N_p(G_t)$ . Indeed,  $\log(N_1(G_t)) - \log(N_p(G_t)) \rightarrow (1+\sigma) \log(p)$  almost-surely as  $t \rightarrow \infty$  under mild assumptions, by the discussion above and the results in [14]. Whence,  $\sigma$  may be estimated by determining the slope of curve  $\log(N_1(G_t)/N_p(G_t))$  as a function of  $p$ <sup>4</sup>.

### 4.2. An estimator and its risk

In view of the heuristic of the previous section, we can construct a simple estimator for  $\sigma$ . For a chosen value of  $p \in (0, 1)$ , we propose to use as an estimator for  $\sigma$ ,

$$\hat{\sigma}_p(G_t) := \begin{cases} \frac{\log N_1(G_t) - \log N_p(G_t)}{-\log p} - 1 & \text{if } N_p(G_t) \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

<sup>4</sup>At least the part of the curve corresponding to  $p$  bounded away from 0, because the asymptotic equivalences are true as  $t \rightarrow \infty$  and  $pt \rightarrow \infty$ .

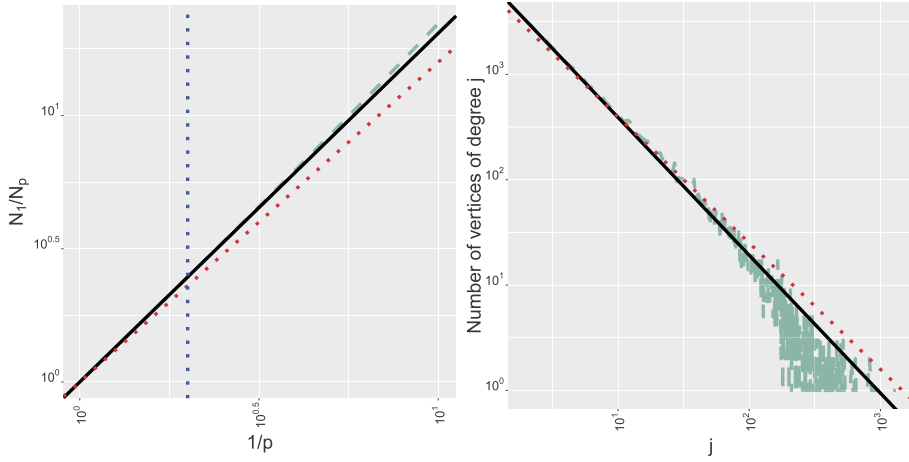


FIG 1. On the left: the plot represent the value of  $N_1(G_t)/N_p(G_t)$  as a function of  $p$  on a log-log scale. Here  $G_t$  is a sample from the GGP model of [13] with parameters  $(\sigma, \tau) = (0.2, 5)$  and  $t = 2000$  (see also example 5). The sample has 26671 vertices and 293923 edges. With these parameters  $\mu$  has a tail-index equal to 0.2. The red dotted line is  $y = (1 + \sigma)x$ , which is the expected asymptotic behaviour of  $\log \frac{N_1(G_t)}{N_p(G_t)}$  as a function of  $-\log p$ . The plain black line is  $y = (1 + \hat{\sigma}_p)x$ , where  $\hat{\sigma}_p$  is the estimated value of  $\sigma$  using  $p = 0.5$  (represented by the blue dotted vertical line). On the right: the green curve represents the degree distribution of the same sample. In plain black (respectively dotted red) is represented  $d_j \propto j^{-1-\hat{\sigma}_p}$  (respectively  $d_j \propto j^{-1-\sigma}$  to illustrate that the degree distribution is also governed by the tail-index.

Notice that  $\hat{\sigma}_p$  is well-defined since  $N_1 \geq N_p$  almost-surely for all  $0 \leq p \leq 1$ . Also note that  $\hat{\sigma}_p$  does not depend explicitly on the sample size  $t$ ; this is important because  $t$  is generally considered to be unknown.

In order to be able to compute the risk of  $\hat{\sigma}_p$ , we require further assumptions on the model. Our first assumption is analogous to the assumptions encountered in [52, 14]. We first introduce the 2-points correlation function  $\nu : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ ,

$$\nu(x, x') := \int_{\mathbb{R}_+} W(x, y)W(y, x') dy.$$

We also make the following assumption on  $\nu$ , which is a classical assumption in the literature [52, 14] to be able to control the variance of certain (non-linear) functionals of graphex processes, such as the number of vertices.

**Assumption 2.** If  $0 \leq \sigma < 1$ , we assume that there is a constant  $\eta > \max(1/2, \sigma)$  and a constant  $C \geq 0$  such that  $\nu(x, x') \leq C\mu(x)^\eta\mu(x')^\eta$  for all  $x, x' \in \mathcal{X}$ . If  $\sigma = 1$ , we assume  $\nu(x, x') \leq C\mu(x)\mu(x')$ , and we set  $\eta = 1$ .

Also, it is well-known [25, 14], that under equation (3.1) the asymptotic equivalence

$$\int_{\mathbb{R}_+} \left(1 - e^{-t\mu(z)}\right) dz \sim \begin{cases} t^\sigma \ell(t) \Gamma(1 - \sigma) & \text{if } 0 \leq \sigma < 1, \\ t \int_t^\infty z^{-1} \ell(z) dz & \text{if } \sigma = 1 \end{cases} \quad (4.1)$$



holds as  $t \rightarrow \infty$ . This relation is already enough to show consistency of  $\hat{\sigma}_p$ . However, further assumptions on  $\mu$  are required to bound the bias of the estimator. In the present paper, we shall assume the following.

**Assumption 3.** *We assume that equation (3.1) holds. We furthermore assume that for all  $p \in (0, 1)$  there is a sequence  $(\Gamma_{p,t})_{t>0}$  such that*

$$\left| \frac{\int_{\mathbb{R}_+} (1 - e^{-pt\mu(z)}) \, dz}{p^\sigma \int_{\mathbb{R}_+} (1 - e^{-t\mu(z)}) \, dz} - 1 \right| \leq \Gamma_{p,t}.$$

Assumption 3 is closely related to second-order regular variation assumptions used in extreme values theory [20]. Its purpose is the following. The integral in equation (4.1) plays a determining role in computing the risk of  $\hat{\sigma}_p$ . Equation (4.1), however, says nothing about the rate at which the integral approaches  $t^\sigma \ell(t)$ . Moreover, we know that  $\ell(pt)/\ell(t) \rightarrow 1$  as  $t \rightarrow \infty$  by the assumptions on  $\ell$ , but again, the rate of convergence is unknown. These two rates are encapsulated in assumption 3, allowing us to quantify the risk of  $\hat{\sigma}_p$ . Bounds on  $\Gamma_{p,t}$  for various examples are given in section 4.3. These examples show that in many cases  $\Gamma_{p,t}$  has polynomial decay.

We are now in position to state the main theorem of this section, whose proof is deferred to appendix A.

**Theorem 1.** *Under assumption 2 and 3, there is a constant  $C'' > 0$  depending only on  $p$  and  $W$  such that for all  $t \geq 1$  it holds*

$$\mathbb{E}_W[(\hat{\sigma}_p(G_t) - \sigma)^2] \leq \frac{\Gamma_{p,t}^2}{(-\log p)^2} + C'' \log(t)^2 \max \left\{ \frac{1}{t^{1+\sigma} \ell(t)}, t^{1-2\eta} \right\}.$$

**Remark 2.** *The computation of the estimator requires to choose a value for  $p$ . Theoretical guidelines are that any value of  $p$  bounded away from zero yields a consistent estimator. But, this leaves unclear how to choose in practice. Optimally, we could build a selection procedure that selects the “optimal” value of  $p$ , in the sense that bias and variance are suitably balanced. This would however require to make finer assumptions on the structure of the bias. We believe the gain would be marginal as in many cases the value of  $\hat{\sigma}_p$  is relatively robust to the choice of  $p$ , see figs. 1 and 4. In practice, we recommend to do a Hill plot to assess the pertinence of estimating  $\sigma$  and to determine a reasonable value of  $p$ , as detailed in section 6.1.*

### 4.3. Discussion about the bias

In [14], the authors suggest to estimate the tail-index using

$$\hat{\sigma}_{\text{CPR}} := \frac{2 \log v(G_t)}{\log e(G_t)} - 1.$$

It is easily seen that  $\mathbb{E}_W[(\hat{\sigma}_{\text{CPR}} - \sigma)^2]$  is in general a  $O(1/\log t)$ . Although the variance of  $\hat{\sigma}_p$  has polynomial decay in  $t$ , the bias is proportional to  $\Gamma_{p,t}$ , whose

decay can be anything. Hence it is not clear if  $\hat{\sigma}_p$  dominates  $\hat{\sigma}_{\text{CPR}}$  asymptotically. Here we present some examples where the bias of  $\hat{\sigma}_p$  has also polynomial decay in  $t$ . We consider the examples which were originally given in [14], because they cover a large spectrum of behaviours.

**Example 1.** *Dense graphs.* Consider the function  $W(x, y) = (1 - x)(1 - y)\mathbf{1}_{x \leq 1}\mathbf{1}_{y \leq 1}$ . Then  $\mu(x) = 0.5(1 - x)\mathbf{1}_{x \leq 1}$  and the graph is almost-surely dense. Assumption 2 is trivially satisfied with  $\eta = 1$ . Also, it follows from appendix C.1 that assumption 3 is satisfied with  $\Gamma_{p,t} \propto t^{-1}$ : the risk is polynomially decreasing on this example.

**Example 2.** *Sparse, almost dense graphs without power law.* Consider the function  $W(x, y) = \exp(-x - y)$ . Then  $\mu(x) = \exp(-x)$  and assumption 2 is trivially satisfied with  $\eta = 1$  and  $\sigma = 0$ . Moreover, it follows from appendix C.2 that assumption 3 is satisfied with  $\Gamma_{p,t} \propto 1/\log(t)$ : the risk is logarithmically decreasing on this example.

**Example 3.** *Sparse graphs with power law, separable.* For  $0 < \sigma < 1$ , consider the function  $W(x, y) = (1 + x)^{-1/\sigma}(1 + y)^{-1/\sigma}$ . Then obviously  $\mu(x) = \sigma(1 + x)^{-1/\sigma}/(1 - \sigma)$  and assumption 2 holds trivially with  $\eta = 1$ . Moreover, from appendix C.3, assumption 3 holds with  $\Gamma_{p,t} \propto t^{-\sigma}$ : the risk is polynomially decreasing on this example.

**Example 4.** *Sparse graphs with power law, non separable.* For  $0 < \sigma < 1$ , consider the function  $W(x, y) = (1 + x + y)^{-1/\sigma - 1}$ . It is shown in [14] that assumption 2 holds with  $\eta = (1 + \sigma)/2$ . We have  $\mu(x) = \sigma(1 + x)^{-1/\sigma}$ , which is up to a constant the same marginal as the previous example, so that assumption 3 is verified with  $\Gamma_{p,t} \propto t^{-\sigma}$ : the risk is polynomially decreasing on this example.

**Example 5.** *GGP model.* For  $0 \leq \sigma < 1$ , let  $\rho_{\sigma,\tau}(dx) = \frac{x^{-1-\sigma} \exp(-x) dx}{\Gamma(1-\sigma)}$ ,  $\bar{\rho}_{\sigma,\tau}(x) := \int_x^\infty \rho_{\sigma,\tau}(dx)$ , and  $W(x, y) = 1 - \exp(-2\bar{\rho}_{\sigma,\tau}^{-1}(x)\bar{\rho}_{\sigma,\tau}^{-1}(y))$ . This corresponds to the model in [13]. It is shown in [14] that assumption 2 holds with  $\eta = 1$ . We establish in appendix C.4 that assumption 3 is satisfied with  $\Gamma_{p,t} \propto t^{-\sigma}$ : the risk is polynomially decreasing on this example.

Although the asymptotic behaviour of the graphs are relatively close for examples 1 and 2, the bias decay is radically different and estimation is harder in the second example. We also estimated  $\mathbb{E}[(\hat{\sigma}_{0.5} - \sigma)^2]^{1/2}$  and  $\mathbb{E}[(\hat{\sigma}_{\text{CPR}} - \sigma)^2]^{1/2}$  on the previous examples using 10000 Monte Carlo samples. In the simulations, we picked  $\sigma = 0.3$  for the sparse separable and sparse non-separable example, while the GGP example corresponds to a choice of  $\sigma = 0.5$ . The raw numbers are given in tables 1 and 2.

We pictured in fig. 2 the risk of  $\hat{\sigma}_{0.5}$  and  $\hat{\sigma}_{\text{CPR}}$  as functions of  $t$  for all the simulated examples. As expected, the risk of  $\hat{\sigma}_{0.5}$  has polynomial decay every-time but the almost dense example, while the risk of  $\hat{\sigma}_{\text{CPR}}$  remains logarithmically decreasing on all examples. We note, however, that even in the situation when the risk has logarithm decay,  $\hat{\sigma}_{0.5}$  seems to perform slightly better than  $\hat{\sigma}_{\text{CPR}}$ .

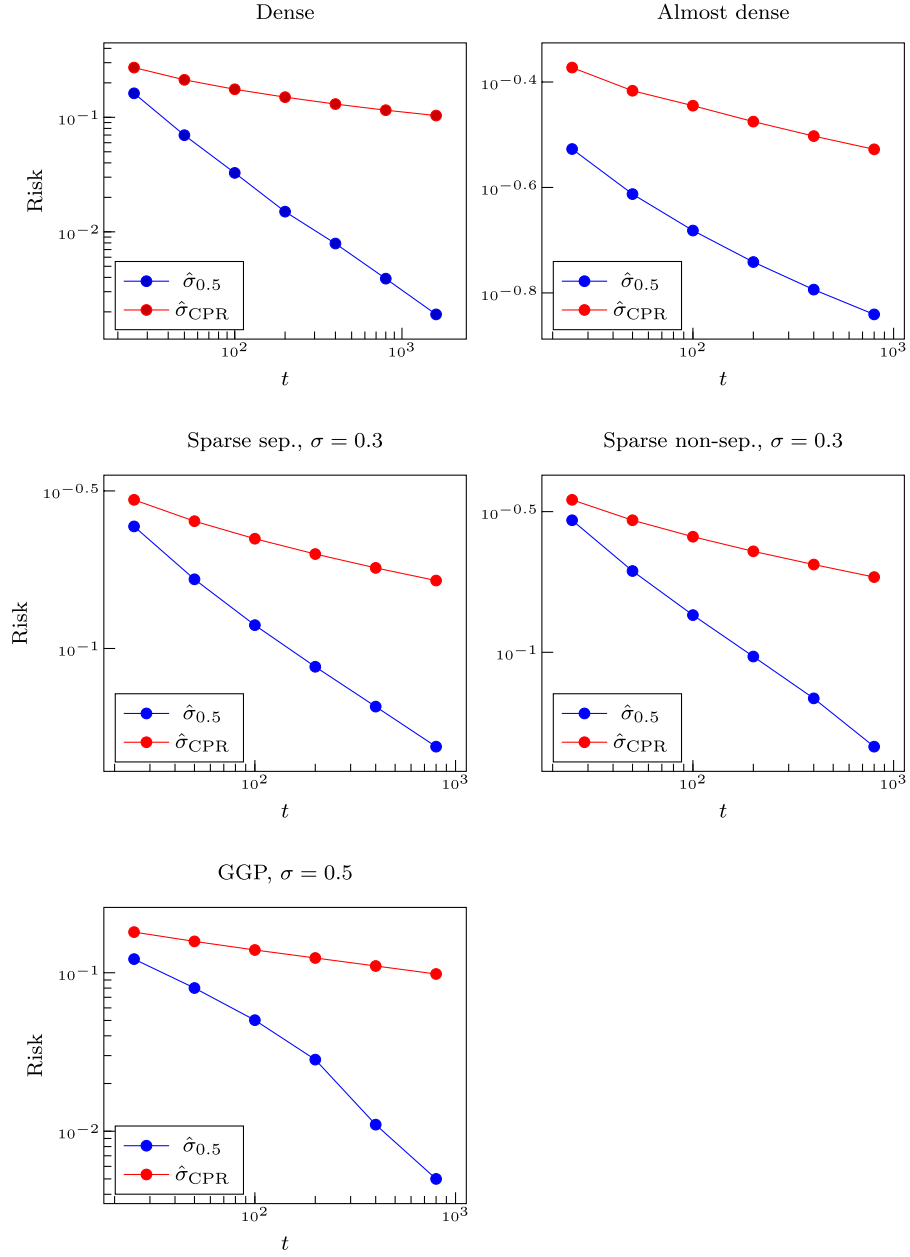


FIG 2. Comparative plot of  $\mathbb{E}[(\hat{\sigma}_{0.5} - \sigma)^2]^{1/2}$  against  $\mathbb{E}[(\hat{\sigma}_{\text{CPR}} - \sigma)^2]^{1/2}$  using 10 000 MC samples for the examples described in section 4.3.

TABLE 1  
 Estimation of  $\mathbb{E}[(\hat{\sigma}_{0.5} - \sigma)^2]^{1/2}$  using 10 000 MC samples for the examples described in section 4.3 and various values of  $t$ .

$t$	25	50	100	200	400	800
Dense	0.1620	0.0698	0.0327	0.0150	0.0079	0.0039
Almost-dense	0.2971	0.2440	0.2081	0.1814	0.1608	0.1443
Sparse sep. ( $\sigma = 0.3$ )	0.2440	0.1658	0.1186	0.0875	0.0654	0.0488
Sparse non-sep. ( $\sigma = 0.3$ )	0.2947	0.1945	0.1356	0.0966	0.0686	0.0462
GGP ( $\sigma = 0.5$ )	0.1219	0.0801	0.0502	0.0283	0.0110	0.0050

TABLE 2  
 Estimation of  $\mathbb{E}[(\hat{\sigma}_{\text{CPR}} - \sigma)^2]^{1/2}$  using 10 000 MC samples for the examples described in section 4.3 and various values of  $t$ .

$t$	25	50	100	200	400	800
Dense	0.2719	0.2125	0.1757	0.1500	0.1305	0.1154
Almost-dense	0.4240	0.3882	0.3590	0.3349	0.3144	0.2967
Sparse sep. ( $\sigma = 0.3$ )	0.2963	0.2535	0.2230	0.1994	0.1802	0.1643
Sparse non-sep. ( $\sigma = 0.3$ )	0.3483	0.2947	0.2575	0.2285	0.2051	0.1851
GGP ( $\sigma = 0.5$ )	0.1805	0.1577	0.1392	0.1239	0.1103	0.0982

#### 4.4. On the controversy about scale-free nature of real graphs

There is no actual consensus on the scale-free nature of empirical graphs<sup>5</sup>. Here, we briefly review both sides of the debate. This is also an opportunity for us to review what  $\sigma$  is and what it isn't.

According to the general opinion, a finite graph  $G$  is called scale-free if there are  $c > 0$ ,  $\gamma > 1$  such that for  $j$  large enough

$$\frac{d_j(G)}{v(G)} \approx c j^{-\gamma}. \tag{4.2}$$

We left the previous definition intentionally non-rigorous to emphasize one of the origin of the controversy: the lack of consensus on the definition of scale-freeness [30, 54].

Barabási and Albert [4] found that for three empirical networks equation (4.2) seemed true, and they attributed this property to a growth mechanism they called *preferential attachment*. Since then, many papers claimed to have discovered new types of power-law networks and new mechanism creating them, see for instance [44, 30]. But not everybody agrees, and some authors believe that power-law degree distribution is rarely encountered in practice [32, 16]. The heated debate has culminated recently with [10], claiming that *scale-free networks are rare* after reviewing a large collection of empirical networks.

The methodology in [10] consists on testing as a null hypothesis a power-law model for the degree distribution against an alternative which is not power-law. On a corpus of 927 empirical networks, which they claim is representative of *real* networks, they found that only a very few failed to reject the null hypothesis.

---

<sup>5</sup>Here empirical refers to typical datasets encountered in the literature. This is of course very subjective and thus one of the source of the controversy.

Since its first appearance as a preprint [10] has generated a lot of activity, whether it be in agreement or disagreement. We share the same opinion as [30, 54] that the source of the debate is the lack of a rigorous definition. [54] makes an effort toward a clean definition and they arrive at the opposite conclusion of [10]. Another argument against [10], which seems to be the most widespread, is that scale-freeness is only well-defined in the infinite-size limit [30]. According to this view, a network is scale-free if its degree distribution approaches a power-law as the network keeps growing following the same mechanisms. In other word, instead of equation (4.2), we shall say that  $\{G_t : t \geq 0\}$  is scale-free if

$$\lim_{j \rightarrow \infty} \lim_{t \rightarrow \infty} \frac{j^\gamma d_j(G_t)}{v(G_t)} = c. \quad (4.3)$$

Then,  $\gamma$  is not defined as property of the sample  $G_t$ , but rather as a joint property of the limiting graph and the growth mechanism, equivalently of the sampling design. Here we insist on the importance of the sampling design in the definition: if  $\{G_t : t \geq 0\}$  is a graphex process, then  $\gamma = 1 + \sigma$  almost-surely under mild conditions [14], though the degree distribution of  $G_t$  at finite  $t$  might have a different index, as illustrated in fig. 3. Another famous example is the *traceroute* sampling which is known to produce samples with power-law degree distributions, even when the population is not power-law [37, 1]. Another example of this type is provided in section 6.3. Hence, if one is ready to accept equation (4.3) as a definition, talking about scale-freeness without knowledge of the sampling design is irrelevant.

One of the criticism of the definition in equation (4.3) is made on the basis that we never observe  $\{G_t : t \geq 0\}$  but only  $G_t$ , and that we need tools adapted to finite networks. We have mixed feelings about this. On one hand, the same criticism could be made about any statistical analysis. Indeed, one of the primary goal of statistics is to infer stuff about a population given a sample. In the case of i.i.d. observations, we have accepted for a while to idealize the population as an abstract probability distribution, and this has been proven to be very effective in summarizing properties of the population as parameters of the distribution. When the observation is a graph, we have no reason to proceed differently (see in particular in [30] the example of epidemic threshold). On the other hand [30] points out the difficulty in determining the right sampling mechanism, and it is true that models of random networks are commonly misspecified. Then, it is less clear that the parameters make sense, albeit this depends on the degree of misspecification and the task to achieve.

Finally, we note that the test in [10] only involves the degree distribution, but graphs are more than just their degree distribution (see in particular section 6.2.2). For these reason, we don't believe in the existence of a universal model that can explain all empirical graphs, and it is the work of the statistician to determine a good model for *a given task*; whence it is useful to study as many models as possible. This opinion seems to be shared by the authors of [10]. We add that in many cases the focus should be put on achieving the task at hand with accuracy rather than trying to explain all the structure of the sample.

Besides, the assertion that graphex processes are not desirable to model empirical graphs because power-law graphs are rare is too simplistic: (i) no serious methodology can objectively prove this rarity (what is a representative set of empirical graphs? what definition of scale-freeness to use?), (ii) it is the degree distribution of graphex processes in the infinite size limit which is power-law<sup>6</sup>, and (iii) there are empirical graphs for which some graphex processes have been found to explain well the degree distribution (see [13, 51] and also section 6). Overall, one should not forget to look at other aspects of the modeling that might be more important than just the degree distribution.

#### 4.5. Relation to Hill estimator

Regardless of the controversy of section 4.4, tail-index of the degree distribution of  $G_t$  is often reported as an interesting summarizing statistics [35, Section 4]. [54] plea in favor of using various estimators as the index is notoriously hard to estimate and different estimators may give very different answers on a given sample. In our opinion, the interest in the tail-index of  $G_t$  is debatable as it might be unrelated with the index of the population, an issue we already discussed in section 4.4, and which we will emphasize here. The most common reported value for the tail-index of the sample is based on the Hill estimator

$$\hat{\gamma}_{\text{Hill}}(G_t) := \frac{\sum_{j=d_{\min}+1}^{\infty} d_j(G_t) \log \frac{j}{d_{\min}}}{\sum_{j=d_{\min}+1}^{\infty} d_j(G_t)},$$

where  $d_{\min} > 0$  is a cutoff determining where the tail of the degree distribution starts. Most of empirical graphs are reported to have  $\hat{\gamma}_{\text{Hill}} \in [2, 3]$ , while graphex processes have almost-surely index  $\gamma = 1 + \sigma \in [1, 2]$ , at least if we rely on the definition of equation (4.3). So one might think graphex processes are unrealistic to model graphs with  $\hat{\gamma}_{\text{Hill}} \in [2, 3]$ . Once again, we have to be careful: it is not obvious that the Hill estimator gives consistent answers for  $\gamma$ . Though it is known from [55] that it is consistent in the preferential attachment model, nothing is known about graphex processes. In particular, in fig. 3 we illustrate what can go wrong: on a simulated sample with true  $\sigma = 0.3$  (hence  $\gamma = 1.3$ ), our estimator gives  $\hat{\sigma}_{0.5} = 0.3335$ , while a non carefully calibrated Hill estimator gives  $\hat{\gamma}_{\text{Hill}} = 2.1181 \in [2, 3]$ . Once again, this shows that one has to be careful when relating degree distribution of the sample and the idealized degree distribution of the population, *i.e.* the sampling design has to be taken into account when inferring properties about the population.

### 5. Count statistics of graphex processes: definition, inference, and hypothesis testing

We briefly discuss how the  $p$ -sampling argument may also be used to estimate other functionals of  $W$ . This essentially extends the work of [7, 5] from graphon

---

<sup>6</sup>And in fact, they can accommodate for other behaviours, by either satisfying  $\sigma = 0$ , or violating assumptions 2 and 3.

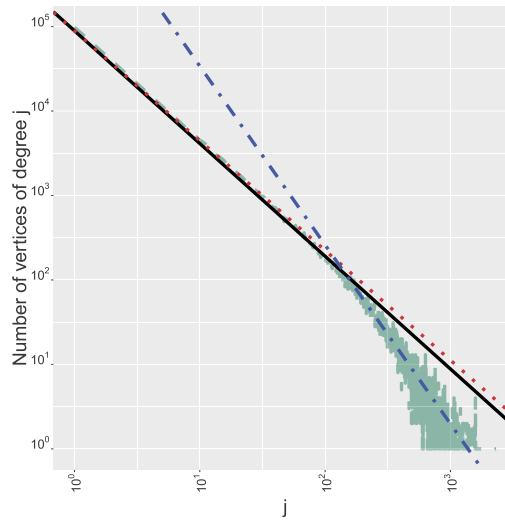


FIG 3. In green is the degree distribution  $G_t$  is a sample from the GGP model of [13] with parameters  $(\sigma, \tau) = (0.3, 10)$  and  $t = 10000$  (see also example 5). The sample has 306791 vertices and 3947805 edges. The red line represents the  $d_j \propto j^{-1-\sigma}$ . The black line is  $d_j \propto j^{-1-\hat{\sigma}_p}$ , where  $\hat{\sigma}_p = 0.333474$  is computed using  $p = 0.75$ . Finally the blue line is  $d_j \propto j^{-\hat{\gamma}_{\text{Hill}}}$ , where  $\hat{\gamma}_{\text{Hill}} = 2.1181$  is computed using  $d_{\min} = 300$ .

to graphex. In particular, using  $p$ -sampling allows to derive a natural notion of bootstrap for networks. Using bootstrap samples, we can then estimate the *count statistics* of the graph (sections 5.1 and 5.2), which can in turn be used to do nonparametric inference on the (rescaled injective) homomorphism densities (section 5.3). We also use count statistics to construct tests for the number of occurrence of certain motifs (section 5.4). Finally, we discuss the relation of these results with some other works (section 5.5).

### 5.1. Count statistics

Count statistics have received a lot of attention in the last few years, for multiple reasons. A first reason is that some motifs are of interest to get insights on structure of the network. For instance, counting the number of triangles permit to determine the *transitivity* of the graph, which is a popular measure of the tendency of the vertices to cluster together [56]. Another reason is that counts of motifs characterize the distribution of the network (see [21] and section 5.3), so that nonparametric inference on the distribution of the network is possible using counts of certain motifs [6, 7, 5]. Other related uses of motifs counts involve hypothesis testing [5] and networks comparison [5, 2].

Count of motifs is notoriously hard, especially when the graph is large and the motif has a large number of vertices. Even counting simple motifs such as triangles might rapidly become intractable on large graphs [48]. Methods

based on subsampling the graph have emerged to bypass the complexity. A non exhaustive list of existing methods includes methods based on vertex sampling [5, 34, 2, 41], the parametric bootstrap [28, 39], or based on edge sampling [3, 23, 26, 27]. Here, we take advantage of the  $p$ -sampling to count motifs. We discuss in section 5.5 how that relates to already existing works, and some of the advantages in doing  $p$ -sampling over other sampling design.

Before going further, we shall first clarify what we mean by count statistics. For a finite graph  $R$ , we define the number of copies of  $R$  in  $G$  as the number of injections  $\phi : V(R) \rightarrow V(G)$  such that if  $\{r_1, r_2\} \in E(R)$  then  $\{\phi(r_1), \phi(r_2)\} \in E(G)$ , *i.e.* the number of adjacency preserving map from  $R$  to  $G$ . We call this number  $n(R, G)$ , the count statistic of  $R$ . If  $\phi : A \rightarrow V(G)$  is injective, we let  $G[\phi]$  denote the graph such that  $V(G[\phi]) = A$  and  $\{v, v'\} \in E(G[\phi])$  if and only if  $\{\phi(v), \phi(v')\} \in E(G)$ . That is,  $G[\phi]$  is isomorphic to the subgraph of  $G$  induced by  $\phi$ . Letting  $\mathcal{A}(R, G)$  denote the set of all injections from  $V(R)$  to  $V(G)$ , we can rewrite

$$n(R, G) = \sum_{\phi \in \mathcal{A}(R, G)} \mathbf{1}\{R \subseteq G[\phi]\}. \tag{5.1}$$

### 5.2. Bootstrap estimators for count statistics: model-free results

We first show how to derive a bootstrap estimate for  $n(R, G)$  using  $p$ -sampling which works for any graph  $G$  (so  $G$  is not necessarily a graphex process). The result is of interest by itself, since it proves a consistent estimator for  $n(R, G)$  regardless of the distribution of  $G$ , and that estimator is parallelizable, in contrast with the computation of  $n(R, G)$  which may be challenging for complex motifs  $R$ . Our estimator is based on the following easy proposition.

**Proposition 1.** *Let  $G \in \mathcal{U}$  and let  $R \in \mathcal{L}_k$  be a connected graph such that  $n(R, G) < \infty$ . Then,*

$$\mathbb{E}[n(R, \text{ps}(G, p)) \mid G] = p^k \cdot n(R, G). \tag{5.2}$$

From the previous proposition, we can obtain a natural estimator of  $n(R, G)$ , which we refer as the *bootstrap* estimator. Indeed, for a fixed  $p \in (0, 1)$  to be chosen, and for some integer  $B \geq 1$ , we let  $\hat{G}_1, \dots, \hat{G}_B$  be i.i.d. copies of  $\text{ps}(G, p) \mid G$ , and we estimate  $n(R, G)$  by

$$\hat{n}(R, G) := \frac{1}{Bp^k} \sum_{b=1}^B n(R, \hat{G}_b).$$

The main advantage is that  $n(R, \hat{G}_b)$  are intrinsically faster to compute than  $n(R, G)$ , especially when  $G$  is sparse, and computations can be run in parallel. Interestingly, the formula (5.2) is *model-free*, in the sense that it is true for any graph  $G$  that has no isolated vertices. It follows that if  $G$  has no isolated vertices,



then  $\hat{n}(R, G)$  is an unbiased estimator of  $n(R, G)$ , comparable to the Horvitz-Thompson estimator constructed in [34]<sup>7</sup>. The next proposition quantify the risk of  $\hat{n}(R, G)$  as an estimator of  $n(R, G)$ .

**Proposition 2.** *Let  $G \in \mathcal{U}$  and let  $R \in \mathcal{L}_k$  be a connected graph such that  $n(R, G) < \infty$ . Then  $\hat{n}(R, G)$  is an unbiased estimator of  $n(R, G)$ , and*

$$\text{var}(\hat{n}(R, G) \mid G) = \frac{\text{var}(n(R, \text{ps}(G, p)) \mid G)}{Bp^{2k}} \leq \frac{k}{Bp} \cdot n(R, G)^2.$$

It is worth pointing out that by the previous,  $\hat{n}(R, G)$  is a consistent estimator of  $n(R, G)$  as  $B \rightarrow \infty$ , and thus with almost no condition on  $G$ . This result is of interest by itself, if one is only interested in counting motifs in  $G$  (see also section 6.2).

**Remark 3.** *In contrast with our estimator for the tail-index, we keep the extra randomness induced by the bootstrap samples, i.e. we have a randomized estimator. But, it is clear from proposition 4 that taking the expectation of  $\hat{n}(R, G)$  conditional on  $G$  is pointless. Moreover, the advantage of keeping the randomized-estimator is firstly to enable tractable computations as bootstrap samples can be made arbitrary small.*

### 5.3. Nonparametric inference

Beyond their interest in describing the structure of the network, count statistics permit to estimate the rescaled injective homomorphism densities [8]. Those are distinguished functionals of  $W$  whose collection entirely determine  $W$  [21, 40, 8]. The rescaled injective homomorphism density of  $R$  in  $W$  is defined as

$$h(R, W) := \frac{\int_{\mathbb{R}_+^{\nu(R)}} \prod_{\{\ell, m\} \in E(R)} W(x_\ell, x_m) dx_1 \dots dx_{\nu(R)}}{\left\{ \int_{\mathbb{R}_+^2} W(x_1, x_2) dx_1 dx_2 \right\}^{\nu(R)/2}}.$$

It follows from [21, 40, 8] that  $W$  can be in principle recovered – up to measure preserving transformations – from the collection of all  $\{h(R, W) : R \in \mathcal{L}\}$ . We shall only consider motifs  $R$  and processes for which the previous quantity is finite. Indeed, we will assume that a certain number of moments of  $W$  are finite.

**Assumption 4.** *A graphex process  $\{G_t : t \geq 0\}$  with graphon function  $W : \mathbb{R}_+^2 \rightarrow [0, 1]$  satisfies this assumption for  $N \geq 1$  integer if,*

$$\max_{n=1, \dots, N} \max_{\substack{R \in \mathcal{L}_n \\ R \text{ connected}}} \int_{\mathbb{R}_+^n} \prod_{\{\ell, m\} \in E(R)} W(x_\ell, x_m) dx_1 \dots dx_n < \infty.$$

Then, we have the following result, borrowed from [8, Proposition 56].

---

<sup>7</sup>However [34] consider the number of induced subgraphs of  $G$  that are isomorphic to  $R$ , which is slightly different from  $n(R, G)$ , but the arguments are comparable.

**Proposition 3.** *Let  $R \in \mathcal{L}_k$  be a connected graph. If  $\{G_t : t \geq 0\}$  is a graphex process satisfying assumption 4 with  $N = k$ , then almost-surely,*

$$\lim_{t \rightarrow \infty} t^{-k} \mathfrak{n}(R, G_t) = \int_{\mathbb{R}_+^k} \prod_{\{\ell, m\} \in E(R)} W(x_\ell, x_m) dx_1 \dots dx_k.$$

The consequence of the last proposition is that  $h(R, W)$  can be consistently estimated as  $t \rightarrow \infty$  by

$$\hat{h}(R, G_t) := \frac{\mathfrak{n}(R, G_t)}{\mathfrak{e}(G_t)^{\nu(R)/2}},$$

where in turn we can leverage the consistency of our bootstrap estimator of  $\mathfrak{n}(R, G_t)$  to make computations tractable. We note that the method of moments of [7] can be adapted to estimate  $W$  from  $\hat{h}(R_1, G_t), \hat{h}(R_2, G_t), \dots$  for clever choice of motifs  $R_1, R_2, \dots$ .

### 5.4. Hypothesis testing

Besides their interest for estimating  $W$ , the count statistics have been used in testing equality of features of networks and finding confidence intervals of the count feature [42, 47]. The authors of [5] have an example where they use counts of certain motifs to test if the dearth of short-cycles in a network is statistically significant. As argued in [5], the main challenge is then the problem of finding estimates of  $s_t^2(R) := \text{var}_W(\mathfrak{n}(R, G_t))$ . Here we show how the  $p$ -sample argument can be used in order to recycle the bootstrap samples to estimate  $s_t^2(R)$ , which provides a somewhat simpler estimator than in [5]. Our estimator is based on the result of the following proposition.

**Proposition 4.** *Let  $\{G_t : t \geq 0\}$  be a graphex process satisfying assumption 4 with  $N = 2k - 1$ . For any connected graph  $R \in \mathcal{L}_k$ , as  $t \rightarrow \infty$*

$$s_t^2(R) = \frac{\mathbb{E}_W[\text{var}(\mathfrak{n}(R, \text{ps}(G_t, p)) \mid G_t)]}{p^{2k-1}(1-p)} + O(t^{2k-2}).$$

This suggests that  $s_t^2(R)$  may be estimated by

$$\hat{s}^2(R, G_t) := \frac{\text{var}(\mathfrak{n}(R, \text{ps}(G_t, p)) \mid G_t)}{p^{2k-1}(1-p)},$$

where in turn the variance in the numerator of the last display can be approximated up to arbitrary precision by recycling the bootstrap samples used for the estimation of  $\mathfrak{n}(R, G_t)$ . The next proposition establishes that  $\hat{s}^2(R)$  is consistent.

**Proposition 5.** *Let  $\{G_t : t \geq 0\}$  be a graphex process satisfying assumption 4 with  $N = 4k - 2$ . For any connected graphs  $R \in \mathcal{L}_k$ , as  $t \rightarrow \infty$*

$$\frac{\hat{s}^2(R, G_t)}{s_t^2(R)} = 1 + O_p(t^{-1/2}).$$

The fact that we are able to consistently estimate  $s_t^2(R)$  permit to perform hypothesis testing for  $\bar{n}(R, W, t) := \mathbb{E}_W[n(R, G_t)]$ . In particular, it is of interest to construct tests for  $H_0 : \bar{n}(R, W, t) = \bar{n}(R, W_0, t_0)$  against  $H_a : \bar{n}(R, W, t) \neq \bar{n}(R, W_0, t_0)$ , see for instance section 6.2.2 and also [5]. We show in the next proposition that asymptotically consistent testing for these hypotheses can be made on the basis of the test statistic

$$T(R, G_t) := \frac{n(R, G_t) - \bar{n}(R, W_0, t_0)}{\sqrt{\hat{s}^2(R, G_t)}}.$$

**Proposition 6.** *Let  $\{G_t : t \geq 0\}$  be a graphex process with graphon function  $W_0$  satisfying assumption 4 with  $N = 4k$ . For any connected graphs  $R \in \mathcal{L}_k$ ,  $T(R, G_{t_0}) \xrightarrow{d} \mathcal{N}(0, 1)$  as  $t_0 \rightarrow \infty$ .*

**Remark 4.** *It would also be of interest to construct a test for the weaker hypothesis  $H_0 : h(R, W) = h(R, W_0)$ . It seems to us that this is a much harder problem. Indeed, it is true that  $\hat{h}(R, G_t) \sim h(R, W_0)$  almost-surely under  $H_0$  by [8], and so we could imagine constructing a test based on  $\hat{h}$ . However, the statistical fluctuations of  $n(R, G_t) - \mathbb{E}_W[n(R, G_t)]$  and of  $e(G_t) - \mathbb{E}_W[e(G_t)]$  are exactly of the same order, which makes the asymptotic normality of  $\hat{h}(R, G_t)$  less evident.*

### 5.5. Relation to other works

The work presented here relates to many of the paper cited previously. In our opinion, the two most related papers are the work of [5] and of [34]. We now discuss the relation between these papers and our results.

In [5], the authors consider nonparametric estimation of certain functionals of the graphon in a *sparse graphon* model, with applications in hypotheses testing and networks comparison. The sparse graphon model and the graphex model are different, though they are tightly related, as explained for instance in [8]. The functionals considered in [5, 7] are related to the expectation of counts of motifs too. We note, however, that they use a slightly different notion of counts than  $n(R, G_t)$ , though this do not fundamentally make difference and their procedures can be easily adapted to  $n(R, G_t)$ . The paper [5] consider two subsampling procedures for counting motifs. Of particular interest for us is their *uniform subsampling bootstrap*. Given a sample  $G_n$  with  $n$  vertices, they consider the subgraph  $\tilde{G}_{n,m}$  which consists on uniformly sampling  $m < n$  vertices without replacement of  $G_n$  and returning the induced subgraph. For moderate  $m$ , this is morally equivalent to  $p$ -sampling. Sampling vertices without replacement is quite natural for *dense* graphon sample, but less evident in the case of a sparse graphon sample. The issue is for instance discussed in [41]: in the classical dense graphon model  $\tilde{G}_{n,m} \stackrel{d}{=} G_m$ , which is not necessarily true in the sparse model. This is not an issue to estimate  $n(R, G_n)$ , but it is if one wants to estimate  $\text{var}(n(R, G_n))$ , to do hypothesis testing for instance. Because the distribution of  $\tilde{G}_{n,m}$  has no distinguished properties, it is unclear that  $\text{var}(n(R, G_n))$  can be

estimated from  $\text{var}(\mathbf{n}(R, \tilde{G}_{n,m}) \mid G_n)$ ; so [5] have to rely on rather complicated procedures to estimate  $\text{var}(\mathbf{s}(R, G_n))$ .

In [34], the authors are also interested in estimating the counts of certain motifs using vertex sampling. They also consider two subsampling schemes. In particular, they consider uniform subsampling of vertices, which is almost  $p$ -sampling. The only difference is that they keep isolated vertices in the subsample while we don't. Since we only consider connected motifs, the two methods are the same, as isolated vertices cannot contribute to the motif. Consequently, our results in section 5.2 are partial restatements of [34], though they also consider a different notion of counts. We note that in [34] the authors are only interested in counting motifs. They don't make probabilistic assumption on the graph and hence do not consider inference on the distribution of the graph.

## 6. Some applications

### 6.1. Assessing the pertinence of sparse graphex models on some real networks

The tail-index of  $\mu$  is a distinguished feature of sparse graphex processes (*i.e.* graphex processes satisfying assumptions 1 to 3 for  $\sigma \neq 0$ ). As such, it can be used to diagnose the pertinence of using such a model on some real datasets. In particular, if  $G_t$  is a sample of a sparse graphex process, then we expect the degree distribution of  $G_t$  to be close to a power-law with index  $1 + \sigma$ , and  $p \mapsto \log \frac{N_1(G_t)}{N_p(G_t)}$  to be closely linear with slope  $1 + \sigma$ , at least for  $p$  bounded away from 0 (the asymptotic being true only in this region). Equivalently, we might look at the plot of  $p \mapsto \hat{\sigma}_p(G_t)$ , which we call a Hill plot in analogy with the classical Hill plots used in extreme value theory to calibrate the Hill estimator [22]. If the model is correct, we expect the Hill plot to be constant, except perhaps at the boundary close to  $p = 0$ . The Hill plot is doubly interesting as it allows for both assessing the pertinence of the model, but also to choose a reasonable value of  $p$  for the estimation of the tail-index (see also remark 2).

We choose to illustrate the diagnostic on two datasets from [43], the Flickr and the Youtube datasets. These datasets are available from the KONECT database [36]. These are social networks datasets where vertices are users and edges represent friendships. We note that the Flickr dataset is a directed graph, but we ignore arrows and see it as an undirected graph. We summarize in table 3 the main statistics of these networks. These two networks are interesting in the

TABLE 3  
Summary of main statistics for the Flickr and Youtube networks.

Network	Number of vertices	Number of edges
Flickr	2302926	22838277
Youtube	3223586	9375375

way they have been sampled. Indeed, [43] crawled these networks by searching for the whole connected component reachable from a random set of users. The

methodology may be summarized as starting from a single user, and in each step, retrieve the list of friends for a not yet visited user and add these users to a list of users to visit. The process is continued until exhaustion of the list. This is known as *breadth-first search* (BFS) sampling. Interestingly, [43] crawled the Flickr and Youtube networks using BFS sampling every day on a large period of time (several weeks). In particular, each day they revisited all the users they had previously discovered, and in addition all the new users that were reachable from the previously known users. By choosing a reasonable sets of users to start with, they have access to the whole largest connected component of these networks. Further, provided the BFS sampling goes well, their crawling mechanism may be viewed as observing snapshots of a graphex process at increasing times.

We plotted side-by-side the Hill plot and the degree distribution, respectively for the Flickr and Youtube networks, in figs. 4 and 5. We plotted the degree

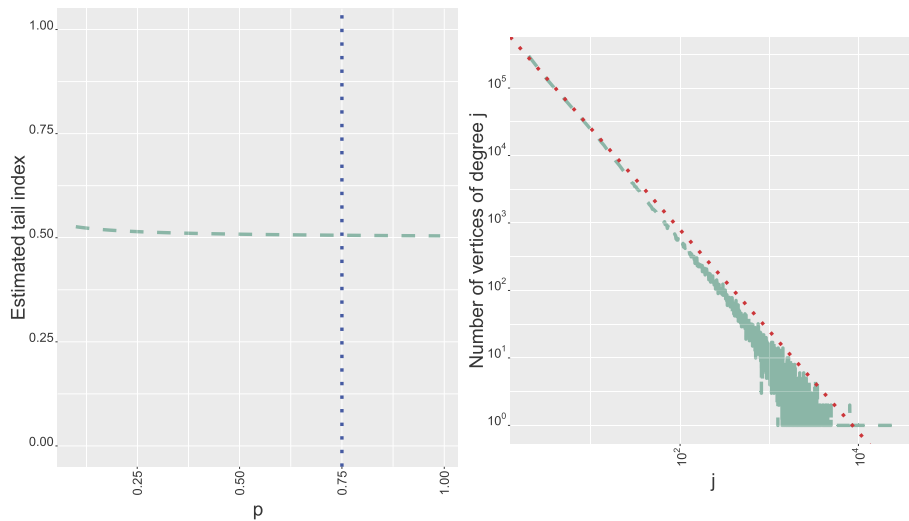


FIG 4. On the left: The Hill plot of  $\hat{\sigma}_p$  estimated on the Flickr dataset, that is the plot of  $p \mapsto \hat{\sigma}_p$ . The vertical blue line at  $p = 0.75$  corresponds to the retained value for the estimation of  $\sigma$ . We notice that the Hill plot is essentially flat in this case, which is the expected behaviour when the model is well-specified. On the right: In green is the degree distribution of the Flickr dataset. In red is the line corresponding to  $d_j \propto j^{-1-\hat{\sigma}_p}$ . We see that both curves align well.

distributions on a log-log scale, and added in red dotted lines a linear curve with slope  $-(1 + \hat{\sigma}_p)$ , where  $\hat{\sigma}_p$  has been computed using  $p = 0.75$ . In case the model is correct, we then expect the degree distribution and the linear curve to align well. We can see in fig. 6 that the Flickr dataset exhibits some features of sparse graphex processes: its Hill plot is relatively flat with value  $\approx 0.5$ , and the slope of its degree distribution on a log-log scale is close to  $\approx -1.5$ . Regarding the Youtube dataset, in fig. 5, this is less evident, even the plots plea in favor of a misspecification of the sparse graphex model. This might be because the model is not reasonable, which is likely to be caused by the crawling procedure

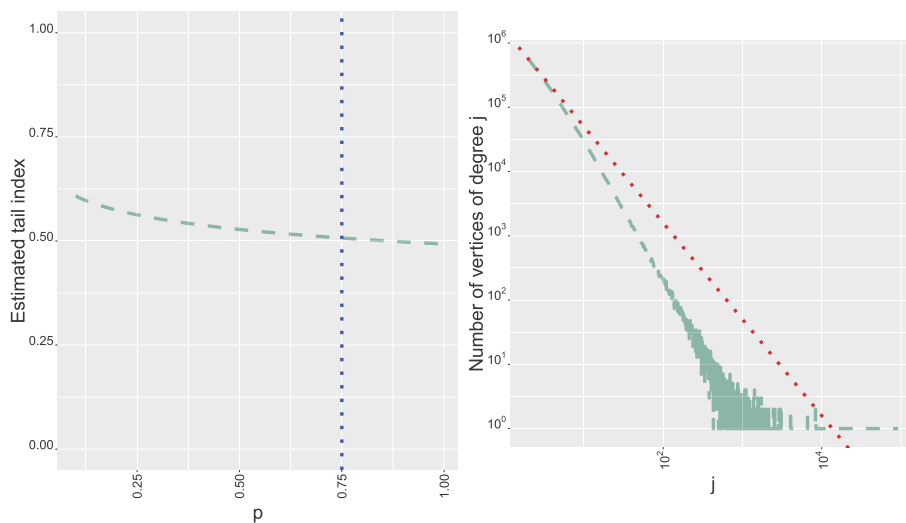


FIG 5. On the left: The Hill plot of  $\hat{\sigma}_p$  estimated on the Youtube dataset, that is the plot of  $p \mapsto \hat{\sigma}_p$ . The vertical blue line at  $p = 0.75$  corresponds to the retained value for the estimation of  $\sigma$ . We notice that the Hill plot is not flat, which should be considered as a warning that the model might be inadequate. On the right: In green is the degree distribution of the Youtube dataset. In red is the line corresponding to  $d_j \propto j^{-1-\hat{\sigma}_p}$ .

used in [43]: the data they have at their disposal make uncertain that all users in Youtube have effectively been sampled [43, Section 4.2.4]<sup>8</sup>.

Finally, we mention that these plots should only be used as visual diagnostics, and it is hard to draw firm conclusions about the validity of the model from these plots. One would need in particular to do a true statistical test for the degree distribution, yet this is unclear how to perform at this time.

## 6.2. Counting motifs

### 6.2.1. Consistent estimation of count statistics

We can also use the results of section 5.2 to count the number of motifs in a network. We illustrate this on the number of triangles in the Youtube network of the previous section. We note here that the number of triangles is equal to  $n(R, G_t)/6$  where  $R$  is a triangle graph. We also note that the method is consistent regardless of the distribution of  $G_t$ , whence it does not matter that we found evidence that the Youtube dataset is unlikely to be a sample of a sparse graphex process. We plot in fig. 6 the estimated number of triangles

<sup>8</sup>To be complete, the same is true for the Flickr network, but in the case of Flickr, the authors of [43] were able to quantify the fraction of missed users [43, Section 4.2.1]. They found that this fraction is likely to be very small. In the case of Youtube, they were unable to quantify the fraction of missing users.

and the estimated standard deviation as a function of the number of bootstrap samples used. The bootstrap samples used are small ( $p = 0.01$ ), implying that in average each sample has  $\approx 12$  triangles, which is fast to count. Yet, with only 2000 samples (which can all be computed in parallel) the accuracy is already quite good.

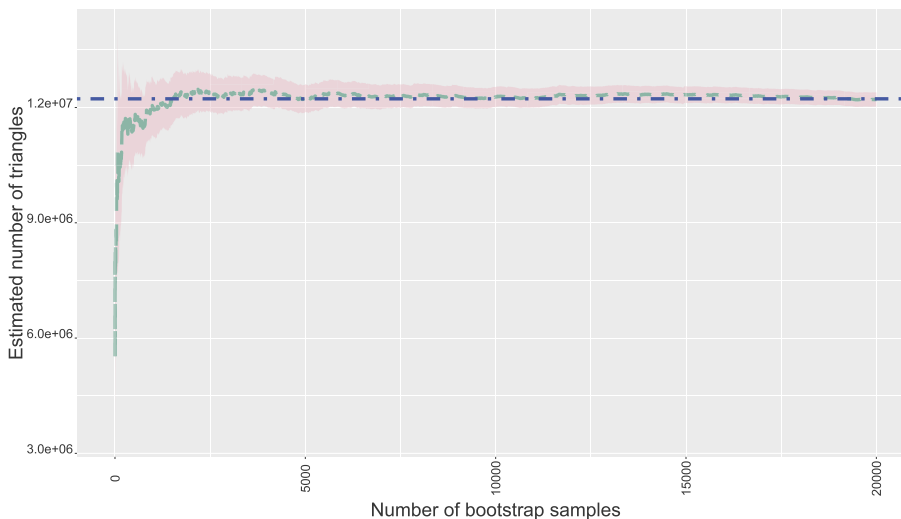


FIG 6. In green is the plot of (one realization of) the estimated number of triangles using  $p$ -sampling of the Youtube dataset, as a function of the number of bootstrap sample used ( $p = 0.01$ ). The horizontal blue line corresponds to the true number of triangles. The pink region is  $[\hat{n}(R, G) \pm 2\delta]$ , where  $\delta^2$  is the empirical estimate of  $\text{var}(\hat{n}(R, G) | G)$ .

### 6.2.2. Testing for count statistics

We have seen in section 6.1 that the degree distribution of the Flickr network was plausibly well explained as a sample from a graphex process. In particular it is found that  $\sigma = 0.5$  does explain quite well the “sparsity” and the degree distribution. The current state of the art algorithm for inference on graphex processes is the algorithm of [13] and its subsequent generalization in [51]. Here we focus on the simpler version of Caron and Fox [13]. Caron and Fox’s algorithm is based on the GGP model previously described in example 5. Hence the model is a parametric family with parameter set  $\{W \equiv W_{\sigma, \tau} : \sigma < 1, \tau > 0\}$ . One of the desirable feature of the GGP model is that the parameter  $\sigma$  is exactly the tail-index of  $\mu$  described in section 3.4<sup>9</sup>. Yet simple, this model is able to accommodate for  $\sigma \in [0, 1)$ , that is from dense to sparse with power-law. The other parameter  $\tau$  acts as a cut-off for the degree distribution, controlling the separation between

<sup>9</sup>To be rigorous, this is true for  $\sigma \in [0, 1)$ , but the GGP model also allows for  $\sigma < 0$ . Then the parameter cannot be interpreted as the tail-index of  $\mu$  anymore (which is zero).

the power-law regime and non power-law regime (corresponding to larger degrees), see [13]. In particular, the GGP model with  $(\sigma, \tau, t) = (0.5, 6.3, 12000)$  satisfies  $\mathbb{E}_W[v(G_t)] \approx 2.3$  millions and  $\mathbb{E}_W[e(G_t)] \approx 23$  millions. So, the degree distribution and the size of Flickr are well explained by the GGP model with  $(\sigma, \tau, t) = (0.5, 6.3, 12000)$ , as emphasized by table 3 and fig. 7. Due to its relative simplicity, this model enables for efficient and tractable computations [13, 51]. Then, it makes sense to determine if other aspects of the structure of Flickr can be as well explained by the same model. Here we test if the number of triangles observed in Flickr is compatible with the GGP model with  $H_0 : (\sigma_0, \tau_0, t_0) = (0.5, 6.3, 12000)$ . Under  $H_0$ , by numerical integration we find that  $\bar{n}(R, W_0, t_0) = 3.75 \cdot 10^8$  when  $R$  is a triangle. Using the bootstrap procedure of section 5 on the Flickr dataset (we used  $p = 0.1$  and 2000 bootstrap samples), we estimated  $\bar{n}(R, G_t) = 5.03 \cdot 10^9$  and  $\sqrt{\hat{s}^2(G_t)} = 1.93 \cdot 10^8$ . This gives an observed value of the test statistic  $T(R, G_t)$  of approximately 24.2, and then a  $p$ -value  $\approx 0$ . Hence  $H_0$  is rejected in this case: without surprise it does not seem reasonable that the Flickr dataset is well-explained by a model as simple as the GGP model. This emphasize the need of more advanced algorithm for inference, in particular for non-parametric estimation.

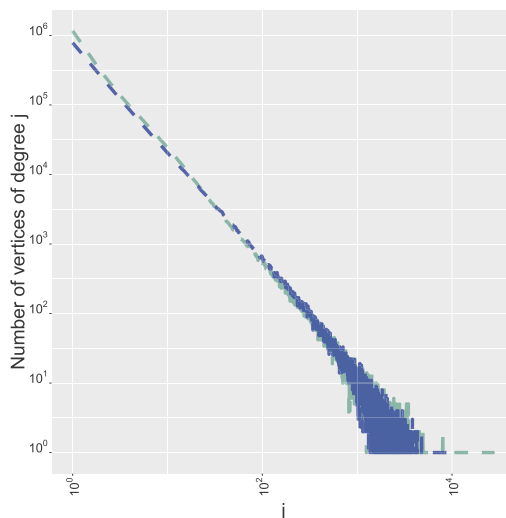


FIG 7. In green is the degree distribution of the Flickr dataset. In blue, the degree distribution of one sample of a GGP model with parameters  $(\sigma, \tau, t) = (0.5, 6.3, 12000)$ .

### 6.3. Test-train split

Model evaluation is a key component of data analysis. Often, this involves randomly splitting the available data into a test set and a training set. There are many seemingly natural ways to partition a graph, so some care is required in



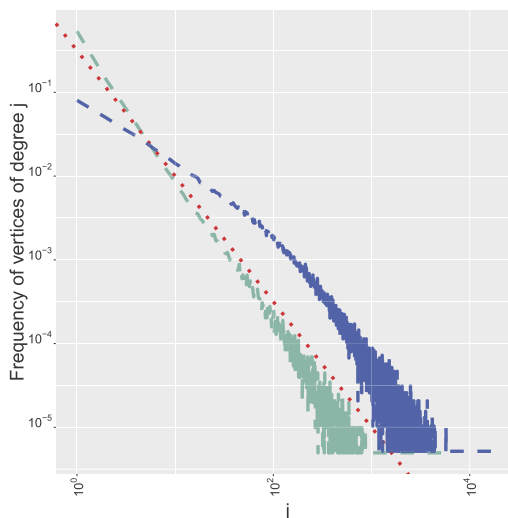


FIG 8. In green: the degree distribution of a  $p$ -sampling of the Flickr dataset with  $p = 0.2$  (the sample has 201633 vertices and 952723 edges). In blue: the degree distribution of a star sampling of the Flickr dataset with  $p' = 0.013\dots$  (the sample has 193075 vertices and 15739418 edges). In red is  $d_j \propto j^{-1.5}$  which was found to be a credible estimate of the degree distribution of the complete Flickr dataset.

splitting the data. The choice of partitioning scheme may induce a significant sampling bias in the test and training sets, impeding evaluation and complicating model comparison. Consider for instance the case of independent and identically distributed observations  $(Y_1, \dots, Y_n) \stackrel{iid}{\sim} P^n$ : it is natural to split the data  $(Y_1, \dots, Y_n)$  by subsampling uniformly without replacement  $0 \leq m \leq n$  observations within the sample. This is because the obtained subsample has the desirable property to be distributed according to  $P^m$ . We argue that in the context of graph processes, the natural way to split the graph is precisely the  $p$ -sampling procedure described in section 3.3, because it is the only procedure leaving (up to rescaling) the distribution of the graph invariant [53, 9].

To illustrate the difficulty with test–train splitting of graphs, we built two training sets from the Flickr dataset: one using the  $p$ -sampling procedure, and the other using the so-called *star sampling* [11, 35]. The star sampling we used consists on selecting a set of vertices by uniform sampling with probability  $p'$  (the stars centers), and for each of these center select all of its neighbors (the star points), and then returning the induced subgraph. We used  $p = 0.2$  and  $p' = 0.013\dots$  to obtain training sets with a comparable number of vertices (respectively 201633 for  $p$ -sampling and 193075 for star sampling). However, the structure of the obtained training sets are very different. The  $p$ -sample training set has 952723 edges, while the star sample has 15739418 edges. Whence the star training set is much denser, as expected. The degree distributions of the two training sets are also quite different, as illustrated in fig. 8. On the same

figure, we drew a line with slope  $-1.5$ , which was found to explain well the degree distribution of the complete dataset. We see that it adjusts also quite well the degree distribution of the  $p$ -sampled training-set. So at least in term of degree distribution, the  $p$ -sampling training set captures better the structure of the complete dataset.

## Acknowledgments

We are thankful to François Caron for feedback on earlier versions of this manuscript and for pointing us many interesting directions and references, leading to a substantially improved document. We also thank the anonymous referees for helpful comments that have also substantially improved the manuscript. This work was supported by U.S. Air Force Office of Scientific Research grant #FA9550-15-1-0074.

## References

- [1] ACHLIOPTAS, D., CLAUSET, A., KEMPE, D. and MOORE, C. (2009). On the Bias of Traceroute Sampling: Or, Power-Law Degree Distributions in Regular Graphs. *Journal of the ACM (JACM)* **56** 1–28. [MR2590823](#)
- [2] ALI, W., WEGNER, A. E., GAUNT, R. E., DEANE, C. M. and REINERT, G. (2016). Comparison of Large Networks With Sub-Sampling Strategies. *Scientific reports* **6** 28955.
- [3] ASSADI, S., KAPRALOV, M. and KHANNA, S. (2018). A Simple Sublinear-Time Algorithm for Counting Arbitrary Subgraphs Via Edge Sampling. *arXiv preprint arXiv:1811.07780*. [MR3899800](#)
- [4] BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of Scaling in Random Networks. *Science* **286** 509–512. [MR2091634](#)
- [5] BHATTACHARYYA, S. and BICKEL, P. J. (2015). Subsampling Bootstrap of Count Features of Networks. *Annals of Statistics* **43** 2384–2411. [MR3405598](#)
- [6] BICKEL, P. J. and CHEN, A. (2009). A Nonparametric View of Network Models and Newman-Girvan and Other Modularities. *Proceedings of the National Academy of Sciences* **106** 21068–21073.
- [7] BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The Method of Moments and Degree Distributions for Network Models. *Annals of Statistics* **39** 2280–2301. [MR2906868](#)
- [8] BORGS, C., CHAYES, J. T., COHN, H. and HOLDEN, N. (2016). Sparse Exchangeable Graphs and Their Limits Via Graphon processes. *ArXiv e-prints*. [MR3827098](#)
- [9] BORGS, C., CHAYES, J., COHN, H. and VEITCH, V. (2017). Sampling perspectives on sparse exchangeable graphs. Preprint. [MR4021236](#)
- [10] BROIDO, A. D. and CLAUSET, A. (2019). Scale-Free Networks Are Rare. *Nature communications* **10** 1–10.

- [11] CAPOBIANCO, M. (1972). Estimating the Connectivity of a Graph. *Lecture Notes in Mathematics* 65–74. [MR0332542](#)
- [12] CARON, F. (2012). Bayesian nonparametric models for bipartite graphs. In *Advances in Neural Information Processing Systems 25* (F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger, eds.) 2051–2059. Curran Associates, Inc.
- [13] CARON, F. and FOX, E. B. (2017). Sparse Graphs Using Exchangeable Random Measures. *J. Royal Statist. Society Series B* **79** 1295–1366. [MR3731666](#)
- [14] CARON, F., PANERO, F. and ROUSSEAU, J. (2017). On Sparsity, Power-Law and Clustering Properties of Graphex Processes. *arXiv preprint arXiv:1708.03120*.
- [15] CARRINGTON, P. J., SCOTT, J. and WASSERMAN, S. (2005). *Models and methods in social network analysis* **28**. Cambridge university press.
- [16] CLAUSET, A., SHALIZI, C. R. and NEWMAN, M. E. (2009). Power-Law Distributions in Empirical Data. *SIAM review* **51** 661–703. [MR2563829](#)
- [17] CRANE, H. and DEMPSEY, W. (2016). Edge Exchangeable Models for Network Data. *arXiv preprint arXiv:1603.04571*. [MR3862359](#)
- [18] DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*, second ed. Springer. [MR1950431](#)
- [19] DALEY, D. J. and VERE-JONES, D. (2007). *An introduction to the theory of point processes: volume II: general theory and structure* **2**. Springer Science & Business Media. [MR2371524](#)
- [20] DE HAAN, L. and STADTMÜLLER, U. (1996). Generalized Regular Variation of Second Order. *Journal of the Australian Mathematical Society* **61** 381–395. [MR1420345](#)
- [21] DIACONIS, P. and JANSON, S. (2008). Graph Limits and Exchangeable Random Graphs. *Rendiconti di Matematica e delle sue Applicazioni. Serie VII* 33–61. [MR2463439](#)
- [22] DREES, H., DE HAAN, L. and RESNICK, S. (2000). How to make a Hill plot. *Annals of Statistics* **28** 254–274. [MR1762911](#)
- [23] EDEN, T., LEVI, A., RON, D. and SESHADHRI, C. (2017). Approximately Counting Triangles in Sublinear Time. *SIAM Journal on Computing* **46** 1603–1646. [MR3709896](#)
- [24] EFRON, B. and TIBSHIRANI, R. J. (1994). *An introduction to the bootstrap*. CRC Press. [MR1270903](#)
- [25] FELLER, W. (1971). *An introduction to probability theory and its applications, volume II* **2**. Wiley, New York. [MR0270403](#)
- [26] GOLDREICH, O. and RON, D. (2008). Approximating Average Parameters of Graphs. *Random Structures & Algorithms* **32** 473–493. [MR2422391](#)
- [27] GONEN, M., RON, D. and SHAVITT, Y. (2011). Counting Stars and Other Small Subgraphs in Sublinear-Time. *SIAM Journal on Discrete Mathematics* **25** 1365–1411. [MR2837605](#)
- [28] GREEN, A. and SHALIZI, C. R. (2017). Bootstrapping Exchangeable Random Graphs. *arXiv preprint arXiv:1711.00813*.

- [29] HERLAU, T., SCHMIDT, M. N. and MØRUP, M. (2016). Completely random measures for modelling block-structured sparse networks. In *Advances in Neural Information Processing Systems 29* (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon and R. Garnett, eds.) 4260–4268. Curran Associates, Inc.
- [30] HOLME, P. (2019). Rare and everywhere: Perspectives on scale-free networks. *Nature communications* **10** 1–3.
- [31] JANSON, S. (2017). On convergence for graphexes. *ArXiv e-prints*.
- [32] JONES, J. H. and HANDCOCK, M. S. (2003). Sexual contacts and epidemic thresholds. *Nature* **423** 605–606.
- [33] KINGMAN, J. F. C. (1993). *Poisson Processes*. Oxford University Press. [MR1207584](#)
- [34] KLUSOWSKI, J. M. and WU, Y. (2018). Counting motifs with graph sampling. *arXiv preprint arXiv:1802.07773*.
- [35] KOLACZYK, E. D. (2009). *Statistical Analysis of Network Data. Methods and models*. Springer. [MR2724362](#)
- [36] KUNEGIS, J. (2013). KONECT. *Proceedings of the 22nd International Conference on World Wide Web - WWW '13 Companion*.
- [37] LAKHINA, A., BYERS, J. W., CROVELLA, M. and XIE, P. (2003). Sampling biases in IP topology measurements. In *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No. 03CH37428)* **1** 332–341. IEEE.
- [38] LESKOVEC, J., KLEINBERG, J. and FALOUTSOS, C. (2005). Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* 177–187.
- [39] LEVIN, K. and LEVINA, E. (2019). Bootstrapping networks with latent space structure. *arXiv preprint arXiv:1907.10821*.
- [40] LOVÁSZ, L. (2013). *Large Networks and Graph Limits* **60**. American Mathematical Society. [MR3012035](#)
- [41] LUNDE, R. and SARKAR, P. (2019). Subsampling sparse graphons under minimal assumptions. *arXiv preprint arXiv:1907.12528*.
- [42] MIDDENDORF, M., ZIV, E. and WIGGINS, C. H. (2005). Inferring network mechanisms: the *Drosophila melanogaster* protein interaction network. *Proceedings of the National Academy of Sciences* **102** 3192–3197.
- [43] MISLOVE, A., MARCON, M., GUMMADI, K. P., DRUSCHEL, P. and BHATTACHARJEE, B. (2007). Measurement and Analysis of Online Social Networks. *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement - IMC '07*.
- [44] NEWMAN, M. E. (2005). Power laws, Pareto distributions and Zipf’s law. *Contemporary physics* **46** 323–351.
- [45] ORBANZ, P. and ROY, D. M. (2015). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE transactions on pattern analysis and machine intelligence* **37** 437–461.
- [46] REITZNER, M. and SCHULTE, M. (2013). Central Limit Theorems for  $U$ -statistics of Poisson Point Processes. *Annals of Probability* **41** 3879–3909.

- [MR3161465](#)
- [47] SHEN-ORR, S. S., MILO, R., MANGAN, S. and ALON, U. (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature genetics* **31** 64–68.
  - [48] SURI, S. and VASSILVITSKII, S. (2011). Counting triangles and the curse of the last reducer. In *Proceedings of the 20th international conference on World wide web* 607–614.
  - [49] THOMPSON, S. K. (2012). *Sampling*, 3rd ed. Wiley. [MR2894042](#)
  - [50] THOMPSON, S. K. and FRANK, O. (2000). Model-based estimation with link-tracing sampling designs. *Survey methodology* **26** 87–98.
  - [51] TODESCHINI, A., MISCOURIDOU, X. and CARON, F. (2016). Exchangeable Random Measures for Sparse and Modular Graphs with Overlapping Communities.
  - [52] VEITCH, V. and ROY, D. M. (2015). The class of random graphs arising from exchangeable random measures. *arXiv preprint arXiv:1512.03099*. [MR3781748](#)
  - [53] VEITCH, V. and ROY, D. M. (2019). Sampling and Estimation for (Sparse) Exchangeable Graphs. *Annals of Statistics* **47** 3274–3299. [MR4025742](#)
  - [54] VOITALOV, I., VAN DER HOORN, P., VAN DER HOFSTAD, R. and KRIKOUKOV, D. (2019). Scale-free networks well done. *Physical Review Research* **1** 033034.
  - [55] WANG, T. and RESNICK, S. I. (2019). Consistency of Hill estimators in a linear preferential attachment model. *Extremes* **22** 1–28. [MR3919737](#)
  - [56] WATTS, D. J. and STROGATZ, S. H. (1998). Collective dynamics of ‘small-world’ networks. *nature* **393** 440–442.

## Appendix A: Proofs related to tail-index estimation

### A.1. Preliminaries

We let  $\mathbf{B} := (B_{i,j})_{(i,j) \in \mathbb{N}^2}$  be the symmetric array of binary variables that indicate that  $v_i$  and  $v_j$  are connected if  $B_{i,j} = 1$ , not connected if  $B_{i,j} = 0$ . All other quantities are defined in the main document. We drop out the subscript  $W$  for convenience, and  $\mathbb{E}$  shall be interpreted as  $\mathbb{E}_W$ . Also, to simplify notations, we write  $N_{1,t} \equiv N_1(G_t)$ ,  $N_{p,t} \equiv N_p(G_t)$ , and  $\hat{\sigma}_{p,t} \equiv \hat{\sigma}_p(G_t)$ .

### A.2. Bias-variance decomposition

The starting point of the proof is to decompose the risk onto a deterministic bias term and some stochastic variance terms. Then, when  $N_{p,t} = 0$  we have

$$(\hat{\sigma}_{p,t} - \sigma)^2 = \sigma^2 \leq 1. \quad (\text{A.1})$$

In the situation where  $N_{p,t} \geq 1$ , we have

$$-\log(p)\hat{\sigma}_{p,t} = \log \frac{\mathbb{E}N_{1,t}}{\mathbb{E}N_{p,t}} + \log p + \log \left( 1 + \frac{N_{1,t} - \mathbb{E}N_{1,t}}{\mathbb{E}N_{1,t}} \right) - \log \left( 1 + \frac{N_{p,t} - \mathbb{E}N_{p,t}}{\mathbb{E}N_{p,t}} \right).$$

To ease notations, we define,

$$b_{\sigma,t} := \frac{\log(\mathbb{E}N_{1,t}/\mathbb{E}N_{p,t})}{-\log p} - 1 - \sigma, \quad Z_p := \frac{N_{p,t} - \mathbb{E}N_{p,t}}{\mathbb{E}N_{p,t}}.$$

Note that  $Z_1, Z_p > -1$  because  $N_{1,t} \geq N_{p,t} \geq 1$ . Then when  $N_{p,t} \geq 1$ ,

$$\begin{aligned} (\hat{\sigma}_{p,t} - \sigma)^2 &= \left( b_{\sigma,t} + \frac{\log(1+Z_1)}{-\log p} - \frac{\log(1+Z_p)}{-\log p} \right)^2 \\ &= b_{\sigma,t}^2 + 2b_{\sigma,t} \frac{\log(1+Z_1)}{-\log p} - 2b_{\sigma,t} \frac{\log(1+Z_p)}{-\log p} \\ &\quad + \frac{\log^2(1+Z_1)}{(-\log p)^2} + \frac{\log^2(1+Z_p)}{(-\log p)^2} - 2 \frac{\log(1+Z_1)\log(1+Z_p)}{(-\log p)^2}. \end{aligned}$$

We now introduce functions  $\varphi_1, \varphi_2 : (-1, \infty) \rightarrow \mathbb{R}_+$  such that for every  $z \neq 0$ ,

$$\varphi_1(z) := \frac{\log(1+z)}{z}, \quad \varphi_2(z) := -\frac{\log(1+z) - z}{z^2}.$$

For  $z = 0$  the functions  $\varphi_1$  and  $\varphi_2$  are extended by continuity. The functions  $\varphi_1$  and  $\varphi_2$  are non-negative and monotonically decreasing on  $(-1, \infty)$ . We can therefore write when  $N_{p,t} \geq 1$ ,

$$\begin{aligned} (\hat{\sigma}_{p,t} - \sigma)^2 &= b_{\sigma,t}^2 + \frac{Z_1^2 \varphi_1(Z_1)^2}{(-\log p)^2} + \frac{Z_p^2 \varphi_1(Z_p)^2}{(-\log p)^2} - 2 \frac{Z_1 Z_p \varphi_1(Z_1) \varphi_1(Z_p)}{(-\log p)^2} \\ &\quad + \frac{2b_{\sigma,t} Z_1}{-\log p} - \frac{2b_{\sigma,t} Z_1^2 \varphi_2(Z_1)}{-\log p} - \frac{2b_{\sigma,t} Z_p}{-\log p} + \frac{2b_{\sigma,t} Z_p^2 \varphi_2(Z_p)}{-\log p}. \quad (\text{A.2}) \end{aligned}$$

But, by Young's inequality we have  $2|Z_1 Z_p \varphi_1(Z_1) \varphi_1(Z_p)| \leq Z_1^2 \varphi_1(Z_1)^2 + Z_p^2 \varphi_1(Z_p)^2$ . Combining the Young inequality estimate with equations (A.1) and (A.2) gives

$$\begin{aligned} \mathbb{E}[(\hat{\sigma}_{p,t} - \sigma)^2] &= \mathbb{E}[(\hat{\sigma}_{p,t} - \sigma)^2 \mathbf{1}_{N_{p,t}=0}] + \mathbb{E}[(\hat{\sigma}_{p,t} - \sigma)^2 \mathbf{1}_{N_{p,t} \geq 1}] \\ &\leq \mathbb{P}(N_{p,t} = 0) + b_{\sigma,t}^2 + \frac{2\mathbb{E}[Z_1^2 \varphi_1(Z_1)^2 \mathbf{1}_{N_{p,t} \geq 1}]}{(-\log p)^2} + \frac{2b_{\sigma,t} \mathbb{E}[Z_1 \mathbf{1}_{N_{p,t} \geq 1}]}{-\log p} \\ &\quad + \frac{2\mathbb{E}[Z_p^2 \varphi_1(Z_p)^2 \mathbf{1}_{N_{p,t} \geq 1}]}{(-\log p)^2} - \frac{2b_{\sigma,t} \mathbb{E}[Z_p \mathbf{1}_{N_{p,t} \geq 1}]}{-\log p} \\ &\quad + \frac{2b_{\sigma,t} \mathbb{E}[Z_p^2 \varphi_2(Z_p) \mathbf{1}_{N_{p,t} \geq 1}]}{-\log p} - \frac{2b_{\sigma,t} \mathbb{E}[Z_1^2 \varphi_2(Z_1) \mathbf{1}_{N_{p,t} \geq 1}]}{-\log p}. \end{aligned}$$

Moreover, because  $\mathbb{E}[Z_1] = \mathbb{E}[Z_p] = 0$  we get by Hölder's inequality the following estimates.

$$\begin{aligned}\mathbb{E}[Z_1 \mathbf{1}_{N_{p,t} \geq 1}] &= \mathbb{E}[Z_1(1 - \mathbf{1}_{N_{p,t}=0})] \leq \mathbb{E}[|Z_1| \mathbf{1}_{N_{p,t}=0}] \leq \sqrt{\mathbb{E}[Z_1^2] \mathbb{P}(N_{p,t} = 0)} \\ \mathbb{E}[Z_p \mathbf{1}_{N_{p,t} \geq 1}] &= \mathbb{E}[Z_p(1 - \mathbf{1}_{N_{p,t} = 0})] = \mathbb{P}(N_{p,t} = 0).\end{aligned}$$

Also, because  $\mathbb{E}[N_{p,t}] > 0$  (see proposition 7 below), by Chebychev inequality, for any  $t < 1$ ,

$$\mathbb{P}(N_{p,t} = 0) \leq \mathbb{P}(N_{p,t} < \mathbb{E}N_{p,t}(1-t)) \leq \mathbb{P}(Z_p < -t) \leq \frac{\mathbb{E}[Z_p^2]}{t^2}.$$

Since this is true for all  $t < 1$ , we certainly have  $\mathbb{P}(N_{p,t} = 0) \leq \mathbb{E}[Z_p^2]$ . Furthermore,  $N_{1,t} \geq N_{p,t} \geq 1$  implies that  $Z_1 \geq -1 + (\mathbb{E}N_{1,t})^{-1}$ , then

$$\varphi_1(Z_1) \mathbf{1}_{N_{p,t} \geq 1} \leq \frac{\log \mathbb{E}N_{1,t}}{1 - (\mathbb{E}N_{1,t})^{-1}}.$$

Also,

$$\varphi_2(Z_1) = \varphi_1(Z_1) - \frac{(1 + Z_1) \log(1 + Z_1) - Z_1}{Z_1^2} \leq \varphi_1(Z_1).$$

Obviously, the same estimates hold for  $Z_p$ . Combining all the previous estimates yield the bound (where  $(x)_+ = x$  if  $x \geq 0$  and  $(x)_+ = 0$  if  $x < 0$ ),

$$\begin{aligned}\mathbb{E}[(\hat{\sigma}_{p,t} - \sigma)^2] &\leq b_{\sigma,t}^2 + \left[1 + \frac{2(-b_{\sigma,t})_+}{-\log p}\right] \mathbb{E}[Z_p^2] + \frac{2(b_{\sigma,t})_+}{-\log p} \sqrt{\mathbb{E}[Z_1^2] \mathbb{E}[Z_p^2]} \\ &+ \left[\frac{\log \mathbb{E}N_{1,t}}{1 - (\mathbb{E}N_{1,t})^{-1}}\right]^2 \frac{2\mathbb{E}[Z_1^2]}{(-\log p)^2} + \left[\frac{\log \mathbb{E}N_{1,t}}{1 - (\mathbb{E}N_{1,t})^{-1}}\right] \frac{2(-b_{\sigma,t})_+ \mathbb{E}[Z_1^2]}{-\log p} \\ &+ \left[\frac{\log \mathbb{E}N_{p,t}}{1 - (\mathbb{E}N_{p,t})^{-1}}\right]^2 \frac{2\mathbb{E}[Z_p^2]}{(-\log p)^2} + \left[\frac{\log \mathbb{E}N_{p,t}}{1 - (\mathbb{E}N_{p,t})^{-1}}\right] \frac{2(b_{\sigma,t})_+ \mathbb{E}[Z_p^2]}{-\log p}.\end{aligned}$$

Then the conclusion of theorem 1 follows from the estimates in sections A.3 and A.4.

### A.3. Study of the bias term $b_{\sigma,t}$

The estimate for the bias term follows from the expectation of  $N_{p,t}$  for  $p \in [0, 1]$  given in the next proposition. Remark that from the definition of  $N_{p,t}$ , recalling  $\mathcal{V} = \{(t_1, x_1), \dots\}$ , we have

$$N_{p,t} \stackrel{d}{=} \begin{cases} p \sum_i \mathbf{1}_{t_i \leq t} \left(1 - (1-p)^{\sum_{j \neq i} B_{i,j} \mathbf{1}_{t_j \leq t}}\right) & \text{if } 0 \leq p < 1, \\ \sum_i \mathbf{1}_{t_i \leq t} \mathbf{1}\{\sum_{j \neq i} B_{i,j} \mathbf{1}_{t_j \leq t} > 0\} & \text{if } p = 1. \end{cases}$$

**Proposition 7.** For any  $p \in [0, 1]$  and  $t > 0$  it holds,

$$\mathbb{E}[N_{p,t}] = pt \int_{\mathbb{R}_+} (1 - e^{-pt\mu(x)}) \, dx.$$

*Proof.* We assume here that  $p \neq 1$ . The case  $p = 1$  follows from similar steps, or by taking cautiously the limit  $p \rightarrow 1$  in the proof. Then we have,

$$\begin{aligned} \mathbb{E}[N_{p,t}] &= p\mathbb{E} \left[ \sum_i \mathbf{1}_{t_i \leq t} \left( 1 - (1-p)^{\sum_{j \neq i} B_{i,j} \mathbf{1}_{t_j \leq t}} \right) \right] \\ &= p\mathbb{E} \left[ \sum_i \mathbf{1}_{t_i \leq t} \left( 1 - \mathbb{E} \left[ (1-p)^{\sum_{j \neq i} B_{i,j} \mathbf{1}_{t_j \leq t}} \mid \mathcal{V} \right] \right) \right] \\ &= p\mathbb{E} \left[ \sum_i \mathbf{1}_{t_i \leq t} \left( 1 - \prod_{j \neq i} \mathbb{E} \left[ (1-p)^{B_{i,j} \mathbf{1}_{t_j \leq t}} \mid \mathcal{V} \right] \right) \right], \end{aligned}$$

where the last line follows from independence and dominated convergence theorem to take care of the infinite product. It is easily seen from the definition of  $\mathbf{B}$  that for all  $i \neq j$ ,

$$\mathbb{E}[(1-p)^{B_{i,j} \mathbf{1}_{t_j \leq t}} \mid \mathcal{V}] = 1 - pW(x_i, x_j) \mathbf{1}_{t_j \leq t}.$$

It follows, invoking the Slivnyak-Mecke formula [19, Chapter 13],

$$\begin{aligned} \mathbb{E}[N_{p,t}] &= p\mathbb{E} \left[ \sum_i \mathbf{1}_{t_i \leq t} \left( 1 - \prod_{j \neq i} (1 - pW(x_i, x_j) \mathbf{1}_{t_j \leq t}) \right) \right] \\ &= pt \int_{\mathbb{R}_+} \left( 1 - \mathbb{E} \left[ \prod_j (1 - pW(x, x_j) \mathbf{1}_{t_j \leq t}) \right] \right) \, dx. \end{aligned}$$

Then the conclusion of the proposition follows by Campbell's formula [33, Section 3.2] because,

$$\begin{aligned} \mathbb{E} \left[ \prod_j (1 - pW(x, x_j) \mathbf{1}_{t_j \leq t}) \right] &= \mathbb{E} \left[ \exp \left\{ \sum_j \mathbf{1}_{t_j \leq t} \log(1 - pW(x, x_j)) \right\} \right] \\ &= \exp \left\{ -pt \int_{\mathbb{R}_+} W(x, x') \, dx' \right\}. \quad \square \end{aligned}$$

It is clear from the result of the previous proposition that we have,

$$\begin{aligned} \log \frac{\mathbb{E}[N_{1,t}]}{\mathbb{E}[N_{p,t}]} &= \log \left[ \frac{\int_{\mathbb{R}_+} (1 - e^{-t\mu(z)}) \, dz}{p \int_{\mathbb{R}_+} (1 - e^{-pt\mu(z)}) \, dz} \right] \\ &= -(1 + \sigma) \log p - \log \left[ \frac{\int_{\mathbb{R}_+} (1 - e^{-pt\mu(z)}) \, dz}{p^\sigma \int_{\mathbb{R}_+} (1 - e^{-t\mu(z)}) \, dz} \right]. \end{aligned}$$

It then follows from assumption 3 that,

$$|b_{\sigma,t}| \leq \frac{\Gamma_{p,t}}{-\log p}.$$



#### A.4. Variance estimates for $N_{p,t}$

We now compute the estimate for  $\mathbb{E}[Z_1^2]$  and  $\mathbb{E}[Z_p^2]$  in the proof of the main theorem. We assume here that  $p \neq 1$ . The case  $p = 1$  follows by taking cautiously the limit  $p \rightarrow 1$  in the proof. To shorten coming equations, we write  $q := 1 - p$ . We then have,

$$\begin{aligned} N_{p,t}^2 &= p^2 \sum_i \mathbf{1}_{t_i \leq t} \left( 1 - q^{\sum_{j \neq i} B_{i,j} \mathbf{1}_{t_j \leq t}} \right)^2 \\ &\quad + p^2 \sum_i \sum_{k \neq i} \mathbf{1}_{t_i \leq t} \mathbf{1}_{t_k \leq t} \left( 1 - q^{\sum_{j \neq i} B_{i,j} \mathbf{1}_{t_j \leq t}} \right) \left( 1 - q^{\sum_{j \neq k} B_{k,j} \mathbf{1}_{t_j \leq t}} \right). \end{aligned}$$

We call  $p^2 S_1$  the first term of the rhs of the previous equation, and  $p^2 S_2$  the second term, so that  $N_{p,t}^2 = p^2 S_1 + p^2 S_2$ . That is,

$$\begin{aligned} S_1 &:= \sum_i \mathbf{1}_{t_i \leq t} \left( 1 - q^{\sum_{j \neq i} B_{i,j} \mathbf{1}_{t_j \leq t}} \right)^2 \\ S_2 &:= \sum_i \sum_{k \neq i} \mathbf{1}_{t_i \leq t} \mathbf{1}_{t_k \leq t} \left( 1 - q^{\sum_{j \neq i} B_{i,j} \mathbf{1}_{t_j \leq t}} \right) \left( 1 - q^{\sum_{j \neq k} B_{k,j} \mathbf{1}_{t_j \leq t}} \right). \end{aligned}$$

We now compute  $\mathbb{E}[S_1]$ . Expanding the square,

$$S_1 = \sum_i \mathbf{1}_{t_i \leq t} \left( 1 - 2q^{\sum_{j \neq i} B_{i,j} \mathbf{1}_{t_j \leq t}} + q^{2 \sum_{j \neq i} B_{i,j} \mathbf{1}_{t_j \leq t}} \right)$$

By independence, and using the dominated convergence theorem to take care of the infinite product,

$$\mathbb{E}[S_1 \mid \mathcal{V}] = \sum_i \mathbf{1}_{t_i \leq t} \left( 1 - 2 \prod_{j \neq i} \mathbb{E} \left[ q^{B_{i,j} \mathbf{1}_{t_j \leq t}} \mid \mathcal{V} \right] + \prod_{j \neq i} \mathbb{E} \left[ q^{2B_{i,j} \mathbf{1}_{t_j \leq t}} \mid \mathcal{V} \right] \right).$$

From the definition of  $\mathbf{B}$  we get that,

$$\mathbb{E}[q^{B_{i,j} \mathbf{1}_{t_j \leq t}} \mid \mathcal{V}] = 1 - pW(x_i, x_j) \mathbf{1}_{t_j \leq t}, \quad (\text{A.3})$$

$$\mathbb{E}[q^{2B_{i,j} \mathbf{1}_{t_j \leq t}} \mid \mathcal{V}] = 1 - p(1 + q)W(x_i, x_j) \mathbf{1}_{t_j \leq t}. \quad (\text{A.4})$$

Therefore,

$$\begin{aligned} \mathbb{E}[S_1] &= \mathbb{E} \left[ \sum_i \mathbf{1}_{t_i \leq t} \left\{ 1 - 2 \prod_{j \neq i} (1 - pW(x_i, x_j) \mathbf{1}_{t_j \leq t}) \right. \right. \\ &\quad \left. \left. + \prod_{j \neq i} (1 - p(1 + q)W(x_i, x_j) \mathbf{1}_{t_j \leq t}) \right\} \right]. \end{aligned}$$

Invoking the Slivnyak-Mecke theorem [19, Chapter 13],

$$\begin{aligned} \mathbb{E}[S_1] &= t \int_{\mathbb{R}_+} \left\{ 1 - 2\mathbb{E}[\prod_j (1 - pW(x, x_j) \mathbf{1}_{t_j \leq t})] \right. \\ &\quad \left. + \mathbb{E}[\prod_j (1 - p(1 + q)W(x, x_j) \mathbf{1}_{t_j \leq t})] \right\} dx. \end{aligned}$$

Therefore, by Campbell's formula [33, Section 3.2] (see also the proof of proposition 7),

$$\mathbb{E}[S_1] = t \int_{\mathbb{R}_+} \left\{ 1 - 2e^{-pt\mu(x)} + e^{-p(1+q)t\mu(x)} \right\} dx.$$

That is,

$$\mathbb{E}[S_1] = t \int_{\mathcal{X}} \left\{ 1 - e^{-pt\mu(x)} \right\}^2 \rho(dx) + t \int_{\mathcal{X}} e^{-p(1+q)t\mu(x)} \left\{ 1 - e^{-p^2t\mu(x)} \right\} \rho(dx).$$

We now compute  $\mathbb{E}[S_2]$ . We start with the expansion of the product in the expression of  $S_2$ ,

$$\begin{aligned} S_2 &= \sum_i \sum_{k \neq i} \mathbf{1}_{t_i \leq t} \mathbf{1}_{t_k \leq t} \left( 1 - q^{\sum_{j \neq i} B_{i,j} \mathbf{1}_{t_j \leq t}} - q^{\sum_{j \neq k} B_{k,j} \mathbf{1}_{t_j \leq t}} \right. \\ &\quad \left. + q^{\sum_{j \neq i} B_{i,j} \mathbf{1}_{t_j \leq t} + \sum_{j \neq k} B_{k,j} \mathbf{1}_{t_j \leq t}} \right) \\ &= \sum_i \sum_{k \neq i} \mathbf{1}_{t_i \leq t} \mathbf{1}_{t_k \leq t} \left( 1 - q^{\sum_{j \neq i} B_{i,j} \mathbf{1}_{t_j \leq t}} - q^{\sum_{j \neq k} B_{k,j} \mathbf{1}_{t_j \leq t}} \right. \\ &\quad \left. + q^{2B_{i,k} + \sum_{j \neq i, j \neq k} B_{i,j} \mathbf{1}_{t_j \leq t} + \sum_{j \neq k, j \neq i} B_{k,j} \mathbf{1}_{t_j \leq t}} \right). \end{aligned}$$

The following is justified by independence and dominated convergence theorem,

$$\begin{aligned} \mathbb{E}[S_2 \mid \mathcal{V}] &= \sum_i \sum_{k \neq i} \mathbf{1}_{t_i \leq t} \mathbf{1}_{t_k \leq t} \left( 1 - \prod_{j \neq i} \mathbb{E}[q^{B_{i,j} \mathbf{1}_{t_j \leq t}} \mid \mathcal{V}] \right. \\ &\quad \left. - \prod_{j \neq k} \mathbb{E}[q^{B_{k,j} \mathbf{1}_{t_j \leq t}} \mid \mathcal{V}] + \mathbb{E}[q^{2B_{i,k}} \mid \mathcal{V}] \prod_{j \neq i, k} \mathbb{E}[q^{(B_{i,j} + B_{k,j}) \mathbf{1}_{t_j \leq t}} \mid \mathcal{V}] \right). \end{aligned}$$

That is, because of equations (A.3) and (A.4),

$$\begin{aligned} \mathbb{E}[S_2 \mid \mathcal{V}] &= \sum_i \sum_{k \neq i} \mathbf{1}_{t_i \leq t} \mathbf{1}_{t_k \leq t} \left( 1 - \prod_{j \neq i} (1 - pW(x_i, x_j) \mathbf{1}_{t_j \leq t}) \right. \\ &\quad \left. - \prod_{j \neq k} (1 - pW(x_k, x_j) \mathbf{1}_{t_j \leq t}) + (1 - p(1+q)W(x_i, x_k)) \right. \\ &\quad \left. \times \prod_{j \neq i, k} (1 - pW(x_i, x_j) \mathbf{1}_{t_j \leq t}) (1 - pW(x_k, x_j) \mathbf{1}_{t_j \leq t}) \right). \end{aligned}$$

We rewrite the previous in a slightly different form in order to be able to use the Slivnyak-Mecke theorem [19, Chapter 13],

$$\begin{aligned} \mathbb{E}[S_2 \mid \mathcal{V}] &= \sum_i \sum_{k \neq i} \mathbf{1}_{t_i \leq t} \mathbf{1}_{t_k \leq t} \left( 1 - \prod_{j \neq i} (1 - pW(x_i, x_j) \mathbf{1}_{t_j \leq t}) \right. \\ &\quad \left. - \frac{1 - pW(x_k, x_i)}{1 - pW(x_k, x_k)} \prod_{j \neq i} (1 - pW(x_k, x_j) \mathbf{1}_{t_j \leq t}) \right. \\ &\quad \left. + (1 - p(1+q)W(x_i, x_k)) \frac{\prod_{j \neq i} (1 - pW(x_i, x_j) \mathbf{1}_{t_j \leq t}) (1 - pW(x_k, x_j) \mathbf{1}_{t_j \leq t})}{(1 - pW(x_i, x_k)) (1 - pW(x_k, x_k))} \right). \end{aligned}$$

Then we can apply the Slivnyak-Mecke theorem to find that,

$$\begin{aligned} \mathbb{E}[S_2] &= t \int_{\mathbb{R}_+} \mathbb{E} \left[ \sum_k \mathbf{1}_{t_k \leq t} \left( 1 - \prod_j (1 - pW(x, x_j) \mathbf{1}_{t_j \leq t}) \right. \right. \\ &\quad \left. \left. - \frac{1 - pW(x_k, x)}{1 - pW(x_k, x_k)} \prod_j (1 - pW(x_k, x_j) \mathbf{1}_{t_j \leq t}) \right) \right. \\ &\quad \left. + (1 - p(1+q)W(x, x_k)) \frac{\prod_j (1 - pW(x, x_j) \mathbf{1}_{t_j \leq t})(1 - pW(x_k, x_j) \mathbf{1}_{t_j \leq t})}{(1 - pW(x, x_k))(1 - pW(x_k, x_k))} \right] dx. \end{aligned}$$

Again, we rewrite the previous in a more convenient form to apply the Slivnyak-Mecke theorem a second time,

$$\begin{aligned} \mathbb{E}[S_2] &= t \int_{\mathbb{R}_+} \mathbb{E} \left[ \sum_k \mathbf{1}_{t_k \leq t} \left( 1 - (1 - pW(x, x_k)) \prod_{j \neq k} (1 - pW(x, x_j) \mathbf{1}_{t_j \leq t}) \right. \right. \\ &\quad \left. \left. - (1 - pW(x_k, x)) \prod_{j \neq k} (1 - pW(x_k, x_j) \mathbf{1}_{t_j \leq t}) \right) \right. \\ &\quad \left. + (1 - p(1+q)W(x, x_k)) \prod_{j \neq k} (1 - pW(x, x_j) \mathbf{1}_{t_j \leq t})(1 - pW(x_k, x_j) \mathbf{1}_{t_j \leq t}) \right] dx. \end{aligned}$$

Applying the Slivnyak-Mecke theorem to the previous,

$$\begin{aligned} \mathbb{E}[S_2] &= t^2 \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \left\{ 1 - (1 - pW(x, x')) \mathbb{E}[\prod_j (1 - pW(x, x_j) \mathbf{1}_{t_j \leq t})] \right. \\ &\quad \left. - (1 - pW(x, x')) \mathbb{E}[\prod_j (1 - pW(x', x_k) \mathbf{1}_{t_j \leq t})] \right\} \\ &\quad \left. + (1 - p(1+q)W(x, x')) \mathbb{E}[\prod_j (1 - pW(x, x_j) \mathbf{1}_{t_j \leq t})(1 - pW(x', x_j) \mathbf{1}_{t_j \leq t})] \right\} \\ &\quad \times dx dx'. \end{aligned}$$

Using Campbell's formula [33, Section 3.2] (see also the proof of proposition 7), and recalling  $\nu(x, x') = \int_{\mathbb{R}_+} W(x, y)W(y, x') dy$ ,

$$\begin{aligned} \mathbb{E}[S_2] &= t^2 \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \left\{ 1 - (1 - pW(x, x')) e^{-pt\mu(x)} - (1 - pW(x, x')) e^{-pt\mu(x')} \right. \\ &\quad \left. + (1 - p(1+q)W(x, x')) e^{-pt\mu(x) - pt\mu(x') + p^2 t \nu(x, x')} \right\} dx dx'. \end{aligned}$$

We are now in position to bound  $\mathbb{E}[Z_p^2]$ , using the expression of  $\mathbb{E}[S_1]$ ,  $\mathbb{E}[S_2]$  and proposition 7. proposition 7 gives

$$\begin{aligned} \mathbb{E}[N_{p,t}]^2 &= (pt)^2 \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \left( 1 - e^{-pt\mu(x)} \right) \left( 1 - e^{-pt\mu(x')} \right) dx dx' \\ &= (pt)^2 \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \left\{ 1 - e^{-pt\mu(x)} - e^{-pt\mu(x')} + e^{-pt\mu(x) - pt\mu(x')} \right\} dx dx'. \end{aligned}$$

Combining this with the expression for  $\mathbb{E}[S_1]$  and  $\mathbb{E}[S_2]$ , we get

$$\begin{aligned} \mathbb{E}[N_{p,t}^2] - \mathbb{E}[N_{p,t}]^2 &= t \int_{\mathbb{R}_+} e^{-p(1+q)t\mu(x)} \left\{ 1 - e^{-p^2t\mu(x)} \right\} dx \\ &\quad + t \int_{\mathbb{R}_+} \left\{ 1 - e^{-pt\mu(x)} \right\}^2 dx + 2p^3t^2 \int_{\mathbb{R}_+} \mu(x)e^{-pt\mu(x)} dx \\ &\quad + p^2t^2 \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} e^{-pt\mu(x)-pt\mu(x')+p^2t\nu(x,x')} \left\{ 1 - e^{-p^2t\nu(x,x')} \right\} dx dx' \\ &\quad - p^3(1+q)t^2 \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} W(x,x')e^{-pt\mu(x)-pt\mu(x')+p^2t\nu(x,x')} dx dx'. \end{aligned}$$

By Hölder's and Young's inequality, we have  $2\nu(x,x') \leq 2\sqrt{\mu(x)\mu(x')} \leq \mu(x) + \mu(x')$ . Moreover,  $p^2 \leq p$ , and  $1 - e^{-x} \leq x$  imply,

$$\begin{aligned} \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} e^{-pt\mu(x)-pt\mu(x')+p^2t\nu(x,x')} \left\{ 1 - e^{-p^2t\nu(x,x')} \right\} dx dx' \\ \leq p^2t \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} e^{-\frac{pt}{2}\mu(x)} e^{-\frac{pt}{2}\mu(x')} \nu(x,x') dx dx'. \end{aligned}$$

If  $\sigma = 1$  it follows from assumption 2 that,

$$\begin{aligned} \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} e^{-pt\mu(x)-pt\mu(x')+p^2t\nu(x,x')} \left\{ 1 - e^{-p^2t\nu(x,x')} \right\} dx dx' \\ \leq Cp^2t \left\{ \int_{\mathbb{R}_+} \mu(x) dx \right\}^2, \end{aligned}$$

which is finite because of assumption 1. Now if  $0 \leq \sigma < 1$ , it follows from assumption 2 and [14, Lemma B.4] that,

$$\begin{aligned} \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} e^{-pt\mu(x)-pt\mu(x')+p^2t\nu(x,x')} \left\{ 1 - e^{-p^2t\nu(x,x')} \right\} dx dx' \\ \leq Cp^2t^{1-2\eta} \left\{ \int_{\mathbb{R}_+} \alpha^\eta \mu(x)^\eta e^{-\frac{pt}{2}\mu(x)} dx \right\}^2 \lesssim t^{1-2\eta+2\sigma} \ell(t)^2. \end{aligned}$$

Moreover, using [14, Lemma B.4], it is easily seen that all other terms involved in  $\mathbb{E}[N_{p,t}^2] - \mathbb{E}[N_{p,t}]^2$  are bounded by a multiple constant of  $t^{1+\sigma}\ell(t)$  for any  $\sigma \in [0, 1]$ . Since we set  $\eta = 1$  when  $\sigma = 1$ , it follows the estimate,

$$\mathbb{E}[N_{p,t}^2] - \mathbb{E}[N_{p,t}]^2 \lesssim t^{1+\sigma} \ell(t) + t^{3-2\eta+2\sigma} \ell(t)^2.$$

The conclusion follows because  $\mathbb{E}[N_{p,t}]^2 \gtrsim t^{2+2\sigma} \ell(t)^2$  by proposition 7.

## Appendix B: Proofs related to count statistics

### B.1. Proof of proposition 1

Let  $(Z_v)_{v \in V(G)}$  be a collection of independent Bernoulli( $p$ ) random variables. Let  $G' \subseteq G$  be the subgraph of  $G$  consisting on vertices such  $v \in V(G)$  such that  $Z_v = 1$ . Then  $\text{ps}(G, p)$  is equal in distribution to  $G'$  without its isolated vertices. Since  $R$  is connected  $n(R, \text{ps}(G, p)) \stackrel{d}{=} n(R, G')$ , and hence by equation (5.1)

$$n(R, \text{ps}(G, p)) \stackrel{d}{=} \sum_{\phi \in \mathcal{A}(R, G')} \mathbf{1}\{R \subseteq G'[\phi]\}.$$

Clearly  $V(G') \subseteq V(G)$ , and can assume wlog that  $\mathcal{A}(R, G') \subseteq \mathcal{A}(R, G)$  using canonical embedding. For simplicity let also  $Z_\phi := \prod_{r \in V(R)} Z_{\phi(r)}$  for any  $\phi \in \mathcal{A}(R, G)$ . Then, we have the equivalence  $\phi \in \mathcal{A}(R, G) \wedge Z_\phi = 1 \Leftrightarrow \phi \in \mathcal{A}(R, G')$ . Therefore,

$$\begin{aligned} n(R, \text{ps}(G, p)) &\stackrel{d}{=} \sum_{\phi \in \mathcal{A}(R, G)} Z_\phi \mathbf{1}\{R \subseteq G'[\phi]\} \\ &= \sum_{\phi \in \mathcal{A}(R, G)} Z_\phi \mathbf{1}\{R \subseteq G[\phi]\}, \end{aligned} \quad (\text{B.1})$$

where the second line follows because we also have the equivalence  $\phi \in \mathcal{A}(R, G) \wedge Z_\phi = 1 \Leftrightarrow G'[\phi] = G[\phi]$ . Since  $\phi$  is injective  $\mathbb{E}[Z_\phi] = p^k$ , whence the result.

### B.2. Proof of proposition 2

The first equality in proposition 2 is obvious. Here we focus on computing and bounding  $\text{var}(n(R, \text{ps}(G, p)) \mid G)$ . But this is almost immediate from equation (B.1). Indeed,  $\text{var}(n(R, \text{ps}(G, p)) \mid G)$  equals

$$\sum_{\phi_1 \in \mathcal{A}(R, G)} \sum_{\phi_2 \in \mathcal{A}(R, G)} \text{cov}(Z_{\phi_1}, Z_{\phi_2}) \mathbf{1}\{R \subseteq G[\phi_1], R \subseteq G[\phi_2]\}.$$

Now, if  $\phi_1(V(R)) \cap \phi_2(V(R)) = \emptyset$ , then  $\text{cov}(Z_{\phi_1}, Z_{\phi_2}) = 0$ . In general if  $|\phi_1(V(R)) \cap \phi_2(V(R))| = m$ , then  $\text{cov}(Z_{\phi_1}, Z_{\phi_2}) \leq p^{2k-m}$ . The result follows.

### B.3. Proof of proposition 4

Let us assume for now that there is a universal  $c > 0$  (i.e. depending only on  $W$ ) such that as  $t \rightarrow \infty$

$$s_t^2(R) = \text{var}_W(n(R, G_t)) = ct^{2k-1} + O(t^{2k-2}). \quad (\text{B.2})$$

We delay the proof of this claim to later. Since  $G_{pt} \stackrel{d}{=} \text{ps}(G_t, p)$  by [9, 53], we have  $\text{var}_W(\mathfrak{n}(R_1, G_{pt})) = \text{var}_W(\mathfrak{n}(R, \text{ps}(G_t, p)))$ , and hence as  $t \rightarrow \infty$

$$p^{2k-1} \text{var}_W(\mathfrak{n}(R, G_t)) = \text{var}_W(\mathfrak{n}(R, \text{ps}(G_t, p))) + O(t^{2k-2}).$$

Then, by the law of total variance,

$$\begin{aligned} & p^{2k-1} \text{var}_W(\mathfrak{n}(R, G_t)) \\ &= \mathbb{E}_W[\text{var}(\mathfrak{n}(R, \text{ps}(G_t, p)) \mid G_t)] + \text{var}_W(\mathbb{E}[\mathfrak{n}(R, \text{ps}(G_t, p)) \mid G_t]) + O(t^{2k-2}) \\ &= \mathbb{E}_W[\text{var}(\mathfrak{n}(R, \text{ps}(G_t, p)) \mid G_t)] + p^{2k} \text{var}_W(\mathfrak{n}(R, G_t)) + O(t^{2k-2}), \end{aligned}$$

where the last line follows from proposition 1. Rearranging the last display gives the proposition.

Hence, it remains to prove the equation (B.2). Let  $\tilde{G} \in \mathcal{L}$  be the graph constructed in section 3.2, and let  $\tilde{G}_t \subseteq \tilde{G}$  be the subgraph consisting only on vertices  $v_i = (t_i, x_i)$  with  $t_i \leq t$ . Likewise,  $G_t$  is (in distribution) the subgraph of  $\tilde{G}_t$  consisting only on non-isolated vertices. But, since  $R$  is connected, we have  $\mathfrak{n}(R, G_t) \stackrel{d}{=} \mathfrak{n}(R, \tilde{G}_t)$ . Write  $\mathcal{V} = \{(t_1, x_1), \dots\}$  so that we can assume wlog that  $V(\tilde{G}_t) = \mathbb{N}$  and  $V(R) = \mathbb{N}_k$ , and we can rewrite,

$$\mathfrak{n}(R, G_t) \stackrel{d}{=} \sum_{\substack{i_1, \dots, i_k \in \mathbb{N} \\ i_1 \neq \dots \neq i_k}} \left( \prod_{j=1}^k \mathbf{1}_{t_{i_j} \leq t} \right) \prod_{\{\ell, m\} \in E(R)} B_{i_\ell, i_m}, \quad (\text{B.3})$$

where  $B_{i,j} = 1$  iff there is an edge between  $(t_i, x_i)$  and  $(t_j, x_j)$  in  $\tilde{G}$ . For simplicity, we define  $\tilde{\mathfrak{n}}_t(R, \mathcal{V}) := \mathbb{E}_W[\mathfrak{n}(R, G_t) \mid \mathcal{V}]$ . Then, by the law of total variance, we can decompose

$$\text{var}_W(\mathfrak{n}(R, G_t)) = \mathbb{E}[\text{var}_W(\mathfrak{n}(R, G_t) \mid \mathcal{V})] + \text{var}(\tilde{\mathfrak{n}}_t(R, \mathcal{V})). \quad (\text{B.4})$$

We now estimate the terms involved in the previous rhs. We write  $A_k := \{(i_1, \dots, i_k) \in \mathbb{N}^k : i_1 \neq \dots \neq i_k\}$ , i.e. the set of  $k$ -arrangements of  $\mathbb{N}$ . We also let  $B_{\mathbf{i}} := \prod_{\{\ell, m\} \in E(R)} B_{i_\ell, i_m} \mathbf{1}_{t_{i_\ell} \leq t} \mathbf{1}_{t_{i_m} \leq t}$ , for any  $\mathbf{i} \in A_k$ . Then, by equation (B.3),

$$\text{var}_W(\mathfrak{n}(R, G_t) \mid \mathcal{V}) \stackrel{d}{=} \sum_{\mathbf{i} \in A_k} \sum_{\mathbf{i}' \in A_k} \text{cov}_W(B_{\mathbf{i}}, B_{\mathbf{i}'} \mid \mathcal{V}). \quad (\text{B.5})$$

We remark that  $\text{cov}_W(B_{\mathbf{i}}, B_{\mathbf{i}'} \mid \mathcal{V}) = 0$  whenever  $\{i_\ell, i_m\}$  and  $\{i'_\ell, i'_m\}$  are all distinct. In other words,  $\text{cov}_W(B_{\mathbf{i}}, B_{\mathbf{i}'} \mid \mathcal{V}) \neq 0$  implies that there exists a pair  $(i_\ell, i_m)$  that is equal to  $(i'_\ell, i'_m)$ . So under assumption 4, we deduce from equation (B.5) and the Slivnyak-Mecke, formula that

$$\mathbb{E}[\text{var}_W(\mathfrak{n}(R, G_t) \mid \mathcal{V})] = O(t^{2k-2}). \quad (\text{B.6})$$

Then, it is enough to establish that  $\text{var}(\tilde{n}_t(R, \mathcal{V})) \sim ct^{2k-1}$  to finish the proof. We note that by equation (B.3)

$$\tilde{n}_t(R, \mathcal{V}) \stackrel{d}{=} \sum_{\mathbf{i} \in A_k} \left( \prod_{j=1}^k \mathbf{1}_{t_{i_j} \leq t} \right) \prod_{\{\ell, m\} \in E(R)} W(x_{i_\ell}, x_{i_m}). \quad (\text{B.7})$$

We let  $N_k : A_k^2 \rightarrow \mathbb{Z}_+$  such that  $N_k(\mathbf{i}, \mathbf{i}')$  counts the number of distinct elements in  $\mathbf{i} \cup \mathbf{i}'$ . Then,

$$\begin{aligned} \tilde{n}_t(R, \mathcal{V})^2 &\stackrel{d}{=} \sum_{m=k}^{2k} \sum_{\mathbf{i} \in A_k} \sum_{\mathbf{i}' \in A_k} \mathbf{1}\{N_k(\mathbf{i}, \mathbf{i}') = m\} \\ &\quad \times \left( \prod_{j=1}^k \mathbf{1}_{t_{i_j} \leq t} \mathbf{1}_{t_{i'_j} \leq t} \right) \prod_{\{\ell, m\} \in E(R)} W(x_{i_\ell}, x_{i_m}) W(x_{i'_\ell}, x_{i'_m}). \end{aligned}$$

All the terms in the previous summation have finite expectation under assumption 4 if  $N \geq 2k$ . Also, using Slivnyak-Mecke's formula, we see that the expectation of the term corresponding to  $m = 2k$  is exactly  $\mathbb{E}[\tilde{n}_t(R, \mathcal{V})]^2$  (so indeed it is enough to have  $N = 2k - 1$ ). Similarly, the expectation of the other terms is seen to be  $O(t^m)$ . It follows that,

$$\begin{aligned} \text{var}(\tilde{n}_t(R, \mathcal{V})) &= O(t^{2k-2}) \\ &+ \mathbb{E} \left[ \sum_{\substack{\mathbf{i}, \mathbf{i}' \in A_k \\ N_k(\mathbf{i}, \mathbf{i}') = 2k-1}} \left( \prod_{j=1}^k \mathbf{1}_{t_{i_j} \leq t} \mathbf{1}_{t_{i'_j} \leq t} \right) \prod_{\{\ell, m\} \in E(R)} W(x_{i_\ell}, x_{i_m}) W(x_{i'_\ell}, x_{i'_m}) \right), \end{aligned}$$

which behave as  $ct^{2k-1}$  when  $t \rightarrow \infty$ , for a constant  $c > 0$  depending only on  $W$  and  $R$ .

#### B.4. Proof of proposition 5

We define for simplicity  $c(G_t) := \text{var}(n(R, \text{ps}(G_t, p)) \mid G_t)$ . So, in view of proposition 4 and its proof, we then have,

$$\frac{\hat{s}_t^2(R, G_t)}{s_t^2(R)} = \frac{c(G_t)}{\mathbb{E}_W[c(G_t)]} \left( 1 + O(t^{-1}) \right).$$

Thus it is enough to show that  $\frac{\text{var}(c(G_t))}{\mathbb{E}_W[c(G_t)]^2} = O(t^{-1})$ . We have obtained in appendix B.3 that  $\mathbb{E}_W[c(G_t)] \asymp t^{2k-1}$ , thus the proposition is proved if we show that  $\text{var}_W(c(G_t)) = O(t^{4k-3})$ . We use the same notations and definitions as in appendix B.3. Then, with the same arguments as in appendices B.2 and B.3 we obtain that

$$c(G_t) \stackrel{d}{=} \sum_{m=k}^{2k-1} p^m (1 - p^{2k-m}) C_m, \quad (\text{B.8})$$

where,

$$C_m := \sum_{\substack{\mathbf{i}, \mathbf{i}' \in A_k \\ N_k(\mathbf{i}, \mathbf{i}') = m}} \left( \prod_{j=1}^k \mathbf{1}_{t_{i_j} \leq t} \mathbf{1}_{t_{i'_j} \leq t} \right) \prod_{\{\ell, m\} \in E(R)} B_{i_\ell, i_m} B_{i'_\ell, i'_m}.$$

Therefore,

$$c(G_t)^2 \stackrel{d}{=} \sum_{m_1=k}^{2k-1} \sum_{m_2=k}^{2k-1} p^{m_1} (1-p)^{2k-m_1} p^{m_2} (1-p)^{2k-m_2} C_{m_1} C_{m_2}.$$

It is easily seen that under assumption 4 with  $N \geq 2k_1 + 2k_2 - 2$ , all the terms involved in equation (B.4) have finite expectations. Also, the terms corresponding to  $m_1 + m_2 \leq 4k - 3$  are seen to have expectation  $O(t^{4k-3})$ . But the only way to have  $m_1 + m_2 > 4k - 3$  is to have  $m_1 = 2k - 1$  and  $m_2 = 2k - 1$ . Hence, we deduce that

$$\mathbb{E}_W[c(G_t)^2] = O(t^{4k-3}) + p^{4k-2} (1-p)^2 \mathbb{E}_W[C_{2k-1}^2].$$

We also note that by equation (B.8) and the same arguments that led to the previous formula,

$$\mathbb{E}_W[c(G_t)]^2 = p^{4k-2} (1-p)^2 \mathbb{E}_W[C_{2k-1}]^2 + O(t^{4k-3}). \tag{B.9}$$

Hence, by combining equations (B.4) and (B.9),

$$\text{var}_W(c(G_t)) \leq \text{var}_W(C_{2k-1}) + O(t^{4k-3}).$$

So in fact it is enough to bound  $\text{var}_W(C_{2k-1})$ . We note that  $N_k(\mathbf{i}, \mathbf{i}') = 2k - 1$  means that there is exactly one of the  $i_1, \dots, i_k$  which is equal to one of the  $i'_1, \dots, i'_k$ . Hence, we can rewrite,

$$C_{2k-1} = \sum_{a_1=1}^k \sum_{a_2=1}^k \tilde{C}_{2k-1}^{a_1, a_2},$$

where,

$$\tilde{C}_{2k-1}^{a_1, a_2} := \sum_{\substack{\mathbf{i}, \mathbf{i}' \in A_k, i_{a_1} = i'_{a_2} \\ N_k(\mathbf{i}, \mathbf{i}') = 2k-1}} \left( \prod_{j=1}^k \mathbf{1}_{t_{i_j} \leq t} \mathbf{1}_{t_{i'_j} \leq t} \right) \prod_{\{\ell, m\} \in E(R)} B_{i_\ell, i_m} B_{i'_\ell, i'_m}.$$

And then,

$$\text{var}_W(C_{2k-1}) \lesssim \sum_{a_1=1}^k \sum_{a_2=1}^k \text{var}_W(\tilde{C}_{2k-1}^{a_1, a_2}).$$

To finish the proof, we note that  $\tilde{C}_{2k-1}^{a_1, a_2} \stackrel{d}{=} n(Q, G_t)$ , where  $Q$  is the connected graph with  $2k - 1$  vertices obtained by joining two copies of  $R$ ,  $R_1$  and  $R_2$ , in such a way that vertex  $a_1$  of  $R_1$  and  $a_2$  of  $R_2$  becomes the same vertex in  $Q$ . Hence  $\text{var}_W(C_{2k-1}) = O(t^{4k-3})$  by equation (B.2), and  $\text{var}_W(c(G_t)) = O(t^{4k-3})$  too.



**B.5. Proof of proposition 6**

We let  $W \equiv W_0$  and  $t \equiv t_0$  for simplicity. We define  $\tilde{n}_t(R, \mathcal{V}) := \mathbb{E}_W[n(R, G_t) \mid \mathcal{V}]$  as in appendix B.3. Then, we rewrite the test statistic  $T \equiv T(R, G_t)$  as

$$T = \underbrace{\frac{\tilde{n}_t(R, \mathcal{V}) - \bar{n}(R, W, t)}{\sqrt{\text{var}(\tilde{n}_t(R, \mathcal{V}))}}}_{=:T_1} \cdot \frac{\sqrt{\text{var}(\tilde{n}_t(R, \mathcal{V}))}}{\sqrt{\hat{s}^2(R, G_t)}} + \underbrace{\frac{n(R, G_t) - \tilde{n}_t(R, \mathcal{V})}{\sqrt{\hat{s}^2(R, G_t)}}}_{=:T_2}.$$

We note that  $\frac{\hat{s}^2(R, G_t)}{s_t^2(R)} = 1 + O_p(t^{-1/2})$  by proposition 5 and  $\frac{\text{var}(\tilde{n}_t(R, \mathcal{V}))}{s_t^2(R)} = 1 + O_p(t^{-1})$  by combining equations (B.4) and (B.6). Thus

$$\frac{\text{var}(\tilde{n}_t(R, \mathcal{V}))}{\hat{s}^2(R, G_t)} = 1 + O_p(t^{-1/2}).$$

Also,  $\tilde{n}_t(R, \mathcal{V}) = \mathbb{E}_W[n(R, G_t \mid \mathcal{V})]$ , and by the same argument as before, it is clear that  $T_2 = O_p(T'_2)$ , where  $T'_2 := \frac{n(R, G_t) - \mathbb{E}_W[n(R, G_t)]}{\sqrt{s_t^2(R)}}$ . Clearly  $\mathbb{E}_W[T'_2] = 0$ , and

$$\text{var}_W(T'_2) = \frac{\mathbb{E}[\text{var}_W(n(R, G_t) \mid \mathcal{V})]}{s_t^2(R)} = O_p(t^{-1}),$$

by equation (B.6) again, and consequently  $T_2 = O_p(t^{-1/2})$ . That entails  $T = T_1 + O_p(t^{-1/2})$ , so the conclusion follows from Slutsky's lemma if we show that  $T_1 \rightarrow \mathcal{N}(0, 1)$ . But, we remark that  $\tilde{n}_t(R, \mathcal{V})$  is a  $U$ -statistic of a Poisson process (see the equation (B.7)). We note that it was not the case for  $n(R, G_t)$  because of the randomness in the edge connections. Under the assumption 4 with  $N \geq 4k$ , we have  $T_1 \xrightarrow{d} \mathcal{N}(0, 1)$  by [46, Theorem 5.2].

**Appendix C: Proof of examples rates bounds**

**C.1. Dense graphs**

For any  $t > 0$ ,

$$\begin{aligned} \int_{\mathbb{R}_+} (1 - e^{-t\mu(x)}) dx - \int_{\mathbb{R}_+} (1 - e^{-pt\mu(x)}) dx \\ = \int_0^1 e^{-ptu/2} \{1 - e^{-(1-p)tu/2}\} du \lesssim t^{-1}. \end{aligned}$$

On the other hand, we have for  $t$  large enough,

$$\int_0^1 (1 - e^{-tu/2}) du \gtrsim 1.$$

Hence,  $\Gamma_{p,t} \propto t^{-1}$ .

### C.2. Sparse, almost dense graphs without power law

For any  $t > 0$ ,

$$\begin{aligned}
 \int_{\mathbb{R}_+} \left(1 - e^{-t\mu(x)}\right) dx - \int_{\mathbb{R}_+} \left(1 - e^{-pt\mu(x)}\right) dx \\
 &= \int_0^1 e^{-ptu} \left\{1 - e^{-(1-p)tu}\right\} \frac{du}{u} \\
 &\leq t(1-p) \int_0^{1/t} e^{-ptu} du + \int_{1/t}^1 e^{-ptu} \frac{du}{u} \\
 &\leq (1-p) + \int_1^t \frac{e^{-pv}}{v} dv \lesssim 1.
 \end{aligned}$$

On the other hand, when  $t > 0$  is large enough,

$$\begin{aligned}
 \int_0^1 (1 - e^{-tu}) \frac{du}{u} &= (1 - e^{-t}) \log(t) - \int_0^t \log(v) e^{-v} dv \\
 &\geq (1 - e^{-t}) \log(pt) - \int_1^t \log(v) e^{-v} dv \\
 &\geq (1 - e^{-t}) \log(t) - \log(t) \int_1^t e^{-v} dv \gtrsim \log(t).
 \end{aligned}$$

Hence  $\Gamma_{p,t} \propto 1/\log t$ .

### C.3. Sparse graphs with power law

Let  $b := 1/(1 - \sigma)$ , then for any  $t > 1$ ,

$$\begin{aligned}
 p^\sigma \int_{\mathbb{R}_+} \left(1 - e^{-t\mu(x)}\right) dx - \int_{\mathbb{R}_+} \left(1 - e^{-pt\mu(x)}\right) dx \\
 &= \sigma \left(\frac{\sigma}{1 - \sigma}\right)^\sigma t^\sigma \left\{ p^\sigma \int_0^{\frac{\sigma t}{1 - \sigma}} (1 - e^{-u}) u^{-1 - \sigma} du - \int_0^{\frac{\sigma t}{1 - \sigma}} (1 - e^{-pu}) u^{-1 - \sigma} du \right\} \\
 &= \sigma \left(\frac{\sigma}{1 - \sigma}\right)^\sigma p^\sigma t^\sigma \int_{\frac{p\sigma t}{1 - \sigma}}^{\frac{\sigma t}{1 - \sigma}} (1 - e^{-u}) u^{-1 - \sigma} du \leq \frac{\sigma(1 - p)}{p}.
 \end{aligned}$$

On the other hand, when  $t$  is large enough,

$$\begin{aligned}
 \int_{\mathbb{R}_+} \left(1 - e^{-t\mu(x)}\right) dx &= t^\sigma \sigma \left(\frac{\sigma}{1 - \sigma}\right)^\sigma \int_0^{\frac{\sigma t}{1 - \sigma}} (1 - e^{-v}) v^{-1 - \sigma} dv \\
 &\geq t^\sigma \sigma \left(\frac{\sigma}{1 - \sigma}\right)^\sigma \int_0^{\frac{\sigma t}{1 - \sigma}} e^{-v} v^{-\sigma} dv \gtrsim t^\sigma.
 \end{aligned}$$

Hence  $\Gamma_{p,t} \propto t^{-\sigma}$ .

**C.4. Generalized gamma process**

In the proof we assume without loss of generality that  $\tau = 1$ . By definition, we can see that the graphex marginal is given by

$$\begin{aligned} \mu(x) &= \int_{\mathbb{R}_+} \left(1 - e^{-2\bar{\rho}_{\sigma,\tau}^{-1}(x')\bar{\rho}_{\sigma,\tau}^{-1}(x)}\right) dx' \\ &= \int_{\mathbb{R}_+} \left(1 - e^{-2y\bar{\rho}_{\sigma,\tau}^{-1}(x)}\right) \rho_{\sigma,\tau}(dy) \\ &= \frac{(1 + 2\bar{\rho}_{\sigma,\tau}^{-1}(x))^\sigma - 1}{\sigma}. \end{aligned}$$

Then, for any  $p \in [0, 1]$ ,

$$\begin{aligned} \int_{\mathbb{R}_+} \left(1 - e^{-pt\mu(z)}\right) dz &= \int_{\mathbb{R}_+} \left(1 - \exp\left(-pt \frac{(1 + 2\bar{\rho}_{\sigma,\tau}^{-1}(x))^\sigma - 1}{\sigma}\right)\right) dx \\ &= \int_{\mathbb{R}_+} \left(1 - \exp\left(-pt \frac{(1 + 2y)^\sigma - 1}{\sigma}\right)\right) \rho_{\sigma,\tau}(dy). \end{aligned}$$

Since  $\tau = 1$  by assumption, the previous can be rewritten as,

$$\begin{aligned} \int_{\mathbb{R}_+} \left(1 - e^{-pt\mu(z)}\right) dz &= \frac{2^\sigma}{pt} \int_{\mathbb{R}_+} (1 - e^{-u}) \left(1 + \frac{\sigma u}{pt}\right)^{-1+1/\sigma} \\ &\quad \times \left(\left(1 + \frac{\sigma u}{pt}\right)^{1/\sigma} - 1\right)^{-1-\sigma} \exp\left\{-\frac{1}{2} \left(\left(1 + \frac{\sigma u}{pt}\right)^{1/\sigma} - 1\right)\right\} du \quad (\text{C.1}) \end{aligned}$$

We will find a bound on  $\Gamma_{p,t}$  by first lower bounding,

$$A := \int_{\mathbb{R}_+} \left(1 - e^{-t\mu(z)}\right) dz,$$

and then upper bounding,

$$B = \left| p^\sigma \int_{\mathbb{R}_+} \left(1 - e^{-t\mu(z)}\right) dz - \int_{\mathbb{R}_+} \left(1 - e^{-pt\mu(z)}\right) dz \right|.$$

From the two bounds computed below, we will conclude that  $\Gamma_{p,t} \propto t^{-\sigma}$ .

*Lower bound on A.* Since the integrand is a positive function, we lower bound  $A$  by integrating on a smaller set. It is clear that when  $u \in [0, t]$ ,

$$\left(1 + \frac{\sigma u}{t}\right)^{-1+1/\sigma} \geq 1, \text{ and } \left(1 + \frac{\sigma u}{t}\right)^{1/\sigma} - 1 \lesssim \frac{u}{t}.$$

Therefore, from equation (C.1) we deduce that,

$$A \gtrsim t^\sigma \int_{[0,t]} (1 - e^{-u}) u^{-1-\sigma} du \gtrsim t^\sigma.$$

Upper bound on  $B$ . We first write,

$$\begin{aligned} \int_{\mathbb{R}_+} (1 - e^{-pt\mu(z)}) \, dz \\ = \int_0^t (1 - e^{-pt\mu(z)}) \, dz + \int_t^\infty (1 - e^{-pt\mu(z)}) \, dz =: C_{p,t} + D_{p,t}. \end{aligned}$$

It is easily seen from equation (C.1) that for some constant  $c > 0$  (eventually depending on  $\sigma$  and  $p$ , but not  $t$ ),

$$\begin{aligned} D_{p,t} &\lesssim \int_t^\infty (1 - e^{-u}) \left(\frac{u}{t}\right)^{-1+1/\sigma} \left(\frac{u}{t}\right)^{(-1-\sigma)/\sigma} e^{-c(u/t)^{1/\sigma}} \frac{du}{t} \\ &= t \int_t^\infty (1 - e^{-u}) u^{-2} e^{-c(u/t)^{1/\sigma}} \, du \\ &\leq \int_1^\infty u^{-2} e^{-cu^{1/\sigma}} \, du \lesssim 1. \end{aligned}$$

It turns out that  $B \lesssim |p^\sigma C_{1,t} - C_{p,t}| + 1$ . We now consider the function,

$$\begin{aligned} F(p, \sigma, t, u) &:= \frac{1}{pt} \left(1 + \frac{\sigma u}{pt}\right)^{-1+1/\sigma} \\ &\times \left( \left(1 + \frac{\sigma u}{pt}\right)^{1/\sigma} - 1 \right)^{-1-\sigma} \exp \left\{ -\frac{1}{2} \left( \left(1 + \frac{\sigma u}{pt}\right)^{1/\sigma} - 1 \right) \right\}. \end{aligned}$$

With a little bit of effort, it is seen that for any  $u \in [0, t]$ ,

$$F(p, \sigma, t, u) = u^{-1-\sigma} (tp)^\sigma + u^{-\sigma} O(t^{-1+\sigma}).$$

Hence,  $p^\sigma F(1, \sigma, t, u) - F(p, \sigma, tu) = u^{-\sigma} O(t^{-1+\sigma})$  whenever  $u \in [0, t]$ . It follows that,

$$\begin{aligned} |p^\sigma C_{1,t} - C_{p,t}| &\lesssim t^{-1+\sigma} \int_0^t (1 - e^{-u}) u^{-\sigma} \, du \\ &= \int_0^1 (1 - e^{-tu}) u^{-\sigma} \, du \leq \frac{1}{1-\sigma}. \end{aligned}$$

Henceforth,  $B \lesssim 1$ .